UNIVERSITEIT GENT

Faculty of Sciences

Department of Applied Mathematics and Computer Science

Chairman: Prof. dr. Willy Govaerts

A comprehensive study of fuzzy rough sets and their application in data reduction

Lynn D'EER

Supervisor: Prof. dr. Chris Cornelis

Co-supervisor: Prof. dr. Lluis Godo

Mentor: Nele Verbiest

Master Thesis submitted to obtain the academic degree of Master of Mathematics, option Applied Mathematics.

Academic year 2012–2013

Faculty of Sciences

Department of Applied Mathematics and Computer Science

Chairman: Prof. dr. Willy Govaerts

A comprehensive study of fuzzy rough sets and their application in data reduction

Lynn D'EER

Supervisor: Prof. dr. Chris Cornelis

Co-supervisor: Prof. dr. Lluis Godo

Mentor: Nele Verbiest

Master Thesis submitted to obtain the academic degree of Master of Mathematics, option Applied Mathematics.

Academic year 2012–2013

Dankwoord

In eerste instantie wil ik graag Prof. Dr. Chris Cornelis en Prof. Dr. Lluis Godo bedanken voor hun steun en inbreng. Zonder hun expertise, tijd en energie was het me niet gelukt om dit werk te schrijven.

Nele Verbiest wil ik graag bedanken voor de begeleiding bij deze thesis, en in het bijzonder voor het nalezen van de tekst en de hulp bij het Engels en de bibliografie.

Vervolgens wil ik mijn ouders bedanken voor hun onvoorwaardelijke steun. Ze hebben me steeds mijn eigen weg laten bewandelen.

Tenslotte wil ik Pieter bedanken voor zijn LaTeX-kennis, zijn eeuwige geduld en zijn liefde.

Toelating tot bruikleen

Lynn D’eer

31 mei 2013

Samenvatting

Datamining en patroonherkenning zijn wetenschappelijke domeinen die patronen herkennen in grote datasets. Toepassingen hiervan zijn bijvoorbeeld symptomen associëren met bepaalde ziektes in de medische wetenschappen en consumentengedrag in de sociale wetenschappen. Grote datasets zijn echter onhandig om mee te werken. We willen deze informatie beperken, maar zodanig dat de resultaten hetzelfde zijn. Gegevensreductie zoekt naar goede algoritmen om dit te doen. We willen een minimale verzameling van relevante attributen verkrijgen. Vaagruwverzamelingen kunnen helpen in het ontwerpen van deze algoritmen.

In ruwverzamelingenleer (Pawlak [50], 1982) benaderen we een onvolledig gekend concept: de onderbenadering bevat deze objecten die zeker aan het concept voldoen, terwijl de bovenbenadering de objecten bevat die mogelijk aan het concept voldoen. Daarnaast is vaagverzamelingenleer (Zadeh [67], 1965) een uitbreiding van de klassieke verzamelingenleer, in die zin dat een object in een zekere mate aan een concept voldoet. Meestal wordt dit beschreven door een getal tussen 0 en 1.

Dubois en Prade ([19, 20], 1990) combineerde als eerste deze twee theorieën. Door de mogelijkheden die vaagruwverzamelingen bieden aan gegevensreductie, winnen ze aan interesse bij onderzoekers. Eén van de uitdagingen is om robuuste modellen te ontwerpen, sinds de data waarmee we werken vaak ruis bevatten.

In deze thesis geven we een overzicht van de verschillende modellen in de literatuur die gebaseerd zijn op vaagruwverzamelingen. We onderzoeken hun eigenschappen en illustreren hoe we ze kunnen gebruiken in gegevensreductie.

In Hoofdstuk 2 bespreken we het model van Pawlak voor een equivalentierelatie en voor een algemene binaire relatie. We bestuderen het variable precision rough set model van Ziarko en de vaagverzamelingenleer van Zadeh. Verder bespreken we vaaglogische operatoren en hun eigenschappen en we vermelden enkele resultaten in verband met vaagrelaties.

In Hoofdstuk 3 geven we een overzicht van de bestaande vaagruwmodellen in de literatuur. We beginnen met het model van Dubois en Prade en geven de werkwijzen van Yao ([65]) en Wu et al. ([62, 63]) die ons meer inzicht geven in het model van Dubois en Prade. Daarna introduceren we een algemeen vaagruwmodel gebaseerd op een implicator en een conjunctor:

Definitie 1. Veronderstel dat A een vaagverzameling is in (U, R), met R een algemene vaagrelatie.

Stel $\mathscr{I}$ een implicator en $\mathscr{C}$ een conjunctor. De $(\mathscr{I},\mathscr{C})$-*vaagruwbenadering* van A is het paar van vaagverzamelingen $(R{\downarrow_{\mathscr{I}}}A, R{\uparrow_{\mathscr{C}}}A)$ zodat voor $x \in U$:

$$(R{\downarrow_{\mathscr{I}}}A)(x) = \inf_{y\in U} \mathscr{I}(R(y,x),A(y)),$$
$$(R{\uparrow_{\mathscr{C}}}A)(x) = \sup_{y\in U} \mathscr{C}(R(y,x),A(y)).$$

Dit model veralgemeent het model van Dubois en Prade en omvat veel vaagruwmodellen uit de literatuur. Vervolgens bestuderen we verfijningen van het algemene vaagruwmodel. Tenslotte bespreken we zes vaagruwmodellen die robust zijn ten opzichte van ruis in de data.

In Hoofdstuk 4 bespreken we de eigenschappen van de modellen uit Hoofdstuk 3. We vragen ons af of de eigenschappen van het scherpe model van Pawlak nog steeds gelden. We willen vooral weten of een model monotoon is wanneer we verschillende relaties beschouwen en of de onderbenadering bevat is in de verzameling zelf. Deze twee eigenschappen zijn belangrijk als we vaagruwmodellen willen gebruiken in gegevensreductie.

In het volgende hoofdstuk bespreken we de benaderingsoperatoren op een axiomatische manier. De operatoren voldoen aan een zeker axioma als en slechts als de relatie waarmee ze gedefinieerd zijn reflexief, symmetrisch of transitief is. Vervolgens bestuderen we duale paren voor een involutive negator en $\mathscr{T}$-gekoppelde paren voor een linkscontinue t-norm. We eindigen met een overzicht van axiomatische werkwijzen in de literatuur.

In Hoofdstuk 6 passen we vaagruwverzamelingenleer toe in gegevensreductie. We bespreken eerst the concepten van gegevensreductie voor modellen gebaseerd op ruwverzamelingenleer, waaronder de algoritmen ‘QuickReduct’ en ‘ReverseReduct’. Daarna breiden we deze concepten uit tot vaagruwverzamelingenleer. We bespreken hoe we positieve gebieden, randgebieden en onderscheidbaarheidsfuncties kunnen gebruiken om beslissingsreducten te vinden. Vervolgens bespreken we twee reductiealgoritmen: één gebaseerd op het model van Dubois en Prade, het andere gebaseerd op het algemene vaagruwmodel met een linkscontinue t-norm en zijn R-implicator. We vermelden ook enkele interessante relaties tussen verschillende reducten. We sluiten dit hoofdstuk af met een kort overzicht uit de literatuur over het gebruik van vaagruwverzamelingen in gegevensreductie.

Conclusies en open problemen worden besproken in Hoofdstuk 7.

Resume

Data mining and pattern recognition are domains in science that want to discover patterns in large datasets. Applications can be found in, for instance, medical science (e.g., what symptoms describe a certain disease) and social sciences (e.g., behaviour of consumers). Large datasets are difficult to work with, we want to reduce the information in such a way that the results are still the same. Feature selection searches for good algorithms to reduce the datasets, i.e., we want to find a minimal set of relevant attributes. Fuzzy rough set theory can help to find such algorithms.

Rough set theory (Pawlak [50], 1982) characterises a concept A by means of a lower and upper approximation. The lower approximation contains those objects that certainly fulfil A, while the upper approximation contains the objects that possibly fulfil A. On the other hand, fuzzy set theory (Zadeh [67], 1965) extends classical set theory in the sense that objects fulfil a concept in a certain degree.

Dubois and Prade ([19, 20], 1990) were the first to combine these two theories and many followed. Due to the potential of fuzzy rough set theory in machine learning and, in particular, feature selection, fuzzy rough set theory gains more and more interest. A big challenge is to find robust fuzzy rough set models that can deal with noise in the data.

In this thesis we give an overview of different fuzzy rough set models in the literature and their properties and we illustrate how we can use them in feature selection.

In the second chapter we recall the rough set model designed by Pawlak for an equivalence relation and a general binary relation. We discuss the variable precision rough set model of Ziarko and fuzzy set theory introduced by Zadeh. Further, we discuss fuzzy logical operators and their properties and we recall some notions about fuzzy relations.

In Chapter 3, we give an overview of existing fuzzy rough set models in the literature. We start with the model designed by Dubois and Prade. The approaches of Yao ([65]) and Wu et al. ([62, 63]) give us more insight in Dubois and Prade's model. Next, we introduce a general fuzzy rough set model based on an implicator and a conjunctor:

Definition 1. Let A be a fuzzy set in a fuzzy approximation space (U, R), with R a general fuzzy relation. Let $\mathscr{I}$ be an implicator and $\mathscr{C}$ a conjunctor. The $(\mathscr{I}, \mathscr{C})$-*fuzzy rough approximation* of A is

the pair of fuzzy sets $(R{\downarrow_{\mathscr{I}}}A, R{\uparrow_{\mathscr{C}}}A)$ such that for $x \in U$:

$$(R{\downarrow_{\mathscr{I}}}A)(x) = \inf_{y \in U} \mathscr{I}(R(y,x), A(y)),$$

$$(R{\uparrow_{\mathscr{C}}}A)(x) = \sup_{y \in U} \mathscr{C}(R(y,x), A(y)).$$

This model extends the model of Dubois and Prade and covers a lot of fuzzy rough set models studied in the literature. We continue with tight and loose approximation operators. They refine the general fuzzy rough set model. To end we discuss six fuzzy rough set models that are designed to deal with noisy data.

In Chapter 4, we discuss the properties of the general fuzzy rough set model, the tight and loose approximation operators and the robust fuzzy rough set models. We study if the properties of Pawlak's rough set model still hold. Among other things, we want to know if a model is monotone with respect to fuzzy relations and if the lower approximation of a set is included in the set itself. These two properties will be important if we want to use fuzzy rough set models in feature selection.

In the next chapter, we characterise an upper and lower approximation operator with axioms. The approximation operators fulfil a certain axiom if and only if a fuzzy relation is reflexive, symmetric or transitive. Next, we study dual pairs with respect to an involutive negator $\mathscr{N}$ and $\mathscr{T}$-coupled pairs with respect to a left-continuous t-norm $\mathscr{T}$. We end with an overview of axiomatic approaches in the literature.

In Chapter 6, we apply fuzzy rough set theory to feature selection. We first recall the concepts of feature selection in crisp rough set analysis. We discuss the QuickReduct and ReverseReduct algorithm. We continue with extending the concepts of feature selection in rough set analysis to fuzzy rough set analysis. We discuss how we can use positive regions, boundary regions and discernibility functions to find decision reducts. Next, we discuss two reduction algorithms based on the model of Dubois and Prade and the general fuzzy rough set model with a t-norm and its R-implicator. We state some interesting relations between different reducts. To end, we give a brief overview of fuzzy rough feature selection in the literature.

We conclude and outline future work in Chapter 7.

Contents

Chapter 1

Introduction

Nowadays, information is everywhere. Due to internet and smartphones, we can search for anything, everywhere. But is all this information relevant?

Not only in everyday life our information pool becomes bigger and bigger, databases in science and technology research also grow. Not only in the rows, i.e., the amount of objects observed, but also in the columns, i.e., the attributes we use to describe the objects. Not all these attributes are relevant. Big datasets are difficult to store and to understand. Feature selection is an important domain in research. The goal is to find good algorithms to select a minimal set of relevant attributes. We want maximal information content and minimal data storage.

Fuzzy rough set theory turns out to be a good technique to develop such algorithms. Since the late 80's, a lot of research on hybridisation of rough sets and fuzzy sets has been carried out.

Rough set theory (Pawlak [50], 1982) is a mathematical theory in which we want to approximate an uncertain concept. The lower approximation of a concept A contains those objects that certainly fulfil the concept, while the upper approximation of A contains the objects that possibly fulfil the concept. We divide the objects by their indiscernibility towards each other. Rough set theory is a common theory used in feature selection. We want to determine one or all decision reducts. A decision reduct is a minimal subset B of attributes such that objects that belong to different decision classes and that are discernible by all the attributes are still discernible by the attributes in B. We discover decision reducts by keeping the positive region of the data invariant or by reducing the discernibility function. To construct the positive region, we use the lower approximation of the decision classes with respect to the B-indiscernibility relation, i.e., an equivalence relation based on the attributes in B.

Problems arise when we have to deal with real-valued or quantitative attributes. Discretising data can lead to information loss. A possible solution is to introduce fuzzy set theory into feature selection.

Fuzzy set theory (Zadeh [67], 1965) is an extension of classical set theory. We use it when we deal with vague infomation. In classical set theory, an object fulfils a concept or it does not fulfil the concept. It is ‘yes’ or ‘no’, ‘1’ or ‘0’. However, in everyday life, nothing is binair. For example, when do you decide a person is old? Or tall? Or beautiful? Fuzzy set theory gives us the possibility to grade objects, i.e., an object belongs to a concept in a certain degree.

Combining these two theories leads to very interesting results that we can use in feature selection. Dubois and Prade ([19, 20], 1990) were the first to construct a fuzzy rough set model and after them, many followed. Since we sometimes deal with data that contains errors, robust models can be very useful. Robust fuzzy rough set models ensure that small changes in the data do not result in big changes in the output. The need for robust crisp rough set models was already stated by Ziarko ([71], 1993).

Feature selection is an important application of this hybrid theory. As in rough set feature selection, we use fuzzy rough set models to construct positive regions and dependency degrees to find one reduct or discernibility functions that gives us all reducts. With these techniques, we can omit irrelevant information and obtain a more workable dataset.

The goal of this thesis is to give an overview of different fuzzy rough set models in the literature and how we can use them for feature selection. We start with some preliminary notions in Chapter 2. In Chapter 3 we give an overview of different fuzzy rough set models and we study their properties in Chapter 4. In Chapter 5, we approach fuzzy rough sets in an axiomatic way. This will give us more insight. In Chapter 6 we illustrate how we can apply some of the models of Chapter 3 in feature selection. Conclusions and future work are stated in Chapter 7.

Chapter 2

Preliminaries

In this chapter we present the two keystones of this work. We start with the study of rough sets proposed by Zdzisław Pawlak, followed by the study of fuzzy sets proposed by Lotfi Zadeh. We also discuss the variable precision rough set model of Ziarko. Further, we study fuzzy logical operators and their properties and fuzzy relations.

2.1 Rough sets

We begin with rough sets introduced by Zdzisław Pawlak (Pawlak [50], 1982). We use them when we deal with insufficient and incomplete information. The basic idea is to construct a lower and an upper approximation of a given subset A of the universe U given an indiscernibility relation R on U. We assume the universe U to be non-empty and finite. If U is infinite, we will explicitly mention it.

We want to study if an element x in U is discernible from the elements in A (see e.g. [13]). This decision is based on the type of indiscernibility relation R on the universe U ($R \subseteq U \times U$). The definitions of the lower and upper approximation of the set A depend on the relation R. The pair (U,R) is called an *approximation space*. Pawlak studied approximations under an equivalence relation. However, his theory can easily be generalised for general binary relations.

Ziarko designed a rough set model that is more robust than the model of Pawlak. As we will see, the model of Pawlak is a special case of the variable precision rough set model of Ziarko.

We begin with the rough set theory of Pawlak.

2.1.1 Pawlak approximation space

When the relation R is an equivalence relation, we call the pair (U,R) a *Pawlak approximation space*.

Definition 2.1.1. An *equivalence relation* R on a universe U is a subset of $U \times U$ such that the following properties are fulfilled:

1. reflexivity, i.e., for all x in U it holds that $(x,x) \in R$,

2. symmetry, i.e., for all x and y in U it holds that $(x,y) \in R \Leftrightarrow (y,x) \in R$,

3. transitivity, i.e., for all x, y and z in U it holds that if $(x,y) \in R$ and $(y,z) \in R$, then $(x,z) \in R$.

With x in U, the subset $[x]_R = \{y \in U \mid (x,y) \in R\}$ of U is called the *equivalence class of x with respect to R*.

Next, we define the lower and upper approximation of a set A in a Pawlak approximation space (U,R) ([50]).

Definition 2.1.2. Let A be a subset in U, R an equivalence relation on U and $x \in U$. We define the *lower approximation $R{\downarrow}A$* of A as

$$\begin{aligned} x \in R{\downarrow}A &\Leftrightarrow [x]_R \subseteq A \\ &\Leftrightarrow (\forall y \in U)\big((x,y) \in R \Rightarrow y \in A\big) \end{aligned}$$

and the *upper approximation $R{\uparrow}A$* of A as

$$\begin{aligned} x \in R{\uparrow}A &\Leftrightarrow [x]_R \cap A \neq \emptyset \\ &\Leftrightarrow (\exists y \in U)\big((x,y) \in R \wedge y \in A\big). \end{aligned}$$

The lower approximation of A contains x if and only if its equivalence class $[x]_R$ is included in A. The upper approximation of A contains x if and only if its equivalence class $[x]_R$ has a non-empty intersection with A. This means that the lower approximation is the set of elements which necessarily satisfy the concept A (strong membership) and the upper approximation is the set of elements which possibly satisfy the concept A (weak membership) (see e.g. [13]). Both the lower approximation and the upper approximation of A are subsets of U.

We give a graphical example. Consider the universe U depicted in Figure 2.1 and a subset $A \subseteq U$. We have a partition of the universe by equivalence classes determined by the equivalence relation R. These equivalence classes are represented by the squares in the grid. The lower approximation is represented by the light grey squares, the upper approximation is the area inside the thick black line.

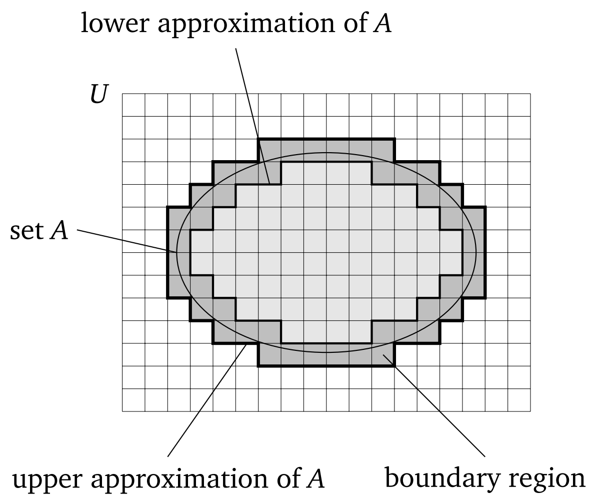

Figure 2.1: The lower and upper approximation of a set A

We now list some properties of rough sets. Every time we consider a new model, we will study which properties still hold in that model, or which assumptions we have to make to fulfil a given property (see Chapter 4).

Proposition 2.1.3. Let A and B be subsets in U and R an equivalence relation on U. Table 2.1 shows which properties are fulfilled.

We see that even in a Pawlak approximation space $R{\uparrow}(A \cap B) = R{\uparrow}A \cap R{\uparrow}B$ and $R{\downarrow}(A \cup B) = R{\downarrow}A \cup R{\downarrow}B$ do not hold. We illustrate this with a graphical example in Figures 2.2 and 2.3. In Figure 2.2, we see that $R{\uparrow}(A \cap B)$ is empty, while $R{\uparrow}A \cap R{\uparrow}B$ is given by the grey area.

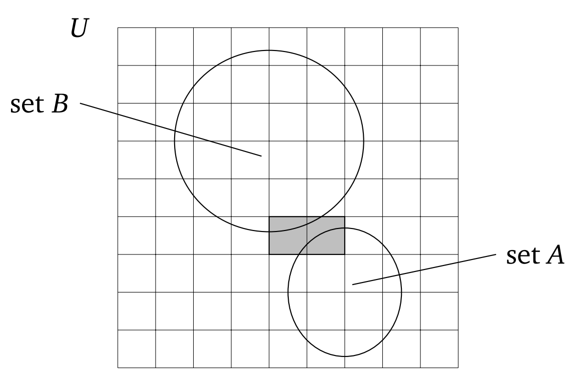

Figure 2.2: $R{\uparrow}(A \cap B) \subsetneq R{\uparrow}A \cap R{\uparrow}B$

In Figure 2.3, the grey area is included in $R{\downarrow}(A \cup B)$, but not in $R{\downarrow}A \cup R{\downarrow}B$.

Name	Property
Duality	$R{\downarrow}A = (R{\uparrow}A^c)^c$ $R{\uparrow}A = (R{\downarrow}A^c)^c$
Inclusion	$R{\downarrow}A \subseteq A$ $A \subseteq R{\uparrow}A$
Monotonicity of sets	$A \subseteq B \Rightarrow R{\downarrow}A \subseteq R{\downarrow}B$ $A \subseteq B \Rightarrow R{\uparrow}A \subseteq R{\uparrow}B$
Monotonicity of relations	$R_1 \subseteq R_2 \Rightarrow R_2{\downarrow}A \subseteq R_1{\downarrow}A$ $R_1 \subseteq R_2 \Rightarrow R_1{\uparrow}A \subseteq R_2{\uparrow}A$
Intersection	$R{\downarrow}(A \cap B) = R{\downarrow}A \cap R{\downarrow}B$ $R{\uparrow}(A \cap B) \subseteq R{\uparrow}A \cap R{\uparrow}B$
Union	$R{\downarrow}(A \cup B) \supseteq R{\downarrow}A \cup R{\downarrow}B$ $R{\uparrow}(A \cup B) = R{\uparrow}A \cup R{\uparrow}B$
Idempotence	$R{\downarrow}(R{\downarrow}A) = R{\downarrow}A$ $R{\uparrow}(R{\uparrow}A) = R{\uparrow}A$
$\emptyset$ and U	$R{\downarrow}\emptyset = \emptyset = R{\uparrow}\emptyset$ $R{\downarrow}U = U = R{\uparrow}U$

Table 2.1: Properties in a Pawlak approximation space

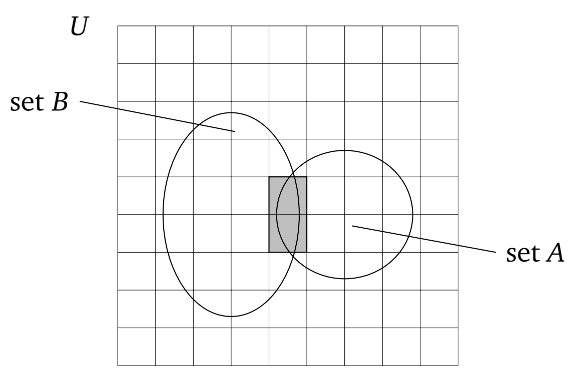

Figure 2.3: $R{\downarrow}(A \cup B) \supsetneq R{\downarrow}A \cup R{\downarrow}B$

2.1.2 Generalised approximation space

Pawlak approximation spaces have been generalised, since in many applications we only have a *binary relation* R on U ($R \subseteq U \times U$), which has fewer properties. When we deal with general binary relations, we do not speak about equivalence classes, but about *R-foresets* and *R-aftersets*.

The R-foreset of an element y in U is the subset

$$Ry = \{x \in U \mid (x, y) \in R\} \subseteq U \tag{2.1}$$

and the R-afterset of an element x in U is the subset

$$xR = \{y \in U \mid (x, y) \in R\} \subseteq U. \tag{2.2}$$

An equivalence relation on the universe U induces a partition of U. This means that two equivalence classes either coincide or are disjoint. If R is not an equivalence relation, it can occur that the R-foresets overlap. Furthermore, it is clear that if R is an equivalence relation, it holds that $Rx = [x]_R = xR$ for all x in U.

We consider some special binary relations besides an equivalence relation: a binary relation R that has the property of being reflexive, is called a *reflexive relation* and a relation R that is both reflexive and symmetric is called a *tolerance relation*.

When R is an arbitrary binary relation, we work in a *generalised approximation space* (U,R) instead of a Pawlak approximation space. Below, we give the definition of the lower and upper approximation of a subset A in a generalised approximation space (U,R). The lower and upper approximation of A are again subsets of U.

Definition 2.1.4. Let A be a subset in U and R a binary relation on U. An element $x \in U$ belongs to the *lower approximation* $R{\downarrow}A$ of A if and only if Rx is a subset of A, i.e.,

$$\begin{aligned} x \in R{\downarrow}A &\Leftrightarrow Rx \subseteq A \\ &\Leftrightarrow (\forall y \in U)((y, x) \in R \Rightarrow y \in A) \end{aligned}$$

and x belongs to the *upper approximation* $R{\uparrow}A$ of A if and only if Rx intersects A, i.e.,

$$\begin{aligned} x \in R{\uparrow}A &\Leftrightarrow Rx \cap A \neq \emptyset \\ &\Leftrightarrow (\exists y \in U)((y, x) \in R \wedge y \in A). \end{aligned}$$

It is clear that when R is an equivalence relation, this definition coincides with Definition 2.1.2.

We study the properties of the lower and upper approximation in a generalised approximation space.

Proposition 2.1.5. Let A and B be subsets in U and R a binary relation on U. The properties of duality, monotonicity of sets, monotonicity of relations, intersection, union, $\emptyset$ and U still hold (see Table 2.1). However, the inclusion property only holds if R is reflexive. For the property of idempotence, we have that $R{\downarrow}(R{\downarrow}A) = R{\downarrow}A$ and $R{\uparrow}(R{\uparrow}A) = R{\uparrow}A$ if R is reflexive and transitive.

To conclude, we list definitions that are applicable in both a Pawlak and a generalised approximation space. We also give a formal definition of a rough set.

Definition 2.1.6. We call a pair (A_1, A_2) in an approximation space (U, R) a *rough set*, if there is a subset A of U such that $R{\downarrow}A = A_1$ and $R{\uparrow}A = A_2$.

If we have the lower and upper approximation of a set A, we can also obtain the boundary region of A. It contains the elements of U for which we cannot say with certainty if they belong to A or to its complement A^c.

Definition 2.1.7. We call the set $R{\uparrow}A \setminus R{\downarrow}A$ the *boundary region* of a set A in (U, R).

The boundary region is marked by the dark grey squares in Figure 2.1. If the boundary region of a set A is empty, we call A a definable set.

Definition 2.1.8. When the lower and upper approximation of a set A in an approximation space (U, R) are the same, i.e., $R{\downarrow}A = R{\uparrow}A$, we call the set A *definable*.

We continue with the variable precision rough set model of Ziarko.

2.1.3 Variable precision rough sets

The original model designed by Pawlak has strict definitions, it does not allow misclassification. Changing one element can lead to drastic changes in the lower and upper approximation. The variable precision rough set model proposed by Ziarko (Ziarko [71], 1993) is designed to include tolerance to noisy data. In this model, we allow some misclassification. To do this, we generalise the standard set inclusion.

Let A and B be non-empty subsets of the universe U. In the classical definition of set inclusion, there is no room for misclassification, i.e., A is only included in B ($A \subseteq B$) if all elements of A belong to B. There is no distinction between sets that are more included in B than others. We introduce the measure to evaluate the relative degree of misclassification of a set A with respect to a set B.

Definition 2.1.9. Let A and B be subsets of U. The measure $c(A, B)$ of the *relative degree of misclassification* of the set A with respect to the set B is defined by

$$c(A,B) = \begin{cases} 1 - \frac{|A \cap B|}{|A|} & \text{if } A \neq \emptyset, \\ 0 & \text{if } A = \emptyset, \end{cases}$$

where $|A|$ denotes the cardinality of the set A.

We also call $c(A, B)$ the relative classification error and $c(A, B) \cdot |A|$ the absolute classification error. The more elements A and B have in common, the lower the relative degree of misclassification. So, if A is included in B according to the classical definition of inclusion, then $c(A, B) = 0$. Based on the measure $c(A, B)$, we can characterise the classical inclusion of A in B without explicitly using a quantifier:

$$A \subseteq B \text{ if and only if } c(A, B) = 0.$$

We can extend this in a natural way to the majority inclusion relation ([71]).

Definition 2.1.10. Given $0 \leq \beta < 0.5$ and $A, B \subseteq U$. We define the *majority inclusion relation* between A and B as

$$A \overset{\beta}{\subseteq} B \text{ if and only if } c(A,B) \leq \beta.$$

We obtain the standard set inclusion (or total inclusion) for $\beta = 0$. We also have the notion of the rough membership function.

Definition 2.1.11. Let R be a binary relation on U. For $A \subseteq U$ and $x \in U$ we define the *rough membership function* R_A of A as

$$R_A(x) = 1 - c(Rx, A) = \begin{cases} \frac{|Rx \cap A|}{|Rx|} & Rx \neq \emptyset \\ 1 & Rx = \emptyset. \end{cases}$$

The rough membership $R_A(x)$ quantifies the degree of inclusion of Rx into A and can be interpreted as the conditional probability that x belongs to A, given its foreset Rx.

Ziarko worked in a Pawlak approximation space, but we can also introduce the model in a generalised approximation space. We work with asymmetric boundaries as proposed by Katzberg and Ziarko ([38]).

Definition 2.1.12. Let A be a subset in U, R a binary relation on U and $x \in U$. With $0 \leq l < u \leq 1$ we define the lower approximation $R{\downarrow_u}A$ of A as

$$x \in R{\downarrow_u}A \Leftrightarrow R_A(x) \geq u$$

and the upper approximation $R{\uparrow_l}A$ of A as

$$x \in R{\uparrow_l}A \Leftrightarrow R_A(x) > l.$$

When $u = 1 - l$, we speak of a symmetric variable precision rough set model (VPRS). The original VPRS model proposed by Ziarko was based on an equivalence relation R and assumed $0 \leq l < 0.5$ and $u = 1 - l$. With $u = 1$ and $l = 0$, we obtain the original rough set model of Definition 2.1.4.

Let us illustrate Ziarko's model ([71]).

Example 2.1.13. Let $U = \{y_1, \ldots, y_{20}\}$ and let R be an equivalence relation on U such that

$$\begin{aligned}
[y_1]_R &= \{y_1, y_2, y_3, y_4, y_5\}, \\
[y_6]_R &= \{y_6, y_7, y_8\}, \\
[y_9]_R &= \{y_9, y_{10}, y_{11}, y_{12}\}, \\
[y_{13}]_R &= \{y_{13}, y_{14}\}, \\
[y_{15}]_R &= \{y_{15}, y_{16}, y_{17}, y_{18}\}, \\
[y_{19}]_R &= \{y_{19}, y_{20}\}.
\end{aligned}$$

Let A be the crisp set $\{y_4, y_5, y_8, y_{14}, y_{16}, y_{17}, y_{18}, y_{19}, y_{20}\}$. We compute the lower approximation of A for $u = 1$ and $u = 0.75$.

Take $x \in U$. If $u = 1$, then $x \in R{\downarrow}_1 A$ if and only if $[x]_R \subseteq A$. This only holds for $[y_{19}]_R$, so we derive that

$$R{\downarrow}_1 A = [y_{19}]_R = \{y_{19}, y_{20}\}.$$

On the other hand, if $u = 0.75$, then $x \in R{\downarrow}_{0.75} A$ if and only if

$$\frac{|[x]_R \cap A|}{|[x]_R|} \geq 0.75.$$

Of course this holds for $[y_{19}]_R$. Let us check this condition for the other equivalence classes:

$$\begin{aligned}
\frac{|[y_1]_R \cap A|}{|[y_1]_R|} &= \frac{2}{5} < 0.75,\\
\frac{|[y_6]_R \cap A|}{|[y_6]_R|} &= \frac{1}{3} < 0.75,\\
\frac{|[y_9]_R \cap A|}{|[y_9]_R|} &= \frac{0}{4} < 0.75,\\
\frac{|[y_{13}]_R \cap A|}{|[y_{13}]_R|} &= \frac{1}{2} < 0.75,\\
\frac{|[y_{15}]_R \cap A|}{|[y_{15}]_R|} &= \frac{3}{4} \geq 0.75.
\end{aligned}$$

We see that the condition also holds for $[y_{15}]_R$. Hence,

$$R{\downarrow}_{0.75} A = [y_{15}]_R \cup [y_{19}]_R = \{y_{15}, y_{16}, y_{17}, y_{18}, y_{19}, y_{20}\}.$$

This lower approximation contains more elements of A than $R{\downarrow}_1 A$.

As previous example already showed, the lower approximation is not necessarily included in A. Next proposition gives the properties that hold in the asymmetric VPRS model.

Proposition 2.1.14. Let A and B be subsets in U and R a binary relation on U. In the model defined in Definition 2.1.12, the monotonicity of sets holds, i.e., if $A \subseteq B$, then

$$\begin{aligned}
R{\downarrow}_u A &\subseteq R{\downarrow}_u B,\\
R{\uparrow}_l A &\subseteq R{\uparrow}_l B.
\end{aligned}$$

Furthermore, it holds that

$$\begin{aligned}
R{\downarrow}_u (A \cap B) &\subseteq R{\downarrow}_u A \cap R{\downarrow}_u B,\\
R{\uparrow}_l (A \cap B) &\subseteq R{\uparrow}_l A \cap R{\uparrow}_l B,\\
R{\downarrow}_u (A \cup B) &\supseteq R{\downarrow}_u A \cup R{\downarrow}_u B,\\
R{\uparrow}_l (A \cup B) &\supseteq R{\uparrow}_l A \cup R{\uparrow}_l B.
\end{aligned}$$

For the empty set and the universe, the following results hold:

$$R{\downarrow}_u\emptyset = \emptyset = R{\uparrow}_l\emptyset,$$
$$R{\downarrow}_u U = U = R{\uparrow}_l U.$$

The other properties of Table 2.1 do not hold in general.

In the special case of Ziarko's original model, some extra properties hold.

Proposition 2.1.15. Let A be a subset in U and R a binary relation on U and assume $l = 1-u$, $0 \leq l < 0.5$. Besides the properties from Proposition 2.1.14, we have the following equalities:

$$R{\downarrow}_u A = (R{\uparrow}_l A^c)^c,$$
$$R{\uparrow}_l A = (R{\downarrow}_u A^c)^c,$$

i.e., the duality property holds. We also have the following inclusions:

$$R{\downarrow}_u A \overset{l}{\subseteq} A,$$
$$R{\downarrow}_u A \subseteq R{\uparrow}_l A.$$

The inclusion $A \overset{u}{\subseteq} R{\uparrow}_l A$ does not hold in general.

Example 2.1.16. Let $U = \{y_1, \ldots, y_{20}\}$ and let R be an equivalence relation on U such that

$$[y_1]_R = \{y_1, y_2, y_3, y_4, y_5\},$$
$$[y_6]_R = \{y_6, y_7, y_8, y_9, y_{10}\},$$
$$[y_{11}]_R = \{y_{11}, y_{12}, y_{13}, y_{14}, y_{15}\},$$
$$[y_{16}]_R = \{y_{16}, y_{17}, y_{18}, y_{19}, y_{20}\}.$$

Let A be the crisp set $\{y_4, y_5, y_7, y_8, y_{14}, y_{16}, y_{17}, y_{18}\}$ and let $l = 0.4$, $u = 0.6$. We compute the upper approximation $R{\uparrow}_{0.4}A$. Since

$$\frac{|[y_1]_R \cap A|}{|[y_1]_R|} = \frac{2}{5} \leq 0.4,$$
$$\frac{|[y_6]_R \cap A|}{|[y_6]_R|} = \frac{2}{5} \leq 0.4,$$
$$\frac{|[y_{11}]_R \cap A|}{|[y_{11}]_R|} = \frac{1}{5} \leq 0.4,$$
$$\frac{|[y_{16}]_R \cap A|}{|[y_{16}]_R|} = \frac{3}{5} > 0.4,$$

the upper approximation of A is $R{\uparrow}_{0.4}A = [y_{16}]_R$. We have that $A \overset{0.6}{\subseteq} R{\uparrow}_{0.4}A$ if and only if

$$1 - \frac{|A \cap R{\uparrow}_{0.4}A|}{|A|} \leq 0.6.$$

Now, because $\frac{|A \cap R\uparrow_{0.4}A|}{|A|} = \frac{3}{8} = 0.375$, we have that $1 - 0.375 = 0.625$, which is greater than 0.6. Hence, $A \overset{0.6}{\nsubseteq} R\uparrow_{0.4}A$.

We continue with fuzzy set theory by Zadeh.

2.2 Fuzzy sets

In this section we recall some notions about fuzzy set theory, developed to model imprecise information and vagueness. Next, we discuss fuzzy logical operators and we end with some notions about fuzzy relations.

2.2.1 Fuzzy sets

Set theory is the basis of (classical) logic. If we work in a universe U, and we have a property A, we may decide for every element x in U whether it satisfies property A or not. For instance, we can say about a piece of fruit if it is an apple or not. Formally, we can denote the property A as a function χ_A from the universe U to the set $\{0,1\}$:

$$\chi_A \colon U \to \{0,1\}.$$

We call A a *crisp set* or an *ordinary set*. The function χ_A is called the *characteristic function* of A, where $\chi_A(x) = 1$ if x belongs to A (x satisfies property A) and $\chi_A(x) = 0$ otherwise. A concept A can be considered as a *subset* of the universe U ($A \subseteq U$). The set of all subsets of U is denoted by $\mathscr{P}(U)$.

In reality however, not everything can be decided in terms of black or white. For instance, consider the linguistic terms which we use to describe the height of a human being. There is no strict way to tell if somebody is tall or not. A man of height 1m80 is taller than a man of height 1m65, but he is not as tall as a man of height 1m95. In general, it is not possible to fix a threshold height for being tall. We cannot describe the property 'tall' with classical set theory.

In 1965, Lotfi Zadeh proposed a solution for this problem: he introduced fuzzy sets (Zadeh [67], 1965).

Definition 2.2.1. A *fuzzy set* A in U is a mapping $\mu_A \colon U \to [0,1]$, which we call the *membership function* of A. The set of fuzzy sets in U is denoted by $\mathscr{F}(U)$. If x is an element of U, we call $\mu_A(x)$ the *membership degree* of x in A.

Note that if A is a crisp set in U (i.e., $A \in \mathscr{P}(U)$), then μ_A is equal to the characteristic function χ_A of A. The set of fuzzy sets $\mathscr{F}(U)$ is therefore a superset of the set of subsets $\mathscr{P}(U)$:

$$\mathscr{P}(U) \subseteq \mathscr{F}(U).$$

Remark 2.2.2. In this work, as in many others, we denote the membership function μ_A by A. We also denote $[0,1]$ by I.

Let $\alpha \in I$. With $\hat{\alpha}$ we denote the constant (fuzzy) set such that $\hat{\alpha}(x) = \alpha$ for all $x \in U$.

When we work with fuzzy sets, we need to provide generalised definitions of the concepts given in classical set theory. For example, we define the cardinality of a fuzzy set A by

$$|A| = \sum_{x \in U} A(x).$$

When A is a crisp set, we obtain the same definition as in classical set theory.

For every fuzzy set, we have the concept of support and kernel. The *support* of a fuzzy set A is the crisp set

$$\mathrm{supp}(A) = \{x \in U \mid A(x) > 0\}.$$

The *kernel* of a fuzzy set A is the crisp set

$$\ker(A) = \{x \in U \mid A(x) = 1\}.$$

We now extend concepts like empty set, union, intersection, ... to fuzzy set theory. We study the extensions proposed by Zadeh.

A fuzzy set A is said to be *empty* if none of the elements of U belong to it, i.e., $A(x) = 0$ for every $x \in U$. We denote the empty set by $\emptyset$.

When we have two fuzzy sets A and B, we can define their union and intersection. We use the classical maximum and minimum operator.

Definition 2.2.3. The membership function of the *union* of two fuzzy sets A and B (denoted by $A \cup B$) is given by

$$\forall x \in U : (A \cup B)(x) = \max\{A(x), B(x)\}$$

with max the classical maximum operator.

Definition 2.2.4. The membership function of the *intersection* of two fuzzy sets A and B (denoted by $A \cap B$) is given by

$$\forall x \in U : (A \cap B)(x) = \min\{A(x), B(x)\}$$

with min the classical minimum operator.

When A and B are crisp sets, we obtain the classical union and intersection: for all x in U it holds that $(A \cup B)(x) = 1$ if and only if $A(x) = 1$ or $B(x) = 1$ (which means that $x \in A$ or $x \in B$) and that $(A \cap B)(x) = 1$ if and only if $A(x) = 1$ and $B(x) = 1$ (which means that $x \in A$ and $x \in B$).

The notion of a subset in fuzzy set theory is an extension of the classical definition.

Definition 2.2.5. We say that a fuzzy set A is *contained* in a fuzzy set B (or A is a *subset* of B, or A is *smaller than or equal to* B) if and only if $A \leq B$, i.e.,

$$\forall x \in U : A(x) \leq B(x).$$

We denote this by $A \subseteq B$.

In fuzzy set theory, the complement of A is defined by means of a decreasing function of the membership function of A. The definition proposed by Zadeh is:

Definition 2.2.6. The *complement* of a fuzzy set A is the fuzzy set A^{c} with membership function defined by

$$\forall x \in U : A^{\mathrm{c}}(x) = 1 - A(x).$$

In the crisp case it holds that the union of A and A^{c} is the entire universe U and the intersection of A and A^{c} is the empty set $\emptyset$. In general, this is not true in fuzzy set theory.

Every fuzzy set A can be associated with two families of crisp sets in U, namely the weak and strong α-level sets.

Definition 2.2.7. Given $\alpha \in I$, the *(weak) α-cut* or *(weak) α-level set* of a fuzzy set A is the crisp set A_α in U defined by

$$A_\alpha = \{x \in U \mid A(x) \geq \alpha\}.$$

Definition 2.2.8. Given $\alpha \in I$, the *strong α-cut* or *strong α-level set* of a fuzzy set A is the crisp set A_{α^+} in U defined by

$$A_{\alpha^+} = \{x \in U \mid A(x) > \alpha\}.$$

Note that the support of A is equal to the strong 0-level set A_{0^+} and that the kernel of A is the weak 1-level set A_1.

When we have a family of weak α-level sets, we can construct the fuzzy set A by

$$A(x) = \sup\{\alpha \mid x \in A_\alpha\} \tag{2.3}$$

for all $x \in U$.

We speak about a family of nested subsets $(A_\alpha)_\alpha$, $\alpha \in I$, if

$$\alpha_1 \leq \alpha_2 \Rightarrow A_{\alpha_2} \subseteq A_{\alpha_1}.$$

We prove the following property of a family of nested subsets.

Proposition 2.2.9. Let $\{\alpha_n \mid n \in \mathbb{N}\}$ be a non-decreasing sequence in I (i.e., $\alpha_i \leq \alpha_j$ for $i \leq j \in \mathbb{N}$) such that $\lim\limits_{n\to+\infty} \alpha_n = \alpha$, then $\bigcap\limits_{n=1}^{\infty} A_{\alpha_n} = A_\alpha$.

Proof. Let $x \in A_\alpha$, then for all $n \in \mathbb{N}$ it holds that $A(x) \geq \alpha \geq \alpha_n$. So, $x \in \bigcap_{n=1}^{\infty} A_{\alpha_n}$. Now, let $x \in \bigcap_{n=1}^{\infty} A_{\alpha_n}$. Then we have

$$\begin{aligned}
&\forall n \in \mathbb{N} \colon x \in A_{\alpha_n}\\
\Rightarrow &\forall n \in \mathbb{N} \colon A(x) \geq \alpha_n\\
\Rightarrow &A(x) \geq \sup\{\alpha_n \mid n \in \mathbb{N}\}\\
\Rightarrow &A(x) \geq \alpha\\
\Rightarrow &x \in A_\alpha.
\end{aligned}$$

This proves the property. □

Next, we discuss fuzzy logical operators.

2.2.2 Fuzzy logical operators

In classical logic, the semantics of the conjunction $\wedge$, disjunction $\vee$, negation $\neg$, implication $\rightarrow$ and coimplication $\nleftarrow$ are given by well-known truth-functions on the binary truth-value set $\{0,1\}$. When we work with truth values in $[0,1]$, we need fuzzy logical operators that extend these logical operators. We introduce in this section conjunctors and triangular norms, disjunctors and triangular conorms, negators, implicators and coimplicators (see e.g. [13, 53]).

Conjunctors and t-norms, disjunctors and t-conorms

The first fuzzy logical operator we discuss, is the conjunctor, an extension of the conjunction.

Definition 2.2.10. A *conjunctor* is a mapping $\mathscr{C} : I^2 \rightarrow I$ which is non-decreasing in both arguments and which satisfies the boundary conditions

$$\mathscr{C}(0,0) = \mathscr{C}(0,1) = \mathscr{C}(1,0) = 0 \text{ and } \mathscr{C}(1,1) = 1.$$

A commutative, associative conjunctor which satisfies $\mathscr{C}(1,a) = a$ for all $a \in I$ is called a t-norm and is denoted by $\mathscr{T}$.

Definition 2.2.11. A *triangular norm*, or t-norm, is a non-decreasing, associative and commutative mapping $\mathscr{T} : I^2 \rightarrow I$ that satisfies the boundary condition

$$\forall a \in I \colon \mathscr{T}(a,1) = a.$$

It holds that $\mathscr{T}(0,0) = \mathscr{T}(0,1) = \mathscr{T}(1,0) = 0$ and $\mathscr{T}(1,1) = 1$ which proves that a t-norm is a conjunctor.

Example 2.2.12. We give some examples of t-norms ($a, b \in I$):

- The standard minimum operator $\mathscr{T}_M(a, b) = \min\{a, b\}$. This is the largest t-norm.
- The product operator $\mathscr{T}_P(a, b) = a \cdot b$.
- The bold intersection or Łukasiewicz t-norm $\mathscr{T}_L(a, b) = \max\{0, a + b - 1\}$.
- The cosine t-norm $\mathscr{T}_{\cos}(a, b) = \max\left\{0, ab - \sqrt{(1 - a^2)(1 - b^2)}\right\}$.
- The drastic t-norm $\mathscr{T}_D$, which is the smallest t-norm and is defined by

$$\mathscr{T}_D(a, b) = \begin{cases} b & \text{if } a = 1 \\ a & \text{if } b = 1 \\ 0 & \text{otherwise.} \end{cases}$$

- The nilpotent minimal t-norm $\mathscr{T}_{nM}$:

$$\mathscr{T}_{nM}(a, b) = \begin{cases} \min\{a, b\} & \text{if } a + b > 1 \\ 0 & \text{otherwise.} \end{cases}$$

For every t-norm $\mathscr{T}$ we have

$$\forall a, b \in I \colon \mathscr{T}_M(a, b) \geq \mathscr{T}(a, b) \geq \mathscr{T}_D(a, b).$$

Because a t-norm is associative, the extension of a t-norm to the n-dimensional case is straightforward. We now introduce the notion of a β-precision quasi-t-norm ([56, 57]).

Definition 2.2.13. Let $\mathscr{T}$ be a t-norm and $\beta \in I$. The corresponding *β-precision quasi-t-norm* $\mathscr{T}_\beta$ is a mapping $\mathscr{T}_\beta \colon I^n \to I$ such that for all $\mathbf{a} = (a_1, \ldots, a_n)$ in I^n it holds that

$$\mathscr{T}_\beta(\mathbf{a}) = \mathscr{T}(b_1, \ldots, b_{n-m})$$

where $b_i = a_j$ if a_j is the ith greatest element of $\mathbf{a}$ and

$$m = \max\left\{ i \in \{0, \ldots, n\} \mid i \leq (1 - \beta) \sum_{j=1}^{n} a_j \right\}.$$

We see that with $\beta = 1$ and $m = 0$ we get the original t-norm $\mathscr{T}$.

When using conjunctors, we can define the $\mathscr{C}$-intersection of two fuzzy sets A and B.

Definition 2.2.14. The *$\mathscr{C}$-intersection* of two fuzzy sets A and B in U is defined by

$$\forall x \in U \colon (A \cap_{\mathscr{C}} B)(x) = \mathscr{C}(A(x), B(x)).$$

We see that the definition of Zadeh is a special case of a $\mathscr{C}$-intersection. He used the t-norm $\mathscr{T}_M = \min$.

Secondly, we give the definition of a disjunctor, an extension of the disjunction.

Definition 2.2.15. A *disjunctor* is a mapping $\mathscr{D}\colon I^2 \to I$ which is non-decreasing in both arguments and which satisfies the boundary conditions

$$\mathscr{D}(1,1) = \mathscr{D}(0,1) = \mathscr{D}(1,0) = 1 \text{ and } \mathscr{D}(0,0) = 0.$$

A commutative, associative disjunctor which satisfies $\mathscr{D}(a,0) = a$ for all $a \in I$ is called a t-conorm and is denoted by $\mathscr{S}$.

Definition 2.2.16. A *triangular conorm*, or t-conorm, is a non-decreasing, associative and commutative mapping $\mathscr{S} : I^2 \to I$ that satisfies the boundary condition

$$\forall a \in I : \mathscr{S}(a,0) = a.$$

Since $\mathscr{S}(0,0) = 0$ and $\mathscr{S}(0,1) = \mathscr{S}(1,0) = \mathscr{S}(1,1) = 1$, we see that a t-conorm is a disjunctor.

Example 2.2.17. We give some examples of t-conorms ($a, b \in I$):

- The standard maximum operator $\mathscr{S}_M(a,b) = \max\{a,b\}$. This is the smallest t-conorm.
- The probabilistic sum $\mathscr{S}_P(a,b) = a + b - a \cdot b$.
- The bounded sum or Łukasiewicz t-conorm $\mathscr{S}_L(a,b) = \min\{1, a+b\}$.
- The cosine t-conorm $\mathscr{S}_{\cos}(a,b) = \min\left\{1, a + b - ab + \sqrt{(2a-a^2)(2b-2b^2)}\right\}$.
- The drastic t-conorm $\mathscr{S}_D$, which is the greatest t-conorm and is defined by

$$\mathscr{S}_D(a,b) = \begin{cases} b & \text{if } a = 0 \\ a & \text{if } b = 0 \\ 1 & \text{otherwise.} \end{cases}$$

For every t-conorm $\mathscr{S}$ we have

$$\forall a, b \in I : \mathscr{S}_M(a,b) \leq \mathscr{S}(a,b) \leq \mathscr{S}_D(a,b).$$

As in the case of t-norms, we can extend t-conorms to the n-dimensional case and define β-precision quasi-t-conorms.

Definition 2.2.18. Let $\mathscr{S}$ be a t-conorm and $\beta \in I$. The corresponding *β-precision quasi-t-conorm* $\mathscr{S}_\beta$ is a mapping $\mathscr{S}_\beta : I^n \to I$ such that for all $\mathbf{a} = (a_1, \ldots, a_n)$ in I^n it holds that

$$\mathscr{S}_\beta(\mathbf{a}) = \mathscr{S}(b_1, \ldots, b_{n-m})$$

where $b_i = a_j$ if a_j is the ith smallest element of $\mathbf{a}$ and

$$m = \max\left\{ i \in \{0, \ldots, n\} \mid i \leq (1-\beta)\sum_{j=1}^{n}(1-a_j) \right\}.$$

With $\beta = 1$, $m = 0$ and we obtain the original t-conorm $\mathscr{S}$.

When using disjunctors, we can define the $\mathscr{D}$-union of two fuzzy sets A and B.

Definition 2.2.19. The *$\mathscr{D}$-union* of two fuzzy sets A and B in U is defined by

$$\forall x \in U : (A \cup_{\mathscr{D}} B)(x) = \mathscr{D}(A(x), B(x)).$$

Again, Zadeh's definition of the union is a special case, he used the t-conorm $\mathscr{S}_M = \max$.

We continue with negators.

Negators

We now consider an extension of the negation.

Definition 2.2.20. A *negator* $\mathscr{N}$ is a non-increasing mapping $\mathscr{N} : I \to I$ satisfying

$$\mathscr{N}(0) = 1 \text{ and } \mathscr{N}(1) = 0.$$

We give two examples of negators.

Example 2.2.21. The negator $\mathscr{N}_S(a) = 1 - a$ with a in I is called the standard negator. Another negator is the Gödel negator

$$\mathscr{N}_G(a) = \begin{cases} 1 & a = 0 \\ 0 & a \in]0, 1]. \end{cases}$$

Definition 2.2.22. A negator $\mathscr{N}$ is called *involutive* if and only if for every $a \in I$:

$$\mathscr{N}(\mathscr{N}(a)) = a.$$

It can be proven that every involutive negator is continuous (see e.g. [53]).

Given a negator $\mathscr{N}$, we can define the $\mathscr{N}$-complement of a fuzzy set A.

Definition 2.2.23. Let A be a fuzzy set of U and $\mathscr{N}$ a negator. We define the $\mathscr{N}$-complement $\mathrm{co}_{\mathscr{N}}$ of A by

$$\forall x \in U: \ \mathrm{co}_{\mathscr{N}}(A)(x) = \mathscr{N}(A(x)).$$

The definition given by Zadeh is a special case of the $\mathscr{N}$-complement, he used $\mathscr{N} = \mathscr{N}_S$.

There are some connections between t-norms, t-conorms and negators. First, in classical logic, we have De Morgan's laws. For all a, b in $\{0, 1\}$:

$$\neg(a \wedge b) = \neg a \vee \neg b,$$
$$\neg(a \vee b) = \neg a \wedge \neg b.$$

The extension of these laws leads us to a special connection between t-norms and t-conorms. This explains why we can talk about dual t-norms and t-conorms.

Definition 2.2.24. Given a negator $\mathscr{N}$, we call a t-norm $\mathscr{T}$ and a t-conorm $\mathscr{S}$ *dual with respect to* $\mathscr{N}$ if and only if De Morgan's laws are satisfied, i.e., for all a, b in I:

$$\mathscr{N}(\mathscr{T}(a,b)) = \mathscr{S}(\mathscr{N}(a), \mathscr{N}(b)),$$
$$\mathscr{N}(\mathscr{S}(a,b)) = \mathscr{T}(\mathscr{N}(a), \mathscr{N}(b)).$$

Secondly, many classical logical equivalences can be extended to fuzzy logic. For example

$$\forall a, b \in I: a \wedge b \longleftrightarrow \neg(\neg a \vee \neg b)$$

is the analogue of the following proposition.

Proposition 2.2.25. Given an involutive negator $\mathscr{N}$ and a t-conorm $\mathscr{S}$. Define

$$\forall a, b \in I: \mathscr{T}_{\mathscr{S},\mathscr{N}}(a,b) = \mathscr{N}\left(\mathscr{S}(\mathscr{N}(a), \mathscr{N}(b))\right),$$

then $\mathscr{T}_{\mathscr{S},\mathscr{N}}$ is a t-norm such that $\mathscr{T}_{\mathscr{S},\mathscr{N}}$ and $\mathscr{S}$ are dual with respect to $\mathscr{N}$.

We now study implicators and coimplicators.

Implicators and coimplicators

We continue with fuzzy logical operators that extend the implication and coimplication.

Definition 2.2.26. An *implicator* $\mathscr{I}$ is a mapping $\mathscr{I}: I^2 \to I$ satisfying

$$\mathscr{I}(1,0) = 0,$$
$$\mathscr{I}(1,1) = \mathscr{I}(0,1) = \mathscr{I}(0,0) = 1$$

and that is non-increasing in the first and non-decreasing in the second argument.

By definition, this is a conservative extension of the implication. Note that for every $a \in I$ we have $\mathscr{I}(0,a) = 1$, since

$$1 = \mathscr{I}(0,0) \leq \mathscr{I}(0,a).$$

We will now introduce some special implicators and their relations with the other fuzzy logical operators.

First, there is a relation between negators and implicators.

Proposition 2.2.27. If $\mathscr{I}$ is a implicator, then the operator $\mathscr{N}_{\mathscr{I}}$ defined by $\mathscr{N}_{\mathscr{I}}(a) = \mathscr{I}(a,0)$ for $a \in I$ is a negator, called the *negator induced by* $\mathscr{I}$.

We illustrate this.

Example 2.2.28. The Łukasiewicz implicator $\mathscr{I}_L(a,b) = \min(1, 1-a+b)$, $a, b \in I$, induces the standard negator $\mathscr{N}_S$:

$$\forall a \in I : \mathscr{N}_{\mathscr{I}_L}(a) = \mathscr{I}_L(a,0) = \min(1, 1-a+0) = 1-a = \mathscr{N}_S(a).$$

Below, we list some properties for implicators ([45]).

Definition 2.2.29. If an implicator $\mathscr{I}$ satisfies the *neutrality principle* (NP):

$$\forall a \in I : \mathscr{I}(1,a) = a,$$

we call $\mathscr{I}$ a *border implicator*.

Definition 2.2.30. If an implicator $\mathscr{I}$ satisfies the *exchange principle* (EP):

$$\forall a, b, c \in I : \mathscr{I}(a, \mathscr{I}(b,c)) = \mathscr{I}(b, \mathscr{I}(a,c)),$$

we call $\mathscr{I}$ an *EP implicator*.

Definition 2.2.31. If an implicator $\mathscr{I}$ satisfies the *confinement principle* (CP):

$$\forall a, b \in I : a \leq b \Leftrightarrow \mathscr{I}(a,b) = 1,$$

we call $\mathscr{I}$ an *CP implicator*.

Definition 2.2.32. Let $\mathscr{N}$ be a negator. If $\mathscr{I}$ satisfies

$$\forall a, b \in I : \mathscr{I}(\mathscr{N}(b), \mathscr{N}(a)) = \mathscr{I}(a,b),$$

we call $\mathscr{I}$ *contrapositive* with respect to $\mathscr{N}$.

We distinguish two important classes of implicators: S-implicators and R-implicators.

Let $\mathscr{T}$, $\mathscr{S}$ and $\mathscr{N}$ be a t-norm, t-conorm and negator respectively. The classical equivalence $a \to b \longleftrightarrow (\neg a) \vee b$ with a and b in $\{0,1\}$ leads to the concept of S-implicators.

Definition 2.2.33. The *S-implicator* $\mathscr{I}_{\mathscr{S},\mathscr{N}}$ based on the t-conorm $\mathscr{S}$ and the negator $\mathscr{N}$ is defined by

$$\forall a, b \in I\colon \mathscr{I}_{\mathscr{S},\mathscr{N}}(a, b) = \mathscr{S}(\mathscr{N}(a), b).$$

The definition of an R-implicator is given as follows:

Definition 2.2.34. The *residual implicator* or *R-implicator* $\mathscr{I}_{\mathscr{T}}$ based on the t-norm $\mathscr{T}$ is defined by

$$\forall a, b \in I\colon \mathscr{I}_{\mathscr{T}}(a, b) = \sup\{\lambda \in I \mid \mathscr{T}(a, \lambda) \leq b\}.$$

Note that if $a \leq b$, then $\mathscr{I}_{\mathscr{T}}(a, b) = 1$.

Proposition 2.2.35. The operators defined in Definitions 2.2.33 and 2.2.34 are border implicators that fulfil the exchange principle.

There is a important connection between a left-continuous[1] t-norm $\mathscr{T}$ and its residual implicator $\mathscr{I}_{\mathscr{T}}$ ([45]).

Proposition 2.2.36. Let $\mathscr{T}$ be a t-norm and $\mathscr{I}_{\mathscr{T}}$ the R-implicator based on $\mathscr{T}$. The pair $(\mathscr{T}, \mathscr{I}_{\mathscr{T}})$ fulfils the residual principle, i.e.,

$$\forall a, b, c \in I\colon \mathscr{T}(a, c) \leq b \Leftrightarrow \mathscr{I}_{\mathscr{T}}(a, b) \geq c,$$

if and only if $\mathscr{T}$ is left-continuous.

This property is sometimes called *Galois correspondance* or *adjunction property*. If $\mathscr{T}$ is left-continuous, then the pair $(\mathscr{T}, \mathscr{I}_{\mathscr{T}})$ has some useful properties ([54]).

Proposition 2.2.37. Let $\mathscr{T}$ be a left-continuous t-norm and $\mathscr{I}_{\mathscr{T}}$ its R-implicator. Let $\mathscr{N}$ be the induced negator by $\mathscr{I}_{\mathscr{T}}$. For $a, b, c, a_j, b_j \in I$, $j \in J$, it holds that

$$\mathscr{T}(a, \mathscr{I}_{\mathscr{T}}(a, b)) \leq b,$$
$$b \leq \mathscr{I}_{\mathscr{T}}(a, \mathscr{T}(a, b)),$$
$$\inf_{j \in J} \mathscr{I}_{\mathscr{T}}(a_j, b) = \mathscr{I}_{\mathscr{T}}(\sup_{j \in J} a_j, b),$$
$$\inf_{j \in J} \mathscr{I}_{\mathscr{T}}(a, b_j) = \mathscr{I}_{\mathscr{T}}(a, \inf_{j \in J} b_j),$$
$$\mathscr{I}_{\mathscr{T}}(a, \mathscr{I}_{\mathscr{T}}(b, c)) = \mathscr{I}_{\mathscr{T}}(\mathscr{T}(a, b), c),$$
$$\mathscr{I}_{\mathscr{T}}(a, \mathscr{N}(b)) = \mathscr{N}(\mathscr{T}(a, b)),$$
$$\mathscr{I}_{\mathscr{T}}(a, b) \leq \mathscr{I}_{\mathscr{T}}(\mathscr{N}(b), \mathscr{N}(a)).$$

[1]A formal definition of left-continuity is given in Definition 2.2.50.

A special group of R-implicators are IMTL-implicators ([21, 24]).

Definition 2.2.38. An involutive monoidal t-norm based logic-implicator or *IMTL-implicator* is an R-implicator based on a left-continuous t-norm $\mathscr{T}$ that has an involutive induced negator.

IMTL-implicators are contrapositive w.r.t. there induced negator, since

$$\mathscr{I}(x,y) \leq \mathscr{I}(\mathscr{N}_{\mathscr{I}}(y), \mathscr{N}_{\mathscr{I}}(x)) \leq \mathscr{I}(\mathscr{N}_{\mathscr{I}}(\mathscr{N}_{\mathscr{I}}(x)), \mathscr{N}_{\mathscr{I}}(\mathscr{N}_{\mathscr{I}}(y))) = \mathscr{I}(x,y)$$

when $\mathscr{I}$ is an R-implicator based on a left-continuous t-norm and $\mathscr{N}_{\mathscr{I}}$ is involutive ([54]).

We give some examples of S-, R- and IMTL-implicators (see [53]).

Example 2.2.39. For $a, b \in I$, three S-implicators are:

- The Kleene-Dienes implicator $\mathscr{I}_{KD}(a,b) = \max\{1-a, b\}$, based on the standard maximum operator $\mathscr{S}_M$ and the standard negator $\mathscr{N}_S$.
- The Kleene-Dienes-Łukasiewicz implicator $\mathscr{I}_{KDL}(a,b) = 1-a+a\cdot b$, based on the probabilistic sum $\mathscr{S}_P$ and the standard negator $\mathscr{N}_S$.
- The Łukasiewicz implicator $\mathscr{I}_L(a,b) = \min\{1, 1-a+b\}$, based on the Łukasiewicz t-conorm $\mathscr{S}_L$ and the standard negator $\mathscr{N}_S$.

Example 2.2.40. For $a, b \in I$, four R-implicators are:

- The Gödel implicator denoted by $\mathscr{I}_G$ and based on the standard minimum operator $\mathscr{T}_M$:

$$\mathscr{I}_G(a,b) = \begin{cases} 1 & \text{if } a \leq b \\ b & \text{if } a > b. \end{cases}$$

- The Gaines implicator denoted by $\mathscr{I}_{GA}$ and based on the product operator $\mathscr{T}_P$:

$$\mathscr{I}_{GA}(a,b) = \begin{cases} 1 & \text{if } a \leq b \\ \frac{b}{a} & \text{if } a > b. \end{cases}$$

- The Łukasiewicz implicator $\mathscr{I}_L(a,b) = \min\{1, 1-a+b\}$, based on the Łukasiewicz t-norm $\mathscr{T}_L$.
- The cosine implicator denoted by $\mathscr{I}_{\cos}$ based on the cosine t-norm $\mathscr{T}_{\cos}$:

$$\mathscr{I}_{\cos}(a,b) = \begin{cases} 1 & \text{if } a \leq b \\ ab + \sqrt{(1-a^2)(1-b^2)} & \text{if } a > b. \end{cases}$$

Example 2.2.41. An example of an IMTL-implicator is the R-implicator $\mathscr{I}_{nM}$ based on the nilpotent minimum t-norm $\mathscr{T}_{nM}$:

$$\forall a, b \in I : \mathscr{I}_{nM}(a, b) = \begin{cases} 1 & \text{if } a \leq b \\ \max\{1 - a, b\} & \text{if } a > b. \end{cases}$$

Just like $\mathscr{C}$-intersections and $\mathscr{D}$-unions, we can define $\mathscr{I}$-implications.

Definition 2.2.42. Let $\mathscr{I}$ be an implicator and A and B fuzzy sets in U. The *$\mathscr{I}$-implication* of A and B is denoted by $\Rightarrow_{\mathscr{I}} (A, B)$ and is defined by

$$\forall x \in U \colon (A \Rightarrow_{\mathscr{I}} B)(x) = \mathscr{I}(A(x), B(x)).$$

Apart from implicators, we also need coimplicators (see e.g. [1]). While implicators are an extension of the implication, coimplicators are an extension of the coimplication $\nleftarrow$, where $p \nleftarrow q$ means 'p is not necessary for q', i.e., $p \nleftarrow q$ only holds if p is false and q is true. We first define a general coimplicator.

Definition 2.2.43. A *coimplicator* $\mathscr{J}$ is a mapping $\mathscr{J} : I^2 \to I$ satisfying

$$\mathscr{J}(0, 1) = 1,$$
$$\mathscr{J}(1, 1) = \mathscr{J}(1, 0) = \mathscr{J}(0, 0) = 0$$

and that is non-increasing in the first and non-decreasing in the second argument.

We mostly work with residual coimplicators, based on a t-conorm $\mathscr{S}$.

Definition 2.2.44. Let $\mathscr{S}$ be a t-conorm. We define the *residual coimplicator* $\mathscr{J}_{\mathscr{S}}$ based on $\mathscr{S}$ by

$$\forall a, b \in I \colon \mathscr{J}_{\mathscr{S}}(a, b) = \inf\{\lambda \in I \mid \mathscr{S}(a, \lambda) \geq b\}.$$

We see that a residual coimplicator is non-increasing in the first and non-decreasing in the second argument and that it satisfies the boundary conditions $\mathscr{J}_{\mathscr{S}}(0, 1) = 1$ and $\mathscr{J}_{\mathscr{S}}(1, 1) = \mathscr{J}_{\mathscr{S}}(1, 0) = \mathscr{J}_{\mathscr{S}}(0, 0) = 0$. Note also that if $a \geq b$, then $\mathscr{J}_{\mathscr{S}}(a, b) = 0$.

Coimplicators are dual operators of implicators in the same way t-conorms are dual operators of t-norms. If $\mathscr{S}$ is the dual t-conorm of $\mathscr{T}$ with respect to a negator $\mathscr{N}$, $\mathscr{J}_{\mathscr{S}}$ is dual to $\mathscr{I}_{\mathscr{T}}$ with respect to $\mathscr{N}$, i.e.,

$$\forall a, b \in I \colon \mathscr{N}(\mathscr{J}_{\mathscr{S}}(a, b)) = \mathscr{I}_{\mathscr{T}}(\mathscr{N}(a), \mathscr{N}(b)).$$

We give some examples of residual coimplicators.

Example 2.2.45. For $a, b \in I$, we have the following residual coimplicators:

- With $\mathscr{S}_M$ the standard maximum operator, we derive the coimplicator $\mathscr{J}_M$ that is defined by

$$\mathscr{J}_M(a,b) = \begin{cases} 0 & \text{if } a \geq b \\ b & \text{if } a < b. \end{cases}$$

- With $\mathscr{S}_P$ the probabilistic sum, we derive the coimplicator $\mathscr{J}_P$ that is defined by

$$\mathscr{J}_P(a,b) = \begin{cases} 0 & \text{if } a \geq b \\ \frac{b-a}{1-a} & \text{if } a < b. \end{cases}$$

- With $\mathscr{S}_L$ the Łukasiewicz t-conorm, we derive the coimplicator $\mathscr{J}_L$ that is defined by

$$\mathscr{J}_L(a,b) = \max\{0, b-a\}.$$

- With $\mathscr{S}_{\cos}$ the cosine t-conorm, we derive the coimplicator $\mathscr{J}_{\cos}$ that is defined by

$$\mathscr{J}_{\cos}(a,b) = \begin{cases} 0 & \text{if } a \geq b \\ a+b-ab-\sqrt{(2a-a^2)(2b-b^2)} & \text{if } a < b. \end{cases}$$

We now connect the notions of coimplicators and conjunctors.

Proposition 2.2.46. Let $\mathscr{N}$ be an involutive negator and $\mathscr{J}$ a coimplicator. The map $\mathscr{C} : I^2 \to I$ defined by

$$\forall a,b \in I : \mathscr{C}(a,b) = \mathscr{J}(\mathscr{N}(a), b)$$

is a conjunctor, but not necessarily a t-norm.

With the four coimplicators defined above and the standard negator $\mathscr{N}_S$, we obtain the following conjunctors:

- The conjunctor based on $\mathscr{J}_M$ and $\mathscr{N}_S$ is

$$\forall a,b \in I : \mathscr{C}(a,b) = \begin{cases} 0 & \text{if } 1-a \geq b \\ b & \text{if } 1-a < b. \end{cases}$$

- The conjunctor based on $\mathscr{J}_P$ and $\mathscr{N}_S$ is

$$\forall a,b \in I : \mathscr{C}(a,b) = \begin{cases} 0 & \text{if } 1-a \geq b \\ \frac{a+b-1}{a} & \text{if } 1-a < b. \end{cases}$$

- The conjunctor based on $\mathscr{I}_L$ and $\mathscr{N}_S$ is

$$\forall a, b \in I \colon \mathscr{C}(a,b) = \max\{0, a+b-1\}.$$

- The conjunctor based on $\mathscr{I}_{\cos}$ and $\mathscr{N}_S$ is

$$\forall a, b \in I \colon \mathscr{C}(a,b) = \begin{cases} 0 & \text{if } 1-a \geq b \\ 1-a+ab-\sqrt{(1-a^2)(2b-b^2)} & \text{if } 1-a < b. \end{cases}$$

The first, second and last conjunctor are not commutative, so they are not a t-norm. The third one is the Łukasiewicz t-norm.

We end this section of fuzzy logical operators by recalling some basic notions of continuity.

Continuity

We recall some definitions about continuity that are used in this dissertation. We first start with the following useful characterisation ([45]).

Proposition 2.2.47. Consider a mapping $F : I^2 \to I$ that is monotonic with respect to one variable. It holds that F is continuous if and only if F is continuous in both variables.

Since all fuzzy logical operators are monotone in both variables, it is enough to define continuity for functions in one variable. We give the definitions of being continuous, lower semicontinuous and left-continuous.

Definition 2.2.48. A function $f : I \to I$ is *continuous* in a point $a \in I$ if

$$(\forall \epsilon > 0)(\exists \delta > 0)(\forall x \in I) : (|x-a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon).$$

A function $f : I \to I$ is continuous if it is continuous in every point of I.

Definition 2.2.49. A function $f : I \to I$ is *lower semicontinuous* in a point $a \in I$ if

$$(\forall \epsilon > 0)(\exists \delta > 0)(\forall x \in I) : (|x-a| < \delta \Rightarrow f(x) \geq f(a) - \epsilon).$$

A function $f : I \to I$ is lower semicontinuous if it is lower semicontinuous in every point of I.

Definition 2.2.50. A function $f : I \to I$ is *left-continuous* in a point $a \in I$ if

$$(\forall \epsilon > 0)(\exists \delta > 0)(\forall x \in I) : (a - \delta < x < a \Rightarrow |f(x) - f(a)| < \epsilon).$$

A function $f : I \to I$ is left-continuous if it is left-continuous in every point of I.

We have a useful connection for t-norms that are left-continuous and that are complete-distributive w.r.t. the supremum.

Definition 2.2.51. A t-norm $\mathcal{T}$ is *complete-distributive w.r.t. the supremum* if for every family $(a_j)_{j\in J}$ in I and for every $b \in I$ it holds that

$$\mathcal{T}\left(\sup_{j\in J} a_j, b\right) = \sup_{j\in J} \mathcal{T}(a_j, b).$$

The next property will be useful in proofs ([45]).

Proposition 2.2.52. A t-norm $\mathcal{T}$ is complete-distributive w.r.t. the supremum if and only if $\mathcal{T}$ is left-continuous.

The residual principle holds for left-continuous t-norms. But sometimes it is enough to have lower semicontinuity, due to the following property and to the fact that a t-norm is non-decreasing in both variables and commutative (see [23]).

Proposition 2.2.53. A t-norm $\mathcal{T}$ is lower semicontinuous if and only if $\mathcal{T}$ is left-continuous in its first component.

To end this chapter, we study fuzzy relations.

2.2.3 Fuzzy relations

In the crisp case, a relation R is a subset of $U \times U$. We now study fuzzy relations that are fuzzy sets in $U \times U$.

Consider a *fuzzy relation* $R \in \mathcal{F}(U \times U)$. We can extend the concept of an R-foreset and R-afterset (see Equations (2.1) and (2.2)): the R-foreset of an element y of U is the fuzzy set $Ry\colon U \to I$ defined by

$$\forall x \in U\colon Ry(x) = R(x, y)$$

and the R-afterset of an element x of U is the fuzzy set $xR\colon U \to I$ defined by

$$\forall y \in U\colon xR(y) = R(x, y).$$

We recall two special types of fuzzy relations.

Definition 2.2.54. A relation R is called a *fuzzy tolerance relation* if it satisfies the following properties:

1. reflexivity, i.e., for all x in U it holds that $R(x, x) = 1$,

2. symmetry, i.e., for all x and y in U it holds that $R(x, y) = R(y, x)$.

Definition 2.2.55. Let $\mathscr{T}$ be a t-norm. If a fuzzy tolerance relation R fulfils the property of being $\mathscr{T}$-transitive, i.e., for all x, y and z in U:

$$\mathscr{T}(R(x,y),R(y,z)) \leq R(x,z),$$

we call R a *fuzzy $\mathscr{T}$-similarity relation*, *fuzzy $\mathscr{T}$-equivalence relation* or *fuzzy $\mathscr{T}$-indistinguishability relation* (see e.g. [66]).

Mostly, we omit the word 'fuzzy'. When $\mathscr{T} = \min$, we shortly speak about a similarity relation. Because the minimum operator is the largest t-norm, we have for every t-norm $\mathscr{T}$ that

$$\mathscr{T}(R(x,y),R(y,z)) \leq \min\{R(x,y),R(y,z)\},$$

which means that if a relation R is min-transitive, it is $\mathscr{T}$-transitive for every t-norm $\mathscr{T}$ and thus, a similarity relation is a $\mathscr{T}$-similarity relation for every t-norm $\mathscr{T}$.

When we have a fuzzy $\mathscr{T}$-similarity relation R, the R-foreset and the R-afterset of x are the same. We call it the *fuzzy similarity class* of x and it will be denoted by Rx, xR or $[x]_R$. The definition of a fuzzy $\mathscr{T}$-similarity relation is a conservative extension of the definition of an equivalence relation in a crisp setting.

If a relation is not $\mathscr{T}$-transitive, one can determine its transitive closure. To do this, we first introduce the round composition of two fuzzy relations ([11]).

Definition 2.2.56. Let $\mathscr{T}$ be a t-norm. The *round composition* of two fuzzy relations R_1 and R_2 in U is the fuzzy relation $R_1 \circ R_2$ in U defined by

$$\forall x,z \in U\colon (R_1 \circ R_2)(x,z) = \sup_{y \in U} \mathscr{T}(R(x,y),R(y,z)).$$

We denote $R^1 = R$ and $R^n = R \circ R^{n-1}$ for a fuzzy relation R and $n \in \mathbb{N} \setminus \{0\}$. If R is $\mathscr{T}$-transitive, then $R \circ R = R$. Now, if R is not $\mathscr{T}$-transitive and if U is finite and $|U| \geq 2$, then the *$\mathscr{T}$-transitive closure of* R is given by $R^{|U|-1}$. This means that $R \circ R^{|U|-1} = R^{|U|-1}$.

When we have a t-norm $\mathscr{T}$, we can define $\mathscr{T}$-partitions on the universe U ([2]). Let $\mathscr{I}_{\mathscr{T}}$ be the R-implicator associated with $\mathscr{T}$, then we have the following fuzzy operator $\mathscr{E}_{\mathscr{T}}$ defined by:

$$\forall a,b \in I\colon \mathscr{E}_{\mathscr{T}}(a,b) = \min\{\mathscr{I}_{\mathscr{T}}(a,b), \mathscr{I}_{\mathscr{T}}(b,a)\} = \mathscr{I}_{\mathscr{T}}(\max\{a,b\}, \min\{a,b\}).$$

With this operator, we can define a $\mathscr{T}$-semipartition.

Definition 2.2.57. Let $\mathscr{T}$ be a t-norm. A collection $\mathscr{P}$ of fuzzy sets in U is called a *$\mathscr{T}$-semipartition* if and only if for every $A,B \in \mathscr{P}$ it holds that

$$\sup_{x \in U} \mathscr{T}(A(x),B(x)) \leq \inf_{x \in U} \mathscr{E}_{\mathscr{T}}(A(x),B(x)).$$

If moreover the kernels of the fuzzy sets in $\mathscr{P}$ forms a crisp partition of U, we speak about a $\mathscr{T}$-partition.

Definition 2.2.58. Let $\mathscr{T}$ be a t-norm. A collection $\mathscr{P}$ of fuzzy sets in U is called a *$\mathscr{T}$-partition* if and only if it is a $\mathscr{T}$-semipartition and if

$$k(\mathscr{P}) = \{\ker(A) \mid A \in \mathscr{A}\}$$

forms a partition of U.

We have a one-to-one correspondance between $\mathscr{T}$-partitions and fuzzy $\mathscr{T}$-similarity relations ([2]).

Proposition 2.2.59. Let $\mathscr{T}$ be a t-norm, then $\mathscr{P}$ is a $\mathscr{T}$-partition of U if and only if there exists a fuzzy $\mathscr{T}$-similarity relation R on U such that

$$\mathscr{P} = \{[x]_R \mid x \in U\}.$$

When we speak about properties of a fuzzy relation R, we mostly refer to reflexivity, symmetry and transitivity. There are also other properties a fuzzy relation can have.

Definition 2.2.60. A fuzzy relation R is *serial* if for every $x \in U$ it holds that $\sup\limits_{y \in U} R(x, y) = 1$.

In an obvious way, we have the property of being inverse serial.

Definition 2.2.61. A fuzzy relation R is *inverse serial* if for every $x \in U$ it holds that $\sup\limits_{y \in U} R(y, x) = 1$.

To end this section, we study some special fuzzy relations based on kernel functions ([28, 30]). We first define a kernel function.

Definition 2.2.62. A real-valued function

$$k\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

is said to be a *kernel function* if it is symmetric and positive semidefinite, i.e., for all $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ and for all complex numbers $\rho_1, \ldots, \rho_n$ it holds that

$$\sum_{i,j=1}^{n} k(x_i - y_j) \cdot \rho_i \cdot \bar{\rho_j} \geq 0$$

where $\bar{\rho_j}$ is the complex conjugate of ρ_j, i.e., if $\rho_j = a + bi$, then $\bar{\rho_j} = a - bi$.

We can see kernel functions as fuzzy relations, if the image of the kernel function is in I, i.e.,

$$k\colon \mathbb{R}^n \times \mathbb{R}^n \to I.$$

Let us assume that $U \subseteq \mathbb{R}^n$. A reflexive kernel function has the following property ([30]):

Proposition 2.2.63. Any kernel function $k\colon U \times U \to I$ with $k(\mathbf{x}, \mathbf{x}) = 1$ is (at least) $\mathscr{T}_{\cos}$-transitive.

Some kernel functions are reflexive, symmetric and $\mathscr{T}_{\cos}$-transitive, thus the relations computed with these kernel functions are fuzzy $\mathscr{T}_{\cos}$-similarity relations.

Recall that the Euclidean distance for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is given by

$$||\mathbf{x}-\mathbf{y}|| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

for $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$.

We give some examples of kernel functions ([28]).

Example 2.2.64. Let $\mathbf{x}$ and $\mathbf{y}$ be elements of U. Every kernel function has a parameter $\delta > 0$ that determines the geometrical structure of the mapped samples in the kernel function space.

1. The Gaussian kernel function: $k_G(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{||\mathbf{x}-\mathbf{y}||^2}{\delta}\right)$.
2. The exponential kernel function: $k_E(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{||\mathbf{x}-\mathbf{y}||}{\delta}\right)$.
3. The rational quadratic kernel function: $k_R(\mathbf{x}, \mathbf{y}) = 1 - \frac{||\mathbf{x}-\mathbf{y}||^2}{||\mathbf{x}-\mathbf{y}||^2+\delta}$.
4. The circular kernel function:

$$k_C(\mathbf{x}, \mathbf{y}) = \frac{2}{\pi}\arccos\left(\frac{||\mathbf{x}-\mathbf{y}||}{\delta}\right) - \frac{2}{\pi}\frac{||\mathbf{x}-\mathbf{y}||}{\delta}\sqrt{1-\left(\frac{||\mathbf{x}-\mathbf{y}||}{\delta}\right)^2}$$

   if $||\mathbf{x}-\mathbf{y}|| < \delta$, and $k_C(\mathbf{x}, \mathbf{y}) = 0$ otherwise.
5. The spherical kernel function:

$$k_S(\mathbf{x}, \mathbf{y}) = 1 - \frac{3}{2}\frac{||\mathbf{x}-\mathbf{y}||}{\delta} + \frac{1}{2}\left(\frac{||\mathbf{x}-\mathbf{y}||}{\delta}\right)^3$$

   if $||\mathbf{x}-\mathbf{y}|| < \delta$, and $k_S(\mathbf{x}, \mathbf{y}) = 0$ otherwise.

Chapter 3

Fuzzy rough sets

In the previous chapter, we studied rough sets and fuzzy sets. We can combine these two essentially different concepts in various ways. Since the first proposal by Dubois and Prade, it was clear that the two theories worked complementary, and not competitive. Using them together, leads to very good models for dealing with uncertain, incomplete and noisy data.

In this chapter, we study constructive approaches, i.e., we start with a fuzzy set A and a fuzzy relation R and we *define* the lower and upper approximation operators based on this data. In Chapter 5, we will study an axiomatic approach to describe fuzzy rough sets.

In Section 3.1, we recall the approach of Dubois and Prade, who constructed the basis of fuzzy rough set theory. In Section 3.2, we generalise the model of Dubois and Prade by using arbitrary implicators and conjunctors. We also give an overview of special cases of this implicator-conjunctor-based fuzzy rough set model. Next, in Section 3.3, we recall a possible way to refine the model introduced in Section 3.2. To end, we study fuzzy rough models designed to deal with noisy data in Section 3.4.

3.1 Hybridisation of rough and fuzzy sets

Hybridisation theory can lead to a rough fuzzy set or a fuzzy rough set. We first recall both concepts. In Section 3.1.3 we explain the difference mathematically.

3.1.1 Rough fuzzy sets and fuzzy rough sets

A *rough fuzzy set* is the pair of the lower and upper approximation of a fuzzy set A in a Pawlak or generalised approximation space (U,R). A *fuzzy rough set* is the pair of the lower and upper approximation of a crisp or fuzzy set A in a fuzzy approximation space (U,R), where a *fuzzy approximation space* is a pair (U,R) with U a universe and R a fuzzy relation.

In most applications, we deal with both a fuzzy set A and a fuzzy relation R. Because a crisp relation is a special type of a fuzzy relation, rough fuzzy sets can be seen as a special case of fuzzy rough sets. The study of fuzzy rough sets is immediately applicable to rough fuzzy sets.

We continue with discussing the fuzzy rough set model of Dubois and Prade.

3.1.2 Fuzzy rough sets by Dubois and Prade

Dubois and Prade laid the foundation of the concept of fuzzy rough sets ([19, 20]). They worked in a universe U with a fuzzy similarity relation R on U. They define a fuzzy rough set as follows:

Definition 3.1.1. Let A be a fuzzy set in a fuzzy approximation space (U,R), where R is a fuzzy similarity relation on U. A *fuzzy rough set* in (U,R) is a pair $(R{\downarrow}A, R{\uparrow}A)$ of fuzzy sets in U that for every x in U are defined by

$$(R{\downarrow}A)(x) = \inf_{y\in U}\{\max\{1-R(y,x),A(y)\}\},$$
$$(R{\uparrow}A)(x) = \sup_{y\in U}\{\min\{R(y,x),A(y)\}\}.$$

Assume now that A is a crisp set in U and R is a crisp equivalence relation on U. For x in U we have that

$$\begin{aligned}
(R{\downarrow}A)(x)=1 &\Leftrightarrow \inf_{y\in U}\{\max\{1-R(y,x),A(y)\}\}=1\\
&\Leftrightarrow \forall y\in U: 1-R(y,x)=1 \vee A(y)=1\\
&\Leftrightarrow \forall y\in U: (y,x)\in R \Rightarrow y\in A\\
&\Leftrightarrow [x]_R \subseteq A,\\
(R{\uparrow}A)(x)=1 &\Leftrightarrow \sup_{y\in U}\{\min\{R(y,x),A(y)\}\}=1\\
&\Leftrightarrow \exists y\in U: R(y,x)=1 \wedge A(y)=1\\
&\Leftrightarrow \exists y\in U: (y,x)\in R \wedge y\in A\\
&\Leftrightarrow [x]_R \cap A \neq \emptyset.
\end{aligned}$$

This shows that Definition 3.1.1 is a conservative extension of Definition 2.1.2.

The definition given by Dubois and Prade is the starting point for research for fuzzy rough sets. They derived these definitions invoking notions of C-calculus and possibility theory which fall outside the scope of this dissertation. In the next section, we provide an alternative justification involving α-level sets proposed by Yao ([65]).

We illustrate Definition 3.1.1 with an example.

Example 3.1.2. Let $U = \{y_1, y_2\}$, A a fuzzy set with $A(y_1) = 0.2$, $A(y_2) = 0.8$ and R a fuzzy similarity relation with $R(y_1, y_2) = 0.5$. We compute the lower and upper approximation of the fuzzy set A:

$$(R{\downarrow}A)(y_1) = \inf\{\max\{1-1, 0.2\}, \max\{1-0.5, 0.8\}\} = \inf\{0.2, 0.8\} = 0.2,$$
$$(R{\downarrow}A)(y_2) = \inf\{\max\{1-0.5, 0.2\}, \max\{1-1, 0.8\}\} = \inf\{0.5, 0.8\} = 0.5,$$

$$(R{\uparrow}A)(y_1) = \sup\{\min\{1, 0.2\}, \min\{0.5, 0.8\}\} = \sup\{0.2, 0.5\} = 0.5,$$
$$(R{\uparrow}A)(y_2) = \sup\{\min\{0.5, 0.2\}, \min\{1, 0.8\}\} = \sup\{0.2, 0.8\} = 0.8.$$

We see that the membership degree of the element y_1 in the lower approximation of A is 0.2 and in the upper approximation of A is 0.5. This means that y_1 necessarily satisfies the concept A with degree 0.2 and possibly satisfies the concept A with degree 0.5.

We now study an approach that has the model of Dubois and Prade as result.

3.1.3 Fuzzy rough sets by Yao

We present the fuzzy rough hybridisation approach as designed by Yao ([65]). It is a constructive approach. A similar approach is due to Liu et al. ([42]). Yao's appraoch is based on the α-level sets introduced in the previous chapter (see Definitions 2.2.7 and 2.2.8). A fuzzy set determines a family of nested subsets of the universe U through weak or strong α-level sets, but here we work only with the weak α-level sets. Wu et al. ([62, 63]) combined both weak and strong α-level sets, their approach will be discussed in the next section.

We first consider a family of α-level sets of a fuzzy set A, together with an equivalence relation R. Next, we consider a crisp set A, together with a family of equivalence relations $(R_\beta)_{\beta\in I}$. Finally, we use this result to give conclusions for a fuzzy set A and a fuzzy relation R.

A fuzzy set and an equivalence relation

We start with the approximation of a fuzzy set A in a Pawlak approximation space (U, R). We have a family of α-level sets $(A_\alpha)_{\alpha\in I}$. We can approximate every A_α: by Definition 2.1.2, we have a rough set $(R{\downarrow}A_\alpha, R{\uparrow}A_\alpha)$ for each $\alpha \in I$. This means that we have a family of lower approximations and one of upper approximations: $(R{\downarrow}A_\alpha)_{\alpha\in I}$ and $(R{\uparrow}A_\alpha)_{\alpha\in I}$. The question is now whether they correspond with two fuzzy sets. To find this out, we use the representation theorem of Negoita and Ralescu ([49]).

Proposition 3.1.3. Let $(A_\alpha)_{\alpha\in I}$ be a family of crisp subsets of U. The necessary and sufficient conditions for the existence of a fuzzy set B such that $B_\alpha = A_\alpha$ for all α in I, are:

(i) if $\alpha_1 \le \alpha_2 \in I$, then $A_{\alpha_2} \subseteq A_{\alpha_1}$,

(ii) let $\{\alpha_n \mid n \in \mathbb{N}\}$ be a non-decreasing sequence in I (i.e., $\alpha_i \le \alpha_j$ for $i \le j \in \mathbb{N}$) such that $\lim_{n\to+\infty} \alpha_n = \alpha$, then $\bigcap_{n=1}^{\infty} A_{\alpha_n} = A_\alpha$.

We need to prove that the family of lower approximations $(R{\downarrow}A_\alpha)_{\alpha\in I}$ and the family of upper approximations $(R{\uparrow}A_\alpha)_{\alpha\in I}$ fulfil conditions (i) and (ii). Since the family $(A_\alpha)_{\alpha\in I}$ is constructed from the fuzzy set A and because of the monotonicity of lower and upper approximation, condition (i) holds, i.e., if $\alpha_1 \leq \alpha_2$, then $A_{\alpha_2} \subseteq A_{\alpha_1}$ and thus

$$\begin{aligned} R{\downarrow}A_{\alpha_2} &\subseteq R{\downarrow}A_{\alpha_1}, \\ R{\uparrow}A_{\alpha_2} &\subseteq R{\uparrow}A_{\alpha_1}. \end{aligned}$$

Both families are also nested and they fulfil condition (ii), because the α-level sets of the fuzzy set A satisfy condition (ii) (see Proposition 2.2.9). So, by Proposition 3.1.3, there are fuzzy sets B_1 and B_2 such that for each α in I it holds that

$$\begin{aligned} (B_1)_\alpha &= R{\downarrow}A_\alpha, \\ (B_2)_\alpha &= R{\uparrow}A_\alpha. \end{aligned} \tag{3.1}$$

We know how these fuzzy sets are defined (see Equation (2.3)): for all $x \in U$ it holds that

$$\begin{aligned} B_1(x) &= \sup\{\alpha \mid x \in (B_1)_\alpha\} \\ &= \sup\{\alpha \mid x \in R{\downarrow}A_\alpha\} \\ &= \sup\{\alpha \mid [x]_R \subseteq A_\alpha\} \\ &= \sup\{\alpha \mid \forall y \in [x]_R : A(y) \geq \alpha\} \\ &= \inf\{A(y) \mid y \in [x]_R\} \\ &= \inf\{A(y) \mid (y,x) \in R\} \\ &= \inf\{\max\{1 - R(y,x), A(y)\} \mid y \in U\} \\ &= (R{\downarrow}A)(x), \\ B_2(x) &= \sup\{\alpha \mid x \in (B_2)_\alpha\} \\ &= \sup\{\alpha \mid x \in R{\uparrow}A_\alpha\} \\ &= \sup\{\alpha \mid [x]_R \cap A_\alpha \neq \emptyset\} \\ &= \sup\{\alpha \mid \exists y \in U : y \in [x]_R \wedge A(y) \geq \alpha\} \\ &= \sup\{A(y) \mid y \in [x]_R\} \\ &= \sup\{A(y) \mid (y,x) \in R\} \\ &= \sup\{\min\{R(y,x), A(y)\} \mid y \in U\} \\ &= (R{\uparrow}A)(x), \end{aligned}$$

where we use Definition 3.1.1 in the last steps.

This means that $(R{\downarrow}A)_\alpha = R{\downarrow}A_\alpha$ and $(R{\uparrow}A)_\alpha = R{\uparrow}A_\alpha$. We conclude that a rough fuzzy set is characterised by a fuzzy set A and a pair of fuzzy sets $(R{\downarrow}A, R{\uparrow}A)$ determined by a crisp relation R.

An α-level set of a rough fuzzy set is a rough set:

$$\begin{aligned}(R{\downarrow}A, R{\uparrow}A)_\alpha &= (R{\downarrow}A_\alpha, R{\uparrow}A_\alpha)\\ &= ((R{\downarrow}A)_\alpha, (R{\uparrow}A)_\alpha).\end{aligned}$$

Next, we consider a crisp set A and a fuzzy similarity relation R.

A crisp set and a fuzzy similarity relation

We now work in a fuzzy approximation space (U, R), with R a similarity relation. As R is a fuzzy set, R can be described with β-level sets: $R = (R_\beta)_{\beta \in I}$. Each R_β is a crisp equivalence relation on U, so we have a family of Pawlak approximation spaces $(U, R_\beta)_{\beta \in I}$.

Let A be a crisp subset of U. For each $\beta \in I$, we have a rough set

$$(R_\beta{\downarrow}A, R_\beta{\uparrow}A).$$

With respect to the fuzzy approximation space (U, R), we have a family of rough sets

$$(R_\beta{\downarrow}A, R_\beta{\uparrow}A)_{\beta \in I}.$$

We need an adapted theorem of Negoita and Ralescu ([55]).

Proposition 3.1.4. Let $\varphi : I \to I$ be a given function and $(A_\alpha)_{\alpha \in I}$ be a family of subsets of U. The necessary and sufficient conditions for the existence of a fuzzy set B such that $B_{\varphi(\alpha)} = A_\alpha$ for all α in I, are:

(i') if $\alpha_1, \alpha_2 \in I$ such that $\varphi(\alpha_1) \leq \varphi(\alpha_2)$, then $A_{\alpha_2} \subseteq A_{\alpha_1}$,

(ii') let $\{\varphi(\alpha_n) \mid n \in \mathbb{N}\}$ be a non-decreasing sequence in I (i.e., $\varphi(\alpha_i) \leq \varphi(\alpha_j)$ for $i \leq j \in \mathbb{N}$) such that $\lim\limits_{n \to +\infty} \varphi(\alpha_n) = \varphi(\alpha)$, then $\bigcap\limits_{n=1}^{\infty} A_{\alpha_n} = A_\alpha$.

If $\beta_2 \leq \beta_1$, then $R_{\beta_1} \subseteq R_{\beta_2}$, i.e., R_{β_1} is a refinement of R_{β_2}:

$$\forall x \in U\colon [x]_{R_{\beta_1}} \subseteq [x]_{R_{\beta_2}}.$$

We need to prove that the families $(R_\beta{\downarrow}A)_{\beta \in I}$ and $(R_\beta{\uparrow}A)_{\beta \in I}$ fulfil conditions (i') and (ii'). Let $\varphi_1(\beta) = 1 - \beta$ in Proposition 3.1.4. If $\varphi_1(\beta_1) \leq \varphi_1(\beta_2)$, then $\beta_2 \leq \beta_1$ and it holds that $R_{\beta_2}{\downarrow}A \subseteq R_{\beta_1}{\downarrow}A$. We need to prove that the family fulfils condition (ii'), i.e., we have to prove that if $\{\varphi_1(\beta_n) \mid n \in \mathbb{N}\}$ is a non-decreasing sequence in I and $\varphi_1(\beta)$ is its supremum, then

$$\bigcap_{n=1}^{\infty} R_{\beta_n}{\downarrow}A = R_\beta{\downarrow}A$$

holds. This follows from the fact that for all $n \in \mathbb{N}$, $\varphi_1(\beta_n) \leq \varphi_1(\beta)$ or $\beta \leq \beta_n$, which means that for all $n \in \mathbb{N}$ and all $x \in U$ it holds that

$$[x]_{R_{\beta_n}} \subseteq [x]_{R_\beta}.$$

We obtain that $R_\beta{\downarrow}A \subseteq R_{\beta_n}{\downarrow}A$, for all $n \in \mathbb{N}$ and

$$\begin{aligned} x \in \bigcap_{n=1}^{\infty} R_{\beta_n}{\downarrow}A &\Leftrightarrow \forall n \in \mathbb{N}\colon x \in R_{\beta_n}{\downarrow}A \\ &\Leftrightarrow \forall n \in \mathbb{N}\colon [x]_{R_{\beta_n}} \subseteq A \\ &\Leftrightarrow \bigcup_{n=1}^{\infty} [x]_{R_{\beta_n}} \subseteq A \end{aligned} \tag{3.2}$$

Now take $y \in [x]_{R_\beta}$, i.e., $R(x, y) \geq \beta$, this means, there is an $n \in \mathbb{N}$ such that $R(x, y) \geq \beta_n$, which means that $y \in [x]_{R_{\beta_n}}$ and thus $y \in A$. This proves that $x \in R_\beta{\downarrow}A$. Thus, the family of lower approximations $(R_\beta{\downarrow}A)_{\beta \in I}$ fulfils conditions (i') and (ii').

In a similar way, with $\varphi_2(\beta) = \beta$, we can derive that the family of upper approximations $(R_\beta{\uparrow}A)_{\beta \in I}$ fulfils conditions (i') and (ii').

So, there are fuzzy sets B_1 and B_2 such that for each $\beta \in I$ it holds that:

$$\begin{aligned} (B_1)_{\varphi_1(\beta)} &= R_\beta{\downarrow}A, \\ (B_2)_{\varphi_2(\beta)} &= R_\beta{\uparrow}A. \end{aligned} \tag{3.3}$$

We derive an explicit expression for both fuzzy sets. Let x be an element of U, then

$$\begin{aligned} B_1(x) &= \sup\{\varphi_1(\beta) \mid x \in (B_1)_{\varphi_1(\beta)}\} \\ &= \sup\{1 - \beta \mid x \in R_\beta{\downarrow}A\} \\ &= \sup\{1 - \beta \mid [x]_{R_\beta} \subseteq A\} \\ &= \sup\{1 - \beta \mid \forall y \in U : R(y, x) \geq \beta \Rightarrow y \in A\} \\ &= \sup\{1 - \beta \mid \forall y \in U : y \notin A \Rightarrow R(y, x) < \beta\} \\ &= \inf\{1 - R(y, x) \mid y \in U \wedge y \notin A\} \\ &= \inf\{\max\{1 - R(y, x), A(y)\} \mid y \in U\} \\ &= (R{\downarrow}A)(x), \\ B_2(x) &= \sup\{\beta \mid x \in (B_2)_{\varphi_2(\beta)}\} \\ &= \sup\{\beta \mid x \in R_\beta{\uparrow}A\} \\ &= \sup\{\beta \mid [x]_{R_\beta} \cap A \neq \emptyset\} \\ &= \sup\{\beta \mid \exists y \in U : R(y, x) \geq \beta \wedge y \in A\} \\ &= \sup\{R(y, x) \mid y \in A\} \\ &= \sup\{\min\{R(y, x), A(y)\} \mid y \in U\} \\ &= (R{\uparrow}A)(x), \end{aligned}$$

where we use Definition 3.1.1 in the last steps.

The pair of fuzzy sets $(R{\downarrow}A, R{\uparrow}A)$ is a fuzzy rough set with reference set the crisp set A determined by a fuzzy relation R. A β-level set of a fuzzy rough set is a rough set in the approximation space (U, R_β):

$$\begin{aligned}(R{\downarrow}A, R{\uparrow}A)_\beta &= (R_\beta{\downarrow}A, R_\beta{\uparrow}A)\\ &= ((R{\downarrow}A)_{1-\beta}, (R{\uparrow}A)_\beta).\end{aligned}$$

We now have the tools for the approach with a fuzzy set and a fuzzy similarity relation.

A fuzzy set and a fuzzy similarity relation

We continue working in the fuzzy approximation space (U, R) with R a fuzzy similarity relation, but now we consider a fuzzy set A instead of a crisp one. We have two families: one of α-level sets representing A and another one of β-level sets representing R (see also [42]).

For a fixed pair (α, β) in $I \times I$, consider the couple consisting of the crisp set A_α and the equivalence relation R_β: this results in a rough set $(R_\beta{\downarrow}A_\alpha, R_\beta{\uparrow}A_\alpha)$. For a fixed β in I, we consider the couple consisting of the fuzzy set $A = ((A_\alpha)_{\alpha \in I})$ and the equivalence relation R_β: this results in a rough fuzzy set $(R_\beta{\downarrow}A, R_\beta{\uparrow}A)$. Finally, with a fixed α in I, we obtain the couple consisting of the crisp set A_α and the fuzzy relation $(R_\beta)_{\beta \in I}$, which results in a fuzzy rough set $(R{\downarrow}A_\alpha, R{\uparrow}A_\alpha)$. In a generalised model, α and β are not fixed.

From Equations 3.1 and 3.3 we derive the following conclusion: for every set A, whether it is crisp or fuzzy, and for every fuzzy similarity relation R, we can describe the lower and upper approximation of A under R as

$$\begin{aligned}(R{\downarrow}A)(x) &= \inf_{y \in U}\{\max\{1 - R(y,x), A(y)\}\},\\ (R{\uparrow}A)(x) &= \sup_{y \in U}\{\min\{R(y,x), A(y)\}\},\end{aligned}$$

with x in U. This scheme is used by Dubois and Prade to define a fuzzy rough set. Note that we can do this whole approach for general fuzzy relations R and R-foresets.

The following approach we study, is the approach of Wu et al., which is based on the approach of Yao.

3.1.4 Fuzzy rough sets by Wu et al.

Another constructive approach to derive fuzzy rough sets is designed by Wu et al. ([62, 63]) and is based on the work of Yao ([65]). The fuzzy rough set they obtain is similar to the one of Dubois and Prade, but their approach is quite different. They work with a general fuzzy relation R from U to W, which we shall restrict in this dissertation to a binary fuzzy relation in U. They consider both weak and strong α-level sets to describe R and a fuzzy set A in (U, R), but the fuzzy rough set they derive is the same for each combination of weak and strong α-level sets, so we only give the approach based on weak α-level sets. The main difference with other approaches is that they work with R-aftersets instead of R-foresets.

We start by defining the lower and upper approximation of a crisp set under a crisp binary relation based on aftersets. Next, we use these approximation operators to define the lower and upper approximation of a fuzzy set in a fuzzy approximation space. We also give a useful characterisation. Finally, we study the approach of Wu et al. with foresets. This will give us Dubois and Prade's model.

We have two families of α-level sets: one that describes a fuzzy set A, i.e., $(A_\alpha)_{\alpha\in I}$, and one that describes a fuzzy relation R, i.e., $(R_\beta)_{\beta\in I}$. We also consider the β-level sets of the R-afterset of an element $x \in U$:

$$(xR)_\beta = \{y \in U \mid R(x, y) \geq \beta\}.$$

We know that for all $\beta \in I$, R_β is a crisp relation. We have a new lower and upper approximation of A_α in the generalised approximation space (U, R_β) for $(\alpha, \beta) \in I \times I$:

$$\begin{aligned}
x \in R_\beta{\downarrow}^* A_\alpha &\Leftrightarrow (xR)_\beta \subseteq A_\alpha \\
&\Leftrightarrow (\forall y \in U)(R(x, y) \geq \beta \Rightarrow A(y) \geq \alpha), \\
x \in R_\beta{\uparrow}^* A_\alpha &\Leftrightarrow (xR)_\beta \cap A_\alpha \neq \emptyset \\
&\Leftrightarrow (\exists y \in U)(R(x, y) \geq \beta \wedge A(y) \geq \alpha).
\end{aligned}$$

We now define the lower and upper approximation of A in (U, R) in this setting.

Definition 3.1.5. Let A be a fuzzy set in a fuzzy approximation space (U, R) and $x \in U$. We define the lower approximation $R{\downarrow}_* A$ of A by

$$(R{\downarrow}_* A)(x) = \sup_{\gamma \in I}\{\min\{\gamma, (R_{1-\gamma}{\downarrow}^* A_\gamma)(x)\}\}$$

and the upper approximation $R{\uparrow}_* A$ of A by

$$(R{\uparrow}_* A)(x) = \sup_{\gamma \in I}\{\min\{\gamma, (R_\gamma{\uparrow}^* A_\gamma)(x)\}\}.$$

We can simplify these expressions.

Proposition 3.1.6. Let A be a fuzzy set in a fuzzy approximation space (U, R). With $R{\downarrow}_* A$ and $R{\uparrow}_* A$ as defined above it holds for all x in U that

$$\begin{aligned}
(R{\downarrow}_* A)(x) &= \inf_{y \in U}\{\max\{1 - R(x, y), A(y)\}\}, \\
(R{\uparrow}_* A)(x) &= \sup_{y \in U}\{\min\{R(x, y), A(y)\}\}.
\end{aligned}$$

Proof. Let A be a fuzzy set of (U,R) and x an element of U. We first observe that $R_{1-\gamma}\downarrow^* A_\gamma$ and $R_\gamma\uparrow^* A_\gamma$ are crisp sets. We have

$$\begin{aligned}
(R\downarrow_* A)(x) &= \sup\{\min\{\gamma, (R_{1-\gamma}\downarrow^* A_\gamma)(x)\} \mid \gamma \in I\} \\
&= \sup\{\gamma \in I \mid (R_{1-\gamma}\downarrow^* A_\gamma)(x) = 1\} \\
&= \sup\{\gamma \in I \mid x \in R_{1-\gamma}\downarrow^* A_\gamma\} \\
&= \sup\{\gamma \in I \mid (xR)_{1-\gamma} \subseteq A_\gamma\} \\
&= \sup\{\gamma \in I \mid \forall y \in U : R(x,y) \geq 1-\gamma \Rightarrow A(y) \geq \gamma\} \\
&= \sup\{\gamma \in I \mid \forall y \in U :\ \max\{1-R(x,y), A(y)\} \geq \gamma\} \\
&= \sup\{\gamma \in I \mid \inf_{y\in U} \max\{1-R(x,y), A(y)\} \geq \gamma\} \\
&= \inf_{y\in U} \max\{1-R(x,y), A(y)\}.
\end{aligned}$$

In a similar way, we derive the other equation:

$$\begin{aligned}
(R\uparrow_* A)(x) &= \sup\{\min\{\gamma, (R_\gamma\uparrow^* A_\gamma)(x)\} \mid \gamma \in I\} \\
&= \sup\{\gamma \in I \mid (R_\gamma\uparrow^* A_\gamma)(x) = 1\} \\
&= \sup\{\gamma \in I \mid x \in R_\gamma\uparrow^* A_\gamma\} \\
&= \sup\{\gamma \in I \mid (xR)_\gamma \cap A_\gamma \neq \emptyset\} \\
&= \sup\{\gamma \in I \mid \exists y \in U : R(x,y) \geq \gamma \wedge A(y) \geq \gamma\} \\
&= \sup\{\gamma \in I \mid \exists y \in U :\ \min\{R(x,y), A(y)\} \geq \gamma\} \\
&= \sup\{\gamma \in I \mid \sup_{y\in U} \min\{R(x,y), A(y)\} \geq \gamma\} \\
&= \sup_{y\in U} \min\{R(x,y), A(y)\}.
\end{aligned}$$

□

We study what happens if we perform this approach with R-foresets, i.e., we change xR by Rx. We obtain that

$$\begin{aligned}
x \in R_\beta\downarrow^{**} A_\alpha &\Leftrightarrow (Rx)_\beta \subseteq A_\alpha, \\
x \in R_\beta\uparrow^{**} A_\alpha &\Leftrightarrow (Rx)_\beta \cap A_\alpha \neq \emptyset,
\end{aligned} \tag{3.4}$$

for all x in U. This is the same as the lower and upper approximation of the set A_α with respect to the binary relation R_β defined in Definition 2.1.4. We define $R\downarrow_{**}A$ and $R\uparrow_{**}A$ in the same way as in Definition 3.1.5, but now with the operators given in Equation (3.4). We can compute that with these operators, we obtain that

$$\begin{aligned}
(R\downarrow_{**} A)(x) &= \inf_{y\in U}\{\max\{1-R(y,x), A(y)\}\}, \\
(R\uparrow_{**} A)(x) &= \sup_{y\in U}\{\min\{R(y,x), A(y)\}\},
\end{aligned}$$

which is the same as the operators defined in Definition 3.1.1. We see that when R is not symmetric, the choice of working with R-foresets or R-aftersets is very important, because it can lead to different approximations. We illustrate this with an example.

Example 3.1.7. Let $U = \{y_1, y_2\}$, A a fuzzy set such that $A(y_1) = 0.4$ and $A(y_2) = 0.6$. We have the fuzzy relation R defined by

$$R(y_1, y_1) = R(y_2, y_2) = 0.5, R(y_1, y_2) = 0.8, R(y_2, y_1) = 0.2.$$

It is clear that R is not symmetric. Let us compute the upper approximation of A in y_1 for both approaches:

$$(R{\uparrow}_{*}A)(y_1) = \sup_{z\in U}\{\min\{R(y_1, z), A(z)\}\} = \sup\{0.4, 0.6\} = 0.6,$$
$$(R{\uparrow}_{**}A)(y_1) = \sup_{z\in U}\{\min\{R(z, y_1), A(z)\}\} = \sup\{0.4, 0.2\} = 0.4.$$

This shows that we obtain different approximations when we work with R-foresets or R-aftersets.

Next, we introduce a general implicator-conjunctor-based fuzzy rough set model.

3.2 General fuzzy rough set model

In this section, we study some types of generalisations of Dubois and Prade's fuzzy rough sets as seen in Definition 3.1.1. We start with introducing a general model, followed by special cases studied in the literature.

When we consider Definition 2.1.2, we see that the definition of the lower approximation contains an implication and the one of the upper approximation contains a conjunction. The extension of these logical operators in a fuzzy setting are implicators and conjunctors. We also consider a general fuzzy relation instead of a similarity relation. With these changes in mind, we introduce a general definition for the lower and upper approximation of a fuzzy set A.

Definition 3.2.1. Let A be a fuzzy set in a fuzzy approximation space (U,R), with R a general fuzzy relation. Let $\mathscr{I}$ be an implicator and $\mathscr{C}$ a conjunctor. The $(\mathscr{I},\mathscr{C})$*-fuzzy rough approximation* of A is the pair of fuzzy sets $(R{\downarrow}_{\mathscr{I}}A, R{\uparrow}_{\mathscr{C}}A)$ such that for $x \in U$:

$$(R{\downarrow}_{\mathscr{I}}A)(x) = \inf_{y\in U} \mathscr{I}(R(y,x), A(y)),$$
$$(R{\uparrow}_{\mathscr{C}}A)(x) = \sup_{y\in U} \mathscr{C}(R(y,x), A(y)).$$

We can now define a general $(\mathscr{I},\mathscr{C})$-fuzzy rough set.

Definition 3.2.2. Let (U,R) be a fuzzy approximation space and $\mathscr{I}$ and $\mathscr{C}$ an implicator and a conjunctor, respectively. A pair (A_1, A_2) of fuzzy sets in U is called a $(\mathscr{I},\mathscr{C})$*-fuzzy rough set* in (U,R) if there is a fuzzy set A in U such that $A_1 = R{\downarrow}_{\mathscr{I}}A$ and $A_2 = R{\uparrow}_{\mathscr{C}}A$ as given in Definition 3.2.1.

We can derive the definition given by Dubois and Prade, when we take for R a similarity relation, for $\mathscr{I}$ the Kleene-Dienes implicator $\mathscr{I}_{KD}$ and for $\mathscr{C}$ the minimum t-norm $\mathscr{T}_M$.

In Table 3.1 we give a chronological overview of special cases of the general model studied in the past.

Wu et al. were the first to consider general fuzzy relations. Mi and Zhang were the first to use conjunctors instead of t-norms. We see that the models of Mi and Zhang, Yeung et al. and Hu et al. are quite similar. In the models of Hu et al., kernels are used as fuzzy relations. The model of Mi and Zhang coincides with the second model of Yeung et al., as we restrict ourselves to fuzzy relations in $U \times U$. In the model of Mi and Zhang, the standard negator is considered, while in the models of Yeung et al., one assumes $\mathscr{N}$ to be involutive. The model of Pei and the model of Liu use the same conjunctor and implicator as Dubois and Prade, but now R is a general fuzzy relation instead of a fuzzy similarity relation.

Remark 3.2.3. We see that most authors assume the considered t-norm to be lower semicontinuous to let the residual principle hold for $(\mathscr{T}, \mathscr{I}_{\mathscr{T}})$. Due to Proposition 2.2.53, this is the same as using a left-continuous t-norm $\mathscr{T}$.

Model	Conjunctor	Implicator
1. Morsi and Yakout ([48], 1998)	lower semicontinuous t-norm	R-implicator based on that t-norm
2. Radzikowska and Kerre ([53], 2002)	t-norm	border implicator[a]
3. Wu et al. ([63], 2003)	standard minimum operator	S-implicator based on the dual t-conorm
4. Mi and Zhang ([44], 2004)	conjunctor based on the dual coimplicator	R-implicator based on a lower semicontinuous t-norm
5. Pei ([51], 2005) and Liu ([40], 2008)	standard minimum operator	Kleene-Dienes implicator
6. Wu et al. ([61], 2005)	continuous t-norm	implicator
7. Yeung et al., model 1 ([66], 2005)	lower semicontinuous t-norm	S-implicator based on the dual t-conorm
8. Yeung et al., model 2 ([66], 2005)	conjunctor based on the dual coimplicator	R-implicator based on a lower semicontinuous t-norm
9. De Cock et al. ([11], 2007)	t-norm	border implicator
10. Mi et al., ([43], 2008)	continuous t-norm	S-implicator based on the dual t-conorm
11. Hu et al., model 1 ([30], 2010 and [28], 2011)	lower semicontinuous t-norm	S-implicator based on the dual t-conorm
12. Hu et al., model 2 ([30], 2010 and [28], 2011)	conjunctor based on the dual coimplicator	R-implicator based on a lower semicontinuous t-norm

Model	Negator	Relations
1. Morsi and Yakout ([48], 1998)	not necessary	fuzzy $\mathscr{T}$-similarity relation
2. Radzikowska and Kerre ([53], 2002)	not necessary	fuzzy similarity relation
3. Wu et al. ([63], 2003)	standard negator	general fuzzy relation
4. Mi and Zhang ([44], 2004)	standard negator	general fuzzy relation[b]
5. Pei ([51], 2005) and Liu ([40], 2008)	not necessary	general fuzzy relation
6. Wu et al. ([61], 2005)	not necessary	general fuzzy relation
7. Yeung et al., model 1 ([66], 2005)	involutive	general fuzzy relation
8. Yeung et al., model 2 ([66], 2005)	involutive	general fuzzy relation
9. De Cock et al. ([11], 2007)	not necessary	general fuzzy relation
10. Mi et al., ([43], 2008)	standard negator	general fuzzy relation
11. Hu et al., model 1 ([30], 2010 and [28], 2011)	involutive	kernel function
12. Hu et al., model 2 ([30], 2010 and [28], 2011)	involutive	kernel function

Table 3.1: Overview of special cases of the general fuzzy rough set model

[a] The implicator is an S-, R- or QL-implicator; QL-implicators are not discussed in this thesis.

[b] Actually, in [43], [44] and [61], a general fuzzy relation from U to W is considered, with both U and W non-empty, finite universes. In this thesis, we restrict ourselves to R in $\mathscr{F}(U \times U)$.

We illustrate Definition 3.2.1 with two examples: first with a fuzzy similarity relation and then with a general fuzzy relation.

Example 3.2.4. Let us take the same U, A and R of Example 3.1.2:$U = \{y_1, y_2\}$, $A(y_1) = 0.2$, $A(y_2) = 0.8$ and R a fuzzy similarity relation with $R(y_1, y_2) = 0.5$. Take the Łukasiewicz implicator and t-norm instead of the Kleene-Dienes implicator and the minimum t-norm. We see that we get other results for the lower and upper approximations of A than in Example 3.1.2:

$$(R{\downarrow}_{\mathscr{I}_L}A)(y_1) = \inf\{0.2, 1\} = 0.2,$$
$$(R{\downarrow}_{\mathscr{I}_L}A)(y_2) = \inf\{0.7, 0.8\} = 0.7,$$

$$(R{\uparrow}_{\mathscr{T}_L}A)(y_1) = \sup\{0.2, 0.3\} = 0.3,$$
$$(R{\uparrow}_{\mathscr{T}_L}A)(y_2) = \sup\{0, 0.8\} = 0.8.$$

Example 3.2.5. Assume $U = \{y_1, y_2\}$ and A a fuzzy set in U such that $A(y_1) = 0.2$ and $A(y_2) = 0.8$, and take the general fuzzy relation R defined by

$$R(y_1, y_1) = R(y_2, y_2) = 0.7, R(y_1, y_2) = 0 \text{ and } R(y_2, y_1) = 0.3.$$

We see that R is not reflexive and not symmetric, but R is min-transitive. We take for $(\mathscr{I}, \mathscr{C})$ again the couple $(\mathscr{I}_L, \mathscr{T}_L)$. We obtain

$$(R{\downarrow}_{\mathscr{I}_L}A)(y_1) = \inf\{0.5, 1\} = 0.5,$$
$$(R{\downarrow}_{\mathscr{I}_L}A)(y_2) = \inf\{1, 1\} = 1,$$

$$(R{\uparrow}_{\mathscr{T}_L}A)(y_1) = \sup\{0, 0.1\} = 0.1,$$
$$(R{\uparrow}_{\mathscr{T}_L}A)(y_2) = \sup\{0, 0.5\} = 0.5.$$

Notice that in this case we have $R{\uparrow}_{\mathscr{T}_L}A \subseteq A \subseteq R{\downarrow}_{\mathscr{I}_L}A$, which is rather counterintuitive. We will take up this setting again when we illustrate the other models of this chapter.

We continue with tight and loose approximations.

3.3 Tight and loose approximations

Some authors define the lower and upper approximation of a set A with R-foresets, instead of with elements of U. For example, when A is a crisp subset of U, they define the rough set $(R{\downarrow}A, R{\uparrow}A)$ in a generalised approximation space (U, R) as

$$R{\downarrow}A = \bigcup\{Rx \mid Rx \subseteq A\},$$
$$R{\uparrow}A = \bigcup\{Rx \mid Rx \cap A \neq \emptyset\}.$$

The sets Rx are often called *(information) granules*. Articles that work with information granules include [28, 66, 68]. For crisp sets, the approximations based on R-foresets coincide with the ones from Chapter 2. In the following section, we ask ourselves what could happen if an element x is contained in more than one granule.

We first consider a crisp subset A in a generalised approximation space (U,R). We assess the inclusion of an R-foreset into A and the overlap of an R-foreset with A. In this way, we are going to 'refine' our generalised model. This idea was first explored by Pomykala ([52]) and further studied by De Cock et al. ([11]) and Cornelis et al. ([13]), who took fuzzy sets and fuzzy relations into account. We give a list of candidate definitions for the lower and upper approximation of a set A in (U,R):

1. The element $x \in U$ belongs to the lower approximation of A if and only if
   (a) all R-foresets containing x are included in A,
   (b) at least one R-foreset containing x is included in A,
   (c) the R-foreset of x is included in A.
2. The element $x \in U$ belongs to the upper approximation of A if and only if
   (a) all R-foresets containing x have a non-empty intersection with A,
   (b) at least one R-foreset containing x has a non-empty intersection with A,
   (c) the R-foreset of x has a non-empty intersection with A.

These candidates result in what is called the tight, loose and (usual) lower and upper approximation of the set A determined by the relation R ([11, 13]). We explain this terminology as follows: 'tight' refers to the fact that we take all R-foresets which contain x into account, while the 'loose' approximation only looks at the 'best' R-foreset. We now paraphrase these expressions. In a generalised approximation space (U,R), we obtain the following definitions.

Definition 3.3.1. Let A be a crisp subset in a generalised approximation space (U,R) and $x \in U$. We define the *tight lower approximation* $R{\downarrow}{\downarrow}A$ of A as

$$x \in R{\downarrow}{\downarrow}A \Leftrightarrow (\forall y \in U)(x \in Ry \Rightarrow Ry \subseteq A)$$

and the *loose lower approximation* $R{\uparrow}{\downarrow}A$ of A as

$$x \in R{\uparrow}{\downarrow}A \Leftrightarrow (\exists y \in U)(x \in Ry \wedge Ry \subseteq A).$$

We define the *tight upper approximation* $R{\downarrow}{\uparrow}A$ of A as

$$x \in R{\downarrow}{\uparrow}A \Leftrightarrow (\forall y \in U)(x \in Ry \Rightarrow Ry \cap A \neq \emptyset)$$

and the *loose upper approximation* $R{\uparrow}{\uparrow}A$ of A as

$$x \in R{\uparrow}{\uparrow}A \Leftrightarrow (\exists y \in U)(x \in Ry \wedge Ry \cap A \neq \emptyset).$$

The usual lower and upper approximation of A, which correspond to option (c) in the list of candidate definitions, are the same as those defined in Definition 2.1.4. All the approximations are crisp subsets of U. It is clear by the definitions that $R{\downarrow}{\downarrow}A \subseteq R{\uparrow}{\downarrow}A$ and that $R{\downarrow}{\uparrow}A \subseteq R{\uparrow}{\uparrow}A$. When R is an equivalence relation, the definitions of the tight and loose approximations coincide with the definition of the usual approximations of A. This is not the case when R is an arbitrary binary relation, we show this with an example.

Example 3.3.2. Let $U = \{y_1, y_2, y_3, y_4, y_5\}$, $A = \{y_1, y_3\}$ and

$$\begin{aligned} R = \{&(y_1,y_1),(y_1,y_3),(y_1,y_5),(y_2,y_1),(y_2,y_2),(y_2,y_4),(y_3,y_1),\\ &(y_3,y_3),(y_3,y_5),(y_4,y_4),(y_4,y_5),(y_5,y_2),(y_5,y_5)\}. \end{aligned}$$

We see that R is not symmetric (for example $(y_1, y_5) \in R$, but $(y_5, y_1) \notin R$), so R is not an equivalence relation. We compute the R-foresets:

$$\begin{aligned} Ry_1 &= \{y_1, y_2, y_3\},\\ Ry_2 &= \{y_2, y_5\},\\ Ry_3 &= \{y_1, y_3\},\\ Ry_4 &= \{y_2, y_4\},\\ Ry_5 &= \{y_1, y_3, y_4, y_5\}. \end{aligned}$$

We can now compute all the approximations of A:

$$\begin{aligned} R{\downarrow}{\downarrow}A &= \emptyset,\\ R{\downarrow}A &= \{y_3\},\\ R{\uparrow}{\downarrow}A &= \{y_1, y_3\},\\ R{\downarrow}{\uparrow}A &= \{y_1, y_3\},\\ R{\uparrow}A &= \{y_1, y_3, y_5\},\\ R{\uparrow}{\uparrow}A &= U. \end{aligned}$$

We see that the tight and loose approximations do not coincide with the usual ones. In this case we have that

$$R{\downarrow}{\downarrow}A \subseteq R{\downarrow}A \subseteq R{\uparrow}{\downarrow}A \subseteq A \subseteq R{\downarrow}{\uparrow}A \subseteq R{\uparrow}A \subseteq R{\uparrow}{\uparrow}A.$$

We now study what happens with a fuzzy set A in a fuzzy approximation space (U, R). When we replace the implications and conjunctions by implicators and conjunctors, we can form a natural generalisation[1].

[1] In [11] and [13] a t-norm was used as conjunctor.

Definition 3.3.3. Let $\mathscr{I}$ be an implicator and $\mathscr{C}$ a conjunctor. Let A be a fuzzy set in a fuzzy approximation space (U,R) and $x \in U$. The *tight lower approximation* of A is the fuzzy set $R{\downarrow_{\mathscr{I}}}{\downarrow_{\mathscr{I}}}A$ defined by

$$(R{\downarrow_{\mathscr{I}}}{\downarrow_{\mathscr{I}}}A)(x) = \inf_{y\in U} \mathscr{I}(R(x,y), \inf_{z\in U} \mathscr{I}(R(z,y),A(z)))$$

and the *loose lower approximation* of A is the fuzzy set $R{\uparrow_{\mathscr{C}}}{\downarrow_{\mathscr{I}}}A$ defined by

$$(R{\uparrow_{\mathscr{C}}}{\downarrow_{\mathscr{I}}}A)(x) = \sup_{y\in U} \mathscr{C}(R(x,y), \inf_{z\in U} \mathscr{I}(R(z,y),A(z))).$$

The *tight upper approximation* of A is the fuzzy set $R{\downarrow_{\mathscr{I}}}{\uparrow_{\mathscr{C}}}A$ defined by

$$(R{\downarrow_{\mathscr{I}}}{\uparrow_{\mathscr{C}}}A)(x) = \inf_{y\in U} \mathscr{I}(R(x,y), \sup_{z\in U} \mathscr{C}(R(z,y),A(z)))$$

and the *loose upper approximation* of A is the fuzzy set $R{\uparrow_{\mathscr{C}}}{\uparrow_{\mathscr{C}}}A$ defined by

$$(R{\uparrow_{\mathscr{C}}}{\uparrow_{\mathscr{C}}}A)(x) = \sup_{y\in U} \mathscr{C}(R(x,y), \sup_{z\in U} \mathscr{C}(R(z,y),A(z))).$$

The usual lower and upper approximation of A are the same as in Definition 3.2.1. The relations between the different approximations will be studied in Chapter 4.

We again illustrate the model by an example.

Example 3.3.4. Let U, A and R be as in Example 3.2.5: $U = \{y_1, y_2\}$, $A(y_1) = 0.2$, $A(y_2) = 0.8$, and R such that

$$R(y_1,y_1) = R(y_2,y_2) = 0.7, R(y_1,y_2) = 0 \text{ and } R(y_2,y_1) = 0.3.$$

Take $(\mathscr{I},\mathscr{C}) = (\mathscr{I}_L,\mathscr{T}_L)$. We first compute

$$\inf_{z\in U} \mathscr{I}_L(R(z,y),A(z)) = \inf_{z\in U} \min\{1, 1 - R(z,y) + A(z)\}$$

for $y \in U$. We obtain

$$\inf_{z\in U} \mathscr{I}_L(R(z,y_1),A(z)) = \inf\{0.5, 1\} = 0.5,$$
$$\inf_{z\in U} \mathscr{I}_L(R(z,y_2),A(z)) = \inf\{1, 1\} = 1.$$

Similarly, with

$$\sup_{z\in U} \mathscr{T}_L(R(z,y),A(z)) = \sup_{z\in U} \max\{0, R(z,y) + A(z) - 1\}$$

we obtain

$$\sup_{z\in U} \mathscr{T}_L(R(z,y_1),A(z)) = \sup\{0, 0.1\} = 0.1,$$
$$\sup_{z\in U} \mathscr{T}_L(R(z,y_2),A(z)) = \sup\{0, 0.5\} = 0.5.$$

We can now compute the four approximations of A in y_1 and y_2:

$$\begin{aligned}
(R{\downarrow}_{\mathscr{I}_L}{\downarrow}_{\mathscr{I}_L}A)(y_1) &= \inf\{0.8, 1\} = 0.8,\\
(R{\downarrow}_{\mathscr{I}_L}{\downarrow}_{\mathscr{I}_L}A)(y_2) &= \inf\{1, 1\} = 1,\\[1em]
(R{\uparrow}_{\mathscr{T}_L}{\downarrow}_{\mathscr{I}_L}A)(y_1) &= \sup\{0.2, 0\} = 0.2,\\
(R{\uparrow}_{\mathscr{T}_L}{\downarrow}_{\mathscr{I}_L}A)(y_2) &= \sup\{0, 0.7\} = 0.7,\\[1em]
(R{\downarrow}_{\mathscr{I}_L}{\uparrow}_{\mathscr{T}_L}A)(y_1) &= \inf\{0.4, 1\} = 0.4,\\
(R{\downarrow}_{\mathscr{I}_L}{\uparrow}_{\mathscr{T}_L}A)(y_2) &= \inf\{0.8, 0.8\} = 0.8,\\[1em]
(R{\uparrow}_{\mathscr{T}_L}{\uparrow}_{\mathscr{T}_L}A)(y_1) &= \sup\{0, 0\} = 0,\\
(R{\uparrow}_{\mathscr{T}_L}{\uparrow}_{\mathscr{T}_L}A)(y_2) &= \sup\{0, 0.2\} = 0.2.
\end{aligned}$$

Together with the result of Example 3.2.5 we obtain that

$$R{\uparrow}_{\mathscr{T}_L}{\uparrow}_{\mathscr{T}_L}A \subseteq R{\uparrow}_{\mathscr{T}_L}A \subseteq R{\uparrow}_{\mathscr{T}_L}{\downarrow}_{\mathscr{I}_L}A \subseteq A \subseteq R{\downarrow}_{\mathscr{I}_L}{\uparrow}_{\mathscr{T}_L}A \subseteq R{\downarrow}_{\mathscr{I}_L}A \subseteq R{\downarrow}_{\mathscr{I}}{\downarrow}_{\mathscr{I}L}A.$$

In this case, the loose approximations are not included in the tight approximations.

We continue with discussing some robust fuzzy rough set models.

3.4 Fuzzy rough set models designed to deal with noisy data

In applications, most classification tasks are described by fuzzy information, which can be noisy. Noise can be come from different sources, e.g., attribute noise and class noise ([70]). Attribute noise are errors introduced in attribute values, e.g., wrong values, missing values, incomplete values, ... This can happen when we acquire data. Class noise is generated by sample mislabelling. It can come from contradictory objects in the sample, i.e., the same object appears more than once and is labeled with different classifications, or misclassifications, i.e., an object is labeled wrong.

Noise is the reason why we want robust fuzzy rough set models, models such that the output does not change drastically if the input changes a little bit. The evolution of these models starts with the variable precision rough set model of Ziarko. An overview of some models is given in [29].

The first model we discuss is the β-precision fuzzy rough set model.

3.4.1 β-precision fuzzy rough sets

We start with the β-precision fuzzy rough set model. This was introduced by Fernández Salido and Murakami to work with numerical attributes, something that is not possible with Ziarko's VPRS model. This model is robust to class noise ([29]).

Fernández Salido and Murakami extended the model designed by Dubois and Prade by extending t-norms and t-conorms to β-precision quasi-t-norms and β-precision quasi-t-conorms ([56, 57]). Although Fernández Salido and Murakami worked with the extension of the maximum and minimum operators $\mathcal{S}_M$ and $\mathcal{T}_M$, Hu et al. ([29]) give a more general β-precision fuzzy rough set model (β-PFRS) that we discuss here:

Definition 3.4.1. Let $\mathcal{N}$ be an involutive negator and $\beta \in I$. Let $\mathcal{T}_\beta$ and $\mathcal{S}_\beta$ be a quasi-t-norm and a quasi-t-conorm based on a t-norm $\mathcal{T}$ and its dual t-conorm $\mathcal{S}$ with respect to $\mathcal{N}$. Let $\mathcal{I}$ be an implicator and $\mathcal{C}$ a conjunctor. We define the *β-precision fuzzy rough set model* as follows: for a fuzzy set A in a fuzzy approximation space (U,R) with R a general fuzzy relation and $x \in U$, we define the *lower approximation* $R{\downarrow}_{\mathcal{I},\mathcal{T}_\beta}A$ of A as

$$(R{\downarrow}_{\mathcal{I},\mathcal{T}_\beta}A)(x) = \underset{y \in U}{\mathcal{T}_\beta}\, \mathcal{I}(R(y,x),A(y)),$$

and the *upper approximation* $R{\uparrow}_{\mathcal{C},\mathcal{S}_\beta}A$ of A as

$$(R{\uparrow}_{\mathcal{C},\mathcal{S}_\beta}A)(x) = \underset{y \in U}{\mathcal{S}_\beta}\, \mathcal{C}(R(y,x),A(y)).$$

Hu et al. used for the pair $(\mathcal{I},\mathcal{C})$ an S-implicator $\mathcal{I}$ based on a t-conorm $\mathcal{S}$ and a t-norm $\mathcal{T}$ which is dual with $\mathcal{S}$ or an R-implicator $\mathcal{I}$ based on a t-norm $\mathcal{T}$ and its dual coimplicator $\mathcal{J}$.

We already know that when $\beta = 1$, we get the original t-norm and t-conorm. In the case studied by Fernández Salido and Murakami this is the infimum and supremum and in this way, we get the general fuzzy rough set model defined in Definition 3.2.1. According to [56, 57], the value of β depends on the application and will typically be high, e.g., 0.95 or 0.99. This means that when computing the lower approximation, we will omit the smallest values and when computing the upper approximation, we will omit the largest values. Outliers will have less impact on the result, which should make the model more robust. Fernández Salido and Murakami called β the precision of the approximations, in a sense that the higher β is, the more elements are taken into account in the computation.

Let us take a look at an example.

Example 3.4.2. We consider the same U, A and R as in Example 3.2.5: $U = \{y_1, y_2\}$, $A(y_1) = 0.2$ and $A(y_2) = 0.8$, and R such that

$$R(y_1,y_1) = R(y_2,y_2) = 0.7, R(y_1,y_2) = 0 \text{ and } R(y_2,y_1) = 0.3.$$

We take $(\mathcal{I},\mathcal{C}) = (\mathcal{I}_L,\mathcal{T}_L)$, $(\mathcal{T},\mathcal{S}) = (\min,\max)$ and $\beta = 0.8$. We obtain for the lower approximation

$$(R{\downarrow}_{\mathcal{I}_L,\min_{0.8}}A)(y_1) = \underset{0.8}{\min}\{0.5, 1\} = 0.5,$$

$$(R{\downarrow}_{\mathcal{I}_L,\min_{0.8}}A)(y_2) = \underset{0.8}{\min}\{1, 1\} = 1,$$

because $(1-0.8)(0.5+1)=0.3$ and $(1-0.8)(1+1)=0.4$ and thus we omit nothing. For the upper approximation, we derive

$$(R\uparrow_{\mathcal{T}_L,\max_{0.8}}A)(y_1)=\max_{0.8}\{0,0.1\}=0.1,$$
$$(R\uparrow_{\mathcal{T}_L,\max_{0.8}}A)(y_2)=\max_{0.8}\{0,0.5\}=0.5,$$

because $(1-0.8)(1-0+1-0.1)=0.38$ and $(1-0.8)(1-0+1-0.5)=0.3$.

This model is more robust than the general fuzzy rough set model. Let us illustrate this with an example.

Example 3.4.3. Let $U=\{y_1,\ldots,y_{100},x\}$, A a fuzzy set in U such that $A(y_i)=\frac{i}{100}$ for all $i\in\{1,\ldots,100\}$ and $A(x)=1$. Let R be a fuzzy relation with $R(y_i,x)=\frac{i}{100}$ for all $i\in\{1,\ldots,100\}$ and $R(x,x)=1$. We compute the lower approximation in x with the general fuzzy rough set model with $\mathcal{I}=\mathcal{I}_L$:

$$\begin{aligned}(R\downarrow_{\mathcal{I}_L}A)(x)&=\inf_{z\in U}\mathcal{I}_L(R(z,x),A(z))\\&=\min\left\{\inf_{i=1}^{100}\mathcal{I}_L\left(\frac{i}{100},\frac{i}{100}\right),\mathcal{I}_L(1,1)\right\}\\&=1.\end{aligned}$$

Now, if $A(y_{100})=0$, i.e., A is different in one point, then

$$\begin{aligned}(R\downarrow_{\mathcal{I}_L}A)(x)&=\min\left\{\inf_{i=1}^{100}\mathcal{I}_L\left(\frac{i}{100},A(y_i)\right),\mathcal{I}_L(1,1)\right\}\\&=\min\left\{\inf_{i=1}^{99}\min\left\{1,1-\frac{i}{100}+\frac{i}{100}\right\},\min\left\{1,1-\frac{100}{100}+0\right\},1\right\}\\&=\min\{1,0,1\}\\&=0.\end{aligned}$$

The difference is very large compared to the small change in A. We study what happens in the β-precision fuzzy rough set. Take $\mathcal{T}=\min$, $\mathcal{I}=\mathcal{I}_L$ and $\beta=0.95$. If $A(y_{100})=1$, we have again that $\mathcal{I}_L(R(z,x),A(z))=1$ for all $z\in U$ and it holds that

$$5\leq(99\cdot1+1\cdot1+1\cdot1)\cdot0.05=5.05,$$

which means we omit the five least values of $\mathcal{I}_L(R(z,x),A(z))$, which are all one. We obtain

$$(R\downarrow_{\mathcal{I}_L,\min_{0.95}}A)(x)=\min_{0.95}\{1,\ldots,1\}=1.$$

Now, if $A(y_{100})=0$, then

$$5\leq(99\cdot1+1\cdot0+1\cdot1)\cdot0.05=5,$$

and we again omit the five smallest values of $\mathcal{I}_L(R(z,x),A(z))$, which means we omit

$$\mathcal{I}_L(R(y_{100},x),A(y_{100}))=0.$$

We obtain again that

$$(R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}A)(x) = \min_{0.95}\{1,\dots,1\} = 1.$$

So, a small change in A does not change the lower approximation in x.

Next, we discuss the variable precision fuzzy rough set model.

3.4.2 Variable precision fuzzy rough sets

Mieszkowicz-Rolka and Rolka ([46, 47]) introduced another fuzzy rough set model to deal with class noise. Their motivation was that the fuzzy rough approximations of Dubois and Prade had the same disadvantages as the original rough set model: just a relatively small inclusion error of a fuzzy similarity class can result in the rejection of that class from the lower approximation, and a small inclusion degree can lead to an excessive increase of the upper approximation. To solve this, they combine the model designed by Dubois and Prade with the model designed by Ziarko to the variable precision fuzzy rough set model (VPFRS) with asymmetric bounds. We study their second model ([47]), since the upper approximation in their first model did not generalise the model of Dubois and Prade.

Before we study their model, we extend the notion of inclusion degree to fuzzy sets. The extension can be done in different ways. Mieszkowicz-Rolka and Rolka use the implication-based inclusion set for the lower approximation and the t-norm-based inclusion set for the upper approximation. We need two different definitions, in order to maintain the compatibility between the VPFRS model and the model designed by Dubois and Prade. We give both concepts.

Definition 3.4.4. Let A and B be fuzzy sets in U and $\mathscr{I}$ an implicator. The *implication-based inclusion set* $\text{Incl}(A,B)$ of A in B is defined by

$$\forall x \in U\colon\ \text{Incl}(A,B)(x) = \mathscr{I}(A(x),B(x)).$$

We need to choose a suitable implicator, because we want that the degree of inclusion with respect to x is 1 if $A(x) \leq B(x)$[2]. Not all implicators satisfy this condition, for example if we take $\mathscr{I} = \mathscr{I}_{KD}$, the condition does not hold. It does hold for R-implicators.

We continue with the t-norm-based inclusion set.

Definition 3.4.5. Let A and B be fuzzy sets in U and $\mathscr{T}$ a t-norm. The *t-norm-based inclusion set* $\text{Incl}'(A,B)$ of A in B is defined by

$$\forall x \in U\colon\ \text{Incl}'(A,B)(x) = \mathscr{T}(A(x),B(x)).$$

As in the model of Ziarko, we need measures for the amount of misclassification we allow, when determining the lower and upper approximation of a fuzzy set. In [47], two inclusion errors based on α-level sets were introduced. The first one is the lower α-inclusion error.

[2] In [47], $\text{Incl}(A,B)(x) = 0$ if $A(x) = 0$, but then the condition does not hold.

Definition 3.4.6. Let $\alpha \in I$ and A and B fuzzy sets in U. The *lower α-inclusion error $e_{l,\alpha}$* of A in B is defined by

$$e_{l,\alpha}(A,B) = 1 - \frac{|A \cap (\mathrm{Incl}(A,B))_\alpha|}{|A|}.$$

The second inclusion error is the upper α-inclusion error.

Definition 3.4.7. Let $\alpha \in I$ and A and B fuzzy sets in U and $\mathscr{N}_S$ the standard negator. The *upper α-inclusion error $e_{u,\alpha}$* of A in B is defined by

$$e_{u,\alpha}(A,B) = 1 - \frac{|A \cap (\mathrm{co}_{\mathscr{N}_S}(\mathrm{Incl}'(A,B)))_\alpha|}{|A|}.$$

With $0 \le l < u \le 1$, we can define the lower and upper approximation of a fuzzy set A. Although Mieszkowicz-Rolka and Rolka worked with fuzzy partitions, we give the definition for R-foresets based on a general fuzzy relation R.

Definition 3.4.8. Let A be a fuzzy set in a fuzzy approximation space (U,R) with R a general fuzzy relation and x an element of U. Let Incl and Incl$'$ be the inclusion sets based on a implicator $\mathscr{I}$ and a t-norm $\mathscr{T}$ respectively, such that $\mathscr{I}$ fulfils the condition

$$\forall B_1, B_2 \in \mathscr{F}(U), \forall x \in U: B_1(x) \le B_2(x) \Rightarrow \mathscr{I}(B_1,B_2)(x) = 1.$$

Let $\mathscr{N}$ be the standard negator. With $0 \le l < u \le 1$ we define the *u-lower approximation $R{\downarrow_{\mathscr{I},u}}A$* of A as

$$(R{\downarrow_{\mathscr{I},u}}A)(x) = \inf_{y \in S_{x,u}} (\mathrm{Incl}(Rx,A))(y)$$

and the *l-upper approximation $R{\uparrow_{\mathscr{T},l}}A$* of A as

$$(R{\uparrow_{\mathscr{T},l}}A)(x) = \sup_{y \in S_{x,l}} (\mathrm{Incl}'(Rx,A))(y)$$

with

$$\begin{aligned}
\alpha_{x,u} &= \sup\{\alpha \in I \mid e_{l,\alpha}(Rx,A) \le 1-u\} \\
&= \sup\left\{\alpha \in I \mid \frac{|Rx \cap (\mathrm{Incl}(Rx,A))_\alpha|}{|Rx|} \ge u\right\}, \\
S_{x,u} &= \mathrm{supp}(Rx) \cap \mathrm{supp}\left(\mathrm{Incl}(Rx,A))_{\alpha_{x,u}}\right) \\
&= \{y \in U \mid R(y,x) > 0 \text{ and } (\mathrm{Incl}(Rx,A))(y) \ge \alpha_{x,u}\}, \\
\alpha_{x,l} &= \sup\{\alpha \in I \mid e_{u,\alpha}(Rx,A) \le l\} \\
&= \sup\left\{\alpha \in I \mid \frac{|Rx \cap (\mathrm{co}_{\mathscr{N}}(\mathrm{Incl}'(Rx,A)))_\alpha|}{|Rx|} \ge 1-l\right\}, \\
S_{x,l} &= \mathrm{supp}(Rx) \cap \mathrm{supp}\left((\mathrm{co}_{\mathscr{N}}(\mathrm{Incl}'(Rx,A)))_{\alpha_{x,l}}\right) \\
&= \{y \in U \mid R(y,x) > 0 \text{ and } (\mathrm{Incl}'(Rx,A))(y) \le 1-\alpha_{x,l}\}.
\end{aligned}$$

With $u = 1$ and $l = 0$, we derive the fuzzy rough set model of Dubois and Prade. This was not the case in the first model from Mieszkowicz-Rolka and Rolka. Note that this holds, although the Kleene-Dienes implicator does not fulfil the condition for Incl.

Proposition 3.4.9. Let $u = 1$ and $l = 0$ and take $\mathscr{I} = \mathscr{I}_{KD}$ and $\mathscr{T} = \min$ to determine Incl and Incl′. With R a fuzzy similarity relation, we obtain the model designed by Dubois and Prade.

Proof. Let A be a fuzzy set in U and $x \in U$. First, we compute the value of $\alpha_{x,1}$:

$$\begin{aligned}\alpha_{x,1} &= \sup\left\{\alpha \in I \mid \frac{|Rx \cap \mathrm{Incl}(Rx,A)_\alpha|}{|Rx|} \geq 1\right\}\\ &= \sup\{\alpha \in I \mid |Rx \cap \mathrm{Incl}(Rx,A)_\alpha| = |Rx|\}\\ &= \sup\{\alpha \in I \mid \forall y \in U : R(y,x) > 0 \Rightarrow \max\{1 - R(y,x), A(y)\} \geq \alpha\}.\end{aligned}$$

Now, since $\max\{1 - R(y,x), A(y)\}$ is also 1 if $R(y,x) = 0$, we obtain that

$$\begin{aligned}\alpha_{x,1} &= \inf_{y \in I} \mathscr{I}_{KD}(R(y,x), A(y))\\ &= \inf_{y \in U} (\mathrm{Incl}(Rx,A))(y).\end{aligned}$$

We continue with $S_{x,1}$:

$$\begin{aligned}S_{x,1} &= \mathrm{supp}(Rx) \cap \mathrm{supp}\left((\mathrm{Incl}(Rx,A))_{\alpha_{x,1}}\right)\\ &= \left\{y \in U \mid R(y,x) > 0 \text{ and } (\mathrm{Incl}(Rx,A))(y) \geq \inf_{z \in U}(\mathrm{Incl}(Rx,A))(z)\right\}\\ &= \mathrm{supp}(Rx).\end{aligned}$$

We can now determine the lower approximation:

$$\begin{aligned}(R{\downarrow}_{\mathscr{I}_{KD},1}A)(x) &= \inf_{y \in S_{x,1}} (\mathrm{Incl}(Rx,A))(y)\\ &= \inf_{y \in \mathrm{supp}(Rx)} \max\{1 - R(y,x), A(y)\}\\ &= \inf_{y \in U} \max\{1 - R(y,x), A(y)\}\\ &= (R{\downarrow}A)(x)\end{aligned}$$

because, if $R(y,x) = 0$, then $\max\{1 - R(y,x), A(y)\} = 1$ and we take the infimum, so these values have no influence. For the upper approximation, we can do something similar. We first start with $\alpha_{x,0}$. Recall that we take the standard negator for $\mathscr{N}$.

$$\begin{aligned}\alpha_{x,0} &= \sup\left\{\alpha \in I \mid \frac{|Rx \cap (\mathrm{co}_{\mathscr{N}}(\mathrm{Incl}'(Rx,A)))_\alpha|}{|Rx|} \geq 1\right\}\\ &= \sup\{\alpha \in I \mid |Rx \cap (\mathrm{co}_{\mathscr{N}}(\mathrm{Incl}'(Rx,A)))_\alpha| = |Rx|\}\\ &= \sup\{\alpha \in I \mid \forall y \in U : R(y,x) > 0 \Rightarrow 1 - \min\{R(y,x), A(y)\} \geq \alpha\}.\end{aligned}$$

Since $1 - \min\{R(y,x), A(y)\}$ is also 1 if $R(y,x) = 0$, we obtain

$$\begin{aligned}\alpha_{x,0} &= \inf_{y\in U}\{1 - \min\{R(y,x),A(y)\}\}\\ &= 1 - \sup_{y\in U}(\mathrm{Incl}'(Rx,A))(y).\end{aligned}$$

With this $\alpha_{x,0}$, we get for $S_{x,0}$ the following:

$$\begin{aligned}S_{x,0} &= \mathrm{supp}(Rx) \cap \mathrm{supp}\left((\mathrm{co}_{\mathcal{N}}(\mathrm{Incl}'(Rx,A)))_{\alpha_{x,0}}\right)\\ &= \left\{y \in U \mid R(y,x) > 0 \text{ and } 1 - (\mathrm{Incl}'(Rx,A))(y) \geq 1 - \sup_{z\in U}(\mathrm{Incl}'(Rx,A))(z)\right\}\\ &= \mathrm{supp}(Rx).\end{aligned}$$

For the upper approximation we derive that

$$\begin{aligned}(R\uparrow_{\mathcal{T}_M,0}A)(x) &= \sup_{y\in S_{x,0}}(\mathrm{Incl}'(Rx,A))(y)\\ &= \sup_{y\in \mathrm{supp}(Rx)} \min\{R(y,x),A(y)\}\\ &= \sup_{y\in U}\min\{R(y,x),A(y)\}\\ &= (R\uparrow A)(x)\end{aligned}$$

because, if $R(y,x) = 0$, then $\min\{R(y,x),A(y)\} = 0$ and this does not influence the supremum. □

Just as the model of Dubois and Prade, Ziarko's VPRS model is a special case of the VPFRS model, i.e., when A and R are crisp, the VPFRS model reduces to Ziarko's model with asymmetric bounds (see Definition 2.1.12).

Proposition 3.4.10. If A is a crisp set in a generalised approximation space (U,R), then the variable precision fuzzy rough set model is exactly the variable precision rough set model with asymmetric bound.

Proof. Take $0 \leq l < u \leq 1$ and $\mathcal{I}$ an implicator and $\mathcal{T}$ a t-norm. Let $\mathcal{N}$ be the standard negator. Because we work with crisp sets, $\alpha_{x,u}$ and $\alpha_{x,l}$ are either 1 or 0.

We start by determining when $\alpha_{x,u}$ is 1. Now, for every $x, y \in U$ we have by the definition of an implicator that

$$y \in (Rx \cap \mathrm{Incl}(Rx,A)) \Leftrightarrow R(y,x) = 1 \text{ and } A(y) = 1 \Leftrightarrow y \in (Rx \cap A).$$

This leads us to:

$$\begin{aligned}\alpha_{x,u} = 1 &\Leftrightarrow \frac{|Rx \cap (\mathrm{Incl}(Rx,A))_1|}{|Rx|} \geq u\\ &\Leftrightarrow \frac{|Rx \cap A|}{|Rx|} \geq u\\ &\Leftrightarrow (R\downarrow_u A)(x) = 1.\end{aligned}$$

If $\alpha_{x,u} = 1$, then $S_{x,u} = Rx \cap A$ and

$$\begin{aligned}(R{\downarrow}_{\mathscr{I},u}A)(x) &= \inf_{y \in S_{x,u}} (\mathrm{Incl}(Rx,A))(y) \\ &= \inf_{y \in S_{x,u}} \mathscr{I}(R(y,x),A(y)) \\ &= \inf_{y \in S_{x,u}} 1 \\ &= 1 \\ &= (R{\downarrow}_u A)(x).\end{aligned}$$

If $\alpha_{x,u} = 0$, then $\frac{|Rx \cap A|}{|Rx|} < u \leq 1$, which means that there is a $y \in U$ such that $y \in Rx$ and $y \notin A$. For this y, we have $\mathscr{I}(R(y,x),A(x)) = 0$. If $\alpha_{x,0} = 0$, then $S_{x,u} = Rx$ and we obtain

$$\begin{aligned}(R{\downarrow}_{\mathscr{I},u}A)(x) &= \inf_{y \in S_{x,u}} (\mathrm{Incl}(Rx,A))(y) \\ &= \inf_{y \in S_{x,u}} \mathscr{I}(R(y,x),A(y)) \\ &= 0 \\ &= (R{\downarrow}_u A)(x).\end{aligned}$$

For both values of $\alpha_{x,u}$ we have $(R{\downarrow}_{\mathscr{I},u}A)(x) = (R{\downarrow}_u A)(x)$.

We do the same thing for the upper approximation. For the t-norm-based inclusion set, we derive for $x, y \in U$ that

$$(\mathrm{co}_{\mathscr{N}}(\mathrm{Incl}'(Rx,A)))(y) \geq 1 \Leftrightarrow \mathscr{T}(R(y,x),A(y)) = 0 \Leftrightarrow y \in ((Rx)^{\mathrm{c}} \cup A^{\mathrm{c}}),$$

and thus

$$\begin{aligned}\alpha_{x,l} = 1 &\Leftrightarrow \frac{|Rx \cap (\mathrm{co}_{\mathscr{N}}(\mathrm{Incl}'(Rx,A)))_1|}{|Rx|} \geq 1 - l \\ &\Leftrightarrow \frac{|Rx \cap (Rx^{\mathrm{c}} \cup A^{\mathrm{c}})|}{|Rx|} \geq 1 - l \\ &\Leftrightarrow \frac{|Rx \cap A^{\mathrm{c}}|}{|Rx|} \geq 1 - l \\ &\Leftrightarrow 1 - \frac{|Rx \cap A|}{|Rx|} \geq 1 - l \\ &\Leftrightarrow \frac{|Rx \cap A|}{|Rx|} \leq l \\ &\Leftrightarrow (R{\uparrow}_l A)(x) = 0.\end{aligned}$$

If $\alpha_{x,l} = 1$, then $S_{x,l} = Rx \cap A^c$ and thus

$$\begin{aligned}(R\uparrow_{\mathcal{T},l}A)(x) &= \sup_{y\in S_{x,l}} (\mathrm{Incl}'(Rx,A))(y)\\ &= \sup_{y\in S_{x,l}} \mathcal{T}(R(y,x),A(y))\\ &= \sup_{y\in S_{x,l}} 0\\ &= 0\\ &= (R\uparrow_l A)(x).\end{aligned}$$

On the other hand, if $\alpha_{x,l} = 0$, then $S_{x,l} = Rx \cap U = Rx$ and $(R\uparrow_l A)(x) = 1$, which means that there is an $y \in U$ such that $y \in (Rx \cap A)$ and thus $\mathcal{T}(R(y,x),A(y)) = 1$. We obtain that

$$\begin{aligned}(R\uparrow_{\mathcal{T},l}A)(x) &= \sup_{y\in S_{x,l}} (\mathrm{Incl}'(Rx,A))(y)\\ &= \sup_{y\in S_{x,l}} \mathcal{T}(R(y,x),A(y))\\ &= 1\\ &= (R\uparrow_l A)(x).\end{aligned}$$

In both cases we have that $(R\uparrow_{\mathcal{T},l}A)(x) = (R\uparrow_l A)(x)$. □

Since Ziarko's model is a special case of the VPFRS model, the properties of this model are very limited. Furher study of this model is required.

We illustrate the model and its robustness.

Example 3.4.11. We consider the same U, A and R as in Example 3.2.5: $U = \{y_1, y_2\}$, $A(y_1) = 0.2$, $A(y_2) = 0.8$ and R such that

$$R(y_1, y_1) = R(y_2, y_2) = 0.7, R(y_1, y_2) = 0 \text{ and } R(y_2, y_1) = 0.3.$$

We take $(\mathcal{I}, \mathcal{T}) = (\mathcal{I}_L, \mathcal{T}_L)$ and $l = 0.1$, $u = 0.6$. We derive the following results:

$$\begin{aligned}&\alpha_{y_1,0.6} = 0.5,\\ &S_{y_1,0.6} = U,\\ &(R\downarrow_{\mathcal{I}_L,0.6}A)(y_1) = \inf\{0.5, 1\} = 0.5,\end{aligned}$$

$$\begin{aligned}&\alpha_{y_2,0.6} = 1,\\ &S_{y_2,0.6} = \{y_2\},\\ &(R\downarrow_{\mathcal{I}_L,0.6}A)(y_2) = \inf\{1\} = 1,\end{aligned}$$

$$\alpha_{y_1,0.4} = 0.9,$$
$$S_{y_1,0.4} = U,$$
$$(R{\uparrow}_{\mathcal{T}_L,0.4}A)(y_1) = \sup\{0, 0.1\} = 0.1,$$

$$\alpha_{y_2,0.4} = 0.5,$$
$$S_{y_2,0.4} = U,$$
$$(R{\uparrow}_{\mathcal{T}_L,0.4}A)(y_2) = \sup\{0, 0.5\} = 0.5.$$

In this case, we have the same results in Example 3.2.5.

To illustrate robustness, we take the same example as in the previous section.

Example 3.4.12. Like in Example 3.4.3, we take $U = \{y_1, \ldots, y_{100}, x\}$, A a fuzzy set in U such that $A(y_i) = \frac{i}{100}$ for all $i \in \{1, \ldots, 100\}$ and $A(x) = 1$. Let R be a fuzzy relation with $R(y_i, x) = \frac{i}{100}$ for all $i \in \{1, \ldots, 100\}$ and $R(x, x) = 1$. Recall that in the general fuzzy rough set model with $\mathcal{I} = \mathcal{I}_L$ we had $(R{\downarrow}_{\mathcal{I}_L}A)(x) = 1$, and we had $(R{\downarrow}_{\mathcal{I}_L}A)(x) = 0$ if $A(y_{100}) = 0$.

We study what happens in the VPFRS model with $\mathcal{I} = \mathcal{I}_L$ and $u = 0.8$. Since $(\mathrm{Incl}(Rx, A))(z) = 1$ for every $z \in U$, we have that $\alpha_{x,0.8} = 1$ and $S_{x,0.8} = U$. Hence, $(R{\downarrow}_{\mathcal{I}_L,0.8}A)(x) = 1$, as in the general fuzzy rough set model. Now, when $A(y_{100}) = 0$, we still have $\alpha_{x,0.8} = 1$, but now $S_{x,0.8} = U \setminus \{y_{100}\}$. Since $\mathcal{I}_L(R(y_{100}, x), A(y_{100})) = 0$ is omitted, we again have $(R{\downarrow}_{\mathcal{I}_L,0.8}A)(x) = 1$, and thus, this model is more robust than the general fuzzy rough set model.

We continue with the vaguely quantified fuzzy rough set model.

3.4.3 Vaguely quantified fuzzy rough sets

In 2007, Cornelis et al. ([12]) introduced vague quantifiers into the existing models. For example, 'most' and 'some' are vague quantifiers. Quantifiers soften the definitions of the lower and upper approximations in the VPRS and the β-PFRS model. The intuition is that an element x belongs to the lower approximation of A if most of the elements related to x are included in A and it belongs to the upper approximation of A if some of the elements related to x are included in A.

We first define the notion of a quantifier.

Definition 3.4.13. A *quantifier* is a mapping $Q\colon I \to I$. We call a quantifier Q *regularly increasing* if it increases and if it satisfies the boundary conditions $Q(0) = 0$ and $Q(1) = 1$.

We give some examples of regularly increasing quantifiers.

Example 3.4.14. Let a be in I and $0 \leq l < u \leq 1$.

1. The existential quantifier:
$$Q_{\exists}(a) = \begin{cases} 0 & a = 0 \\ 1 & a > 0 \end{cases}$$

2. The universal quantifier:
$$Q_{\forall}(a) = \begin{cases} 0 & a < 1 \\ 1 & a = 1 \end{cases}$$

3. Quantifier with boundary l:
$$Q_{>l}(a) = \begin{cases} 0 & a \leq l \\ 1 & a > l \end{cases}$$

4. Quantifier with boundary u:
$$Q_{\geq u}(a) = \begin{cases} 0 & a < u \\ 1 & a \geq u \end{cases}$$

The examples above are all crisp quantifiers, but there also exist fuzzy quantifiers.

Example 3.4.15. Let a be in I and $0 \leq \alpha < \beta \leq 1$, we define the quantifier $Q_{(\alpha,\beta)}$ as

$$Q_{(\alpha,\beta)}(a) = \begin{cases} 0 & a \leq \alpha \\ \frac{2(a-\alpha)^2}{(\beta-\alpha)^2} & \alpha \leq a \leq \frac{\alpha+\beta}{2} \\ 1 - \frac{2(a-\beta)^2}{(\beta-\alpha)^2} & \frac{\alpha+\beta}{2} \leq a \leq \beta \\ 1 & \beta \leq a. \end{cases}$$

We can use $Q_s = Q_{(0.1,0.6)}$ and $Q_m = Q_{(0.2,1)}$ to reflect the vague quantifiers 'some' and 'most' ([12]).

Given fuzzy sets A_1 and A_2 in U and a fuzzy quantifier Q, we can compute the truth value of the statement " 'Q' A_1's are also in A_2" by the formula

$$Q\left(\frac{|A_1 \cap A_2|}{|A_1|}\right).$$

Recall that in the fuzzy case $(A_1 \cap A_2)(x) = \min\{A_1(x), A_2(x)\}$ and $|A| = \sum_{x \in U} A(x)$.

Once we have fixed a couple (Q_u, Q_l) of fuzzy quantifiers, we can formally define the vaguely quantified fuzzy rough set model (VQFRS).

Definition 3.4.16. Let A be a fuzzy set in a fuzzy approximation space (U,R) and $x \in U$. For the couple (Q_u, Q_l) of fuzzy quantifiers we can define the Q_u-*lower approximation* $R{\downarrow_{Q_u}}A$ of A as

$$(R{\downarrow_{Q_u}}A)(x) = \begin{cases} Q_u\left(\frac{|Rx \cap A|}{|Rx|}\right) & Rx \neq \emptyset \\ Q_u(1) & Rx = \emptyset \end{cases}$$

and the Q_l-*upper approximation* $R{\uparrow_{Q_l}}A$ of A as

$$(R{\uparrow_{Q_l}}A)(x) = \begin{cases} Q_l\left(\frac{|Rx \cap A|}{|Rx|}\right) & Rx \neq \emptyset \\ Q_l(1) & Rx = \emptyset. \end{cases}$$

It is easy to verify that with $(Q_\forall, Q_\exists)$ we derive Definition 2.1.4 and with $(Q_{\geq u}, Q_{>l})$ we derive Definition 2.1.12. When A and R are crisp, we call this model the vaguely quantified rough set model (VQRS). We see that in the VQFRS model, we do not use conjunctors and implicators.

Remark 3.4.17. There are other possible cardinalities besides $|A|$ which can be used to define fuzzy rough sets such as done in Fan et al. ([22]).

We give an example of the VQFRS model.

Example 3.4.18. We take U, A and R as in Example 3.2.5: $U = \{y_1, y_2\}$, $A(y_1) = 0.2$, $A(y_2) = 0.8$ and R such that

$$R(y_1, y_1) = R(y_2, y_2) = 0.7, R(y_1, y_2) = 0 \text{ and } R(y_2, y_1) = 0.3.$$

We take $(Q_u, Q_l) = (Q_m, Q_s)$. We compute that $\frac{|Ry_1 \cap A|}{|Ry_1|} = \frac{1}{2}$ and that $\frac{|Ry_2 \cap A|}{|Ry_2|} = 1$. With these values, we can compute the lower and upper approximation of A:

$$(R{\downarrow_{Q_m}}A)(y_1) = Q_m\left(\frac{1}{2}\right) = 0.28125,$$
$$(R{\downarrow_{Q_m}}A)(y_2) = Q_m(1) = 1,$$

$$(R{\uparrow_{Q_s}}A)(y_1) = Q_s\left(\frac{1}{2}\right) = 0.92,$$
$$(R{\uparrow_{Q_s}}A)(y_2) = Q_s(1) = 1.$$

These results are different from the results in Example 3.2.5.

We illustrate the robustness of the VQFRS model.

Example 3.4.19. We consider the same U, A and R as in Example 3.4.3: $U = \{y_1, \ldots, y_{100}, x\}$, A a fuzzy set in U such that $A(y_i) = \frac{i}{100}$ for all $i \in \{1, \ldots, 100\}$ and $A(x) = 1$ and R a fuzzy relation with $R(y_i, x) = \frac{i}{100}$ for all $i \in \{1, \ldots, 100\}$ and $R(x, x) = 1$. We have seen that with $A(y_{100}) = 1$ we have $(R{\downarrow_{\mathscr{I}_L}}A)(x) = 1$ and with $A(y_{100}) = 0$ we have $(R{\downarrow_{\mathscr{I}_L}}A)(x) = 0$. Let us compute the lower appromation in the VQFRS model with $Q_u = Q_m = Q_{(0.2,1)}$. We have for all $z \in U$ that $Rx(z) = R(z, x) = 1$, which means that $|Rx| = 101$. With $A(y_{100}) = 1$, we derive that $\frac{|Rx \cap A|}{|Rx|} = \frac{99+1+1}{101} = 1$ and thus that $(R{\downarrow_{Q_m}}A)(x) = 1$, as $Q_m(1) = 1$. Now, let $A(y_{100}) = 0$, then $\frac{|Rx \cap A|}{|Rx|}(x) = \frac{99+0+1}{101} = \frac{100}{101}$ and because $Q_m\left(\frac{100}{101}\right) = 0.9997$, we have that $(R{\downarrow_{Q_m}}A)(x) = 0.9997$, which is only a small change from 1.

We continue with the fuzzy variable precision rough set model.

3.4.4 Fuzzy variable precision rough sets

In this model designed by Zhao et al. ([68]), we again work with fuzzy logical operators and a general fuzzy relation R. It will be effective if we just consider attribute noise ([29]).

In the fuzzy variable precision rough set model (FVPRS), we define a fuzzy lower and upper approximation with variable precision α, with $\alpha \in [0, 1[$. For computing the lower approximation, we only take into account the values $A(y)$ which are greater than α, for the upper approximation we only consider the values $A(y)$ which are smaller than $\mathscr{N}(\alpha)$ for a certain negator $\mathscr{N}$. This means that we omit values which are too small, respectively too big.

Definition 3.4.20. Let $\mathscr{N}$ be a negator, $\mathscr{I}$ an implicator and $\mathscr{C}$ a conjunctor. Let A be a fuzzy set in a fuzzy approximation space (U, R) with R a general fuzzy relation and $x \in U$. Let $\alpha \in [0, 1[$. The *lower approximation with variable precision* α of A, $R{\downarrow_{\mathscr{I},\alpha}}A$, is defined by

$$(R{\downarrow_{\mathscr{I},\alpha}}A)(x) = \inf_{y \in U} \mathscr{I}(R(y, x), \max\{\alpha, A(y)\}),$$

and the *upper approximation with variable precision* α of A, $R{\uparrow_{\mathscr{C},\alpha}}A$, is defined by

$$(R{\uparrow_{\mathscr{C},\alpha}}A)(x) = \sup_{y \in U} \mathscr{C}(R(y, x), \min\{\mathscr{N}(\alpha), A(y)\}).$$

With $\alpha = 0$, we obtain the general fuzzy rough set model of Definition 3.2.1. In most cases, α will be small. When we have a big α (i.e., close to 1), the values of the lower approximation of A in the different elements of U will be close to 1, and the values of the upper approximation of A will be close to 0. Note also that we always have the following connection between this model and the general fuzzy rough set model:

$$R{\downarrow_{\mathscr{I},\alpha}}A = R{\downarrow_{\mathscr{I}}}(A \cup \hat{\alpha}),$$
$$R{\uparrow_{\mathscr{C},\alpha}}A = R{\uparrow_{\mathscr{C}}}(A \cap \widehat{1-\alpha}),$$

for every fuzzy set A and every $\alpha \in I$.

If A is a crisp set and $x \in U$, Definition 3.4.20 becomes the following:

$$(R{\downarrow}_{\mathscr{I},\alpha}A)(x) = \inf_{A(y)=0} \mathscr{I}((R(y,x)),\alpha),$$
$$(R{\uparrow}_{\mathscr{C},\alpha}A)(x) = \sup_{A(y)=1} \mathscr{C}(R(y,x),\mathscr{N}(\alpha)).$$

We again apply this model to our standard setting.

Example 3.4.21. Let U, A and R be as in Example 3.2.5: $U = \{y_1, y_2\}$, $A(y_1) = 0.2$, $A(y_2) = 0.8$ and R such that

$$R(y_1, y_1) = R(y_2, y_2) = 0.7, R(y_1, y_2) = 0 \text{ and } R(y_2, y_1) = 0.3.$$

Let $(\mathscr{I},\mathscr{C}) = (\mathscr{I}_L, \mathscr{T}_L)$ and $\alpha = 0.3$. We obtain with this model:

$$(R{\downarrow}_{\mathscr{I}_L,0.3}A)(y_1) = \inf\{0.6, 1\} = 0.6,$$
$$(R{\downarrow}_{\mathscr{I}_L,0.3}A)(y_2) = \inf\{1, 1\} = 1,$$

$$(R{\uparrow}_{\mathscr{T}_L,0.3}A)(y_1) = \sup\{0, 0\} = 0,$$
$$(R{\uparrow}_{\mathscr{T}_L,0.3}A)(y_2) = \sup\{0, 0.4\} = 0.4.$$

The values for the lower approximation are slightly larger than the values obtained by the general fuzzy rough set model. On the other hand, the values for the upper approximation are slightly smaller in this case.

The following example shows that the fuzzy variable precision rough set model is a robust model.

Example 3.4.22. We again consider the setting of Example 3.4.3. Let $U = \{y_1, \ldots, y_{100}, x\}$, A a fuzzy set in U such that $A(y_i) = \frac{i}{100}$ for all $i \in \{1, \ldots, 100\}$ and $A(x) = 1$. Let R be a fuzzy relation with $R(y_i, x) = \frac{i}{100}$ for all $i \in \{1, \ldots, 100\}$ and $R(x, x) = 1$. Let us take $\mathscr{I} = \mathscr{I}_L$ and $\alpha = 0.2$, then we have that

$$\mathscr{I}_L(R(y_i, x), \max\{0.2, A(y_i)\}) = \min\left\{1, 1 - \frac{i}{100} + 0.2\right\} = 1$$

if $i \leq 20$, and for $20 < i$, we have

$$\mathscr{I}_L(R(y_i, x), \max\{0.2, A(y_i)\}) = \min\left\{1, 1 - \frac{i}{100} + \frac{i}{100}\right\} = 1.$$

We obtain that the lower approximation of A in x is

$$\begin{aligned}(R{\downarrow}_{\mathscr{I}_L,0.2}A)(x) &= \inf_{z \in U} \mathscr{I}_L(R(z, x), \max\{0.2, A(z)\}) \\ &= 1.\end{aligned}$$

If $A(y_{100}) = 0$, then we obtain that

$$(R{\downarrow}_{\mathscr{I}_L,0.2}A)(x) = \mathscr{I}_L(1, 0.2) = 0.2,$$

which means that the lower approximation in the FVPRS model changes less than the lower approximation in the general fuzzy rough set model. However, the change is very large compared to other robust models.

The next model we discuss, is the soft fuzzy rough set model.

3.4.5 Soft fuzzy rough sets

Another robust fuzzy rough set model was introduced by Hu et al. ([26], [27]). As they use a soft threshold to compute the lower and upper approximation, it is called the soft fuzzy rough set model. We will show that this model is not well-defined.

We start with defining the soft distance between an element and a set.

Definition 3.4.23. Let A be a crisp set in U and $x \in U$. The *soft distance* between x and A is defined by

$$SD(x, A) = \arg_{d(x,y)} \sup_{y \in A} \{d(x, y) - \beta m_{x,y}\}$$

where d is a distance function, $\beta > 0$ is a penalty factor and

$$m_{x,y} = |\{y_i \in U \mid d(x, y_i) < d(x, y)\}|.$$

We already encounter a problem in this definition due to the use of the function $\arg_{d(x,y)}$. When U is infinite, the value of the supremum may not be reached for any y. A more serious problem occurs when the value of the supremum is reached for different values of y. Let us illustrate this with an example.

Example 3.4.24. Let $U = \{x, y_1, y_2, y_3\}$, $A = \{y_1, y_2, y_3\}$, $\beta = 0.1$ and

$$d(x, y_1) = 0.2, d(x, y_2) = 0.3, d(x, y_3) = 0.4.$$

Because $d(x, y_1) - \beta m_{x,y_1} = d(x, y_2) - \beta m_{x,y_2} = d(x, y_3) - \beta m_{x,y_3} = 0.2$, $SD(x, A)$ could be either 0.2, 0.3 or 0.4.

Based on this soft distance, Hu et al. define the soft fuzzy rough set model with distance function $d(x, y) = 1 - R(y, x)$ for all $x, y \in U$.

Definition 3.4.25. Let A be a fuzzy set in a fuzzy approximation space (U, R) with R a general fuzzy relation. With $x \in U$, define the *soft fuzzy lower approximation* $R{\downarrow}_S A$ of A as

$$(R{\downarrow}_S A)(x) = 1 - R\left(x, \arg_y \sup_{A(y) \leq A(y_L)} \{1 - R(y, x) - \beta m_{x,y}^{y_L}\}\right),$$

and the *soft fuzzy upper approximation* $R\uparrow_S A$ of A as

$$(R\uparrow_S A)(x) = R\left(x, \arg_y \inf_{A(y)\geq A(y_U)} \{R(y,x) + \beta n_{x,y}^{y_U}\}\right),$$

where

$$y_L = \arg_y \inf_{y\in U} \max\{1 - R(y,x), A(y)\},$$
$$m_{x,y}^{y_L} = |\{y_i \in U \mid A(y_i) \leq A(y_L) \wedge R(y_i,x) > R(y,x)\}|,$$
$$y_U = \arg_y \sup_{y\in U} \min\{R(y,x), A(y)\},$$
$$n_{x,y}^{y_U} = |\{y_i \in U \mid A(y_i) \geq A(y_U) \wedge R(y_i,x) > R(y,x)\}|.$$

and $\beta > 0$ a penalty factor.

We illustrate that this model is not well-defined.

Example 3.4.26. Let $U = \{x, y_1, y_2\}$, $A(y_1) = 0.1$, $A(y_2) = 0$, $A(x) = 0.5$, $R(y_1,x) = 0.95$, $R(y_2,x) = 0.9$, $R(x,x) = 1$, $\beta = 0.06$.

In this case, y_L could be equal to either y_1 or y_2, because for both of these values of y, $\max(1 - R(y,x), A(y)) = 0.1$.

If $y_L = y_1$, then $(R\downarrow_S A)(x) = 0.05$, because $m_{x,y_1}^{y_L} = 0$ and $m_{x,y_2}^{y_L} = 1$, and $1 - R(y_1,x) - \beta m_{x,y_1}^{y_L} = 0.05 > 1 - R(x,y_2) - \beta m_{x,y_2}^{y_L} = 0.04$.

On the other hand, if $y_L = y_2$, then $(R\downarrow_S A)(x) = 0.1$. This gives us two different values for the soft lower approximation of A in x.

As this model is not well-defined, we will not study the properties of this model in Chapter 4.

The last model we study, is the ordered weighted average-based fuzzy rough set model.

3.4.6 Ordered weighted average-based fuzzy rough sets

We continue with the model based on ordered weighted average (OWA) operators (Cornelis et al. [16]). Traditionally, the lower and upper approximation of a set A in U are determined by the worst, respectively best performing object. As we have seen, this leads to approximations which are sensitive to noisy data. OWA-based fuzzy rough sets are a possible solution for this problem.

The approximations are computed by an aggregation process, which is similar to the vaguely quantified fuzzy rough set approach, but the OWA-based approach has some advantages. First, it is monotonous with regard to the fuzzy relation R, as we will show in the next chapter. Secondly, the traditional fuzzy rough approximations can be recovered by choosing a particular OWA-operator. Finally, we can maintain the VQFRS rationale by introducing vague quantifiers into the OWA model.

Let us start with defining an OWA-operator.

Definition 3.4.27. Given a sequence D of n scalar values and a weight vector $W = \langle w_1, \ldots, w_n\rangle$ of length n, such that $w_i \in I$ for all $i \in \{1, \ldots, n\}$, and $\sum\limits_{i=1}^{n} w_i = 1$. Let σ be the permutation on $\{1, \ldots, n\}$ such that $d_{\sigma(i)}$ is the i^{th} largest value of D. The *OWA-operator* acting on D gives the value:

$$\mathrm{OWA}_W(D) = \sum_{i=1}^{n} w_i d_{\sigma(i)}.$$

The main strength of the OWA-operator is its flexibility. We can model a wide range of aggregation strategies, such as the maximum, the minimum and the average.

Example 3.4.28.

1. When we take $W_{\max} = \langle w_i\rangle$ with $w_1 = 1$ and $w_i = 0, i \neq 1$, we have $\mathrm{OWA}_{W_{\max}}(D) = \max\limits_{i=1}^{n}\{d_i\}$.
2. When we take $W_{\min} = \langle w_i\rangle$ with $w_n = 1, w_i = 0, i \neq n$, we have $\mathrm{OWA}_{W_{\min}}(D) = \min\limits_{i=1}^{n}\{d_i\}$.
3. When we take $W_{\text{avg}} = \langle w_i\rangle$ with $w_i = \frac{1}{n}, \forall i \in \{1, \ldots, n\}$, we have $\mathrm{OWA}_{W_{\text{avg}}}(D) = \frac{1}{n}\sum\limits_{i=1}^{n} d_i$.

There are several measures to analyse the OWA-operator, we give two of them: the orness- and the andness-degree. These measures compute how similar the OWA-operator is to the classical max-operator, respectively min-operator.

Definition 3.4.29. Let W be a weight vector of length n. The orness- and andness-degree of W are defined by

$$\mathrm{orness}(W) = \frac{1}{n-1}\sum_{i=1}^{n}((n-i)\cdot w_i),$$
$$\mathrm{andness}(W) = 1 - \mathrm{orness}(W).$$

As $\mathrm{orness}(W_{\max}) = 1$ and $\mathrm{andness}(W_{\min}) = 1$, we see that these measures indeed compute the similarity with the classical max-operator, respectively min-operator.

Now we can define the OWA-based lower and upper approximation of a fuzzy set A in a fuzzy approximation space (U, R).

Definition 3.4.30. Let A be a fuzzy set in a fuzzy approximation space (U, R), with $U = \{y_1, \ldots, y_n\}$ and R a general fuzzy relation. Given an implicator $\mathcal{I}$ and a conjunctor $\mathcal{C}$[3], and weight vectors W_1 and W_2 of length n, the OWA-based lower and upper approximation $R{\downarrow_{\mathcal{I},W_1}}A$ and $R{\uparrow_{\mathcal{C},W_2}}A$ of A are defined by

$$(R{\downarrow_{\mathcal{I},W_1}}A)(x) = \mathop{\mathrm{OWA}_{W_1}}_{y\in U}\langle \mathcal{I}(R(y,x), A(y))\rangle,$$
$$(R{\uparrow_{\mathcal{C},W_2}}A)(x) = \mathop{\mathrm{OWA}_{W_2}}_{y\in U}\langle \mathcal{C}(R(y,x), A(y))\rangle.$$

for all $x \in U$.

[3]In [16], t-norms instead of conjunctors were used.

To distinguish the behaviour of the lower and upper approximation, we enforce the conditions andness$(W_1) > 0.5$ and orness$(W_2) > 0.5$. When we take $W_1 = W_{\min}$ and $W_2 = W_{\max}$, we retrieve the traditional lower and upper approximation as in Definition 3.2.1.

Another possible pair of weight vectors (W_1, W_2) that fulfils the conditions andness$(W_1) > 0.5$ and orness$(W_2) > 0.5$ is given by

$$(W_1)_{n+1-i} = \begin{cases} \frac{2^{m-i}}{2^m-1} & i = 1, \dots, m \\ 0 & i = m+1, \dots, n \end{cases}$$

$$(W_2)_i = \begin{cases} \frac{2^{m-i}}{2^m-1} & i = 1, \dots, m \\ 0 & i = m+1, \dots, n \end{cases}$$

with $m \leq n$.

Let us study an example.

Example 3.4.31. Let U, A and R be as in Example 3.2.5: $U = \{y_1, y_2\}$, $A(y_1) = 0.2$, $A(y_2) = 0.8$ and R such that

$$R(y_1, y_1) = R(y_2, y_2) = 0.7, R(y_1, y_2) = 0 \text{ and } R(y_2, y_1) = 0.3.$$

Let $(\mathcal{I}, \mathcal{C}) = (\mathcal{I}_L, \mathcal{T}_L)$ and take $W_1 = \langle \frac{1}{3}, \frac{2}{3} \rangle$ and $W_2 = \langle \frac{2}{3}, \frac{1}{3} \rangle$, then andness$(W_1) > 0.5$ and orness$(W_2) > 0.5$. We obtain

$$\begin{aligned} (R{\downarrow}_{\mathcal{I}_L, W_1} A)(y_1) &= \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 0.5 = \frac{2}{3}, \\ (R{\downarrow}_{\mathcal{I}_L, W_1} A)(y_2) &= \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 1 = 1 \end{aligned}$$

$$\begin{aligned} (R{\uparrow}_{\mathcal{T}_L, W_2} A)(y_1) &= \frac{2}{3} \cdot 0.1 + \frac{1}{3} \cdot 0 = \frac{1}{15}, \\ (R{\uparrow}_{\mathcal{T}_L, W_2} A)(y_2) &= \frac{2}{3} \cdot 0.5 + \frac{1}{3} \cdot 0 = \frac{1}{3}. \end{aligned}$$

We illustrate that the OWA-based FRS model is more robust then the general fuzzy rough set model.

Example 3.4.32. Take U, A and R as in Example 3.4.3. Let $U = \{y_1, \dots, y_{100}, x\}$, i.e., $n = 101$, A a fuzzy set in U such that $A(y_i) = \frac{i}{100}$ for all $i \in \{1, \dots, 100\}$ and $A(x) = 1$. Let R be a fuzzy relation with $R(y_i, x) = \frac{i}{100}$ for all $i \in \{1, \dots, 100\}$ and $R(x, x) = 1$. When $A(y_{100}) = 0$ instead of 1, the lower approximation of A in x changes drastically from 1 to 0, if we apply the general fuzzy rough set model. Now, take $\mathcal{I} = \mathcal{I}_l$ and W_1 the weight vector

$$W_1 = \left\langle \frac{1}{102}, \dots, \frac{1}{102}, \frac{1}{102}, \frac{2}{102} \right\rangle,$$

then we have that andness$(W_1) = 0.505 > 0.5$. If $A(y_{100}) = 1$, then we have for all $z \in U$ that

$$\mathscr{I}_L(R(z,x),A(z)) = 1,$$

and so we have that

$$(R{\downarrow}_{\mathscr{I}_L,W_1}A)(x) = 100 \cdot \frac{1}{102} \cdot 1 + \frac{2}{102} \cdot 1 = 1.$$

If $A(y_{100}) = 0$, then

$$(R{\downarrow}_{\mathscr{I}_L,W_1}A)(x) = 100 \cdot \frac{1}{102} \cdot 1 + \frac{2}{102} \cdot 0 = \frac{100}{102},$$

which illustrates that the OWA-based FRS model is more robust than the general fuzzy rough set model.

By defining a weight vector W based on a quantifier Q, we maintain the VQFRS rationale. Yager ([64]) gave a lot of connections between weight vectors W and quantifiers Q. For example, with Q_u and Q_l regularly increasing fuzzy quantifiers, we can define weight vectors W_1 for the lower approximation and W_2 for the upper approximation as

$$(W_1)_i = Q_u\left(\frac{i}{n}\right) - Q_u\left(\frac{i-1}{n}\right),$$
$$(W_2)_i = Q_l\left(\frac{i}{n}\right) - Q_l\left(\frac{i-1}{n}\right),$$

for all $i \in \{1, \dots, n\}$. For example, with $(Q_u, Q_l) = (Q_\forall, Q_\exists)$ we obtain the weight vectors $W_{\min}$ and $W_{\max}$. Recall that not every quantifier is suitable, since the weight vectors W_1 and W_2 have to fulfil the conditions andness$(W_1) > 0.5$ and orness$(W_2) > 0.5$.

To end, we show that fuzzy rough sets based on robust nearest neighbour are a special case of OWA-based fuzzy rough sets.

Fuzzy rough sets based on robust nearest neighbour

Hu et al. ([29]) do not only give an overview of different fuzzy rough set models, they also introduce a new fuzzy rough set model based on the robust nearest neighbour. Because they focus on classification tasks, they only consider crisp subsets of U. They work with a kernel function R. However, their model turns out to be a special case of the OWA-model, where they use the weight vectors $W = \langle w_1, \dots, w_n \rangle$ which are shown in Table 3.2. The first three weight vectors are used to define a lower approximation of a subset A, the last three for an upper approximation. For the pair $(\mathscr{I}, \mathscr{C})$ they used the pairs $(\mathscr{I}_L, \mathscr{T}_M)$ and $(\mathscr{I}_{\cos}, \mathscr{C}_{\cos})$ with $\mathscr{C}_{\cos}(a,b) = \mathscr{J}_{\cos}(1-a, b)$, for all $a, b \in I$.

When we use their models, we expect to reduce the variation of approximations due to outliers, which means that the models are robust.

In the next chapter, we will study the properties of some of the models that we have discussed in this chapter.

	OWA weight vector
k-trimmed minimum	$w_i = \begin{cases} 1 & \text{if } i = k+1 \\ 0 & \text{otherwise} \end{cases}$
k-mean minimum	$w_i = \begin{cases} \frac{1}{k} & \text{if } i < k+1 \\ 0 & \text{otherwise} \end{cases}$
k-median minimum	$w_i = \begin{cases} 1 & \text{if } k \text{ odd, } i = \frac{k+1}{2} \\ \frac{1}{2} & \text{if } k \text{ even, } i = \frac{k}{2} \text{ or } i = \frac{k}{2} - 1 \\ 0 & \text{otherwise} \end{cases}$
k-trimmed maximum	$w_i = \begin{cases} 1 & \text{if } i = n-k-1 \\ 0 & \text{otherwise} \end{cases}$
k-mean maximum	$w_i = \begin{cases} \frac{1}{k} & \text{if } i > n-k-1 \\ 0 & \text{otherwise} \end{cases}$
k-median maximum	$w_i = \begin{cases} 1 & \text{if } k \text{ odd, } i = n - \frac{k+1}{2} \\ \frac{1}{2} & \text{if } k \text{ even, } i = n - \frac{k}{2} \text{ or } i = n - \frac{k}{2} + 1 \\ 0 & \text{otherwise} \end{cases}$

Table 3.2: Correspondence between robust nearest neighbour fuzzy rough sets and OWA fuzzy rough sets

Chapter 4

Properties of fuzzy rough sets

In this chapter we study the different properties given in Table 2.1 for some of the models discussed in Chapter 3. In this chapter, we consider all the constant sets $\hat{\alpha}$ for $\alpha \in I$ and not only for 0 and 1. Given a model, a fuzzy relation R and a finite universe U, we study which properties hold and which do not hold.

We start with the general fuzzy rough set model. Next, we discuss the properties of the tight and loose approximations. Further, we study the properties of the β-precision fuzzy rough set model, the vaguely quatified fuzzy rough set model, the fuzzy variable fuzzy rough set model and finally, the OWA-based fuzzy rough set model.

4.1 The general fuzzy rough set model

We start with the general model given in Definition 3.2.1. We first examine which properties hold when R is a general fuzzy relation and then which properties hold when R is a fuzzy similarity relation. We end this section with a brief overview of the properties of the original model of Dubois and Prade.

General fuzzy relation

The first property we study is the duality property. We show that this property holds for an implicator and a conjunctor based on its dual coimplicator and for an S-implicator based on a t-conorm and its dual t-norm. The duality property holds for an R-implicator based on a t-norm and this t-norm under extra conditions. We also show that the choice of negator is important: the negator has to be involutive and the implicator and conjunctor have to be dual with respect to this negator.

Proposition 4.1.1. Let $\mathcal{N}$ be an involutive negator and A a fuzzy set in a fuzzy approximation space (U,R) with R a general fuzzy relation. If the pair $(\mathcal{I},\mathcal{C})$ consists of an implicator $\mathcal{I}$ and a conjunctor $\mathscr{C}$ defined by the dual coimplicator $\mathscr{J}$ of $\mathscr{I}$ w.r.t. $\mathscr{N}$, then the duality property holds, i.e.,

$$R{\downarrow}_{\mathscr{I}}A = \mathrm{co}_{\mathscr{N}}(R{\uparrow}_{\mathscr{C}}(\mathrm{co}_{\mathscr{N}}(A))),$$
$$R{\uparrow}_{\mathscr{C}}A = \mathrm{co}_{\mathscr{N}}(R{\downarrow}_{\mathscr{I}}(\mathrm{co}_{\mathscr{N}}(A))).$$

Proof. Let $\mathscr{N}$ be an involutive negator and R a general fuzzy relation. Let us assume that $(\mathscr{I}, \mathscr{C})$ is such a pair, i.e., $\mathscr{I}$ is an implicator and $\mathscr{C}$ is a conjunctor based on the dual coimplicator $\mathscr{J}$ of $\mathscr{I}$ w.r.t $\mathscr{N}$, then by definition of having a dual implicator and coimplicator we have that

$$\forall a, b \in I : \mathscr{N}(\mathscr{C}(a, \mathscr{N}(b))) = \mathscr{N}(\mathscr{J}(\mathscr{N}(a), \mathscr{N}(b))) = \mathscr{I}(a, b)$$

and on the other hand, we have

$$\forall a, b \in I : \mathscr{N}(\mathscr{I}(a, \mathscr{N}(b))) = \mathscr{J}(\mathscr{N}(a), b)) = \mathscr{C}(a, b).$$

Now, let $A \in \mathscr{F}(U)$, $x \in U$. We obtain

$$\begin{aligned}
(\mathrm{co}_{\mathscr{N}}(R{\uparrow}_{\mathscr{C}}(\mathrm{co}_{\mathscr{N}}A)))(x) &= \mathscr{N}\left(\sup_{y\in U} \mathscr{C}(R(y,x), \mathscr{N}(A(y)))\right) \\
&= \inf_{y\in U} \mathscr{N}\left(\mathscr{C}(R(y,x), \mathscr{N}(A(y)))\right) \\
&= \inf_{y\in U} \mathscr{I}(R(y,x), A(y)) \\
&= (R{\downarrow}_{\mathscr{I}}A)(x).
\end{aligned}$$

In a similar way, we obtain

$$\begin{aligned}
(\mathrm{co}_{\mathscr{N}}(R{\downarrow}_{\mathscr{I}}(\mathrm{co}_{\mathscr{N}}A)))(x) &= \mathscr{N}\left(\inf_{y\in U} \mathscr{I}(R(y,x), \mathscr{N}(A(y)))\right) \\
&= \sup_{y\in U} \mathscr{N}\left(\mathscr{I}(R(y,x), \mathscr{N}(A(y)))\right) \\
&= \sup_{y\in U} \mathscr{C}(R(y,x), A(y)) \\
&= (R{\uparrow}_{\mathscr{C}}A)(x).
\end{aligned}$$

□

This property also holds for an S-implicator $\mathscr{I}$ based on a t-conorm $\mathscr{S}$ and a t-norm $\mathscr{T}$ dual to $\mathscr{S}$ w.r.t. an involutive negator $\mathscr{N}$, as shown in the next corollary.

Corollary 4.1.2. Let $\mathscr{N}$ be an involutive negator and $\mathscr{T}$ and $\mathscr{S}$ a dual t-norm and t-conorm with respect to $\mathscr{N}$. Let A be a fuzzy set in a fuzzy approximation space (U, R) with R a general fuzzy relation. If the pair $(\mathscr{I}, \mathscr{C})$ consists of the S-implicator based on $\mathscr{S}$ and the t-norm $\mathscr{T}$, then the duality principle holds, i.e.,

$$R{\downarrow}_{\mathscr{I}}A = \mathrm{co}_{\mathscr{N}}(R{\uparrow}_{\mathscr{C}}(\mathrm{co}_{\mathscr{N}}(A))),$$
$$R{\uparrow}_{\mathscr{C}}A = \mathrm{co}_{\mathscr{N}}(R{\downarrow}_{\mathscr{I}}(\mathrm{co}_{\mathscr{N}}(A))).$$

It also holds for a left-continuous t-norm $\mathscr{T}$ and its R-implicator $\mathscr{I}_{\mathscr{T}}$, but only when the involutive negator is the negator induced by $\mathscr{I}_{\mathscr{T}}$ (see [54]).

Corollary 4.1.3. Let $\mathscr{N}$ be an involutive negator and $\mathscr{T}$ a left-continuous t-norm. Let $\mathscr{I}_{\mathscr{T}}$ be the R-implicator based on $\mathscr{T}$. Let A be a fuzzy set in a fuzzy approximation space (U,R) with R a general fuzzy relation. If the pair $(\mathscr{I},\mathscr{C})$ consists of the R-implicator based on $\mathscr{T}$ and the left-continuous t-norm $\mathscr{T}$ and the negator $\mathscr{N}$ is the negator induced by $\mathscr{I}_{\mathscr{T}}$, then the duality principle holds, i.e.,

$$R{\downarrow}_{\mathscr{I}}A = \mathrm{co}_{\mathscr{N}}(R{\uparrow}_{\mathscr{C}}(\mathrm{co}_{\mathscr{N}}(A))),$$
$$R{\uparrow}_{\mathscr{C}}A = \mathrm{co}_{\mathscr{N}}(R{\downarrow}_{\mathscr{I}}(\mathrm{co}_{\mathscr{N}}(A))).$$

The duality property does not necessarily holds for other choices of fuzzy logical operators. Let us illustrate this with a counterexample.

Example 4.1.4. Let $\mathscr{N}$ be a negator defined by

$$\mathscr{N}(a) = \begin{cases} 1-a & 0 \le a \le \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} < a \le \frac{2}{3} \\ 1-a & \frac{2}{3} \le a \le 1. \end{cases}$$

We see that $\mathscr{N}$ is not involutive, since $\mathscr{N}\left(\mathscr{N}\left(\frac{1}{2}\right)\right) = \mathscr{N}\left(\frac{1}{3}\right) = \frac{2}{3}$. Let us define a t-norm $\mathscr{T}$ by

$$\mathscr{T}(a,b) = \begin{cases} 0 & a \le \mathscr{N}(b) \\ \min\{a,b\} & \mathscr{N}(b) < a \end{cases}$$

for all $a,b \in I$, then $\mathscr{T}$ is left-continuous. The R-implicator based on $\mathscr{T}$ is given by

$$\mathscr{I}(a,b) = \begin{cases} 1 & a \le b \\ \max\{\mathscr{N}(a),b\} & b < a \end{cases}$$

for all $a,b \in I$. The negator induced by this $\mathscr{I}$ is the negator defined above, i.e., for all $a \in I$ we have that $\mathscr{N}(a) = \mathscr{I}(a,0)$. We compute $\mathscr{N}\left(\mathscr{T}\left(\frac{2}{3},\mathscr{N}\left(\frac{1}{2}\right)\right)\right)$ and $\mathscr{I}\left(\frac{2}{3},\frac{1}{2}\right)$:

$$\mathscr{N}\left(\mathscr{T}\left(\frac{2}{3},\mathscr{N}\left(\frac{1}{2}\right)\right)\right) = \mathscr{N}\left(\mathscr{T}\left(\frac{2}{3},\frac{2}{3}\right)\right) = \mathscr{N}\left(\frac{2}{3}\right) = \frac{1}{3},$$
$$\mathscr{I}\left(\frac{2}{3},\frac{1}{2}\right) = \max\left\{\mathscr{N}\left(\frac{2}{3}\right),\frac{1}{2}\right\} = \max\left\{\frac{1}{3},\frac{1}{2}\right\} = \frac{1}{2},$$

which is not the same. This means that we have found an a and b in I such that

$$\mathscr{N}(\mathscr{T}(a,\mathscr{N}(b))) \neq \mathscr{I}(a,b).$$

This means that

$$\mathrm{co}_{\mathscr{N}}(R\uparrow_{\mathscr{C}}(\mathrm{co}_{\mathscr{N}}(A))) = R\downarrow_{\mathscr{I}}A$$

not necessarily holds for this choice of $\mathscr{N}$, $\mathscr{I}$ and $\mathscr{C}$. For example for $U = \{x, y\}$, A such that $A(x) = 1$ and $A(y) = \frac{1}{2}$ and R such that $R(y,x) = R(x,y) = \frac{2}{3}$ and $R(x,x) = R(y,y) = 1$.

It is not only important that the negator is involutive, it is also important that the negator is equal to the negator induced by $\mathscr{I}$, which is the same as assuming that $\mathscr{I}$ and $\mathscr{C}$ are dual with respect to that specific negator.

Example 4.1.5. Let $\mathscr{N}$ be the standard negator $\mathscr{N}_S$, $\mathscr{I}$ the Gödel implicator $\mathscr{I}_G$ and $\mathscr{C}$ the minimum t-norm $\mathscr{T}_M$. The negator induced by $\mathscr{I}$ is the Gödel negator $\mathscr{N}_G$ and thus not $\mathscr{N}_S$. It also holds that $\mathscr{I}_G$ and $\mathscr{T}_M$ are not dual with respect to $\mathscr{N}_S$, since

$$\mathscr{N}_S(\mathscr{T}_M(0.5, \mathscr{N}_S(0.5))) = 1 - \min\{0.5, 1 - 0.5\} = 1 - 0.5 = 0.5,$$

and $\mathscr{I}_G(0.5, 0.5) = 1$. For this triple $(\mathscr{N}_S, \mathscr{I}_G, \mathscr{T}_M)$ the duality property will not hold, although $\mathscr{N}_S$ is involutive.

We continue with the monotonicity properties. We show that the monotonicity of sets and the monotonicity of relations hold in this model. Especially the monotonicity of relations will be important to have when dealing with feature selection, an important application of fuzzy rough sets. Note that the monotonicity properties do not depend on properties of the fuzzy relation.

Proposition 4.1.6. Let A and B be fuzzy sets in (U, R) with R a general fuzzy relation. Let $\mathscr{I}$ be an implicator and $\mathscr{C}$ a conjunctor. If $A \subseteq B$, then we have that

$$R\downarrow_{\mathscr{I}}A \subseteq R\downarrow_{\mathscr{I}}B,$$
$$R\uparrow_{\mathscr{C}}A \subseteq R\uparrow_{\mathscr{C}}B.$$

Proof. This follows from that fact that both an implicator and a conjunctor are non-decreasing in the second argument. □

Proposition 4.1.7. Let R_1 and R_2 be fuzzy relations on U and A a fuzzy set in U. Let $\mathscr{I}$ be an implicator and $\mathscr{C}$ a conjunctor. If $R_1 \subseteq R_2$, then we have that

$$R_2\downarrow_{\mathscr{I}}A \subseteq R_1\downarrow_{\mathscr{I}}A,$$
$$R_1\uparrow_{\mathscr{C}}A \subseteq R_2\uparrow_{\mathscr{C}}A.$$

Proof. This follows from the fact that an implicator is non-increasing and a conjunctor is non-decreasing in the first argument. □

When we look at the minimum and maximum operator, the properties of ‘Intersection’ and ‘Union’ still hold.

Proposition 4.1.8. Let A and B be fuzzy sets in (U,R) with R a general fuzzy relation. Let $\mathscr{I}$ be an implicator and $\mathscr{C}$ a conjunctor. We have that

$$\begin{aligned}
R{\downarrow_{\mathscr{I}}}(A\cap B) &= R{\downarrow_{\mathscr{I}}}A\cap R{\downarrow_{\mathscr{I}}}B,\\
R{\uparrow_{\mathscr{C}}}(A\cap B) &\subseteq R{\uparrow_{\mathscr{C}}}A\cap R{\uparrow_{\mathscr{C}}}B,\\
R{\downarrow_{\mathscr{I}}}(A\cup B) &\supseteq R{\downarrow_{\mathscr{I}}}A\cup R{\downarrow_{\mathscr{I}}}B,\\
R{\uparrow_{\mathscr{C}}}(A\cup B) &= R{\uparrow_{\mathscr{C}}}A\cup R{\uparrow_{\mathscr{C}}}B.
\end{aligned}$$

Proof. Let A and B be fuzzy sets in U. Based on the monotonicity properties proved in Proposition 4.1.6 and

$$A\cap B\subseteq A,B\subseteq A\cup B,$$

the second and third property are fulfilled. With $x\in U$, we have that

$$\begin{aligned}
&(R{\downarrow_{\mathscr{I}}}A\cap R{\downarrow_{\mathscr{I}}}B)(x)\\
&\quad=\min\{R{\downarrow_{\mathscr{I}}}A(x),R{\downarrow_{\mathscr{I}}}B(x)\}\\
&\quad=\min\left\{\inf_{y\in U}\mathscr{I}(R(y,x),A(y)),\ \inf_{y\in U}\mathscr{I}(R(y,x),B(y))\right\}\\
&\quad=\min\Big\{\inf_{y\in U}\mathscr{I}(R(y,x),\min\{A(y),B(y)\}),\\
&\qquad\qquad\inf_{y\in U}\mathscr{I}(R(y,x),\max\{A(y),B(y)\})\Big\}\\
&\quad=\inf_{y\in U}\mathscr{I}(R(y,x),\min\{A(y),B(y)\})\\
&\quad=R{\downarrow_{\mathscr{I}}}(A\cap B)(x).
\end{aligned}$$

The third step holds, since

$$\begin{aligned}
&\inf_{y\in U}\mathscr{I}(R(y,x),A(y))\\
&=\min\Bigg\{\inf_{\substack{y\in U\\ A(y)\le B(y)}}\mathscr{I}(R(y,x),\min\{A(y),B(y)\}),\\
&\qquad\qquad\inf_{\substack{y\in U\\ B(y)\le A(y)}}\mathscr{I}(R(y,x),\max\{A(y),B(y)\})\Bigg\}
\end{aligned}$$

and

$$\begin{aligned}
&\inf_{y\in U}\mathscr{I}(R(y,x),B(y))\\
&=\min\Bigg\{\inf_{\substack{y\in U\\ B(y)\le A(y)}}\mathscr{I}(R(y,x),\min\{A(y),B(y)\}),\\
&\qquad\qquad\inf_{\substack{y\in U\\ A(y)\le B(y)}}\mathscr{I}(R(y,x),\max\{A(y),B(y)\})\Bigg\}
\end{aligned}$$

and hence

$$
\begin{aligned}
&\min\left\{\inf_{y\in U}\mathscr{I}(R(y,x),A(y)),\inf_{y\in U}\mathscr{I}(R(y,x),B(y))\right\}\\
&=\min\left\{\min\left\{\inf_{\substack{y\in U\\A(y)\leq B(y)}}\mathscr{I}(R(y,x),\min\{A(y),B(y)\}),\right.\right.\\
&\qquad\left.\inf_{\substack{y\in U\\B(y)\leq A(y)}}\mathscr{I}(R(y,x),\max\{A(y),B(y)\})\right\},\\
&\qquad\min\left\{\inf_{\substack{y\in U\\B(y)\leq A(y)}}\mathscr{I}(R(y,x),\min\{A(y),B(y)\}),\right.\\
&\qquad\left.\left.\inf_{\substack{y\in U\\A(y)\leq B(y)}}\mathscr{I}(R(y,x),\max\{A(y),B(y)\})\right\}\right\}\\
&=\min\left\{\inf_{y\in U}\mathscr{I}(R(y,x),\min\{A(y),B(y)\}),\right.\\
&\qquad\left.\inf_{y\in U}\mathscr{I}(R(y,x),\max\{A(y),B(y)\})\right\}.
\end{aligned}
$$

We prove the last property.

$$
\begin{aligned}
&(R{\uparrow_{\mathscr{C}}}A\cup R{\uparrow_{\mathscr{C}}}B)(x)\\
&\qquad=\max\left\{R{\uparrow_{\mathscr{C}}}A(x),R{\uparrow_{\mathscr{C}}}B(x)\right\}\\
&\qquad=\max\left\{\sup_{y\in U}\mathscr{C}(R(y,x),A(y)),\sup_{y\in U}\mathscr{C}(R(y,x),B(y))\right\}\\
&\qquad=\max\left\{\sup_{y\in U}\mathscr{C}(R(y,x),\min\{A(y),B(y)\}),\right.\\
&\qquad\qquad\left.\sup_{y\in U}\mathscr{C}(R(y,x),\max\{A(y),B(y)\})\right\}\\
&\qquad=\sup_{y\in U}\mathscr{C}(R(y,x),\max\{A(y),B(y)\})\\
&\qquad=R{\uparrow_{\mathscr{C}}}(A\cup B)(x).
\end{aligned}
$$

□

We also always have that $R{\uparrow_{\mathscr{C}}}\emptyset=\emptyset$ and $R{\downarrow_{\mathscr{I}}}U=U$, because for all conjunctor $\mathscr{C}$ it holds that $\mathscr{C}(a,0)\leq\mathscr{C}(1,0)=0$, for all $a\in I$, and for all implicators $\mathscr{I}$ it holds that $\mathscr{I}(a,1)\geq\mathscr{I}(1,1)=1$, for all $a\in I$. The other properties do not hold for general fuzzy relations. For example, the inclusion property only holds when the relation R is reflexive, as we will show in Chapter 5 and now illustrate with an example.

Example 4.1.9. Consider the universe $U=\{y_1,y_2\}$, A a fuzzy set such that $A(y_1)=0.5$ and $A(y_2)=1$ and R the general fuzzy relation such that $R(x,z)=0.5$, for all $x,z\in U$. Let us take the

Łukasiewicz implicator and t-norm $(\mathscr{I}_L, \mathscr{T}_L)$. Then we have that

$$\begin{aligned}(R\downarrow_{\mathscr{I}_L} A)(y_1) &= \inf_{z\in U} \min\{1, 1 - R(z, y_1) + A(z)\}\\ &= \inf_{z\in U} \min\{1, 1/2 + A(z)\}\\ &= \min\{1, 1\}\\ &> A(y_1)\end{aligned}$$

and thus we have that $R\downarrow_{\mathscr{I}_L} A \nsubseteq A$. Similarly, we obtain

$$\begin{aligned}(R\uparrow_{\mathscr{T}_L} A)(y_2) &= \sup_{z\in U} \max\{0, R(z, y_2) + A(z) - 1\}\\ &= \sup_{z\in U} \max\{0, A(z) - 0.5\}\\ &= \max\{0, 0.5\}\\ &< A(y_2)\end{aligned}$$

and thus $A \nsubseteq R\uparrow_{\mathscr{T}_L} A$.

We study now which properties hold when R is a fuzzy similarity relation.

Fuzzy similarity relation

Recall that if R is a fuzzy similarity relation, then it is a fuzzy $\mathscr{T}$-similarity relation for every t-norm $\mathscr{T}$. We start with the inclusion property, i.e., we prove that the lower approximation of A is contained in A and that A is contained in the upper approximation of A.

Proposition 4.1.10. Let A be a fuzzy set in a fuzzy approximation space (U, R) with R a fuzzy similarity relation. If $\mathscr{I}$ is a border implicator and if $\mathscr{C}$ is a conjunctor that satisfies the condition $\mathscr{C}(1, a) = a$ for all $a \in I$, then we have

$$\begin{aligned}&R\downarrow_{\mathscr{I}} A \subseteq A,\\ &A \subseteq R\uparrow_{\mathscr{C}} A.\end{aligned}$$

Proof. Let $\mathscr{I}$ be a border implicator, $\mathscr{C}$ a conjunctor such that $\mathscr{C}(1, a) = a$ for all $a \in I$ and R a fuzzy similarity relation. Let A be a fuzzy set in U and $x \in U$, then it holds that

$$\begin{aligned}(R\downarrow_{\mathscr{I}} A)(x) &= \inf_{y\in U} \mathscr{I}(R(y, x), A(y))\\ &\leq \mathscr{I}(R(x, x), A(x))\\ &= \mathscr{I}(1, A(x))\\ &= A(x),\end{aligned}$$

and it holds that

$$\begin{aligned}(R{\uparrow_{\mathscr{C}}}A)(x) &= \sup_{y\in U} \mathscr{C}(R(y,x),A(y)) \\ &\geq \mathscr{C}(R(x,x),A(x)) \\ &= \mathscr{C}(1,A(x)) \\ &= A(x).\end{aligned}$$

□

Note that when $\mathscr{C}$ is a t-norm, the condition for $\mathscr{C}$ is satisfied. The inclusion property also holds for relations that are only reflexive. If the inclusion property holds, then we have that $R{\downarrow_{\mathscr{I}}}\emptyset = \emptyset$ and $R{\uparrow_{\mathscr{C}}}U = U$.

When we work with fuzzy sets, we can generalise the property $R{\downarrow_{\mathscr{I}}}\emptyset = \emptyset = R{\uparrow_{\mathscr{C}}}\emptyset$ to all constant sets.

Proposition 4.1.11. Let (U,R) be a fuzzy approximation space with R a fuzzy similarity relation. Let $\hat{\alpha}$ be the constant α-set, with $\alpha \in I$. If $\mathscr{I}$ is a border implicator and if $\mathscr{C}$ a conjunctor that satisfies the condition $\mathscr{C}(1,a) = a$ for all $a \in I$, then we have

$$\begin{aligned}R{\downarrow_{\mathscr{I}}}\hat{\alpha} &= \hat{\alpha}, \\ R{\uparrow_{\mathscr{C}}}\hat{\alpha} &= \hat{\alpha}.\end{aligned}$$

Proof. Let R be a fuzzy similarity relation and $\alpha \in I$. Let $\mathscr{I}$ be a border implicator and $\mathscr{C}$ a conjunctor that satisfies the condition $\mathscr{C}(1,a) = a$ for all $a \in I$. Since the inclusion property holds, we have that $R{\downarrow_{\mathscr{I}}}\hat{\alpha} \subseteq \hat{\alpha}$ and $\hat{\alpha} \subseteq R{\uparrow_{\mathscr{C}}}\hat{\alpha}$. Take $x \in U$. Due to the monotonicity of an implicator, we have for all $y \in U$ that

$$\alpha = \mathscr{I}(1,\alpha) \leq \mathscr{I}(R(y,x),\alpha),$$

which means that

$$(R{\downarrow_{\mathscr{I}}}\hat{\alpha})(x) = \inf_{y\in U} \mathscr{I}(R(y,x),\alpha) \geq \alpha = \hat{\alpha}(x).$$

We obtain that $R{\downarrow_{\mathscr{I}}}\hat{\alpha} = \hat{\alpha}$. Similarly, because for all $y \in U$ it holds that

$$\mathscr{C}(R(y,x),\alpha) \leq \mathscr{C}(1,\alpha) = \alpha,$$

and thus

$$(R{\uparrow_{\mathscr{C}}}\hat{\alpha})(x) = \sup_{y\in U} \mathscr{C}(R(y,x),\alpha) \leq \alpha = \hat{\alpha}(x),$$

we obtain that $R{\uparrow_{\mathscr{C}}}\hat{\alpha} = \hat{\alpha}$. □

Note that this property holds for all reflexive relations R, but not if R is a general fuzzy relation. We give a counterexample.

Example 4.1.12. Let $R(x,y) = 0.5$, for all x, y in U. R is not reflexive, and thus no similarity relation. We take the Łukasiewicz implicator and t-norm as implicator and conjunctor of the model. Consider the fuzzy set $\hat{\alpha}(x) = 0.5$, for all $x \in U$. For $x \in U$ we have

$$(R{\downarrow_{\mathscr{I}_L}}\hat{\alpha})(x) = \inf_{y\in U} \min\{1, 1 - R(y,x) + 0.5\} = \inf_{y\in U} \min\{1,1\} = 1$$

which is greater than 0.5. We also have that

$$(R{\uparrow_{\mathscr{T}_L}}\hat{\alpha})(x) = \sup_{y\in U} \max\{0, R(y,x) + 1 - 0.5\} = \sup_{y\in U} \max\{0,1\} = 1$$

which is greater than 0.5. This proves that Proposition 4.1.11 does not hold in general.

We end with the idempotence property, i.e., doing the same approximation twice gives the same result as doing the approximation only once.

Proposition 4.1.13. Let $\mathscr{C}$ be a left-continuous t-norm $\mathscr{T}$ and $\mathscr{I}_{\mathscr{T}}$ the R-implicator based on $\mathscr{T}$. Let A be a fuzzy set in a fuzzy approximation space (U,R) with R a fuzzy $\mathscr{T}$-similarity relation, then we have that

$$R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}(R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A) = R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A,$$
$$R{\uparrow_{\mathscr{T}}}(R{\uparrow_{\mathscr{T}}}A) = R{\uparrow_{\mathscr{T}}}A.$$

Proof. Since a t-norm fulfils the equation $\mathscr{T}(1,a) = a$ for all $a \in I$ and since an R-implicator is a border implicator (see Proposition 2.2.35), the inclusion property holds. This means that

$$R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}(R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A) \subseteq R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A,$$
$$R{\uparrow_{\mathscr{T}}}A \subseteq R{\uparrow_{\mathscr{T}}}(R{\uparrow_{\mathscr{T}}}A),$$

for all $A \in \mathscr{F}(U)$. Since $\mathscr{T}$ is left-continuous and R is $\mathscr{T}$-transitive, we have for $x \in U$ that

$$\begin{aligned}
(R{\uparrow_{\mathscr{T}}}(R{\uparrow_{\mathscr{T}}}A))(x) &= \sup_{y\in U} \mathscr{T}\left(R(y,x), \sup_{z\in U} \mathscr{T}(R(z,y),A(z))\right)\\
&= \sup_{y\in U}\sup_{z\in U} \mathscr{T}(R(y,x), \mathscr{T}(R(z,y),A(z)))\\
&= \sup_{z\in U}\sup_{y\in U} \mathscr{T}(\mathscr{T}(R(z,y),R(y,x)),A(z)))\\
&= \sup_{z\in U} \mathscr{T}(\sup_{y\in U} \mathscr{T}(R(z,y),R(y,x)),A(z)))\\
&\leq \sup_{z\in U} \mathscr{T}(R(z,x),A(z))\\
&= (R{\uparrow_{\mathscr{T}}}A)(x),
\end{aligned}$$

and thus $R{\uparrow_{\mathscr{T}}}(R{\uparrow_{\mathscr{T}}}A) = R{\uparrow_{\mathscr{T}}}A$. For the other equality, recall the following properties for $\mathscr{I}_{\mathscr{T}}$ and $\mathscr{T}$ (see [54]):

$$\mathscr{I}_{\mathscr{T}}(\sup_{j\in J} b_j, a) = \inf_{j\in J} \mathscr{I}_{\mathscr{T}}(b_j, a),$$
$$\mathscr{I}_{\mathscr{T}}(a, \inf_{j\in J} b_j) = \inf_{j\in J} \mathscr{I}_{\mathscr{T}}(a, b_j),$$
$$\mathscr{I}_{\mathscr{T}}(a, \mathscr{I}_{\mathscr{T}}(b,c)) = \mathscr{I}_{\mathscr{T}}(\mathscr{T}(a,b),c),$$

for $a, b_j, b, c \in I$ and J a set of indices. Since R is $\mathscr{T}$-transitive we obtain for $x \in U$ that

$$\begin{aligned}
(R\downarrow_{\mathscr{I}_\mathscr{T}} A)(x) &= \inf_{y\in U} \mathscr{I}_\mathscr{T}(R(y,x), A(y)) \\
&\leq \inf_{y\in U} \mathscr{I}_\mathscr{T}\left(\sup_{z\in U} \mathscr{T}(R(y,z), R(z,x)), A(y)\right) \\
&= \inf_{y\in U} \inf_{z\in U} \mathscr{I}_\mathscr{T}(\mathscr{T}(R(z,x), R(y,z)), A(y)) \\
&= \inf_{y\in U} \inf_{z\in U} \mathscr{I}_\mathscr{T}(R(z,x), \mathscr{I}_\mathscr{T}(R(y,z), A(y)) \\
&= \inf_{z\in U} \mathscr{I}_\mathscr{T}(R(z,x), \inf_{y\in U} \mathscr{I}_\mathscr{T}(R(y,z), A(y)) \\
&= \inf_{z\in U} \mathscr{I}_\mathscr{T}(R(z,x), (R\downarrow_{\mathscr{I}_\mathscr{T}} A)(z)) \\
&= (R\downarrow_{\mathscr{I}_\mathscr{T}}(R\downarrow_{\mathscr{I}_\mathscr{T}} A))(x).
\end{aligned}$$

This completes the proof. □

This property also holds for relations that are reflexive and $\mathscr{T}$-transitive. It is important that $\mathscr{I}$ is the R-implicator of $\mathscr{T}$. We illustrate this with an example.

Example 4.1.14. Take the implicator $\mathscr{I}(a,b) = \max\{1-a, b^2\}$, $a, b \in I$. This is not an R-implicator. Let us look at the universe U with one element $\{y\}$, the fuzzy set A such that $A(y) = 0.2$ and the relation $R(y,y) = 1$. Then $(R\downarrow_\mathscr{I} A)(y) = \mathscr{I}(1, 0.2) = 0.04$ and $(R\downarrow_\mathscr{I}(R\downarrow_\mathscr{I} A))(y) = \mathscr{I}(1, 0.04) = 0.0016$. The idempotence property does not hold.

We can conclude that, under certain conditions, all the properties that hold in a Pawlak approximation space, still hold for the general fuzzy rough set model.

Next, we study the properties of the model of Dubois and Prade.

Dubois and Prade's model

We briefly discuss which properties hold in the model designed by Dubois and Prade, i.e., R is a fuzzy min-similarity relation, $\mathscr{I}$ is the Kleene-Dienes implicator $\mathscr{I}_{KD}$ and $\mathscr{C}$ is the minimum t-norm $\mathscr{T}_M$.

It is obvious that the inclusion property and the monotonicity properties hold. The duality property with $\mathscr{N} = \mathscr{N}_S$ also holds, since $\mathscr{I}_{KD}$ is the S-implicator based on $\mathscr{S}_M$, the t-conorm dual to $\mathscr{T}_M$ with respect to $\mathscr{N}_S$. The intersection property and union property hold for the intersection and union defined by Zadeh. We also have that

$$R\downarrow_{\mathscr{I}_{KD}} \hat{\alpha} = \hat{\alpha} = R\uparrow_{\mathscr{T}_M} \hat{\alpha}$$

holds for all $\alpha \in I$ and thus also for $\emptyset$ and U. Less obvious is the idempotence property. The Kleene-Dienes implicator is an S-implicator, but not an R-implicator. We prove that the property holds for Dubois and Prade's model.

Proposition 4.1.15. The idempotence property holds for the model designed by Dubois and Prade.

Proof. Let A be a fuzzy set in (U,R) with R a fuzzy similarity relation. As the inclusion property holds, we have that $R{\downarrow}_{\mathscr{I}_{KD}}(R{\downarrow}_{\mathscr{I}_{KD}}A) \subseteq R{\downarrow}_{\mathscr{I}_{KD}}A$ and $R{\uparrow}_{\mathscr{T}_M}(R{\uparrow}_{\mathscr{T}_M}A) \supseteq R{\uparrow}_{\mathscr{T}_M}A$. We know that the minimum operator is left-continuous and thus complete-distributive w.r.t. the supremum. We also know that the minimum t-norm is associative and that R is min-transitive. Now let x be an element of U, we have:

$$\begin{aligned}
(R{\uparrow}_{\mathscr{T}_M}(R{\uparrow}_{\mathscr{T}_M}A))(x) &= \sup_{y\in U}\min\left\{R(y,x), \sup_{z\in U}\min\{R(z,y),A(z)\}\right\}\\
&= \sup_{y\in U}\sup_{z\in U}\min\left\{R(y,x),\min\{R(z,y),A(z)\}\right\}\\
&= \sup_{z\in U}\min\left\{\sup_{y\in U}\min\{R(z,y),R(y,x)\},A(z)\right\}\\
&\leq \sup_{z\in U}\min\{R(z,x),A(z)\}\\
&= (R{\uparrow}_{\mathscr{T}_M}A)(x).
\end{aligned}$$

So we have that $R{\uparrow}_{\mathscr{T}_M}(R{\uparrow}_{\mathscr{T}_M}A) = R{\uparrow}_{\mathscr{T}_M}A$. Since the duality property holds with $\mathscr{N} = \mathscr{N}_S$, we have that

$$\begin{aligned}
R{\downarrow}_{\mathscr{I}_{KD}}A &= \mathrm{co}_{\mathscr{N}_S}(R{\uparrow}_{\mathscr{T}_M}(\mathrm{co}_{\mathscr{N}_S}(A)))\\
&= \mathrm{co}_{\mathscr{N}_S}(R{\uparrow}_{\mathscr{T}_M}(R{\uparrow}_{\mathscr{T}_M}\,\mathrm{co}_{\mathscr{N}_S}(A)))\\
&= R{\downarrow}_{\mathscr{I}_{KD}}(\mathrm{co}_{\mathscr{N}_S}(R{\uparrow}_{\mathscr{T}_M}(\mathrm{co}_{\mathscr{N}_S}(A))))\\
&= R{\downarrow}_{\mathscr{I}_{KD}}(R{\downarrow}_{\mathscr{I}_{KD}}(\mathrm{co}_{\mathscr{N}_S}(\mathrm{co}_{\mathscr{N}_S}(A))))\\
&= R{\downarrow}_{\mathscr{I}_{KD}}(R{\downarrow}_{\mathscr{I}_{KD}}A)
\end{aligned}$$

and thus $R{\downarrow}_{\mathscr{I}_{KD}}(R{\downarrow}_{\mathscr{I}_{KD}}A) = R{\downarrow}_{\mathscr{I}_{KD}}A$. This completes the proof. □

In the next section, we discuss the properties of tight and loose approximations.

4.2 Tight and loose approximations

We continue with the properties of the model defined in Definition 3.3.3. We again start with considering a general fuzzy relation. A lot of properties were studied in [11, 13]. As the traditional lower and upper approximation were already discussed in the previous section, we only focus on the tight and loose approximations in this section.

General fuzzy relation

We start again with the duality propery. This holds for the same combinations of $\mathscr{I}$ and $\mathscr{C}$ as we saw before.

Proposition 4.2.1. Let $\mathscr{N}$ be an involutive negator and A a fuzzy set in a fuzzy approximation space (U,R) with R a general fuzzy relation. If the pair $(\mathscr{I},\mathscr{C})$ consists of an implicator $\mathscr{I}$ and a conjunctor $\mathscr{C}$ defined by the dual coimplicator $\mathscr{J}$ of $\mathscr{I}$ w.r.t. $\mathscr{N}$, then the duality property holds, i.e.,

$$R{\downarrow_{\mathscr{I}}}{\downarrow_{\mathscr{I}}}A = \mathrm{co}_{\mathscr{N}}(R{\uparrow_{\mathscr{C}}}{\uparrow_{\mathscr{C}}}(\mathrm{co}_{\mathscr{N}}(A))),$$
$$R{\uparrow_{\mathscr{C}}}{\uparrow_{\mathscr{C}}}A = \mathrm{co}_{\mathscr{N}}(R{\downarrow_{\mathscr{I}}}{\downarrow_{\mathscr{I}}}(\mathrm{co}_{\mathscr{N}}(A))),$$

$$R{\uparrow_{\mathscr{C}}}{\downarrow_{\mathscr{I}}}A = \mathrm{co}_{\mathscr{N}}(R{\downarrow_{\mathscr{I}}}{\uparrow_{\mathscr{C}}}(\mathrm{co}_{\mathscr{N}}(A))),$$
$$R{\downarrow_{\mathscr{I}}}{\uparrow_{\mathscr{C}}}A = \mathrm{co}_{\mathscr{N}}(R{\uparrow_{\mathscr{C}}}{\downarrow_{\mathscr{I}}}(\mathrm{co}_{\mathscr{N}}(A))).$$

Proof. The proof of the proposition is similar to the one of the general fuzzy rough set model (see Proposition 4.1.1). □

Again, this also holds for an S-implicator $\mathscr{I}$ based on a t-conorm $\mathscr{S}$ and its dual t-norm $\mathscr{T}$ with respect to an involutive negator $\mathscr{N}$ and for a left-continuous t-norm $\mathscr{T}$ and its R-implicator $\mathscr{I}_{\mathscr{T}}$ if $\mathscr{N} = \mathscr{N}_{\mathscr{I}_{\mathscr{T}}}$ is involutive.

The monotonicity of sets still holds.

Proposition 4.2.2. Let A and B be fuzzy sets in (U,R) with R a general fuzzy relation. Let $\mathscr{I}$ be an implicator and $\mathscr{C}$ a conjunctor. If $A \subseteq B$, then we have that

$$R{\downarrow_{\mathscr{I}}}{\downarrow_{\mathscr{I}}}A \subseteq R{\downarrow_{\mathscr{I}}}{\downarrow_{\mathscr{I}}}B,$$
$$R{\uparrow_{\mathscr{C}}}{\downarrow_{\mathscr{I}}}A \subseteq R{\uparrow_{\mathscr{C}}}{\downarrow_{\mathscr{I}}}B,$$
$$R{\downarrow_{\mathscr{I}}}{\uparrow_{\mathscr{C}}}A \subseteq R{\downarrow_{\mathscr{I}}}{\uparrow_{\mathscr{C}}}B,$$
$$R{\uparrow_{\mathscr{C}}}{\uparrow_{\mathscr{C}}}A \subseteq R{\uparrow_{\mathscr{C}}}{\uparrow_{\mathscr{C}}}B.$$

Proof. This follows from that fact that both an implicator and a conjunctor are non-decreasing in the second argument. □

The property of monotonicity of relations holds for the tight lower approximation and the loose upper approximation.

Proposition 4.2.3. Let R_1 and R_2 be fuzzy relations on U and A a fuzzy set in U. Let $\mathscr{I}$ be an implicator and $\mathscr{C}$ a conjunctor. If $R_1 \subseteq R_2$, then we have that

$$R_2{\downarrow_{\mathscr{I}}}{\downarrow_{\mathscr{I}}}A \subseteq R_1{\downarrow_{\mathscr{I}}}{\downarrow_{\mathscr{I}}}A,$$
$$R_1{\uparrow_{\mathscr{C}}}{\uparrow_{\mathscr{C}}}A \subseteq R_2{\uparrow_{\mathscr{C}}}{\uparrow_{\mathscr{C}}}A.$$

Proof. This follows from the fact that an implicator is non-increasing and a conjunctor is non-decreasing in the first argument. □

We cannot give such a property for the loose lower approximation and the tight upper approximation. We illustrate this with an example.

Example 4.2.4. Let us take $U = \{y_1, y_2\}$, R_1 a general fuzzy relation such that

$$R_1(y_1, y_1) = 1, R_1(y_1, y_2) = R_1(y_2, y_1) = 0.3, R_1(y_2, y_2) = 0.5,$$

and R_2 a general fuzzy relation such that

$$R_2(y_1, y_1) = 1, R_2(y_1, y_2) = R_2(y_2, y_1) = 0.7, R_2(y_2, y_2) = 1.$$

This means that $R_1 \subseteq R_2$. Let A be a fuzzy set such that $A(y_1) = 0.2$ and $A(y_2) = 0.8$. Let $\mathscr{I}$ be the Gödel implicator and $\mathscr{C}$ the minimum t-norm. We obtain for R_1 that

$$\begin{aligned}
(R_1{\downarrow_{\mathscr{I}_G}}{\uparrow_{\mathscr{T}_M}}A)(y_1) &= \min\Big\{\mathscr{I}_G(1, \max\{\min(1, 0.2), \min(0.3, 0.8)\},\\
&\qquad \mathscr{I}_G(0.3, \max\{\min(0.3, 0.2), \min(0.5, 0.8)\})\Big\}\\
&= \min\{\mathscr{I}_G(1, 0.3), \mathscr{I}_G(0.3, 0.5)\}\\
&= \min\{0.3, 1\}\\
&= 0.3,\\
(R_1{\downarrow_{\mathscr{I}_G}}{\uparrow_{\mathscr{T}_M}}A)(y_2) &= \min\Big\{\mathscr{I}_G(0.3, \max\{\min(1, 0.2), \min(0.3, 0.8)\},\\
&\qquad \mathscr{I}_G(0.5, \max\{\min(0.3, 0.2), \min(0.5, 0.8)\})\Big\}\\
&= \min\{\mathscr{I}_G(0.3, 0.3), \mathscr{I}_G(0.5, 0.5)\}\\
&= \min\{1, 1\}\\
&= 1.
\end{aligned}$$

On the other hand, for R_2, we have that

$$\begin{aligned}
(R_2{\downarrow_{\mathscr{I}_G}}{\uparrow_{\mathscr{T}_M}}A)(y_1) &= \min\Big\{\mathscr{I}_G(1, \max\{\min(1, 0.2), \min(0.7, 0.8)\},\\
&\qquad \mathscr{I}_G(0.7, \max\{\min(0.7, 0.2), \min(1, 0.8)\})\Big\}\\
&= \min\{\mathscr{I}_G(1, 0.7), \mathscr{I}_G(0.7, 0.8)\}\\
&= \min\{0.7, 1\}\\
&= 0.7,\\
(R_2{\downarrow_{\mathscr{I}_G}}{\uparrow_{\mathscr{T}_M}}A)(y_2) &= \min\Big\{\mathscr{I}_G(0.7, \max\{\min(1, 0.2), \min(0.7, 0.8)\},\\
&\qquad \mathscr{I}_G(1, \max\{\min(0.7, 0.2), \min(1, 0.8)\}))\Big\}\\
&= \min\{\mathscr{I}_G(0.7, 0.7), \mathscr{I}_G(1, 0.8)\}\\
&= \min\{1, 0.8\}\\
&= 0.8.
\end{aligned}$$

This means that

$$(R_1\downarrow_{\mathscr{I}_G}\uparrow_{\mathscr{T}_M}A)(y_1) \leq (R_2\downarrow_{\mathscr{I}_G}\uparrow_{\mathscr{T}_M}A)(y_1),$$
$$(R_2\downarrow_{\mathscr{I}_G}\uparrow_{\mathscr{T}_M}A)(y_2) \leq (R_1\downarrow_{\mathscr{I}_G}\uparrow_{\mathscr{T}_M}A)(y_2).$$

A similar counterexample can be constructed for the loose lower approximation.

We now study the tight and loose approximations of the intersection and the union.

Proposition 4.2.5. Let A and B be fuzzy sets in (U,R) with R a general fuzzy relation and $\mathscr{I}$ an implicator and $\mathscr{C}$ a conjunctor. Then it holds that

$$\begin{aligned}
R\downarrow_{\mathscr{I}}\downarrow_{\mathscr{I}}(A\cap B) &= R\downarrow_{\mathscr{I}}\downarrow_{\mathscr{I}}A\cap R\downarrow_{\mathscr{I}}\downarrow_{\mathscr{I}}B,\\
R\uparrow_{\mathscr{C}}\downarrow_{\mathscr{I}}(A\cap B) &\subseteq R\uparrow_{\mathscr{C}}\downarrow_{\mathscr{I}}A\cap R\uparrow_{\mathscr{C}}\downarrow_{\mathscr{I}}B,\\
R\downarrow_{\mathscr{I}}\uparrow_{\mathscr{C}}(A\cap B) &\subseteq R\downarrow_{\mathscr{I}}\uparrow_{\mathscr{C}}A\cap R\downarrow_{\mathscr{I}}\uparrow_{\mathscr{C}}B,\\
R\uparrow_{\mathscr{C}}\uparrow_{\mathscr{C}}(A\cap B) &\subseteq R\uparrow_{\mathscr{C}}\uparrow_{\mathscr{C}}A\cap R\uparrow_{\mathscr{C}}\uparrow_{\mathscr{C}}B,
\end{aligned}$$

$$\begin{aligned}
R\downarrow_{\mathscr{I}}\downarrow_{\mathscr{I}}(A\cup B) &\supseteq R\downarrow_{\mathscr{I}}\downarrow_{\mathscr{I}}A\cup R\downarrow_{\mathscr{I}}\downarrow_{\mathscr{I}}B,\\
R\uparrow_{\mathscr{C}}\downarrow_{\mathscr{I}}(A\cup B) &\supseteq R\uparrow_{\mathscr{C}}\downarrow_{\mathscr{I}}A\cup R\uparrow_{\mathscr{C}}\downarrow_{\mathscr{I}}B,\\
R\downarrow_{\mathscr{I}}\uparrow_{\mathscr{C}}(A\cup B) &\supseteq R\downarrow_{\mathscr{I}}\uparrow_{\mathscr{C}}A\cup R\downarrow_{\mathscr{I}}\uparrow_{\mathscr{C}}B,\\
R\uparrow_{\mathscr{C}}\uparrow_{\mathscr{C}}(A\cup B) &= R\uparrow_{\mathscr{C}}\uparrow_{\mathscr{C}}A\cup R\uparrow_{\mathscr{C}}\uparrow_{\mathscr{C}}B.
\end{aligned}$$

Proof. Let A, B be fuzzy sets in (U,R). Since Proposition 4.2.2 holds and

$$A\cap B \subseteq A, B, \subseteq A\cup B$$

the equations holds except for the first and the last one. To prove the first and the last equation, note that for $x\in U$ it holds that

$$\begin{aligned}
(R\downarrow_{\mathscr{I}}\downarrow_{\mathscr{I}}A)(x) &= \inf_{y\in U}\mathscr{I}(R(x,y),\inf_{z\in U}\mathscr{I}(R(z,y),A(z)))\\
&= \inf_{y\in U}\mathscr{I}(R(x,y),(R\downarrow A)(y)),
\end{aligned}$$

and

$$\begin{aligned}
(R\uparrow_{\mathscr{C}}\uparrow_{\mathscr{C}}A)(x) &= \sup_{y\in U}\mathscr{C}(R(x,y),\sup_{z\in U}\mathscr{C}(R(z,y),A(z)))\\
&= \sup_{y\in U}\mathscr{C}(R(x,y),(R\uparrow A)(y)).
\end{aligned}$$

The rest of the prove is similar to the proof of Proposition 4.1.8, but with $R\downarrow A$ and $R\downarrow B$ and $R\uparrow A$ and $R\uparrow B$ instead of A and B. □

Since the other properties do not hold for the usual lower and upper approximation and a general fuzzy relation, they also do not hold for the tight and loose lower and upper approximations. We now study which properties hold when we consider a fuzzy similarity relation R.

Fuzzy similarity relation

We have a useful property when R is a fuzzy similarity relation.

Proposition 4.2.6. Let A be a fuzzy set in a fuzzy approximation space (U,R) with R a fuzzy similarity relation. Let $\mathscr{I}$ be an implicator and $\mathscr{C}$ a conjunctor. We have that

$$\begin{aligned}
R{\downarrow_{\mathscr{I}}}{\downarrow_{\mathscr{I}}}A &= R{\downarrow_{\mathscr{I}}}(R{\downarrow_{\mathscr{I}}}A),\\
R{\uparrow_{\mathscr{C}}}{\downarrow_{\mathscr{I}}}A &= R{\uparrow_{\mathscr{C}}}(R{\downarrow_{\mathscr{I}}}A),\\
R{\downarrow_{\mathscr{I}}}{\uparrow_{\mathscr{C}}}A &= R{\downarrow_{\mathscr{I}}}(R{\uparrow_{\mathscr{C}}}A),\\
R{\uparrow_{\mathscr{C}}}{\uparrow_{\mathscr{C}}}A &= R{\uparrow_{\mathscr{C}}}(R{\uparrow_{\mathscr{C}}}A).
\end{aligned}$$

Proof. This follows immediately from Definitions 3.2.1 and 3.3.3 and by the fact that R is symmetric. □

This characterisation leads to very easy proofs, as we can just apply the properties studied in Section 4.1. For example, the inclusion property for tight and loose approximations follows immediately from Proposition 4.1.10.

Proposition 4.2.7. Let A be a fuzzy set in a fuzzy approximation space (U,R) with R a fuzzy similary relation. If $\mathscr{I}$ is a border implicator and $\mathscr{C}$ a conjunctor such that $\mathscr{C}(1,a)=a$, for all a in I, then we have that

$$\begin{gathered}
R{\downarrow_{\mathscr{I}}}{\downarrow_{\mathscr{I}}}A \subseteq R{\downarrow_{\mathscr{I}}}A \subseteq A \subseteq R{\uparrow_{\mathscr{C}}}A \subseteq R{\uparrow_{\mathscr{C}}}{\uparrow_{\mathscr{C}}}A,\\
R{\downarrow_{\mathscr{I}}}A \subseteq R{\uparrow_{\mathscr{C}}}{\downarrow_{\mathscr{I}}}A \subseteq R{\uparrow_{\mathscr{C}}}A,\\
R{\downarrow_{\mathscr{I}}}A \subseteq R{\downarrow_{\mathscr{I}}}{\uparrow_{\mathscr{C}}}A \subseteq R{\uparrow_{\mathscr{C}}}A.
\end{gathered}$$

Proof. This follows immediately from Propositions 4.2.6 and 4.1.10. □

The idempotence property holds.

Proposition 4.2.8. Let $\mathscr{C}$ be a left-continuous t-norm $\mathscr{T}$ and $\mathscr{I}_{\mathscr{T}}$ its R-implicator. Let A be a fuzzy set in a fuzzy approximation space (U,R) with R a fuzzy $\mathscr{T}$-similartiy relation. Then it holds that

$$\begin{gathered}
R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}{\downarrow_{\mathscr{I}_{\mathscr{T}}}}(R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A) = R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A,\\
R{\uparrow_{\mathscr{T}}}{\downarrow_{\mathscr{I}_{\mathscr{T}}}}(R{\uparrow_{\mathscr{T}}}{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A) = R{\uparrow_{\mathscr{T}}}{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A,\\
R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}{\uparrow_{\mathscr{T}}}(R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}{\uparrow_{\mathscr{T}}}A) = R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}{\uparrow_{\mathscr{T}}}A,\\
R{\uparrow_{\mathscr{T}}}{\uparrow_{\mathscr{T}}}(R{\uparrow_{\mathscr{T}}}{\uparrow_{\mathscr{T}}}A) = R{\uparrow_{\mathscr{T}}}{\uparrow_{\mathscr{T}}}A.
\end{gathered}$$

The following property for constant sets can easily be derived from Proposition 4.2.6 and Proposition 4.1.11.

Proposition 4.2.9. Let (U,R) be a fuzzy approximation space with R a fuzzy similarity relation. Let $\mathscr{I}$ be a border implicator and $\mathscr{C}$ be a conjunctor such that $\mathscr{C}(1,a)=a$ for all $a \in I$. Let $\hat{\alpha}$ be the constant α-set for $\alpha \in I$. Then it holds that

$$R{\downarrow_{\mathscr{I}}}{\downarrow_{\mathscr{I}}}\hat{\alpha} = \hat{\alpha},$$
$$R{\uparrow_{\mathscr{C}}}{\downarrow_{\mathscr{I}}}\hat{\alpha} = \hat{\alpha},$$
$$R{\downarrow_{\mathscr{I}}}{\uparrow_{\mathscr{C}}}\hat{\alpha} = \hat{\alpha},$$
$$R{\uparrow_{\mathscr{C}}}{\uparrow_{\mathscr{C}}}\hat{\alpha} = \hat{\alpha}.$$

And this holds of course for $\emptyset = \hat{0}$ and $U = \hat{1}$.

We have some extra properties that hold for this model. For example, if we have an extra connection between $\mathscr{I}$ and $\mathscr{C}$, we can say the following:

Proposition 4.2.10. Let $\mathscr{T}$ be a left-continuous t-norm and $\mathscr{I}_{\mathscr{T}}$ its R-implicator. Let A be a fuzzy set in a fuzzy approximation space (U,R) with R a $\mathscr{T}$-fuzzy similarity relation, then we have that

$$R{\uparrow_{\mathscr{T}}}{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A \subseteq R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A,$$
$$R{\uparrow_{\mathscr{T}}}A \subseteq R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}{\uparrow_{\mathscr{T}}}A.$$

Proof. Let A be a fuzzy set, R a fuzzy $\mathscr{T}$-similarity relation and $\mathscr{I}_{\mathscr{T}}$ an R-implicator based on a left-continuous t-norm $\mathscr{T}$. Recall that for $\mathscr{I}_{\mathscr{T}}$ and $\mathscr{T}$ it holds that

$$\forall a,b,c \in I: \mathscr{I}_{\mathscr{T}}(a, \mathscr{I}_{\mathscr{T}}(b,c)) = \mathscr{I}_{\mathscr{T}}(\mathscr{T}(a,b),c)$$

and that for every index set J

$$\forall a \in I, b_j \in I, j \in J: \inf_{j\in J} \mathscr{I}_{\mathscr{T}}(a,b_j) = \mathscr{I}_{\mathscr{T}}(a, \inf_{j\in J} b_j).$$

It also holds for all $a,b \in I$ that

$$\mathscr{T}(a, \mathscr{I}_{\mathscr{T}}(a,b)) \leq b \text{ and } b \leq \mathscr{I}_{\mathscr{T}}(a, \mathscr{T}(a,b)).$$

We base our proof on the proof of proposition 12 in [54]. For all $x,y \in U$, we have that

$$\begin{aligned}
(R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A)(y) &= \inf_{z\in U} \mathscr{I}_{\mathscr{T}}(R(z,y), A(z)) \\
&\leq \inf_{z\in U} \mathscr{I}_{\mathscr{T}}(\mathscr{T}(R(z,x), R(x,y)), A(z)) \\
&= \inf_{z\in U} \mathscr{I}_{\mathscr{T}}(\mathscr{T}(R(x,y), R(z,x)), A(z)) \\
&= \inf_{z\in U} \mathscr{I}_{\mathscr{T}}(R(x,y), \mathscr{I}_{\mathscr{T}}(R(z,x), A(z))) \\
&= \mathscr{I}_{\mathscr{T}}(R(x,y), \inf_{z\in U} \mathscr{I}_{\mathscr{T}}(R(z,x), A(z))) \\
&= \mathscr{I}_{\mathscr{T}}(R(x,y), (R{\downarrow_{\mathscr{I}_{\mathscr{T}}}}A)(x)).
\end{aligned}$$

Then it holds that

$$\begin{aligned}(R\uparrow_{\mathscr{T}}\downarrow_{\mathscr{I}_{\mathscr{T}}}A)(x) &= \sup_{y\in U}\mathscr{T}(R(y,x),(R\downarrow_{\mathscr{I}_{\mathscr{T}}}A)(y))\\ &\le \sup_{y\in U}\mathscr{T}(R(x,y),\mathscr{I}_{\mathscr{T}}(R(x,y),(R\downarrow_{\mathscr{I}_{\mathscr{T}}}A)(x)))\\ &\le (R\downarrow_{\mathscr{I}_{\mathscr{T}}}A)(x).\end{aligned}$$

In a similar way, we can obtain the second equation. □

In general, this property does not hold for other combinations for $(\mathscr{I},\mathscr{C})$.

Example 4.2.11. Let $U=\{y_1,y_2\}$, R a fuzzy similarity relation with $R(y_1,y_2)=0.3$ and A a fuzzy set such that $A(y_1)=1$ and $A(y_2)=0.7$. Take the minimum t-norm and the implicator based on the maximum t-conorm $\mathscr{I}_{KD}(a,b)=\max\{1-a,b\}$ for all $a,b\in I$. Then we have that $(R\uparrow_{\mathscr{T}_M}A)(y_1)=1$ and $(R\uparrow_{\mathscr{T}_M}A)(y_2)=0.7$ and thus that

$$\begin{aligned}(R\downarrow_{\mathscr{I}_{KD}}\uparrow_{\mathscr{T}_M}A)(y_1) &= (R\downarrow_{\mathscr{I}_{KD}}(R\uparrow_{\mathscr{T}_M}A))(y_1)\\ &= \min_{y\in U}\max\{1-R(y_1,y),R\uparrow_{\mathscr{T}_M}A(y)\}\\ &= \min\{\max\{0,1\},\max\{0.7,0.7\}\}\\ &= 0.7\\ &< (R\uparrow_{\mathscr{T}_M}A)(y_1).\end{aligned}$$

Due to Proposition 4.2.10, we have the following for a left-continuous t-norm $\mathscr{T}$ and its R-implicator:

Proposition 4.2.12. Let A be a fuzzy set in a fuzzy approximation space (U,R) with R a fuzzy similarity relation. Let $\mathscr{C}$ be a left-continuous t-norm $\mathscr{T}$ and $\mathscr{I}$ the R-implicator based on $\mathscr{T}$. We have that

$$R\downarrow_{\mathscr{I}}\downarrow_{\mathscr{I}}A=R\downarrow_{\mathscr{I}}A=R\uparrow_{\mathscr{T}}\downarrow_{\mathscr{I}}A\subseteq A\subseteq R\downarrow_{\mathscr{I}}\uparrow_{\mathscr{T}}A=R\uparrow_{\mathscr{T}}A=R\uparrow_{\mathscr{T}}\uparrow_{\mathscr{T}}A.$$

Proof. This follows from Propositions 4.1.13, 4.2.6, 4.2.7 and 4.2.10. □

Again we see that under certain conditions, all properties of Table 2.1 hold for the tight and loose approximation operators, except for the monotonicity of relations. This property only holds for the tight lower and the loose upper approximation operator.

We continue to examine models that deal with noisy data and that were discussed in Chapter 3.

4.3 Fuzzy rough set models designed to deal with noisy data

We shall not discuss all the robust models from Section 3.4. As the soft fuzzy rough set model is ill-defined, we do not discuss its properties. The fuzzy rough set model based on the robust nearest neighbor is a special case of the OWA-based fuzzy rough set model and is not discussed separately. The variable precision fuzzy rough set model will also not be discussed. Further studies are necessary to fully understand this model.

We begin with the β-precision fuzzy rough set model.

4.3.1 β-precision fuzzy rough sets

We study the β-precision fuzzy rough set model, given in Definition 3.4.1. We again make the distinction between general fuzzy relations and fuzzy similarity relations.

General fuzzy relation

We start by studying which properties hold when R is a general fuzzy relation. The duality property holds, if $\mathcal{T}_\beta$ and $\mathcal{S}_\beta$ are dual w.r.t. the standard negator $\mathcal{N}_S$ and if $\mathcal{I}$ and $\mathcal{C}$ are dual w.r.t. $\mathcal{N}_S$.

Proposition 4.3.1. Let $\mathcal{N}_S$ be the standard negator and A a fuzzy set in a fuzzy approximation space (U,R) with R a general fuzzy relation. Let $\mathcal{T}$ be a t-norm and $\mathcal{S}$ its dual t-conorm w.r.t. $\mathcal{N}_S$ and $\beta \in I$. If the pair $(\mathcal{I},\mathcal{C})$ consists of an implicator $\mathcal{I}$ and a conjunctor $\mathcal{C}$ defined by the dual coimplicator $\mathcal{J}$ of $\mathcal{I}$ w.r.t. $\mathcal{N}_S$, then the duality property holds, i.e.,

$$R{\downarrow}_{\mathcal{I},\mathcal{T}_\beta}A = \mathrm{co}_{\mathcal{N}_S}(R{\uparrow}_{\mathcal{C},\mathcal{S}_\beta}(\mathrm{co}_{\mathcal{N}_S}(A))),$$
$$R{\uparrow}_{\mathcal{C},\mathcal{S}_\beta}A = \mathrm{co}_{\mathcal{N}_S}(R{\downarrow}_{\mathcal{I},\mathcal{T}_\beta}(\mathrm{co}_{\mathcal{N}_S}(A))).$$

Proof. We only need to prove that $\mathcal{T}_\beta$ and $\mathcal{S}_\beta$ are also dual w.r.t. $\mathcal{N}_S$, because if this holds, then the rest of the proof is completely similar to the proof of Proposition 4.1.1. Let us take $(a_1,\ldots a_n) \in I^n$ and σ the permutation on $\{1,\ldots,n\}$ such that $a_{\sigma(i)}$ is the i^{th} biggest element of $(a_1,\ldots a_n)$. Let $m \in \mathbb{N}$ be such that

$$m = \max\left\{ j \in \{0,\ldots,n\} \mid j \leq (1-\beta)\cdot\sum_{i=1}^{n} a_i \right\}.$$

Now, since

$$j \leq (1-\beta)\cdot\sum_{i=1}^{n} a_i \Leftrightarrow j \leq (1-\beta)\cdot\sum_{i=1}^{n} 1-(1-a_i),$$

we omit m values to calculate $\mathcal{T}_\beta$ and we omit m values to calculate $\mathcal{S}_\beta$. Hence,

$$\begin{aligned}
\mathcal{N}_S(\mathcal{T}_\beta(a_1,\ldots,a_n)) &= \mathcal{N}_S\left(\mathcal{T}(a_{\sigma(1)},\ldots,a_{\sigma(n-m)})\right)\\
&= \mathcal{S}(1-a_{\sigma(1)},\ldots,1-a_{\sigma(n-m)})\\
&= \mathcal{S}_\beta(1-a_1,\ldots,1-a_n)\\
&= \mathcal{S}_\beta(\mathcal{N}_S(a_1),\ldots,\mathcal{N}_S(a_n)).
\end{aligned}$$

In a similar way we obtain that

$$\mathscr{N}_S(\mathscr{S}_\beta(a_1,\ldots,a_n)) = \mathscr{T}_\beta(\mathscr{N}_S(a_1),\ldots,\mathscr{N}_S(a_n)).$$

□

Proposition 4.3.1 also holds for an S-implactor $\mathscr{I}$ based on $\mathscr{S}$ and the dual t-norm $\mathscr{T}$ w.r.t. the standard negator $\mathscr{N}_S$ and for a left-continuous t-norm $\mathscr{T}$ and its R-implicator $\mathscr{I}_\mathscr{T}$ if $\mathscr{N}_{\mathscr{I}_\mathscr{T}} = \mathscr{N}_S$. This latter holds, for example, for the couple $(\mathscr{T}_{nM}, \mathscr{I}_{nM})$.

Remark 4.3.2. This property only holds if $\mathscr{N}$ is the standard negator, otherwise it does not hold that $\mathscr{T}_\beta$ and $\mathscr{S}_\beta$ are dual to each other, since we do not necessarily omit the same amount of values.

The monotonicity properties hold due to the monotonicity properties of implicators and conjunctors.

Proposition 4.3.3. Let A and B be fuzzy sets in (U,R) with R a general fuzzy relation. Let $\mathscr{T}$ be a t-norm, $\mathscr{S}$ a t-conorm, $\mathscr{I}$ an implicator, $\mathscr{C}$ a conjunctor and $\beta \in I$. If $A \subseteq B$, then we have that

$$R{\downarrow}_{\mathscr{I},\mathscr{T}_\beta} A \subseteq R{\downarrow}_{\mathscr{I},\mathscr{T}_\beta} B,$$
$$R{\uparrow}_{\mathscr{C},\mathscr{S}_\beta} A \subseteq R{\uparrow}_{\mathscr{C},\mathscr{T}_\beta} B.$$

Proof. This follows from that fact that both an implicator and a conjunctor are non-decreasing in the second argument and from the fact that if we have $\mathbf{a},\mathbf{b} \in I^n$ such that $\mathbf{a} \leq \mathbf{b}$, i.e.,

$$\forall i \in \{1,\ldots,n\} \colon a_i \leq b_i$$

then for the permutation σ on $\{1,\ldots,n\}$ such that $a_{\sigma(i)}$ is the i^{th} biggest element of $\mathbf{a}$ we have

$$\forall i \in \{1,\ldots,n\} \colon a_{\sigma(i)} \leq b_{\sigma(i)}.$$

Note also that if $\mathbf{a} \leq \mathbf{b}$, then

$$\max\left\{\forall j \in \{0,\ldots,n\} \mid j \leq (1-\beta)\cdot\sum_{i=1}^{n} a_i\right\} \leq \max\left\{\forall j \in \{0,\ldots,n\} \mid j \leq (1-\beta)\cdot\sum_{i=1}^{n} b_i\right\}.$$

□

Proposition 4.3.4. Let R_1 and R_2 be fuzzy relations on U, and A a fuzzy set in U. Let $\mathscr{T}$ be a t-norm, $\mathscr{S}$ a t-conorm, $\mathscr{I}$ an implicator, $\mathscr{C}$ a conjunctor and $\beta \in I$. If $R_1 \subseteq R_2$, then we have that

$$R_2{\downarrow}_{\mathscr{I},\mathscr{T}_\beta} A \subseteq R_1{\downarrow}_{\mathscr{I},\mathscr{T}_\beta} A,$$
$$R_1{\uparrow}_{\mathscr{C},\mathscr{S}_\beta} A \subseteq R_2{\uparrow}_{\mathscr{C},\mathscr{S}_\beta} A.$$

Proof. This follows from the fact that an implicator is non-increasing and a conjunctor is non-decreasing in the first argument and from the facts we stated in the proof of Proposition 4.3.3. □

For the intersection and the union property we have the inclusions that follow from the monotonicity of sets, but we do not have equalities.

Proposition 4.3.5. Let A and B be fuzzy sets in (U,R) with R a general fuzzy relation. Let $\mathscr{T}$ be a t-norm, $\mathscr{S}$ a t-conorm, $\mathscr{I}$ an implicator, $\mathscr{C}$ a conjunctor and $\beta \in I$. We have that

$$\begin{aligned}
&R{\downarrow}_{\mathscr{I},\mathscr{T}_\beta}(A\cap B) \subseteq R{\downarrow}_{\mathscr{I},\mathscr{T}_\beta}A \cap R{\downarrow}_{\mathscr{I},\mathscr{T}_\beta}B,\\
&R{\uparrow}_{\mathscr{C},\mathscr{S}_\beta}(A\cap B) \subseteq R{\uparrow}_{\mathscr{C},\mathscr{S}_\beta}A \cap R{\uparrow}_{\mathscr{C},\mathscr{S}_\beta}B,\\
&R{\downarrow}_{\mathscr{I},\mathscr{T}_\beta}(A\cup B) \supseteq R{\downarrow}_{\mathscr{I},\mathscr{T}_\beta}A \cup R{\downarrow}_{\mathscr{I},\mathscr{T}_\beta}B,\\
&R{\uparrow}_{\mathscr{C},\mathscr{S}_\beta}(A\cup B) \supseteq R{\uparrow}_{\mathscr{C},\mathscr{S}_\beta}A \cup R{\uparrow}_{\mathscr{C},\mathscr{S}_\beta}B.
\end{aligned}$$

We give a counterexample which illustrates that the first and last equation are not necessarily equalities.

Example 4.3.6. Let $U = \{y_0,\dots,y_{10}\}$ and R a fuzzy similarity relation such that $R(x,z) = 1$ for all $x,z \in U$. Let $\mathscr{T}$ be the minimum t-norm and let β be 0.8. Let $\mathscr{I}$ be a border implicator. Let us consider the following fuzzy sets A and B:

$$\begin{aligned}
&A(y_0) = 1, B(y_0) = 0,\\
&A(y_i) = 1, B(y_i) = \frac{i}{10}, \text{ for } i \text{ even}, i \neq 0,\\
&A(y_i) = \frac{i}{10}, B(y_i) = 1, \text{ for } i \text{ odd}, i \neq 0.
\end{aligned}$$

Then for all $i \in \{0,\dots,10\}$ we have that $(A\cap B)(y_i) = \frac{i}{10}$. We have for $x \in U$ that

$$\begin{aligned}
(R{\downarrow}_{\mathscr{I},\min_{0.8}}A)(x) &= \min_{\substack{0.8\\ y\in U}} \mathscr{I}(1, A(y))\\
&= \min_{0.8}\{1,1,1,1,1,1,0.9,0.7,0.5,0.3,0.1\}\\
&= 0.3,
\end{aligned}$$

since $(1-0.8)\cdot \sum_{i=0}^{10} A(y_i) = 0.2 \cdot 8.5 = 1.7$. For B we have that

$$\begin{aligned}
(R{\downarrow}_{\mathscr{I},\min_{0.8}}B)(x) &= \min_{\substack{0.8\\ y\in U}} \mathscr{I}(1, B(y))\\
&= \min_{0.8}\{1,1,1,1,1,1,0.8,0.6,0.4,0.2,0\}\\
&= 0.2,
\end{aligned}$$

since $(1-0.8)\cdot\sum_{i=0}^{10} B(y_i) = 0.2\cdot 8 = 1.6$. On the other hand, we have that

$$\begin{aligned}(R\downarrow_{\mathscr{I},\min_{0.8}}(A\cap B))(x) &= \min_{\substack{0.8\\ y\in U}} \mathscr{I}(1,(A\cap B)(y))\\ &= \min_{0.8}\{1,0.9,0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1,0\}\\ &= 0.1,\end{aligned}$$

since $(1-0.8)\cdot\sum_{i=0}^{10}\frac{i}{10} = 0.2\cdot 5.5 = 1.1$. This means that for all $x\in U$

$$(R\downarrow_{\mathscr{I},\min_{0.8}}(A\cap B))(x) < (R\downarrow_{\mathscr{I},\min_{0.8}}A)(x)\cap(R\downarrow_{\mathscr{I},\min_{0.8}}B)(x).$$

A similar counterexample can be constructed to prove that

$$(R\uparrow_{\mathscr{C},\mathscr{S}_\beta}(A\cup B))\subseteq(R\uparrow_{\mathscr{C},\mathscr{S}_\beta}A)\cup(R\uparrow_{\mathscr{C},\mathscr{S}_\beta}B)$$

not always holds.

The other properties do not hold for general fuzzy relations. We study which properties require a fuzzy similarity relation.

Fuzzy similarity relation

In contrast to the previous two models, the inclusion does not hold, even when R is a fuzzy similarity relation.

Example 4.3.7. Let $\mathscr{I}$ be a border implicator and $\mathscr{C}$ a conjunctor such that $\mathscr{C}(1,a)=a$ for all $a\in I$. Let $U=\{y_0,\ldots,y_{10}\}$, A a fuzzy set such that $A(y_i)=\frac{i}{10}$ for all $i\in\{0,\ldots,10\}$ and R a fuzzy similarity relation with $R(y_i,y_j)=1$ for all $i,j\in\{0,\ldots,10\}$. Let $(\mathscr{T}_\beta,\mathscr{S}_\beta)$ be $(\min_\beta,\max_\beta)$ and $\beta=0.8$.

As $R(y_i,y_j)=1$, we have that $\mathscr{I}(R(z,x),A(z))=A(z)=\mathscr{C}(R(z,x),A(z))$ for all $x,z\in U$. We also have that $\sum_{i=0}^{10}A(y_i)=5.5$. As $\beta=0.8$, $1\le 5.5\cdot 0.2=1.1$, so in the lower approximation the lowest value will be omited. We obtain for the lower approximation of A in $x\in U$ that

$$\begin{aligned}\min_{\substack{0.8\\ y\in U}}\mathscr{I}(R(y,x),A(y)) &= \min_{\substack{0.8\\ y\in U}}A(y)\\ &= \min_{0.8}\{1,0.9,\ldots,0.1,0\}\\ &= 0.1,\end{aligned}$$

which means that $(R{\downarrow}_{\mathscr{I},\min_{0.8}}A)(y_0) > A(y_0)$. Since $1 \leq (11-5.5)\cdot 0.2 = 1.1$, we obtain for $x \in U$ that

$$\begin{aligned}\max_{\substack{0.8\\y\in U}} \mathscr{C}(R(y,x),A(y)) &= \max_{\substack{0.8\\y\in U}} A(y)\\ &= \max_{0.8}\{0,0.1,\ldots,0.9,1\}\\ &= 0.9,\end{aligned}$$

and so $(R{\uparrow}_{\mathscr{C},\max_{0.8}}A)(y_{10}) < A(y_{10})$.

The constant α-set property does also not hold.

Example 4.3.8. Let $U = \{y_1,\ldots,y_{100}\}$ and R a fuzzy similarity relation such that $R(x,x) = 1$ and $R(x,z) = 0.5$ for $x \neq z \in U$. Let $\mathscr{I}$ be the Łukasiewicz implicator $\mathscr{I}_L$ and let β be 0.95. Let $\mathscr{T}$ be the minimum t-norm For $x \in U$ we obtain that

$$\begin{aligned}(R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}\emptyset)(x) &= \min_{\substack{0.95\\y\in U}} \mathscr{I}_L(R(y,x),0)\\ &= 0.5\\ &> 0,\end{aligned}$$

since $0.05\cdot(99\cdot 0.5 + 1\cdot 0) = 0.05\cdot 49.5 = 2.475$, which means we omit one 0 and one 0.5 in the second step. This means that $(R{\downarrow}_{\mathscr{I},\mathscr{T}_\beta}\hat{0}) \neq \hat{0}$. Now, let $\mathscr{C}$ be any t-norm $\mathscr{T}$ and $\mathscr{S}$ the maximum operator. For $x \in U$ we obtain that

$$\begin{aligned}(R{\uparrow}_{\mathscr{T},\max_{0.95}}U)(x) &= \max_{\substack{0.95\\y\in U}} \mathscr{T}(R(y,x),1)\\ &= 0.5\\ &< 1,\end{aligned}$$

since $0.05\cdot(99\cdot 0.5 + 1\cdot 1) = 0.05\cdot 50.5 = 2.525$, which means we omit one 1 and one 0.5 in the second step. This means that $(R{\uparrow}_{\mathscr{C},\mathscr{S}_\beta}\hat{1}) \neq \hat{1}$.

Note that we always have $R{\uparrow}_{\mathscr{C},\mathscr{S}_\beta}\emptyset = \emptyset$ and $R{\downarrow}_{\mathscr{I},\mathscr{T}_\beta}U = U$. The last property we study is the idempotence property: this property does not hold in general.

Example 4.3.9. If we take the same setting as in the previous example, we have obtained for every $x \in U$ that $(R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}\emptyset)(x) = 0.5$. Note that the R we use is a fuzzy similiarty relation and $\mathscr{I}_L$ is the R-implicator of $\mathscr{T}_L$. We now compute $(R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}(R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}\emptyset))(x)$:

$$\begin{aligned}(R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}(R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}\emptyset))(x) &= \min_{\substack{0.95\\y\in U}} \mathscr{I}_L(R(y,x),(R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}\emptyset)(y))\\ &= \min_{\substack{0.95\\y\in U}} \min\{1, 1-R(y,x)+0.5\}.\end{aligned}$$

Since $0.05 \cdot (99 \cdot 1 + 1 \cdot 0.5) = 4.975$, we omit one 0.5 and three 1's. This means that for every $x \in U$: $(R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}(R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}\emptyset))(x) = 1$ and thus

$$\forall x \in U: (R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}(R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}\emptyset))(x) > (R{\downarrow}_{\mathscr{I}_L,\min_{0.95}}\emptyset)(x).$$

We also derived that for every t-norm $\mathscr{T}$, the upper approximation of U based on $\mathscr{T}$ and $\max_\beta$ was 0.5. Now take $\mathscr{T}$ the product t-norm $\mathscr{T}_P$ and $x \in U$. We obtain

$$\begin{aligned}(R{\uparrow}_{\mathscr{T}_P,\max_{0.95}}(R{\uparrow}_{\mathscr{T}_P,\max_{0.95}}U))(x) &= \max_{\substack{0.95\\ y\in U}} \mathscr{T}_P(R(y,x),(R{\uparrow}_{\mathscr{T}_P,\max_{0.95}}U)(y))\\ &= \max_{\substack{0.95\\ y\in U}} R(y,x)\cdot 0.5.\end{aligned}$$

Because $0.05 \cdot (99 \cdot 0.25 + 1 \cdot 0.5) = 1.2625$, we omit the 0.5 and so

$$(R{\uparrow}_{\mathscr{T}_P,\max_{0.95}}(R{\uparrow}_{\mathscr{T}_P,\max_{0.95}}U))(x) = 0.25,$$

which is strictly smaller than $(R{\uparrow}_{\mathscr{T}_P,\max_{0.95}}U)(x)$.

In contrast to the previous to models, some properties no longer holds. This is a price we have to pay for having a more robust model. We continue with the vaguely quantified fuzzy rough set model.

4.3.2 Vaguely quantified fuzzy rough sets

We study the model given in Definition 3.4.16. We saw earlier that the asymmetric VPRS model (Definition 2.1.12) can be derived from the VQFRS model. So, if a property does not hold in the asymmetric VPRS model, it will not hold in the VQFRS model, because a counterexample for the VPRS model is also a counterexample for the more general VQFRS model. This immediately gives us that the properties of 'Duality', 'Inclusion', 'Monotonicity of relations' and 'Idempotence' do not hold in the VQFRS model. We study the other properties.

The monotonicity of sets holds for a general fuzzy relation R.

Proposition 4.3.10. Let A and B be fuzzy sets in a fuzzy approximation space (U,R) with R a general fuzzy relation and Q_u and Q_l regularly increasing quantifiers. If $A \subseteq B$, then it holds that

$$\begin{aligned}R{\downarrow}_{Q_u}A &\subseteq R{\downarrow}_{Q_u}B,\\ R{\uparrow}_{Q_l}A &\subseteq R{\uparrow}_{Q_l}B.\end{aligned}$$

Proof. If $A \subseteq B$, then for all $x \in U$ it holds that

$$\frac{|Rx \cap A|}{|Rx|} \leq \frac{|Rx \cap B|}{|Rx|},$$

if Rx is not empty. The property follows from the fact that Q_u and Q_l are increasing. If Rx is empty, then we have that

$$R{\downarrow}_{Q_u}A = R{\downarrow}_{Q_u}B = 1$$

and

$$R{\uparrow}_{Q_l}A = R{\uparrow}_{Q_l}B = 1.$$

□

Because this property holds, we have the following for the 'Intersection' and 'Union' property.

Proposition 4.3.11. Let A and B be fuzzy sets in a fuzzy approximation space (U, R) with R a general fuzzy relation and Q_u and Q_l regularly increasing quantifiers. It holds that

$$\begin{aligned}
R{\downarrow}_{Q_u}(A \cap B) &\subseteq R{\downarrow}_{Q_u}A \cap R{\downarrow}_{Q_u}B,\\
R{\uparrow}_{Q_l}(A \cap B) &\subseteq R{\uparrow}_{Q_l}A \cap R{\uparrow}_{Q_l}B,\\
R{\downarrow}_{Q_u}(A \cup B) &\supseteq R{\downarrow}_{Q_u}A \cup R{\downarrow}_{Q_u}B,\\
R{\uparrow}_{Q_l}(A \cup B) &\supseteq R{\uparrow}_{Q_l}A \cup R{\uparrow}_{Q_l}B.
\end{aligned}$$

Other inclusions do not hold, since they also do not hold in the VPRS model.

For a fuzzy similarity relation R, we have that the constant set property holds for $\emptyset$ and U, but not for other α's.

Proposition 4.3.12. Let R be a fuzzy similarity relation and Q_u and Q_l regularly increasing quantifiers. We have that

$$\begin{aligned}
R{\downarrow}_{Q_u}\emptyset &= \emptyset = R{\uparrow}_{Q_l}\emptyset,\\
R{\downarrow}_{Q_u}U &= U = R{\uparrow}_{Q_l}U.
\end{aligned}$$

Proof. Since $x \in Rx$, we have for all $x \in U$ that

$$\frac{|Rx \cap \emptyset|}{|Rx|} = 0$$

and

$$\frac{|Rx \cap U|}{|Rx|} = 1.$$

The property follows from the fact that Q_u and Q_l are regularly increasing quantifiers, and this means that $Q_u(0) = Q_l(0) = 0$ and $Q_u(1) = Q_l(1) = 1$. □

The property for U also holds for general fuzzy relation R and it holds for $\emptyset$ if the relation R is serial. We illustrate that it not necessarily holds for $\alpha \in]0, 1[$.

Example 4.3.13. Let $U = \{y_1, y_2, y_3\}$ and let R be a fuzzy similarity relation such that $R(y_i, y_j) = 1$ for $i, j \in \{1,2,3\}$. Take for the couple (Q_u, Q_l) the quantifiers for 'Most' and 'Some' as defined in Section 3.4.3, i.e., $(Q_m, Q_s) = (Q_{(0.2,1)}, Q_{(0.1,0.6)})$ and take $\alpha = 0.1$. We derive that

$$\begin{aligned}(R{\downarrow}_{Q_m}\hat{\alpha})(y_1) &= Q_m\left(\frac{|Ry_1 \cap \hat{\alpha}|}{|Ry_1|}\right)\\ &= Q_m\left(\frac{0.1+0.1+0.1}{3}\right)\\ &= Q_m(0.1)\\ &= 0\end{aligned}$$

which is strictly smaller than $\alpha = 0.1$. Similarly, we derive that

$$R{\uparrow}_{Q_s}\hat{\alpha}(y_1) = Q_s(0.1) = 0.$$

Again, not all the properties hold. Due to the fact that the monotonicity of relations not hold, this model will not be interesting to use in feature selection. The following model we study is the fuzzy variable precision rough set model.

4.3.3 Fuzzy variable precision rough sets

The FVPRS model, given in Definition 3.4.20, is similar to the general fuzzy rough set model, only the second argument of the implicator and conjunctor are different. Recall that

$$\begin{aligned}R{\downarrow}_{\mathscr{I},\alpha}A &= R{\downarrow}_{\mathscr{I}}(A \cup \hat{\alpha}),\\ R{\uparrow}_{\mathscr{C},\alpha}A &= R{\uparrow}_{\mathscr{C}}(A \cap \widehat{1-\alpha}),\end{aligned}$$

for every fuzzy set A, every $\alpha \in I$ and every choice of the pair $(\mathscr{I}, \mathscr{C})$. We shall see that most properties hold in this model and the proofs are similar to the ones in Section 4.1.

General fuzzy relation

We start with the properties that hold for a general fuzzy relation R. We begin with the duality property.

Proposition 4.3.14. Let $\mathscr{N}$ be an involutive negator and A a fuzzy set in a fuzzy approximation space (U, R) with R a general fuzzy relation. If the pair $(\mathscr{I}, \mathscr{C})$ consists of an implicator $\mathscr{I}$ and a conjunctor $\mathscr{C}$ defined by the dual coimplicator $\mathscr{J}$ of $\mathscr{I}$ w.r.t. $\mathscr{N}$, then the duality property holds, i.e., for every $\alpha \in I$ it holds that

$$\begin{aligned}R{\downarrow}_{\mathscr{I},\alpha}A &= \mathrm{co}_{\mathscr{N}}(R{\uparrow}_{\mathscr{C},\alpha}(\mathrm{co}_{\mathscr{N}}(A))),\\ R{\uparrow}_{\mathscr{C},\alpha}A &= \mathrm{co}_{\mathscr{N}}(R{\downarrow}_{\mathscr{I},\alpha}(\mathrm{co}_{\mathscr{N}}(A))).\end{aligned}$$

Proof. This is completely similar to the proof of Proposition 4.1.1, as we have that

$$\mathscr{N}(\min\{\mathscr{N}(\alpha),\mathscr{N}(A(y))\}) = \max\{\alpha, A(y)\}$$

and

$$\mathscr{N}(\max\{\alpha,\mathscr{N}(A(y))\}) = \min\{\mathscr{N}(\alpha), A(y)\}$$

for all involutive negators $\mathscr{N}$, all $\alpha \in I$ and all $A \in \mathscr{F}(U)$. □

This property also holds if we have an S-implicator $\mathscr{I}$ based on a t-conorm $\mathscr{S}$ and a t-norm $\mathscr{T}$ which is dual to $\mathscr{S}$ with respect to the involutive negator $\mathscr{N}$ and if we have a left-continuous t-norm $\mathscr{T}$ and its R-implicator $\mathscr{I}_{\mathscr{T}}$ such that $\mathscr{N} = \mathscr{N}_{\mathscr{I}_{\mathscr{T}}}$ is involutive.

Completely similar with the general fuzzy rough set model, the monotonicity properties hold, just as the properties 'Intersection' and 'Union'.

Proposition 4.3.15. Let A and B be fuzzy sets in (U,R) with R a general fuzzy relation and $\alpha \in I$. If $A \subseteq B$, then we have that

$$R{\downarrow}_{\mathscr{I},\alpha}A \subseteq R{\downarrow}_{\mathscr{I},\alpha}B,$$
$$R{\uparrow}_{\mathscr{C},\alpha}A \subseteq R{\uparrow}_{\mathscr{C},\alpha}B.$$

Proposition 4.3.16. Let R_1 and R_2 be fuzzy relations on U, A a fuzzy set in U and $\alpha \in I$. If $R_1 \subseteq R_2$, then we have that

$$R_2{\downarrow}_{\mathscr{I},\alpha}A \subseteq R_1{\downarrow}_{\mathscr{I},\alpha}A,$$
$$R_1{\uparrow}_{\mathscr{C},\alpha}A \subseteq R_2{\uparrow}_{\mathscr{C},\alpha}A.$$

Proposition 4.3.17. Let A and B be fuzzy sets in (U,R) with R a general fuzzy relation and $\alpha \in I$. We have that

$$R{\downarrow}_{\mathscr{I},\alpha}(A\cap B) = R{\downarrow}_{\mathscr{I},\alpha}A \cap R{\downarrow}_{\mathscr{I},\alpha}B,$$
$$R{\uparrow}_{\mathscr{C},\alpha}(A\cap B) \subseteq R{\uparrow}_{\mathscr{C},\alpha}A \cap R{\uparrow}_{\mathscr{C},\alpha}B,$$
$$R{\downarrow}_{\mathscr{I},\alpha}(A\cup B) \supseteq R{\downarrow}_{\mathscr{I},\alpha}A \cup R{\downarrow}_{\mathscr{I},\alpha}B,$$
$$R{\uparrow}_{\mathscr{C},\alpha}(A\cup B) = R{\uparrow}_{\mathscr{C},\alpha}A \cup R{\uparrow}_{\mathscr{C},\alpha}B.$$

The other properties do not hold for general fuzzy relations.

Fuzzy similarity relation

When R is a fuzzy $\mathscr{T}$-similarity relation based on a left-continuous t-norm $\mathscr{T}$, we also have the 'Idempotence' property.

Proposition 4.3.18. If $\mathscr{C}$ is a left-continuous t-norm $\mathscr{T}$, $\mathscr{I}_{\mathscr{T}}$ its R-implicator and R a fuzzy $\mathscr{T}$-similarity relation, then we have for A a fuzzy set in a fuzzy approximation space (U,R) and for all $\alpha \in I$ that

$$R{\downarrow}_{\mathscr{I}_{\mathscr{T}},\alpha}(R{\downarrow}_{\mathscr{I}_{\mathscr{T}},\alpha}A) = R{\downarrow}_{\mathscr{I}_{\mathscr{T}},\alpha}A,$$
$$R{\uparrow}_{\mathscr{T},\alpha}(R{\uparrow}_{\mathscr{T},\alpha}A) = R{\uparrow}_{\mathscr{T},\alpha}A.$$

Proof. Again, this proof is similar to that of proposition 4.1.13. □

This property holds for relations which are reflexive and $\mathcal{T}$-transitive. The inclusion property and the relation

$$R{\downarrow}_{\mathcal{I},\alpha}\hat{\beta} = \hat{\beta} = R{\uparrow}_{\mathcal{C},\alpha}\hat{\beta},$$

for all $\alpha, \beta \in I$, do not hold, not even when R is a similarity relation. We illustrate this

Example 4.3.19. Let $U = \{y_1, y_2, y_3\}$ and let R be a fuzzy similarity relation such that $R(y_i, y_j) = 1$ for $i, j \in \{1, 2, 3\}$. Let $\hat{\beta}$ be a fuzzy set with $\beta = 0.6$ and let $\alpha = 0.7$. We take the standard negator $\mathcal{N}_S$, the Łukasiewicz implicator $\mathcal{I}_L$ and the Łukasiewicz t-norm $\mathcal{T}_L$. We obtain for $x \in U$ that

$$\begin{aligned}
(R{\downarrow}_{\mathcal{I}_L,0.7}\hat{\beta})(x) &= \inf_{y\in U} \mathcal{I}_L(R(y,x), \max\{\alpha, \hat{\beta}(y)\}) \\
&= \inf_{y\in U} \min\{1, 1-1+\max\{0.7, 0.6\}\} \\
&= 0.7 \\
&> \hat{\beta}(x), \\
(R{\uparrow}_{\mathcal{T}_L,0.7}\hat{\beta})(x) &= \sup_{y\in U} \mathcal{T}_L(R(y,x), \min\{1-\alpha, \hat{\beta}(y)\}) \\
&= \sup_{y\in U} \max\{0, 1+\min\{0.3, 0.6\}-1\} \\
&= 0.3 \\
&< \hat{\beta}(x).
\end{aligned}$$

This gives a counterexample for the both the inclusion property and the constant α-set property. Note that these properties do also not hold for $\emptyset$ and U.

Remark 4.3.20. We can prove that $R{\downarrow}_{\mathcal{I},\alpha}\hat{\beta} = \hat{\beta}$ holds if $\alpha \leq \beta$ and that $R{\uparrow}_{\mathcal{C},\alpha}\hat{\beta} = \hat{\beta}$ holds if $\beta \leq 1-\alpha$.

We see that indeed a lot of properties hold. Still, the fact that the inclusion property does not hold is a problem, since lower approximations can be bigger than the set itself. The last model we study, is the OWA-based fuzzy rough set model.

4.3.4 OWA-based fuzzy rough sets

The last model we study is the OWA-based fuzzy rough set model, given in Definition 3.4.30. As in the VQFRS model, not all properties will hold, but the main advantage of this model compared to the VQFRS model is that monotonicity of relations does hold.

We start with the monotonicity properties.

Proposition 4.3.21. Let A and B be fuzzy sets in (U,R) with R a general fuzzy relation. Let $\mathscr{I}$ be an implicator, $\mathscr{C}$ a conjunctor and W_1 and W_2 weightvectors such that there length is equal to $|U|$, then we have that

$$R{\downarrow}_{\mathscr{I},W_1}A \subseteq R{\downarrow}_{\mathscr{I},W_1}B,$$
$$R{\uparrow}_{\mathscr{C},W_2}A \subseteq R{\uparrow}_{\mathscr{C},W_2}B.$$

Proof. Let U have n elements. If we have $\mathbf{a},\mathbf{b} \in I^n$ such that $\mathbf{a} \leq \mathbf{b}$, i.e.,

$$\forall i \in \{1,\ldots,n\}\colon a_i \leq b_i,$$

then for the permutation σ on $\{1,\ldots,n\}$ such that $a_{\sigma(i)}$ is the i^{th} biggest element of $\mathbf{a}$ we have that

$$\forall i \in \{1,\ldots,n\}\colon a_{\sigma(i)} \leq b_{\sigma(i)}$$

and thus also that

$$\sum_{i=1}^{n} a_{\sigma(i)} \cdot w_i \leq \sum_{i=1}^{n} b_{\sigma(i)} \cdot w_i,$$

since $w_i > 0$ for all i. The property follows from that fact that both an implicator and a conjunctor are non-decreasing in the second argument. □

Proposition 4.3.22. Let R_1 and R_2 be fuzzy relations on U, and A a fuzzy set in U. Let $\mathscr{I}$ be an implicator, $\mathscr{C}$ a conjunctor and W_1 and W_2 weightvectors such that there length is equal to $|U|$. If $R_1 \subseteq R_2$, then we have that

$$R_2{\downarrow}_{\mathscr{I},W_1}A \subseteq R_1{\downarrow}_{\mathscr{I},W_1}A,$$
$$R_1{\uparrow}_{\mathscr{C},W_2}A \subseteq R_2{\uparrow}_{\mathscr{C},W_2}A.$$

Proof. This follows from that fact that an implicator is non-increasing and a conjunctor is non-decreasing in the first argument and from the facts we stated in the proof of Proposition 4.3.21. □

Since the monotonicity of sets holds, we have the following inclusions.

Proposition 4.3.23. Let A and B be fuzzy sets in (U,R) with R a general fuzzy relation. Let $\mathscr{I}$ be an implicator, $\mathscr{C}$ a conjunctor and W_1 and W_2 weightvectors such that there length is equal to $|U|$. We have that

$$R{\downarrow}_{\mathscr{I},W_1}(A \cap B) \subseteq R{\downarrow}_{\mathscr{I},W_1}A \cap R{\downarrow}_{\mathscr{I},W_1}B,$$
$$R{\uparrow}_{\mathscr{C},W_2}(A \cap B) \subseteq R{\uparrow}_{\mathscr{C},W_2}A \cap R{\uparrow}_{\mathscr{C},W_2}B,$$
$$R{\downarrow}_{\mathscr{I},W_1}(A \cup B) \supseteq R{\downarrow}_{\mathscr{I},W_1}A \cup R{\downarrow}_{\mathscr{I},W_1}B,$$
$$R{\uparrow}_{\mathscr{C},W_2}(A \cup B) \supseteq R{\uparrow}_{\mathscr{C},W_2}A \cup R{\uparrow}_{\mathscr{C},W_2}B.$$

In the next example, we show that the other inclusions do not hold.

Example 4.3.24. Let $U = \{y_0, \ldots, y_{10}\}$ and R a fuzzy similarity relation such that $R(x,z) = 1$ for all $x, z \in U$. Let us consider the following fuzzy sets A and B:

$$\begin{aligned} &A(y_0) = 1, B(y_0) = 0, \\ &A(y_i) = 1, B(y_i) = \frac{i}{10}, \text{ for } i \text{ even}, i \neq 0, \\ &A(y_i) = \frac{i}{10}, B(y_i) = 1, \text{ for } i \text{ odd}, i \neq 0, \end{aligned}$$

then $(A \cap B)(y_i) = \frac{i}{10}$ and $(A \cup B)(y_i) = 1$ for all $i \in \{0, \ldots, 10\}$. Take $\mathscr{I} = \mathscr{I}_L$ and

$$W_1 = \left\langle \frac{10}{100}, \frac{9}{100}, \frac{8}{100}, \frac{7}{100}, \frac{6}{100}, \frac{5}{100}, \frac{4}{100}, \frac{3}{100}, \frac{2}{100}, \frac{1}{100}, \frac{45}{100} \right\rangle.$$

Note that andness$(W_1) = 0.615 > 0.5$. We compute $R{\downarrow}_{\mathscr{I},W_1}(A \cap B)$, $R{\downarrow}_{\mathscr{I},W_1}A$ and $R{\downarrow}_{\mathscr{I},W_1}B$. We obtain for $x \in U$ that

$$\begin{aligned} (R{\downarrow}_{\mathscr{I},W_1}(A \cap B))(x) &= \mathrm{OWA}_{W_1} \left\langle \frac{0}{10}, \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}, \frac{5}{10}, \frac{6}{10}, \frac{7}{10}, \frac{8}{10}, \frac{9}{10}, \frac{10}{10} \right\rangle \\ &= 0.385, \\ (R{\downarrow}_{\mathscr{I},W_1}A)(x) &= \mathrm{OWA}_{W_1} \left\langle 1, \frac{1}{10}, 1, \frac{3}{10}, 1, \frac{5}{10}, 1, \frac{7}{10}, 1, \frac{9}{10}, 1 \right\rangle \\ &= 0.565, \\ (R{\downarrow}_{\mathscr{I},W_1}B)(x) &= \mathrm{OWA}_{W_1} \left\langle \frac{0}{10}, 1, \frac{2}{10}, 1, \frac{4}{10}, 1, \frac{6}{10}, 1, \frac{8}{10}, 1, \frac{10}{10} \right\rangle \\ &= 0.51, \end{aligned}$$

which means that for every $x \in U$

$$(R{\downarrow}_{\mathscr{I},W_1}(A \cap B))(x) < (R{\downarrow}_{\mathscr{I},W_1}A)(x) \cap (R{\downarrow}_{\mathscr{I},W_1}B)(x).$$

Now, with

$$W_2 = \left\langle \frac{45}{100}, \frac{10}{100}, \frac{9}{100}, \frac{8}{100}, \frac{7}{100}, \frac{6}{100}, \frac{5}{100}, \frac{4}{100}, \frac{3}{100}, \frac{2}{100}, \frac{1}{100} \right\rangle$$

we obtain for every $x \in U$ that

$$1 = (R{\uparrow}_{\mathscr{C},W_2}(A \cup B))(x) > (R{\uparrow}_{\mathscr{C},W_2}A)(x) \cup (R{\uparrow}_{\mathscr{C},W_2}B)(x) = \max\{0.945, 0.93\}.$$

The other properties do not hold in this model. For example, we give a counterexample for the duality property.

Example 4.3.25. Let $U = \{y_1, y_2\}$, A a fuzzy set such that $A(y_1) = 0.2$ and $A(y_2) = 0.2$, R a fuzzy similarity relation with $R(x,z) = 0.5$ and $R(x,x) = 1$ for $x \neq z \in U$. Let $\mathscr{I} = \mathscr{I}_L$, $\mathscr{C} = \mathscr{T}_L$, $\mathscr{N} = \mathscr{N}_S$, $W_1 = \langle \frac{1}{3}, \frac{2}{3} \rangle$ and $W_2 = \langle \frac{3}{4}, \frac{1}{4} \rangle$, then andness$(W_1) > 0.5$ and orness$(W_2) > 0.5$. We have for $y_1 \in U$ that

$$(R{\downarrow}_{\mathscr{I}_L,W_1}A)(y_1) = \frac{1}{3} \cdot 0.7 + \frac{2}{3} \cdot 0.2 = \frac{11}{30}.$$

On the other hand we have that

$$\begin{aligned}(\mathrm{co}_{\mathcal{N}_S}(R\uparrow_{\mathcal{T}_L,W_2}(\mathrm{co}_{\mathcal{N}_S}A)))(y_1) &= 1-\mathrm{OWA}_{W_2}\langle\max\{0,R(y,y_1)+1-A(y)-1\}\rangle\\ &= 1-\mathrm{OWA}_{W_2}\langle 0.8,0.3\rangle\\ &= 1-(0.75\cdot 0.8+0.25\cdot 0.3)\\ &= \frac{13}{40}\end{aligned}$$

which is not the same. The other equality also does not hold, since $(R\uparrow_{\mathcal{T}_L,W_2}A)(y_1)=0.15$ and

$$(\mathrm{co}_{\mathcal{N}_S}(R\downarrow_{\mathcal{I}_L,W_1}(\mathrm{co}_{\mathcal{N}_S}A)))(y_1)=1-\left(\frac{1}{3}\cdot 0.7+\frac{2}{3}\cdot 0.2\right)=\frac{19}{30}.$$

Also, the inclusion property and constant α-set property do not hold.

Example 4.3.26. Let $U=\{y_1,y_2\}$, $A=\emptyset$ and R a similarity relation such that

$$\begin{aligned}R(y_1,y_1)&=R(y_2,y_2)=1,\\ R(y_1,y_2)&=R(y_2,y_1)=0.5.\end{aligned}$$

We take the weight vector $W_1=\langle\frac{1}{3},\frac{2}{3}\rangle$. The andness of W_1 is $\frac{2}{3}$, which is larger than 0.5. Take $\mathcal{I}=\mathcal{I}_L$. We compute the lower approximation of A in $x\in U$:

$$\begin{aligned}(R\downarrow_{\mathcal{I}_L,W_1}\emptyset)(x) &= (W_1)_1\cdot\mathcal{I}_L(0.5,0)+(W_1)_2\cdot\mathcal{I}_L(1,0)\\ &= \frac{1}{3}\cdot\frac{1}{2}\\ &= \frac{1}{6}\\ &> \emptyset(x).\end{aligned}$$

This means that the lower approximation of a fuzzy set not necessarily is included in the set.

Let us now take the same U and R, but $A=U$ and $\mathcal{C}$ a t-norm $\mathcal{T}$. We take the weight vector $W_2=\langle\frac{2}{3},\frac{1}{3}\rangle$, then the orness of W_2 is $\frac{2}{3}$. We obtain for the upper approximation of A in $x\in U$ that

$$\begin{aligned}(R\uparrow_{\mathcal{T},W_2}U)(x) &= (W_2)_1\cdot\mathcal{T}(1,1)+(W_2)_2\cdot\mathcal{T}(0.5,1)\\ &= \frac{2}{3}\cdot 1+\frac{1}{3}\cdot\frac{1}{2}\\ &= \frac{5}{6}\\ &< U(x).\end{aligned}$$

So a fuzzy set is not always included in its upper approximation.

Note that we do always have that $R\uparrow_{\mathcal{C},W_2}\emptyset=\emptyset$ and $R\downarrow_{\mathcal{I},W_1}U=U$. To end this section, we illustrate that the idempotence property does not hold.

Example 4.3.27. Consider the same setting as in the previous example. Note that R is a fuzzy similarity relation. We have

$$\begin{aligned}(R{\downarrow}_{\mathscr{I}_L,W_1}(R{\downarrow}_{\mathscr{I}_L,W_1}\emptyset))(x) &= (W_1)_1 \cdot \mathscr{I}_L\left(\frac{1}{2},\frac{1}{6}\right) + (W_1)_2 \cdot \mathscr{I}_L\left(1,\frac{1}{6}\right)\\ &= \frac{1}{3}\cdot\frac{2}{3}+\frac{2}{3}\cdot\frac{1}{6}\\ &= \frac{1}{3}\\ &> (R{\downarrow}_{\mathscr{I}_L,W_1}\emptyset)(x)\end{aligned}$$

and

$$\begin{aligned}(R{\uparrow}_{\mathscr{T},W_2}(R{\uparrow}_{\mathscr{T},W_2}U))(x) &= (W_2)_1 \cdot \mathscr{T}\left(1,\frac{5}{6}\right) + (W_2)_2 \cdot \mathscr{T}\left(0.5,\frac{5}{6}\right)\\ &= \frac{2}{3}\cdot\frac{5}{6}+\frac{1}{3}\cdot\frac{1}{3}\\ &= \frac{4}{6}\\ &< (R{\uparrow}_{\mathscr{T},W_2}U)(x).\end{aligned}$$

We see that we have to give in properties for having more robust models. Finding a robust model that is monotone w.r.t. relations and has the inclusion property is an open problem.

In the next chapter, we study axiomatic approaches for fuzzy rough sets. We will see why some properties only hold under certain conditions for the fuzzy relation R.

Chapter 5

Axiomatic approach of fuzzy rough sets

In the previous two chapters, we studied constructive approaches to design fuzzy rough set models. We recalled the definitions of some fuzzy rough set models and studied their properties. In this chapter, we do the opposite. We start with unary operators and some axioms to obtain a fuzzy relation R such that the operators work as approximation operators with respect to R. Axiomatic approaches are not used in applications, but are rather used to get more insight in the logical structure of fuzzy rough sets. Note that in this chapter, we can work with infinite universes.

We study the axiomatic approach developed by Wu et al. ([61]), as they characterise the general fuzzy rough set model with an EP implicator $\mathscr{I}$ that is left-continuous in the first argument and such that $\mathscr{I}(\cdot,0)$, i.e., the induced negator, is continuous and a left-continuous t-norm $\mathscr{T}$[1]. They give axioms to characterise the lower and upper approximation operator separately, while other authors use dual operators. When the operators are not dual, we do not necessarily get the same fuzzy relation.

Other papers that describe an axiomatic approach are [48, 62, 63, 44, 51, 66, 40, 41]. We will shortly discuss their approaches at the end of this chapter.

The axioms the authors use to characterise the lower and upper approximation operators, are based on properties of fuzzy relations (see e.g. [54]). The choice of axioms depends on which model we want to derive. For example, as we will see in the next section, Wu et al. use a t-norm and an implicator to derive the general fuzzy rough set model. If we use max and min instead, we would obtain the model designed by Dubois and Prade. Although the axioms to characterise the fuzzy rough set model are different in the papers, the axioms needed to obtain reflexivity, symmetry or transitivity are quite similar.

We begin with the axiomatic characterisation of an upper approximation operator and a lower approximation operator. Next, we study two interesting pairs of operators: dual and $\mathscr{T}$-coupled pairs. We end with a short overview of other axiomatic approaches in the literature.

[1]They assumed a continuous implicator and continuous t-norm, but we were able to prove that these conditions can be weakened.

5.1 Axiomatic characterisation of $\mathscr{T}$-upper fuzzy approximation operators

Wu et al. ([61]) discuss the axiomatic characterisation of $(\mathscr{I},\mathscr{T})$-fuzzy rough sets, i.e., the general fuzzy rough set model defined in Definition 3.2.1 with $\mathscr{C}$ a left-continuous t-norm $\mathscr{T}$ and $\mathscr{I}$ an EP implicator that is left-continuous in the first argument and of which the induced negator is continuous. The approach does not work for more general conjunctors, since we need the properties that t-norms are commutative and assiociative.

We use a fuzzy set-valued operator H to characterise the upper approximation operator.

Definition 5.1.1. Let $H: \mathscr{F}(U) \to \mathscr{F}(U)$ be an operator and let $\mathscr{T}$ be a left-continuous t-norm. H is called a *$\mathscr{T}$-upper fuzzy approximation operator* if and only if it satisfies the following axioms:

(H1) $\forall A \in \mathscr{F}(U), \forall \alpha \in I : H(\hat{\alpha} \cap_{\mathscr{T}} A) = \hat{\alpha} \cap_{\mathscr{T}} H(A)$,

(H2) $\forall A_j \in \mathscr{F}(U), j \in J : H\left(\bigcup_{j\in J} A_j\right) = \bigcup_{j\in J} H(A_j)$,

with $\hat{\alpha}(x) = \alpha$ for all $x \in U$ as before[2].

If H is a $\mathscr{T}$-upper fuzzy approximation operator on $\mathscr{F}(U)$, we define the fuzzy relation $\mathrm{Rel}(H)$ on $U \times U$ as

$$\forall (x,y) \in U \times U:\ \mathrm{Rel}(H)(x,y) = H(\{x\})(y). \tag{5.1}$$

Remark 5.1.2. In [62], $\mathrm{Rel}(H)$ is defined by $\mathrm{Rel}(H)(x,y) = H(\{y\})(x)$, since they work with the model

$$(R{\downarrow_{\mathscr{I}}}A)(x) = \inf_{y\in U} \mathscr{I}(R(x,y), A(y)),$$
$$(R{\uparrow_{\mathscr{T}}}A)(x) = \sup_{y\in U} \mathscr{T}(R(x,y), A(y)).$$

We see that the operator $R{\uparrow_{\mathscr{T}}}$ is a $\mathscr{T}$-upper fuzzy approximation operator: the first axiom is fulfilled by the fact that a left-continuous t-norm is associative and complete-distributive w.r.t. the supremum. Due to the latter, the second axiom is fulfilled by extension of Proposition 4.1.8. We have the following connection between $\mathrm{Rel}(R{\uparrow_{\mathscr{T}}})$ and R:

Lemma 5.1.3. Let $R \in \mathscr{F}(U \times U)$. We have that $\mathrm{Rel}(R{\uparrow_{\mathscr{T}}}) = R$.

Proof. This holds because for all $(x,y) \in U \times U$, we have that

$$(\mathrm{Rel}(R{\uparrow_{\mathscr{T}}}))(x,y) = R{\uparrow_{\mathscr{T}}}(\{x\})(y) = \sup_{z\in U} \mathscr{T}(R(z,y), \{x\}(z)) = R(x,y),$$

as for all t-norms and for all $a \in I$ we have that $\mathscr{T}(a,0) = 0$ and $\mathscr{T}(a,1) = a$. □

[2] Wu et al. used an operator $H: \mathscr{F}(W) \to \mathscr{F}(U)$, but as before, we restrict ourselves to fuzzy relations $R \in \mathscr{F}(U \times U)$.

We have the following relation between $(\mathrm{Rel}(H))\uparrow_{\mathscr{T}}$ and H:

Lemma 5.1.4. Let $\mathscr{T}$ be a left-continuous t-norm and H a $\mathscr{T}$-upper fuzzy approximation operator, then we have that $(\mathrm{Rel}(H))\uparrow_{\mathscr{T}} = H$, i.e., for all $A \in \mathscr{F}(U)$ it holds that $(\mathrm{Rel}(H))\uparrow_{\mathscr{T}} A = H(A)$.

Proof. Recall that with $\widehat{A(y)}$ we denote the constant $A(y)$-set. If A is a fuzzy set in U, then we can write

$$A = \bigcup_{y \in U} (\widehat{A(y)} \cap_{\mathscr{T}} \{y\})$$

because for all $x \in U$ it holds that

$$\begin{aligned}
\left(\bigcup_{y \in U} (\widehat{A(y)} \cap_{\mathscr{T}} \{y\}) \right)(x) &= \sup_{y \in U} \mathscr{T}(\widehat{A(y)}(x), \{y\}(x)) \\
&= \max \left\{ \sup_{y \neq x} \mathscr{T}(A(y), 0), \mathscr{T}(A(x), 1) \right\} \\
&= \max\{0, A(x)\} \\
&= A(x).
\end{aligned}$$

We obtain for $x \in U$ that

$$\begin{aligned}
(\mathrm{Rel}(H))\uparrow_{\mathscr{T}} A(x) &= \sup_{y \in U} \mathscr{T}(\mathrm{Rel}(H)(y,x), A(y)) \\
&= \sup_{y \in U} \mathscr{T}(H(\{y\})(x), A(y)) \\
&= \sup_{y \in U} (H(\{y\}) \cap_{\mathscr{T}} \widehat{A(y)})(x) \\
&= \sup_{y \in U} (\widehat{A(y)} \cap_{\mathscr{T}} H(\{y\}))(x) \\
&= \sup_{y \in U} H(\widehat{A(y)} \cap_{\mathscr{T}} \{y\})(x) \\
&= H\left(\bigcup_{y \in U} (\widehat{A(y)} \cap_{\mathscr{T}} \{y\} \right)(x) \\
&= H(A)(x)
\end{aligned}$$

which proves that $(\mathrm{Rel}\, H)\uparrow_{\mathscr{T}} = H$. We have used (H1) in step 5 and (H2) in step 6. □

These two lemmas lead to the desired theorem.

Theorem 5.1.5. Let $\mathscr{T}$ be a left-continuous t-norm. An operator $H : \mathscr{F}(U) \to \mathscr{F}(U)$ is a $\mathscr{T}$-upper fuzzy approximation operator if and only if there exists a general fuzzy relation R on $U \times U$ such that $H = R\uparrow_{\mathscr{T}}$, i.e., for all $A \in \mathscr{F}(U)$:

$$H(A) = R\uparrow_{\mathscr{T}} A.$$

Proof. $\Rightarrow$ Take $R = \mathrm{Rel}\, H$. It follows immediately from Lemma 5.1.4.

$\Leftarrow$ $R{\uparrow}_{\mathscr{T}}$ is a $\mathscr{T}$-upper fuzzy approximation operator.

□

If we add axioms to the ones in Definition 5.1.1, we can obtain extra properties of the relation R.

Proposition 5.1.6. Let $\mathscr{T}$ be a left-continuous t-norm and H a $\mathscr{T}$-fuzzy approximation operator. Then there exists an inverse serial fuzzy relation R such that $H = R{\uparrow}_{\mathscr{T}}$ if and only if H satisfies the axiom $H(U) = U$.

Proof. By Theorem 5.1.5 we know that there is a fuzzy relation R such that $H = R{\uparrow}_{\mathscr{T}}$.

First suppose R is inverse serial, then we have that $\sup\limits_{y\in U} R(y,x) = 1$, for all $x \in U$. We have for $x \in U$ that

$$\begin{aligned}(R{\uparrow}_{\mathscr{T}} U)(x) &= \sup_{y\in U} \mathscr{T}(R(y,x),1)\\ &= \sup_{y\in U} R(y,x)\\ &= 1\\ &= U(x).\end{aligned} \tag{5.2}$$

Now assume that $R{\uparrow}_{\mathscr{T}} U = U$. We deduce from Equation (5.2) that $\sup\limits_{y\in U} R(y,x) = 1$, for all $x \in U$, i.e., R is inverse serial. This completes the proof. □

The axiom $H(U) = U$ is equivalent with the axiom $\forall \alpha \in I : H(\hat{\alpha}) = \hat{\alpha}$. This follows from the fact that for each fuzzy relation R and for each $\alpha \in I$ it holds that

$$R{\uparrow}_{\mathscr{T}}(\hat{\alpha}) = \hat{\alpha} \Leftrightarrow R{\uparrow}_{\mathscr{T}} U = U.$$

Let us check this condition:

$\Leftarrow$ Assume $R{\uparrow}_{\mathscr{T}} U = U$, then for $\alpha \in I$ it holds that

$$\begin{aligned}R{\uparrow}_{\mathscr{T}} \hat{\alpha} &= R{\uparrow}_{\mathscr{T}}(\hat{\alpha} \cap_{\mathscr{T}} \hat{1})\\ &= R{\uparrow}_{\mathscr{T}}(\hat{\alpha} \cap_{\mathscr{T}} U)\\ &= \hat{\alpha} \cap_{\mathscr{T}} R{\uparrow}_{\mathscr{T}} U\\ &= \hat{\alpha} \cap_{\mathscr{T}} U\\ &= \hat{\alpha} \cap_{\mathscr{T}} \hat{1}\\ &= \hat{\alpha}.\end{aligned}$$

$\Rightarrow$ Take $\alpha = 1$.

We now characterise the properties of being reflexive, symmetric and $\mathscr{T}$-transitive.

Proposition 5.1.7. Let $\mathscr{T}$ be a left-continuous t-norm and H a $\mathscr{T}$-upper fuzzy approximation operator. There exists a fuzzy relation R such that $H = R\uparrow_{\mathscr{T}}$ that is

1. reflexive $\Leftrightarrow \forall A \in \mathscr{F}(U)\colon A \subseteq H(A)$,
2. symmetric $\Leftrightarrow \forall (x,y) \in U \times U\colon H(\{x\})(y) = H(\{y\})(x)$,
3. $\mathscr{T}$-transitive $\Leftrightarrow \forall A \in \mathscr{F}(U)\colon H(H(A)) \subseteq H(A)$.

So, H fulfils the three above axioms if and only if R is a $\mathscr{T}$-similarity relation.

Proof. By Theorem 5.1.5 we know that there is a fuzzy relation R such that $H = R\uparrow_{\mathscr{T}}$.

1. Let R be reflexive, then for all $A \in \mathscr{F}(U)$ and for all $x \in U$ we have that
$$\begin{aligned}(R\uparrow_{\mathscr{T}} A)(x) &= \sup_{y \in U} \mathscr{T}(R(y,x), A(y)) \\ &\geq \mathscr{T}(R(x,x), A(x)) \\ &= \mathscr{T}(1, A(x)) \\ &= A(x),\end{aligned}$$
i.e., $A \subseteq R\uparrow_{\mathscr{T}} A = H(A)$. Now assume that H fulfils the first axiom. We have
$$R(x,x) = (R\uparrow_{\mathscr{T}}\{x\})(x) \geq \{x\}(x) = 1,$$
and thus R is reflexive.

2. This follows immediately from the fact that for all $x, y \in U$ we have that
$$\begin{aligned}(R\uparrow_{\mathscr{T}}\{y\})(x) &= \sup_{z \in U} \mathscr{T}(R(z,x), \{y\}(z)) \\ &= R(y,x),\end{aligned}$$
as $\mathscr{T}(R(z,x),0) = 0$ and $\mathscr{T}(R(y,x),1) = R(y,x)$.

3. Assume that R is $\mathscr{T}$-transitive. For all $A \in \mathscr{F}(U)$ and for all $x \in U$ it holds that
$$\begin{aligned}(R\uparrow_{\mathscr{T}}(R\uparrow_{\mathscr{T}} A))(x) &= \sup_{y \in U} \mathscr{T}(R(y,x), \sup_{z \in U} \mathscr{T}(R(z,y), A(z))) \\ &= \sup_{y \in U} \sup_{z \in U} \mathscr{T}(R(y,x), \mathscr{T}(R(z,y), A(z))) \\ &= \sup_{y \in U} \sup_{z \in U} \mathscr{T}(\mathscr{T}(R(z,y), R(y,x)), A(z)) \\ &\leq \sup_{y \in U} \sup_{z \in U} \mathscr{T}(R(z,x), A(z)) \\ &= \sup_{z \in U} \mathscr{T}(R(z,x), A(z)) \\ &= (R\uparrow_{\mathscr{T}} A)(x)\end{aligned}$$

so $H(H(A)) \subseteq H(A)$. Conversely, assume H fulfils the third axiom. For all $x, z \in U$ we have that

$$\begin{aligned} R(x,z) &= (R\uparrow_{\mathcal{T}}\{x\})(z) \\ &\geq (R\uparrow_{\mathcal{T}}(R\uparrow_{\mathcal{T}}\{x\}))(z) \\ &= \sup_{y\in U} \mathcal{T}(R(y,z),(R\uparrow_{\mathcal{T}}\{x\})(y)) \\ &= \sup_{y\in U} \mathcal{T}(R(y,z),R(x,y)) \\ &= \sup_{y\in U} \mathcal{T}(R(x,y),R(y,z)). \end{aligned}$$

This proves that R is $\mathcal{T}$-transitive.

□

Proposition 5.1.7 explains why some properties do not hold for general fuzzy relations. For example, the inclusion property only holds for reflexive fuzzy relations. On the other hand, the idempotence property can only hold when we have a reflexive fuzzy relation that is $\mathcal{T}$-transitive.

We continue with an axiomatic characterisation of the lower approximation.

5.2 Axiomatic characterisation of $\mathcal{I}$-lower fuzzy approximation operators

Throughout this section, we assume $\mathcal{I}$ to be an EP implicator on I such that $\mathcal{I}$ is left-continuous in the first argument and such that $\mathcal{N}_{\mathcal{I}}$ is continuous. We shall refer to this three conditions as 'the standard conditions' on $\mathcal{I}$. S-implicators and IMTL-implicators are examples of implicators that fulfil the standard conditions.

We start by defining an $\mathcal{I}$-lower fuzzy approximation operator.

Definition 5.2.1. Let $L\colon \mathcal{F}(U) \to \mathcal{F}(U)$ be an operator and $\mathcal{I}$ an implicator that satisfies the standard conditions. L is called an *$\mathcal{I}$-lower fuzzy approximation operator* if and only if it satisfies the following axioms:

(L1) $\forall A \in \mathcal{F}(U), \forall \alpha \in I : L(\hat{\alpha} \Rightarrow_{\mathcal{I}} A) = \hat{\alpha} \Rightarrow_{\mathcal{I}} L(A)$,

(L2) $\forall A_j \in \mathcal{F}(U), j \in J : L\left(\bigcap_{j\in J} A_j\right) = \bigcap_{j\in J} L(A_j)$,

with $\hat{\alpha}(x) = \alpha$ for all $x \in U$[3].

[3]Wu et al. used an operator $L\colon \mathcal{F}(W) \to \mathcal{F}(U)$, but as before, we restrict ourselves to fuzzy relations $R \in \mathcal{F}(U \times U)$.

Now, let L be an $\mathscr{I}$-lower fuzzy approximation operator on $\mathscr{F}(U)$, then we can define a relation $\mathrm{Rel}(L)$ on $U \times U$ as

$$\forall (x,y) \in U \times U\colon\ \mathrm{Rel}(L)(x,y) = \sup\{\alpha \in I \mid \mathscr{I}(\alpha,0) = L(U \setminus \{x\})(y)\}. \tag{5.3}$$

Note that by standard conditions of $\mathscr{I}$ we have that

$$\mathscr{I}(\mathrm{Rel}(L)(x,y),0) = L(U \setminus \{x\})(y).$$

If $\mathscr{N}_{\mathscr{I}}$ is the negator induced by $\mathscr{I}$ we obtain that

$$\mathscr{N}_{\mathscr{I}}(\mathrm{Rel}(L)(x,y)) = L(U \setminus \{x\})(y).$$

Just like $R{\uparrow}_{\mathscr{T}}$ a $\mathscr{T}$-upper fuzzy approximation operator is, is $R{\downarrow}_{\mathscr{I}}$ an $\mathscr{I}$-lower fuzzy approximation operator. We obtain the first axiom by

$$\begin{aligned}
(R{\downarrow}_{\mathscr{I}}(\hat{\alpha} \Rightarrow_{\mathscr{I}} A))(x) &= \inf_{y\in U} \mathscr{I}(R(y,x),(\hat{\alpha} \Rightarrow_{\mathscr{I}} A)(y)) \\
&= \inf_{y\in U} \mathscr{I}(R(y,x),\mathscr{I}(\alpha,A(y))) \\
&= \inf_{y\in U} \mathscr{I}(\alpha,\mathscr{I}(R(y,x),A(y))) \\
&= \mathscr{I}\left(\alpha, \inf_{y\in U} \mathscr{I}(R(y,x),A(y))\right) \\
&= \mathscr{I}(\alpha,(R{\downarrow}_{\mathscr{I}}A)(x)) \\
&= (\hat{\alpha} \Rightarrow_{\mathscr{I}} R{\downarrow}_{\mathscr{I}}A)(x)
\end{aligned}$$

for all $x \in U$. The second axiom is fulfilled by extension of Proposition 4.1.8 and the fact that $\mathscr{I}$ is left-continuous in the first argument.

We study the relation between $\mathrm{Rel}(R{\downarrow}_{\mathscr{I}})$ and R.

Lemma 5.2.2. If (U,R) is a fuzzy approximation space with R a general fuzzy relation and $\mathscr{I}$ is an implicator that satisfies the standard conditions, then for all $x,y \in U$ it holds that

$$\mathscr{I}((\mathrm{Rel}(R{\downarrow}_{\mathscr{I}}))(x,y),0) = \mathscr{I}(R(x,y),0).$$

Proof. This follows from the fact that for all $x,y \in U$

$$\begin{aligned}
\mathscr{I}((\mathrm{Rel}(R{\downarrow}_{\mathscr{I}}))(x,y),0) &= (R{\downarrow}_{\mathscr{I}}(U \setminus \{x\}))(y) \\
&= \min\left\{\inf_{z\neq x} \mathscr{I}(R(z,y),1), \mathscr{I}(R(x,y),0)\right\} \\
&= \mathscr{I}(R(x,y),0),
\end{aligned}$$

as for all $z \in U$ it holds that

$$\mathscr{I}(R(z,y),1) \geq \mathscr{I}(0,1) = 1.$$

□

We have the following relation between $(\mathrm{Rel}(L))\downarrow_{\mathscr{I}}$ and L:

Lemma 5.2.3. Let $\mathscr{I}$ be an implicator that satisfies the standard conditions and L an $\mathscr{I}$-lower fuzzy approximation operator, then $(\mathrm{Rel}(L))\downarrow_{\mathscr{I}} = L$.

Proof. We can write $A \in \mathscr{F}(U)$ as

$$A = \bigcap_{y \in U} (\widehat{A(y)} \cup U \setminus \{y\})$$

because for $x \in U$ it holds that

$$\begin{aligned}\left(\bigcap_{y \in U} (\widehat{A(y)} \cup U \setminus \{y\})\right)(x) &= \inf_{y \in U} \max\{A(y), (U \setminus \{y\})(x), \} \\ &= \min\left\{\inf_{y \neq x} \max\{A(y), 1\}, \max\{A(x), 0\}\right\} \\ &= \min\{1, A(x)\} \\ &= A(x).\end{aligned}$$

Since we assumed $\mathscr{I}(\cdot, 0)$ to be continuous, for all $y \in U$ there exists a $b_y \in I$ such that

$$A(y) = \mathscr{I}(b_y, 0),$$

and thus

$$(\widehat{A(y)} \cup U \setminus \{y\}) = (\widehat{b_y} \Rightarrow_{\mathscr{I}} U \setminus \{y\}).$$

Recall that

$$\mathscr{I}(\mathrm{Rel}(L)(x, y), 0) = L(U \setminus \{x\})(y).$$

We can now prove the lemma: for all $x \in U$ it holds that

$$\begin{aligned}(\mathrm{Rel}(L)\downarrow_{\mathscr{I}} A)(x) &= \inf_{y \in U} \mathscr{I}(\mathrm{Rel}(L)(y, x), A(y)) \\ &= \inf_{y \in U} \mathscr{I}(\mathrm{Rel}(L)(y, x), \mathscr{I}(b_y, 0)) \\ &= \inf_{y \in U} \mathscr{I}(b_y, \mathscr{I}(\mathrm{Rel}(L)(y, x), 0)) \\ &= \inf_{y \in U} \mathscr{I}(b_y, L(U \setminus \{y\})(x)) \\ &= \inf_{y \in U} L(\widehat{b_y} \Rightarrow_{\mathscr{I}} U \setminus \{y\})(x) \\ &= \inf_{y \in U} L(\widehat{A(y)} \cup U \setminus \{y\})(x) \\ &= L\left(\bigcap_{y \in U} (\widehat{A(y)} \cup U \setminus \{y\})\right)(x) \\ &= L(A)(x),\end{aligned}$$

where we use (L1) in step 5 and (L2) in step 7. □

We again obtain the desired theorem.

Theorem 5.2.4. Let $\mathscr{I}$ be an implicator that satisfies the standard conditions. An operator $L: \mathscr{F}(U) \to \mathscr{F}(U)$ is an $\mathscr{I}$-lower fuzzy approximation operator if and only if there exists a general fuzzy relation R on $U \times U$ such that $L = R\downarrow_{\mathscr{I}}$, i.e., for all $A \in \mathscr{F}(U)$:

$$L(A) = R\downarrow_{\mathscr{I}} A.$$

Proof. This follows immediately from Equation (5.3), Lemma 5.2.2 and Lemma 5.2.3 with $R = \operatorname{Rel} L$. □

Just like we have done with the $\mathscr{T}$-upper fuzzy approximation operator, we want to describe which axioms an $\mathscr{I}$-lower fuzzy approximation operator has to fulfil to obtain certain properties of the relation R. We start with an inverse serial relation.

Proposition 5.2.5. Let $\mathscr{I}$ be a border implicator that fulfils the standard conditions and the following condition for $a, b \in I$:

$$a \leq b \Leftrightarrow \forall c \in I : \mathscr{I}(a, c) \leq \mathscr{I}(b, c). \tag{5.4}$$

Let L be an $\mathscr{I}$-lower fuzzy approximation operator. Then there exists an inverse serial fuzzy relation R on U such that $L = R\downarrow_{\mathscr{I}}$ if and only if L satisfies the axiom

$$\forall \alpha \in I : L(\hat{\alpha}) = \hat{\alpha}.$$

Proof. Due to Theorem 5.2.4, we have a relation R such that $L = R\downarrow_{\mathscr{I}}$. Suppose R is inverse serial, then we have for all $x \in U$ that

$$\begin{aligned}
(R\downarrow_{\mathscr{I}} \hat{\alpha})(x) &= \inf_{y \in U} \mathscr{I}(R(y, x), \alpha) \\
&= \mathscr{I}\left(\sup_{y \in U} R(y, x), \alpha\right) \\
&= \mathscr{I}(1, \alpha) \\
&= \alpha \\
&= \hat{\alpha}(x).
\end{aligned}$$

On the other hand, assume L fulfils the axiom. Since $\mathscr{I}$ satisfies condition (5.4), $\mathscr{I}$ satisfies also

$$a = b \Leftrightarrow \forall c \in I : \mathscr{I}(a, c) = \mathscr{I}(b, c).$$

We have for every $\alpha \in I$ that

$$\begin{aligned}
\mathscr{I}\left(\sup_{y \in U} R(y, x), \alpha\right) &= (R\downarrow_{\mathscr{I}} \hat{\alpha})(x) \\
&= \hat{\alpha}(x) \\
&= \alpha \\
&= \mathscr{I}(1, \alpha),
\end{aligned}$$

which implies that $\sup_{y\in U} R(y,x) = 1$ and thus that R is inverse serial. $\square$

For example, IMTL-implicators fulfil the extra conditions on $\mathscr{I}$.

If $\mathscr{I}$ is a CP implicator, then $\mathscr{I}$ satisfies condition (5.4). Let us prove this. If $a \leq b$, then of course $\mathscr{I}(a,c) \geq \mathscr{I}(b,c)$ for every $c \in I$ since an implicator is non-increasing in the first argument. Suppose $\mathscr{I}(a,c) \geq \mathscr{I}(b,c)$ for all $c \in I$ and $a > b$. Then we have that $\mathscr{I}(a,c) \leq \mathscr{I}(b,c)$ for all $c \in I$ and thus

$$\forall c \in I: \mathscr{I}(a,c) = \mathscr{I}(b,c).$$

In particular, $\mathscr{I}(a,a) = \mathscr{I}(b,a)$ and $\mathscr{I}(a,b) = \mathscr{I}(b,b)$. Since $\mathscr{I}$ is a CP implicator, this means that $\mathscr{I}(b,a) = 1$ and $\mathscr{I}(a,b) = 1$ or $b \leq a$ and $a \leq b$. This contradicts the assumption $a > b$. We conclude that

$$a \leq b \Leftrightarrow \forall c \in I : \mathscr{I}(a,c) \geq \mathscr{I}(b,c).$$

We now characterise the properties of being reflexive, symmetric and $\mathscr{T}$-transitive.

Proposition 5.2.6. Let $\mathscr{I}$ be a border implicator that fulfils the standard conditions and condition 5.4. Let $\mathscr{T}$ be a t-norm and L an $\mathscr{I}$-lower fuzzy approximation operator. There exists a fuzzy relation R such that $L = R{\downarrow_{\mathscr{I}}}$ that is

1. reflexive $\Leftrightarrow \forall A \in \mathscr{F}(U): L(A) \subseteq A$,
2. symmetric $\Leftrightarrow \forall (x,y) \in U \times U, \forall \alpha \in I$:

$$L(\{x\} \Rightarrow_{\mathscr{I}} \hat{\alpha})(y) = L(\{y\} \Rightarrow_{\mathscr{I}} \hat{\alpha})(x),$$

3. $\mathscr{T}$-transitive $\Leftrightarrow \forall A \in \mathscr{F}(U): L(A) \subseteq L(L(A))$ and if $\mathscr{I}$ satisfies

$$\forall a,b,c \in I: \mathscr{I}(a, \mathscr{I}(b,c)) = \mathscr{I}(\mathscr{T}(a,b),c).$$

So, L fulfils the three above axioms if and only if R is a $\mathscr{T}$-similarity relation.

Proof. By Theorem 5.2.4, we know that there exists a relation R such that $L = R{\downarrow_{\mathscr{I}}}$.

1. Let R be reflexive. For all $A \in \mathscr{F}(U)$ and $x \in U$ it holds that

$$\begin{aligned}(R{\downarrow_{\mathscr{I}}}A)(x) &= \inf_{y\in U} \mathscr{I}(R(y,x),A(y))\\ &\leq \mathscr{I}(R(x,x),A(x))\\ &= \mathscr{I}(1,A(x))\\ &= A(x),\end{aligned}$$

which means that $L(A) \subseteq A$. Now assume that L fulfils the first axiom. Note that for all $x, y \in U$ and $\alpha \in I$ it holds that $\mathscr{I}(\{x\}(y), \alpha) = 1$ if $y \neq x$ and that it is equal to α if $y = x$. We obtain

$$\begin{aligned}\mathscr{I}(R(x,x),\alpha) &= R{\downarrow_{\mathscr{I}}}(\{x\} \Rightarrow_{\mathscr{I}} \hat{\alpha})(x)\\ &\leq (\{x\} \Rightarrow_{\mathscr{I}} \hat{\alpha})(x)\\ &= \mathscr{I}(1,\alpha).\end{aligned}$$

This means by condition (5.4) that $R(x,x) = 1$ and thus that R is reflexive.

2. This follows immediately from the fact that

$$\forall (x,y) \in U \times U, \forall \alpha \in I : (R{\downarrow_{\mathscr{I}}}(\{y\} \Rightarrow_{\mathscr{I}} \hat{\alpha}))(x) = \mathscr{I}(R(y,x),\alpha).$$

and condition (5.4). Let us prove the above equation: take $x, y \in U$ and $\alpha \in I$, we obtain

$$\begin{aligned}(R{\downarrow_{\mathscr{I}}}(\{y\} \Rightarrow_{\mathscr{I}} \hat{\alpha}))(x) &= \inf_{z\in U} \mathscr{I}(R(z,x),(\{y\} \Rightarrow_{\mathscr{I}} \hat{\alpha})(z))\\ &= \inf_{z\in U} \mathscr{I}(R(z,x),\mathscr{I}(\{y\}(z),\alpha))\\ &= \min\left\{\inf_{z\neq y} \mathscr{I}(R(z,x),\mathscr{I}(\{y\}(z),\alpha)),\mathscr{I}(R(y,x),\mathscr{I}(\{y\}(y),\alpha))\right\}\\ &= \min\left\{\inf_{z\neq a} \mathscr{I}(R(z,x),\mathscr{I}(0,\alpha)),\mathscr{I}(R(y,x),\mathscr{I}(1,\alpha))\right\}\\ &= \min\left\{\inf_{z\neq y} \mathscr{I}(R(z,x),1),\mathscr{I}(R(y,x),\alpha)\right\}\\ &= \mathscr{I}(R(y,x),\alpha),\end{aligned}$$

since $\mathscr{I}$ is non-increasing in the first argument and non-decreasing in the second argument.

3. Assume $\mathscr{I}$ satisfies

$$\forall a,b,c \in I : \mathscr{I}(a,\mathscr{I}(b,c)) = \mathscr{I}(\mathscr{T}(a,b),c).$$

Let R be $\mathscr{T}$-transitive. Then we have for all $A \in \mathscr{F}(U)$ and $x \in U$ that

$$\begin{aligned}(R{\downarrow_{\mathscr{I}}}A)(x) &= \inf_{z\in U} \mathscr{I}(R(z,x),A(z))\\ &\leq \inf_{z\in U} \mathscr{I}\left(\sup_{y\in U} \mathscr{T}(R(z,y),R(y,x)),A(z)\right)\\ &= \inf_{z\in U}\inf_{y\in U} \mathscr{I}(\mathscr{T}(R(y,x),R(z,y)),A(z))\\ &= \inf_{y\in U}\inf_{z\in U} \mathscr{I}(R(y,x),\mathscr{I}(R(z,y),A(z)))\\ &= \inf_{y\in U} \mathscr{I}\left(R(y,x),\inf_{z\in U} \mathscr{I}(R(z,y),A(z))\right)\\ &= (R{\downarrow_{\mathscr{I}}}(R{\downarrow_{\mathscr{I}}}A))(x).\end{aligned}$$

Thus, we obtain $L(A) \subseteq L(L(A))$. On the other hand, assume that L satisfies the third axiom. For all $x, z \in U$ and $\alpha \in I$ we have that

$$\begin{aligned}
\mathscr{I}(R(y,x),\alpha) &= (R{\downarrow_{\mathscr{I}}}(\{y\} \Rightarrow_{\mathscr{I}} \hat{\alpha}))(x) \\
&\leq (R{\downarrow_{\mathscr{I}}}(R{\downarrow_{\mathscr{I}}}(\{y\} \Rightarrow_{\mathscr{I}} \hat{\alpha})))(x) \\
&= \inf_{z\in U} \mathscr{I}(R(z,x),(R{\downarrow_{\mathscr{I}}}(\{y\} \Rightarrow_{\mathscr{I}} \hat{\alpha}))(z)) \\
&= \inf_{z\in U} \mathscr{I}(R(z,x),\mathscr{I}(R(y,z),\alpha)) \\
&= \inf_{z\in U} \mathscr{I}(\mathscr{T}(R(z,x),R(y,z)),\alpha) \\
&= \mathscr{I}\left(\sup_{z\in U} \mathscr{T}(R(y,z),R(z,x)),\alpha\right).
\end{aligned}$$

By applying Equation (5.4) we obtain

$$R(y,x) \geq \sup_{z\in U} \mathscr{T}(R(y,z),R(z,x)),$$

i.e., R is $\mathscr{T}$-transitive.

□

The extra condition on $\mathscr{I}$ to obtain $\mathscr{T}$-transitivity is fulfilled by R-implicators based on a left-continuous t-norm and thus in particular by IMTL-implicators. An example of an implicator which fulfils all the conditions is the Łukasiewicz implicator.

The axiomatic approach gives us more insight in the logical structure of the general fuzzy rough set model. For example, we saw that the inclusion property only holds if the relation is reflexive, so this never can hold in general for a general fuzzy relation.

We now discuss some interactions between a $\mathscr{T}$-upper fuzzy approximation operator and an $\mathscr{I}$-lower fuzzy approximation operator.

5.3 Dual and $\mathscr{T}$-coupled pairs

In the previous two sections, we gave axioms to describe an upper and a lower approximation operator separately. We discuss now some interesting relations between an upper and lower approximation operator. The first pair we study is a dual pair.

With the right choices for $\mathscr{T}$ and $\mathscr{I}$, there is a duality between $\mathscr{T}$-upper and $\mathscr{I}$-lower fuzzy approximation operators.

Definition 5.3.1. Let $L, H\colon \mathscr{F}(U) \to \mathscr{F}(U)$ be two operators and $\mathscr{N}$ an involutive negator. We call L and H *dual operators with respect to* $\mathscr{N}$ if for all $A \in \mathscr{F}(U)$ we have:

$$L(A) = \text{co}_{\mathscr{N}}(H(\text{co}_{\mathscr{N}}(A))),$$
$$H(A) = \text{co}_{\mathscr{N}}(L(\text{co}_{\mathscr{N}}(A))).$$

If we have dual operators, we only need to define one operator and then derive the other operator by the duality relation. Furthermore, we can obtain the axioms for the corresponding operator from the axioms for the defined operator. We have dual operators if we work for example with a t-norm $\mathscr{T}$ and the S-implicator $\mathscr{I}_{\mathscr{S}}$ based on the dual t-conorm $\mathscr{S}$ of $\mathscr{T}$ with respect to $\mathscr{N}$.

If H is a $\mathscr{T}$-fuzzy approximation operator and L is an $\mathscr{I}$-fuzzy approximation operator and if H and L are dual operators, it holds that $\mathrm{Rel}(H) = \mathrm{Rel}(L)$, i.e., we obtain the same relation R in Theorems 5.1.5 and 5.2.4.

Lemma 5.3.2. Let $\mathscr{T}$ be a left-continuous t-norm, $\mathscr{I}$ an implicator that satisfies the standard conditions and $\mathscr{N}_{\mathscr{I}}$ the negator induced by $\mathscr{I}$. Let H be a $\mathscr{T}$-upper fuzzy approximation operator and L an $\mathscr{I}$-lower fuzzy approximation operator. If H and L are dual to $\mathscr{N}_{\mathscr{I}}$ and $\mathscr{N}_{\mathscr{I}}$ is involutive, then for all $(x, y) \in U \times U$ it holds that

$$\mathrm{Rel}(L)(x, y) = \mathrm{Rel}(H)(x, y).$$

Proof. Since $\mathscr{N}_{\mathscr{I}}$ is involutive and induced by $\mathscr{I}$ we have that for all $(x, y) \in U \times U$:

$$\begin{aligned}
\mathrm{Rel}(L)(x, y) &= \mathscr{N}_{\mathscr{I}}(L(U \setminus \{x\})(y)) \\
&= H(\mathscr{N}_{\mathscr{I}}(U \setminus \{x\}))(y) \\
&= H(\{x\})(y) \\
&= \mathrm{Rel}(H)(x, y).
\end{aligned}$$

□

We can see that the duality between L and H is analogous to the duality properties studied in Chapter 4.

Next, we discuss a $\mathscr{T}$-coupled pair, i.e., a pair consisting of a left-continuous t-norm and its R-implicator. This can be useful, because not every negator induced by an implicator is involutive, for example, the Gödelnegator is induced by the Gödelimplicator, but it is not involutive.

Definition 5.3.3. Let $\mathscr{T}$ be a left-continuous t-norm and let $\mathscr{I}_{\mathscr{T}}$ be its R-implicator. Let

$$H, L : \mathscr{F}(U) \to \mathscr{F}(U)$$

be two operators. We say that (H, L) is a *$\mathscr{T}$-coupled pair* of approximation operators if the following conditions hold:

(H1,H2) H is a $\mathscr{T}$-upper fuzzy approximation operator,

(L2) $\forall A_j \in \mathscr{F}(U), j \in J : L\left(\bigcap_{j\in J} A_j\right) = \bigcap_{j\in J} L(A_j)$,

(HL) $\forall A \in \mathscr{F}(U), \forall \alpha \in I : L(A \Rightarrow_{\mathscr{I}_{\mathscr{T}}} \hat{\alpha}) = H(A) \Rightarrow_{\mathscr{I}_{\mathscr{T}}} \hat{\alpha}$, where $\mathscr{I}$ is the R-implicator of $\mathscr{T}$, and with $\hat{\alpha}(x)=\alpha$ for all $x \in U$, $\alpha \in I$.

We have the following characterisation for a $\mathscr{T}$-coupled pair.

Theorem 5.3.4. Let $\mathscr{T}$ be a left-continuous t-norm. A pair of operators (H, L) is $\mathscr{T}$-coupled pair of approximation operators if and only if there exists a general fuzzy relation R on $U \times U$ such that $H = R{\uparrow}_{\mathscr{T}}$ and $L = R{\downarrow}_{\mathscr{I}_{\mathscr{T}}}$, i.e., for all $A \in \mathscr{F}(U)$:

$$H(A) = R{\uparrow}_{\mathscr{T}}A \text{ and } L(A) = R{\downarrow}_{\mathscr{I}_{\mathscr{T}}}A.$$

Proof. It is clear that $R{\uparrow}_{\mathscr{T}}$ and $R{\downarrow}_{\mathscr{I}_{\mathscr{T}}}$ satisfy (H1, H2) and (L2) respectively. Let us show that they also satisfy (HL). Recall the following properties for $\mathscr{I}_{\mathscr{T}}$ and $\mathscr{T}$ ([54]):

$$\mathscr{I}_{\mathscr{T}}(a, \mathscr{I}_{\mathscr{T}}(b, c)) = \mathscr{I}_{\mathscr{T}}(\mathscr{T}(a, b), c),$$
$$\mathscr{I}_{\mathscr{T}}(\sup_{j\in J} b_j, a) = \inf_{j\in J} \mathscr{I}_{\mathscr{T}}(b_j, a).$$

Take $x \in U$ and $\alpha \in I$, then

$$\begin{aligned}
\left(R{\downarrow}_{\mathscr{I}_{\mathscr{T}}}(A \Rightarrow_{\mathscr{I}_{\mathscr{T}}} \hat{\alpha})\right)(x) &= \inf_{y\in U} \mathscr{I}_{\mathscr{T}}(R(y,x), (A \Rightarrow_{\mathscr{I}_{\mathscr{T}}} \hat{\alpha})(y)) \\
&= \inf_{y\in U} \mathscr{I}_{\mathscr{T}}(R(y,x), \mathscr{I}_{\mathscr{T}}(A(y), \alpha)) \\
&= \inf_{y\in U} \mathscr{I}_{\mathscr{T}}(\mathscr{T}(R(y,x), A(y)), \alpha) \\
&= \mathscr{I}_{\mathscr{T}}\left(\sup_{y\in U} \mathscr{T}(R(y,x), A(y)), \alpha\right) \\
&= \mathscr{I}_{\mathscr{T}}((R{\uparrow}_{\mathscr{T}}A)(x), \alpha) \\
&= (R{\uparrow}_{\mathscr{T}}A \Rightarrow_{\mathscr{I}_{\mathscr{T}}} \hat{\alpha})(x).
\end{aligned}$$

Hence, $R{\uparrow}_{\mathscr{T}}$ and $R{\downarrow}_{\mathscr{I}_{\mathscr{T}}}$ fulfil (HL).

Conversely, assume (H, L) is a $\mathscr{T}$-coupled pair. By (H1, H2), H is a $\mathscr{T}$-upper fuzzy approximation operator, and by Theorem 5.1.5 we have a general fuzzy relation $R = \text{Rel}(H)$ such that for all $A \in \mathscr{F}(U)$ we have that

$$H(A) = R{\uparrow}_{\mathscr{T}}A.$$

We have the following representation for a fuzzy set A:

$$A = \bigcap_{y\in U} (\{y\} \Rightarrow_{\mathscr{I}_{\mathscr{T}}} \widehat{A(y)}).$$

Take $x \in U$, then

$$\begin{aligned}
\left(\bigcap_{y\in U} (\{y\} \Rightarrow_{\mathscr{I}_{\mathscr{T}}} \widehat{A(y)})\right)(x) &= \inf_{y\in U} \mathscr{I}_{\mathscr{T}}(\{y\}(x), A(y)) \\
&= \min\left\{\mathscr{I}_{\mathscr{T}}(1, A(x)), \inf_{y\neq x} \mathscr{I}_{\mathscr{T}}(0, A(y))\right\} \\
&= \min\{A(x), 1\} \\
&= A(x),
\end{aligned}$$

since R-implicators are border implicators. Because L satisfies (L2), we have that

$$L(A) = \bigcap_{y \in U} L(\{y\} \Rightarrow_{\mathscr{I}_{\mathscr{T}}} \widehat{A(y)})$$

and by (HL) we derive

$$L(A) = \bigcap_{y \in U} H(\{y\}) \Rightarrow_{\mathscr{I}_{\mathscr{T}}} \widehat{A(y)}.$$

For $x \in U$ we obtain:

$$\begin{aligned} L(A)(x) &= \inf_{y \in U} \mathscr{I}_{\mathscr{T}}(H\{y\}(x), \widehat{A(y)}(x)) \\ &= \inf_{y \in U} \mathscr{I}_{\mathscr{T}}(R(y,x), A(y)) \\ &= (R{\downarrow}_{\mathscr{I}_{\mathscr{T}}} A)(x), \end{aligned}$$

where we have used Equation (5.1) in the second step. This proves the theorem. □

If we take $\alpha = 0$ in (HL), then we obtain

$$\forall A \in \mathscr{F}(U) : L(\mathrm{co}_{\mathscr{N}}(A)) = \mathrm{co}_{\mathscr{N}}(H(A))$$

with $\mathscr{N} = \mathscr{N}_{\mathscr{I}}$. This is another form of duality where $\mathscr{N}$ is not necessarily involutive. If $\mathscr{N}_{\mathscr{I}}$ is involutive (as it is the case of $\mathscr{T}$ being the Łukasiewicz t-norm or in general any IMTL t-norm[4]), then a $\mathscr{T}$-coupled pair (H, L) is also dual in the sense of Definition 5.3.1.

We now characterise the properties of being inverse serial, reflexive, symmetric and $\mathscr{T}$-transitive.

Proposition 5.3.5. Let $\mathscr{T}$ be a left-continuous t-norm and let (H, L) be a $\mathscr{T}$-coupled pair of approximation operators. Then there exists a fuzzy relation R on $U \times U$ such that $H = R{\uparrow}_{\mathscr{T}}$ and $L = R{\downarrow}_{\mathscr{I}_{\mathscr{T}}}$ that is:

1. inverse serial $\Leftrightarrow H(U) = U$
   $\Leftrightarrow \forall A \in \mathscr{F}(U) \colon L(A) \subseteq H(A)$,
2. reflexive $\Leftrightarrow \forall A \in \mathscr{F}(U) \colon L(A) \subseteq A$,
3. symmetric $\Leftrightarrow \forall x, y \in U \colon H(\{x\})(y) = H(\{y\})(x)$
   $\Leftrightarrow \forall A \in \mathscr{F}(U) \colon H(L(A)) \subseteq A$
   $\Leftrightarrow \forall A \in \mathscr{F}(U) \colon A \subseteq L(H(A))$,
4. $\mathscr{T}$-transitive $\Leftrightarrow \forall A \in \mathscr{F}(U) \colon L(A) \subseteq L(L(A))$
   $\Leftrightarrow \forall A \in \mathscr{F}(U) \colon H(H(A)) \subseteq H(A)$.

So, H and L fulfil the last three axioms if and only if R is a $\mathscr{T}$-similarity relation.

[4]An IMTL t-norm is a t-norm of which its R-implicator $\mathscr{I}$ is contrapositive w.r.t. $\mathscr{N}_{\mathscr{I}}$ (see [21, 33]).

Proof. By Theorem 5.3.4 we know that there exists a relation R such that $H = R{\uparrow}_{\mathscr{T}}$ and $L = R{\downarrow}_{\mathscr{I}_{\mathscr{T}}}$. Then we can use results from [54] in the frame of fuzzy modal logics that we can adapt to our framework of a $\mathscr{T}$-coupled pair of approximation operators.

1. The equivalence that R is inverse serial if and only if $L(A) \subseteq H(A)$ for all $A \in \mathscr{F}(U)$ corresponds to [54, Proposition 4]. The equivalence with the condition $H(U) = U$ can easily be proved as follows:
$$\begin{aligned} H(U)(x) &= \sup_{y\in U} \mathscr{T}(R(y,x), U(y)) \\ &= \sup_{y\in U} \mathscr{T}(R(y,x), 1) \\ &= \sup_{y\in U} R(y,x). \end{aligned}$$
Hence, $U = H(U)$ if and only if $H(U)(x) = 1$ for all $x \in U$, i.e., if and only if $\sup\limits_{y\in U} R(y,x) = 1$ for all $x \in U$.

2. The characterisation of the reflexivity of R by the conditions $L(A) \subseteq A$ for all $A \in \mathscr{F}(U)$, or $A \subseteq H(A)$ for all $A \in \mathscr{F}(U)$, corresponds to [54, Proposition 5].

3. The characterisation of the symmetry of R by the conditions $H(L(A)) \subseteq A$ for all $A \in \mathscr{F}(U)$, or $A \subseteq L(H(A))$ for all $A \in \mathscr{F}(U)$, corresponds to [54, Proposition 9]. The equivalence with the condition $H(\{x\})(y) = H(\{y\})(x)$ for all $x, y \in U$ is proved in Proposition 5.1.7.

4. The characterisation of the $\mathscr{T}$-transitivity of R by the conditions $L(A) \subseteq L(L(A))$ for all $A \in \mathscr{F}(U)$, or $H(H(A)) \subseteq H(A)$ for all $A \in \mathscr{F}(U)$, corresponds to [54, Proposition 13].

□

To end this chapter, we provide a brief overview of other axiomatic characterisations that can be found in the literature.

5.4 A chronological overview of axiomatic approaches

In this section, we will give a more detailed overview of axiomatic approaches in the literature.

Morsi and Yakout ([48]) were the first to approach lower and upper approximations in a more axiomatic way, but not yet in the way we have seen it in Sections 5.1 and 5.2. They were the first to study the properties and other authors used their results. The model Morski and Yakout used is the general fuzzy rough set model with a left-continuous t-norm $\mathscr{T}$, its R-implicator $\mathscr{I}_{\mathscr{T}}$ and a $\mathscr{T}$-similarity relation R.

Wu et al. ([62, 63]) used the model of Dubois and Prade with a general fuzzy relation $R \subseteq \mathscr{F}(U \times W)$, which we shall restrict to relations from U to U. Wu et al. worked with finite universes. We have the following theorem.

Theorem 5.4.1. Let $H, L : \mathscr{F}(U) \to \mathscr{F}(U)$ be two dual operators, i.e., for a fuzzy set A in U:

$$L(A) = \mathrm{co}_{\mathscr{N}}(H(\mathrm{co}_{\mathscr{N}}(A))),$$
$$H(A) = \mathrm{co}_{\mathscr{N}}(L(\mathrm{co}_{\mathscr{N}}(A))).$$

for a given involutive negator $\mathscr{N}$. Then there exists a general fuzzy relation R such that $L = R{\downarrow}$ and $H = R{\uparrow}$ if and only if L and H satisfy the following axioms:

$$(L1') \;\; \forall A \in \mathscr{F}(U), \forall \alpha \in I : L(\hat{\alpha} \cup A) = \hat{\alpha} \cup L(A),$$
$$(L2') \;\; \forall A, B \in \mathscr{F}(U) : L(A \cap B) = L(A) \cap L(B),$$
$$(H1') \;\; \forall A \in \mathscr{F}(U), \forall \alpha \in I : H(\hat{\alpha} \cap A) = \hat{\alpha} \cap H(A),$$
$$(H2') \;\; \forall A, B \in \mathscr{F}(U) : H(A \cup B) = H(A) \cup H(B).$$

This was done by defining $R(x, y) = H(\{x\})(y)$ for $x, y \in U$. To characterise that R is reflexive, symmetric or transitive, the same axioms were used as in Proposition 5.1.7 and Proposition 5.2.6, only to characterise symmetry with the operator L, they used the following axiom:

$$\forall x, y \in U : L(U \setminus \{x\})(y) = L(U \setminus \{y\})(x).$$

Mi and Zhang ([44]) used the general fuzzy rough set model with an R-implicator $\mathscr{I}$ and its dual coimplicator $\mathscr{J}$ with respect to the standard negator $\mathscr{N}_S$ and a general fuzzy relation $R \subseteq \mathscr{F}(U \times W)$. They worked with dual operators. We give the approach for the operator H.

Theorem 5.4.2. Let $H : \mathscr{F}(U) \to \mathscr{F}(U)$ be an operator and let $\mathscr{C}$ be the conjunctor based on $\mathscr{J}$ and $\mathscr{N}_S$. Then there exists a general fuzzy relation R such that $H = R{\uparrow}_{\mathscr{C}}$ if and only if H satisfies the following axioms[5]:

$$(H1) \;\; \forall A \in \mathscr{F}(U), \forall \alpha \in I : H(\hat{\alpha} \cap_{\mathscr{C}} A) = \hat{\alpha} \cap_{\mathscr{C}} H(A),$$
$$(H2) \;\; \forall A_j \in \mathscr{F}(U), j \in J : H\left(\bigcup_{j \in J} A_j\right) = \bigcup_{j \in J} H(A_j).$$

The relation we obtain based on H is the following:

$$\begin{aligned} \forall x, y \in U : R(x, y) &= 1 - \sup_{\alpha \in I} \mathscr{C}(1 - H(\hat{\alpha} \cap_{\mathscr{C}} \{x\})(x), \alpha) \\ &= \inf_{\alpha \in I} \mathscr{C}(H(\hat{\alpha} \cap_{\mathscr{C}} \{x\})(x), 1 - \alpha) \\ &= \inf_{\alpha \in I} \mathscr{C}(\mathscr{C}(\alpha, H(\{x\})(x)), 1 - \alpha). \end{aligned}$$

The axioms to derive a reflexive or transitive relation are the same as in Proposition 5.1.7, but to characterise a symmetric relation, they used the following axiom:

$$\forall x, y \in U, \forall \alpha \in I : \mathscr{C}(\alpha, H\{x\}(y)) = \mathscr{C}(\alpha, H\{y\}(x)).$$

[5] In [44] finite unions were used, but since they worked in an infinite universe, infinite unions have to be used.

Pei ([51]) used Dubois and Prade's model with a general fuzzy relation R. He worked with dual operators.

Theorem 5.4.3. Let $H, L : \mathscr{F}(U) \to \mathscr{F}(U)$ be two dual operators, i.e., for a fuzzy set A in U:

$$L(A) = \mathrm{co}_{\mathscr{N}}(H(\mathrm{co}_{\mathscr{N}}(A))),$$
$$H(A) = \mathrm{co}_{\mathscr{N}}(L(\mathrm{co}_{\mathscr{N}}(A))).$$

for a given involutive negator $\mathscr{N}$. Then there exists a general fuzzy relation R such that $L = R\!\downarrow$ and $H = R\!\uparrow$ if and only if L and H satisfy the following axioms:

$$(L1')\ \ \forall A \in \mathscr{F}(U), \forall \alpha \in I : L(\hat{\alpha} \cup A) = \hat{\alpha} \cup L(A),$$
$$(L2)\ \ \forall A_j \in \mathscr{F}(U), j \in J : L\left(\bigcap_{j\in J} A_j\right) = \bigcap_{j\in J} L(A_j),$$
$$(H1')\ \ \forall A \in \mathscr{F}(U), \forall \alpha \in I : H(\hat{\alpha} \cap A) = \hat{\alpha} \cap H(A),$$
$$(H2)\ \ \forall A_j \in \mathscr{F}(U), j \in J : H\left(\bigcup_{j\in J} A_j\right) = \bigcup_{j\in J} H(A_j).$$

Again this was done by defining $R(x, y) = H(\{x\})(y)$ for $x, y \in U$. To characterise that R is reflexive, symmetric or transitive, the same axioms as in [62, 63] were used.

Yeung et al. ([66]) used the general fuzzy rough set model with a left-continuous t-norm and an S-implicator based on the dual t-conorm and the general fuzzy rough set model with an R-implicator based on a left-continuous t-norm and its dual coimplicator. The negator is an arbitrary involutive negator and the relation is a general fuzzy relation. We will only discuss the model based on a left-continuous t-norm and an S-implicator.

Theorem 5.4.4. Let $H : \mathscr{F}(U) \to \mathscr{F}(U)$ be an operator and let $\mathscr{T}$ be a left-continuous t-norm. Then there exists a general fuzzy relation R such that $H = R\!\uparrow_{\mathscr{T}}$ if and only if H satisfies the following axioms:

$$(H1)\ \ \forall A \in \mathscr{F}(U), \forall \alpha \in I : H(\hat{\alpha} \cap_{\mathscr{T}} A) = \hat{\alpha} \cap_{\mathscr{T}} H(A),$$
$$(H2)\ \ \forall A_j \in \mathscr{F}(U), j \in J : H\left(\bigcup_{j\in J} A_j\right) = \bigcup_{j\in J} H(A_j).$$

Again, we obtain this result by setting $R(x, y) = H(\{x\})(y)$ for all $x, y \in U$. For the lower approximation operator we have:

Theorem 5.4.5. Let $L : \mathscr{F}(U) \to \mathscr{F}(U)$ be an operator and $\mathscr{S}$ the t-conorm dual to $\mathscr{T}$ w.r.t. an involutive negator $\mathscr{N}$. Then there exists a general fuzzy relation R such that $L = R\!\downarrow_{\mathscr{I}_{\mathscr{S}}}$ if and only if L satisfies the following axioms:

$$(L1')\ \ \forall A \in \mathscr{F}(U), \forall \alpha \in I : L(\hat{\alpha} \cup_{\mathscr{S}} A) = \hat{\alpha} \cup_{\mathscr{S}} L(A),$$

$$(L2)\ \ \forall A_j \in \mathscr{F}(U), j \in J : L\left(\bigcap_{j\in J} A_j\right) = \bigcap_{j\in J} L(A_j).$$

This result is obtained by setting $R(x,y) = \mathrm{co}_{\mathscr{N}}(L(U \setminus \{x\}))(y)$ for $x, y \in U$. If L and H are dual to the same involutive negator as $\mathscr{T}$ and $\mathscr{S}$, then the two relations are the same, i.e.,

$$\forall x, y \in U : \mathrm{co}_{\mathscr{N}}(L(U \setminus \{x\}))(y) = H(\{x\})(y).$$

The axioms to characterise reflexivity, symmetry and transitivity are the same as in [62, 63] were used.

Liu ([40]) also used the model designed by Dubois and Prade with a general fuzzy relation R. He used the operator L.

Theorem 5.4.6. Let $L : \mathscr{F}(U) \to \mathscr{F}(U)$ be an operator. Then there exists a general fuzzy relation R such that $L = R\downarrow$ if and only if L satisfies the following axioms:

$$(L1')\ \ \forall A \in \mathscr{F}(U), \forall \alpha \in I : L(\hat{\alpha} \cup A) = \hat{\alpha} \cup L(A),$$

$$(L2)\ \ \forall A_j \in \mathscr{F}(U), j \in J : L\left(\bigcap_{j\in J} A_j\right) = \bigcap_{j\in J} L(A_j).$$

This was done by setting $R(x,y) = 1 - L(U \setminus \{x\})(y)$ for $x, y \in U$. The axioms to characterise a reflexive or transitive relation R are the same as in Proposition 5.2.6. The axiom to characterise a symmetric relation is:

$$\forall A, B \in \mathscr{F}(U) : [A, L(B)] = [B, L(A)]$$

where $[A, B]$ denotes the outer product of A and B. This is defined by

$$[A, B] = \inf_{x\in U} \max\{A(x), B(x)\}.$$

The characterisation of a fuzzy similarity relation by an operator H was derived by dual results.

Next, we discuss an important application of fuzzy rough sets: feature selection.

Chapter 6

Application of fuzzy rough sets: feature selection

In this chapter, we discuss an application of fuzzy rough sets: attribute selection or feature subset selection. This is a common problem in data mining, machine learning and pattern recognition. For example, which symptoms determine a certain disease? And is it possible to do easy tests for those symptoms instead of advanced ones?

Nowadays, databases expand not only in the rows, i.e., the objects we observe (the elements of the universe), but also in the columns, i.e., the attributes or features we use to describe the objects. Not all these attributes are relevant. Too much data can lead to big training and test times and can make data understanding very difficult.

A challenge is to find good strategies to select a minimal subset of relevant attributes, i.e., a decision reduct. We want to say as much as possible with as little as possible. Features can be misleading of they can be redundant, i.e., they do not add extra information. To find such a decision reduct, we can start with the whole set and then omit irrelevant attributes or we can start with the empty set and add relevant attributes.

To do this within the context of rough set theory, we can use positive regions and dependency degrees to find a decision superreduct, i.e., a set that contains a decision reduct, or we can use discernibility matrices and functions to determine all decision reducts. Both strategies will be discussed. We study some theoretical approaches to determine decision reducts and describe algorithms to do this in practice. We will illustrate the algorithms and techniques with an artificial example.

The structure of this chapter is as follows: in Section 6.1, we start with studying feature selection in rough set analysis, where we define all concepts. In Section 6.2, we extend the crisp concepts in an intuitive way to fuzzy rough analysis. We study the approaches of Cornelis et al. ([15]), where a new definition of positive region is introduced, and Jensen and Shen ([37]). Next, in Section 6.3, we will use the general fuzzy rough set model to find decision reducts. Tsang et al.

([60]) propose a method to find all decision reducts using the fuzzy rough set model designed by Dubois and Prade. Chen et al. ([6, 7]) do something similar, but they use the general fuzzy rough set model with a left-continuous t-norm $\mathscr{T}$ and its R-implicator $\mathscr{I}_{\mathscr{T}}$. Zhao and Tsang ([69]) study relations between different types of decision reducts. We discuss these three approaches. To end, we give in Section 6.4 an overview of approaches to fuzzy rough feature selection in the literature.

6.1 Feature selection in rough set analysis

We start by introducing the concepts we need in feature selection (see e.g., [15]). In rough set analysis, data is represented as an *information system* $(U, \mathscr{A})$ with U a finite, non-empty universe of objects and $\mathscr{A}$ a finite, non-empty set of attributes. Each attribute a in $\mathscr{A}$ corresponds to a mapping $a\colon U \to V_a$, where V_a is the value set of a over U. Note that V_a is a finite set. For each subset B of $\mathscr{A}$, we define the *B-indiscernibility relation* R_B as

$$R_B = \{(x, y) \in U^2 \mid \forall a \in B\colon a(x) = a(y)\}. \tag{6.1}$$

When B is a singleton $\{a\}$, we write R_a instead of $R_{\{a\}}$. It is clear that R_B is an equivalence relation on $U \times U$. If $B \subseteq \mathscr{A}$ is a subset such that $R_B = R_{\mathscr{A}}$, then we call B a *superreduct*. If B is a superreduct and for all $B' \subsetneq B$ it holds that $R_{B'} \neq R_{\mathscr{A}}$, then we call B a *reduct*.

A *decision system* $(U, \mathscr{A} \cup \{d\})$ is an information system such that the attribute $d \notin \mathscr{A}$. We call the elements of $\mathscr{A}$ *conditional attributes* and we call d the *decision attribute*. Given a subset B of $\mathscr{A}$, the *B-positive region* POS_B contains those objects from U for which the values of B allow to predict the decision class unequivocally, i.e.,

$$\mathrm{POS}_B = \bigcup_{y \in U} R_B{\downarrow}[y]_{R_d},$$

where the lower approximation operator is the one defined in Definition 2.1.2. Some authors also use the boundary region of a subset B to determine decision reducts (e.g., [37]). The *B-boundary region* of $B \subseteq \mathscr{A}$ is given by

$$\mathrm{BNR}_B = \left(\bigcup_{y \in U} R_B{\uparrow}[y]_{R_d}\right) \setminus \left(\bigcup_{y \in U} R_B{\downarrow}[y]_{R_d}\right).$$

If an element x is in BNR_B then there is a $y \in U$ such that $[x]_{R_B} \cap [y]_{R_d} \neq \emptyset$, but for all $z \in U$ it holds that $[x]_{R_B} \nsubseteq [z]_{R_d}$. The element x can not be classified in a decision class $[z]_{R_d}$ by the information in B.

The *degree of dependency of d on B*, denoted by γ_B, measures the predictive ability w.r.t. d of the attributes in B:

$$\gamma_B = \frac{|\mathrm{POS}_B|}{|U|}.$$

A decision system is called *consistent* if $\gamma_{\mathscr{A}} = 1$. A subset B of $\mathscr{A}$ is called a *decision superreduct* if $\mathrm{POS}_B = \mathrm{POS}_{\mathscr{A}}$ and it is called a *decision reduct* if it is a decision superreduct and if there is no proper subset B' of B such that $\mathrm{POS}_{B'} = \mathrm{POS}_{\mathscr{A}}$, i.e., B is minimal for the condition $\mathrm{POS}_B = \mathrm{POS}_{\mathscr{A}}$.

Feature selection can have different goals, e.g.,

- find all decision reducts,
- find one decision reduct,
- find one decision superreduct,
- find all decision superreducts,
- find a global minimal decision reduct, i.e., the smallest possible decision reducts over all decision reducts.

Finding all the decision reducts is an NP-problem, but mostly it is enough to generate a subset of decision reducts, or to generate decision superreducts. We will concentrate ourselves on the first three goals. The QuickReduct algorithm (Algorithm 1) finds a single decision superreduct of the decision system based on the degree of dependency. The ReverseReduct algorithm (Algorithm 2) always finds a decision reduct ([14]). Sometimes it can be practical to first determine a decision superreduct $S \subseteq \mathscr{A}$ with QuickReduct and then apply ReverseReduct to S to make it minimal, i.e., take $B = S$ instead of $B = \mathscr{A}$ in the first step of Algorithm 2.

Algorithm 1 QuickReduct

$B \leftarrow \{\}$
do
$T \leftarrow B$
for each $a \in (\mathscr{A} \setminus B)$
if $\gamma_{B\cup\{a\}} > \gamma_T$
$T \leftarrow B \cup \{a\}$
$B \leftarrow T$
until $\gamma_B = \gamma_{\mathscr{A}}$
return B

Let us illustrate the concepts and algorithms we saw above in an artificial example. In Table 6.1, we consider a decision system[1] with seven objects ($U = \{y_1, \ldots, y_7\}$) and eight conditional attributes that are all quantitive ($\mathscr{A} = \{a_1, \ldots, a_8\}$). We have one qualitative decision attribute d.

We see that we have two decision classes: $[y_1]_{R_d}$ contains all $y \in U$ such that $d(y) = 0$ and $[y_2]_{R_d}$ contains all $y \in U$ such that $d(y) = 1$.

[1]This is a sample taken from the Pima Indians Diabetes data set located at the UCI Machine Learning repository, availabe at `http://www.ics.uci.edu/~mlearn/MLRepository.html` and was also given in [15].

Algorithm 2 ReverseReduct

$B \leftarrow \mathscr{A}$
do
$T \leftarrow \emptyset$
for each $a \in B$
if $\gamma_{B\setminus\{a\}} = \gamma_{\mathscr{A}}$
$T \leftarrow B \setminus \{a\}$
if $T \neq \emptyset$
$B \leftarrow T$
until $T = \emptyset$
return B

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	d
y_1	1	101	50	15	36	24.2	0.526	26	0
y_2	8	176	90	34	300	33.7	0.467	58	1
y_3	7	150	66	42	342	34.7	0.718	42	0
y_4	7	187	68	39	304	37.7	0.254	41	1
y_5	0	100	88	60	110	46.8	0.962	31	0
y_6	0	105	64	41	142	41.5	0.173	22	0
y_7	1	95	66	13	38	19.6	0.334	25	0

Table 6.1: Decision system $(U, \mathscr{A} \cup \{d\})$

Since we only work with crisp sets, we need to discretise the data. A possible way to discretise the data is given in Table 6.2. We first prove that the system is consistent. Since no two objects

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	d
y_1	0	0	0	0	0	0	2	0	0
y_2	1	2	2	1	1	1	1	1	1
y_3	1	1	1	1	1	2	2	1	0
y_4	1	2	1	1	1	2	0	1	1
y_5	0	0	2	1	0	3	2	1	0
y_6	0	0	1	1	0	3	0	0	0
y_7	0	0	1	0	0	0	1	0	0

Table 6.2: Discretised data

have the same value for all conditional attributes, we have that $[y]_{R_{\mathscr{A}}} = \{y\}$, and thus $\mathrm{POS}_{\mathscr{A}} = U$, which means the system is consistent, i.e., $\gamma_{\mathscr{A}} = 1$.

Let $B = \{a_4, a_5\}$. We want to compute the positive region of B. Let us do this by first calculating the lower approximation of $[y_1]_{R_d}$ and $[y_2]_{R_d}$ for the B-indiscernibility relation R_B:

$$\begin{aligned} R_B{\downarrow}[y_1]_{R_d} &= \{y_1, y_5, y_6, y_7\}, \\ R_B{\downarrow}[y_2]_{R_d} &= \emptyset. \end{aligned}$$

This means that $\mathrm{POS}_B = \{y_1, y_5, y_6, y_7\}$ and the degree of dependency of d on B is $\gamma_B = \frac{4}{7}$. The upper approximation for the B-indiscernibility relation R_B is U for $[y_1]_{R_d}$ and $\{y_2, y_3, y_4\}$ for $[y_2]_{R_d}$. The boundary region of B is then:

$$\mathrm{BNR}_B = U \setminus \{y_1, y_5, y_6, y_7\} = \{y_2, y_3, y_4\}.$$

Let us apply QuickReduct and ReverseReduct to these discretised data. It can be checked that $\mathrm{POS}_{a_2} = U$, therefore QuickReduct terminates after the first iteration, yielding the decision reduct $\{a_2\}$.

ReverseReduct will take more work. Since $\mathrm{POS}_{\mathscr{A}\setminus\{a_1\}} = U$, we can omit a_1. Since $\mathrm{POS}_{\mathscr{A}\setminus\{a_1,a_2\}} = U$, we can also omit a_2. We can do the same with a_3, a_4, a_5 and a_6, since $\mathrm{POS}_{\mathscr{A}\setminus\{a_1,\ldots,a_6\}} = U$. We cannot omit a_7 or a_8, since $\mathrm{POS}_{a_7} = \{y_1, y_3, y_5\}$ and $\mathrm{POS}_{a_8} = \{y_1, y_6, y_7\}$. ReverseReduct gives us the decision reduct $\{a_7, a_8\}$.

Both algorithms give us one decision reduct, and the output is different for both algorithms.

A possible technique to generate all decision reducts is using the discernibility matrix and function. The *discernibility matrix* O of $(U, \mathscr{A} \cup \{d\})$ is the $n \times n$-matrix (with $n = |U|$) such that $\forall i, j \in \{1, \ldots, n\}$:

$$O_{ij} = \begin{cases} \emptyset & \text{if } d(y_i) = d(y_j) \\ \{a \in \mathscr{A} \mid a(y_i) \neq a(y_j)\} & \text{otherwise} \end{cases}$$

with $y_i, y_j \in U$. The *discernibility function* of $(U, \mathscr{A} \cup \{d\})$ is the mapping $f : \{0,1\}^m \to \{0,1\}$ (with $m = |\mathscr{A}|$) such that

$$f(a_1^*, \ldots, a_m^*) = \wedge \left\{ \vee O_{ij}^* \mid 1 \leq i < j \leq n, O_{ij} \neq \emptyset \right\} \tag{6.2}$$

with $O_{ij}^* = \{a^* \mid a \in O_{ij}\}$ and a^* the Boolean variable corresponding with the attribute a. We denote $\mathscr{A}^* = \{a_1^*, \ldots, a_m^*\}$. Let F be the disjunctive normal form of f, i.e., there is an $l \in \mathbb{N}$ and there are $B_k^* \subseteq \mathscr{A}$, $1 \leq k \leq l$, such that

$$F(a_1^*, \ldots, a_m^*) = (\wedge B_1^*) \vee \ldots \vee (\wedge B_l^*),$$

then the set of decision reducts is $\{B_1, \ldots, B_l\}$ with each B_k a set of attributes of $\mathscr{A}$ ([59]).

We can also use the valuation function to determine decision superreducts. If $B \subseteq \mathscr{A}$, then the *valuation function corresponding to B*, denoted by $\mathscr{V}_B$, is defined by $\mathscr{V}_B(a^*) = 1$ if and only if $a \in B$. We can extend this valuation to arbitrary Boolean formulas such that

$$\mathscr{V}_B(f(a_1^*, \ldots, a_m^*)) = f(\mathscr{V}_B(a_1^*), \ldots, \mathscr{V}_B(a_m^*)).$$

This expresses whether the attributes in B preserve the discernibility of $(U, \mathscr{A} \cup \{d\})$. If the decision system is consistent, we only have that $\mathscr{V}_B(f(a_1^*, \dots, a_m^*)) = 1$ if for every i and j in $\{1, \dots, n\}$ such that $d(y_i) \neq d(y_j)$ there is an $a \in B$ such that $a(y_i) \neq a(y_j)$. This means that there is an attribute in B that distinguishes y_i and y_j if $d(y_i) \neq d(y_j)$ ([59]).

Let us illustrate how O and f find all decision reducts. We take again the discretised data of Table 6.2. Note that O is a symmetric matrix, so we only give the lower triangular matrix. Since for all $i \in \{1, \dots, n\}$, we have $O_{ii} = \emptyset$, we can also omit the diagonal (see [15]):

$$O = \begin{bmatrix} & & & & & \\ \mathscr{A} & & & & & \\ \emptyset & \{a_2, a_3, a_6, a_7\} & & & & \\ \mathscr{A} & \emptyset & \{a_2, a_7\} & & & \\ \emptyset & \{a_1, a_2, a_5, a_6, a_7\} & \emptyset & \{a_1, a_2, a_3, a_5, a_6, a_7\} & & \\ \emptyset & \{a_1, a_2, a_3, a_5, a_6, a_7, a_8\} & \emptyset & \{a_1, a_2, a_5, a_6, a_8\} & \emptyset & \\ \emptyset & \{a_1, a_2, a_3, a_4, a_5, a_6, a_8\} & \emptyset & \{a_1, a_2, a_4, a_5, a_6, a_7, a_8\} & \emptyset & \emptyset \end{bmatrix}.$$

From this, we want to construct the discernibility function. We use the following properties of $\vee$ and $\wedge$:

$$a^* \wedge (a^* \vee b^*) = a^*,$$
$$a^* \vee (a^* \wedge b^*) = a^*,$$

with a^* and b^* Boolean variables. We obtain

$$f(a_1^*, \dots, a_8^*) = (a_2^* \vee a_7^*) \wedge (a_1^* \vee a_2^* \vee a_5^* \vee a_6^* \vee a_8^*).$$

Now, if we reduce f to its disjunctive normal form, we get

$$F(a_1^*, \dots, a_8^*) = (a_2^*) \vee (a_1^* \wedge a_7^*) \vee (a_5^* \wedge a_7^*) \vee (a_6^* \wedge a_7^*) \vee (a_8^* \wedge a_7^*).$$

The set of all decision reducts is

$$\{\{a_2\}, \{a_1, a_7\}, \{a_5, a_7\}, \{a_6, a_7\}, \{a_7, a_8\}\}.$$

It is easy to see that $\{a_2\}$ is a global minimal decision reduct. So, if we take $B_1 = \{a_1, a_7\}$, then

$$\begin{aligned} \mathscr{V}_{B_1}(f(a_1^*, \dots, a_8^*)) &= f(\mathscr{V}_{B_1}(a_1^*), \dots, \mathscr{V}_{B_1}(a_8^*)) \\ &= f(1, 0, 0, 0, 0, 0, 1, 0) \\ &= (0 \vee 1) \wedge (1 \vee 0 \vee 0 \vee 0 \vee 0) \\ &= 1 \end{aligned}$$

but with $B_2 = \{a_4, a_5\}$ we have

$$\begin{aligned} \mathscr{V}_{B_2}(f(a_1^*, \dots, a_8^*)) &= f(\mathscr{V}_{B_2}(a_1^*), \dots, \mathscr{V}_{B_2}(a_8^*)) \\ &= f(0, 0, 0, 1, 1, 0, 0, 0) \\ &= (0 \vee 0) \wedge (0 \vee 0 \vee 1 \vee 0 \vee 0) \\ &= 0. \end{aligned}$$

We see that B_1 is a decision reduct and B_2 is not.

Let us extend these concepts to a fuzzy rough setting.

6.2 Feature selection in fuzzy rough set analysis

We have seen above that when we work in rough set analysis, we need to discretise the data. This leads to information loss. This information loss is one of the main reasons why we introduce fuzzy sets into the models and why fuzzy rough sets are so interesting for feature selection: rough sets let us deal with imprecision, vagueness and uncertainty in the data, while fuzzy sets give us the opportunity to work with real-valued attributes, as we can construct fuzzy similarity relations to model the discernibility between objects.

In this section we discuss the approaches of Cornelis et al. ([15]) and Jensen and Shen ([37]). We extend the concepts we defined in Section 6.1. We again work in a decision system $(U, \mathscr{A} \cup \{d\})$[2] and we assume that $U = \{y_1, \dots, y_n\}$ and $\mathscr{A} = \{a_1, \dots, a_m\}$. In most applications, we work with a fuzzy tolerance relation R. Some authors will also impose $\mathscr{T}$-transitivity (e.g., [37]).

For a subset B of $\mathscr{A}$ and a t-norm $\mathscr{T}$, the *fuzzy B-indiscernibility relation* R_B is defined by

$$\forall x, y \in U : R_B(x, y) = \mathscr{T}(R_a(x, y))$$

where we take the t-norm over all attributes $a \in B$. When all $a \in B$ are qualititive, we obtain the traditional indiscernibility relation defined in Equation (6.1). Jensen and Shen used the minimum t-norm for $\mathscr{T}$, while Cornelis et al. used arbitrary t-norms.

We give an example of a fuzzy tolerance relation that we can use in feature selection ([15]). Let a be a quantitative attribute in $\mathscr{A} \cup \{d\}$ and $x, y \in U$, then $R_a(x, y)$ can be given by

$$R_a(x, y) = \max\left\{0, \min\left\{\frac{a(y) - a(x) + \sigma_a}{\sigma_a}, \frac{a(x) - a(y) + \sigma_a}{\sigma_a}\right\}\right\} \tag{6.3}$$

with σ_a the standard deviation of a, i.e.,

$$\sigma_a = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(a(y_i) - \bar{a})^2}$$

with $\bar{a} = \frac{1}{n}\sum\limits_{i=1}^{n} a(y_i)$. If a is qualitative (or nominal) then $R_a(x, y) = 1$ if $a(x) = a(y)$ and $R_a(x, y) = 0$ otherwise. Possible fuzzy $\mathscr{T}$-similarity relations are given in the following example ([37]).

Example 6.2.1. Let $\mathscr{T}$ be a t-norm, $x, y \in U$, $a \in \mathscr{A}$ and σ_a the standard deviation of a. Possible $\mathscr{T}$-similarity relations to use in feature selection are:

[2]Jensen and Shen considered a set of decision attributes $\mathscr{D}$, but we will not discuss this.

- $R_a(x,y) = 1 - \frac{|a(x)-a(y)|}{\max(a)-\min(a)}$,
- $R_a(x,y) = \exp\left(-\frac{(a(x)-a(y))^2}{2\sigma_a^2}\right)$,
- $R_a(x,y) = \max\left\{0, \min\left\{\frac{a(y)-(a(x)-\sigma_a)}{a(x)-(a(x)-\sigma_a)}, \frac{(a(x)+\sigma_a)-a(y)}{(a(x)+\sigma_a)-a(x)}\right\}\right\}$.

If a choice for R_a is not $\mathscr{T}$-transitive, then the fuzzy transitive closure can be computed for each attribute, i.e., R_a^{n-1} with $n = |U|$ (see Section 2.2.3).

To derive good algorithms, we first need to define the concept of a decision reduct in a fuzzy rough setting ([15]).

Definition 6.2.2. Let $\mathscr{M}$ be a monotone $\mathscr{P}(\mathscr{A}) \to I$ mapping such that $\mathscr{M}(\mathscr{A}) = 1$. Let $B \subseteq \mathscr{A}$ and $0 < \alpha \leq 1$. B is a *fuzzy $\mathscr{M}$-decision superreduct to degree α* if $\mathscr{M}(B) \geq \alpha$ and B is a *fuzzy $\mathscr{M}$-decision reduct to degree α* if moreover for all $B' \subsetneq B$, $\mathscr{M}(B') < \alpha$.

We discuss three approaches to determine decision reducts. Herefore we use fuzzy positive regions, fuzzy boundary regions and fuzzy discernibility functions.

6.2.1 Feature selection based on fuzzy positive regions

We recall the definition of a B-positive region ([15]).

Definition 6.2.3. Let $\mathscr{I}$ be an implicator, $B \subseteq \mathscr{A}$ and R_B a fuzzy B-indiscernibility relation, then the *fuzzy B-positive region* for $x \in U$ is

$$\text{POS}_B(x) = \sup_{y \in U} (R_B \downarrow_{\mathscr{I}} R_d y)(x) \tag{6.4}$$

where d is the decision attribute and where we take the lower approximation of $R_d y$ as in Definition 3.2.1.

If R_d is a crisp relation, then we have that $\text{POS}_B(x) = (R_B \downarrow_{\mathscr{I}} R_d x)(x)$:

$$\begin{aligned}
\text{POS}_B(x) &= \sup_{y \in U} (R_B \downarrow_{\mathscr{I}} R_d y)(x) \\
&= \max\left\{ \sup_{y \in R_d x} (R_B \downarrow_{\mathscr{I}} R_d y)(x), \sup_{y \notin R_d x} (R_B \downarrow_{\mathscr{I}} R_d y)(x) \right\} \\
&= \max\left\{ \sup_{y \in R_d x} \inf_{z \in U} \mathscr{I}(R_B(z,x), R_d(z,y)), 0 \right\} \\
&= \sup_{y \in R_d x} \inf_{z \in U} \mathscr{I}(R_B(z,x), R_d(z,x)) \\
&= \inf_{z \in U} \mathscr{I}(R_B(z,x), R_d(z,x)) \\
&= (R_B \downarrow_{\mathscr{I}} R_d x)(x),
\end{aligned}$$

since $\inf_{z \in U} \mathscr{I}(R_B(z,x), R_d(z,y)) \leq \mathscr{I}(R_B(x,x), R_d(x,y)) = \mathscr{I}(1,0) = 0$. If d is quantitive, then this does not longer hold in general, but we do have $\text{POS}_B(x) \geq (R_B{\downarrow_{\mathscr{I}}} R_d x)(x)$ when R_d is a fuzzy tolerance relation. This leads to another possible way of defining the fuzzy positive region ([15]).

Definition 6.2.4. Let $\mathscr{I}$ be an implicator, $B \subseteq \mathscr{A}$ and R_B a fuzzy B-indiscernibility relation, then we define for $x \in U$

$$\text{POS}'_B(x) = (R_B{\downarrow_{\mathscr{I}}} R_d x)(x)$$

where d is the decision attribute and where we take the lower approximation of $R_d x$ as in Definition 3.2.1.

As explained above, we always have $\text{POS}'_B(x) \leq \text{POS}_B(x)$, so the new definition results in smaller positive regions, i.e., we have less objects we can classify based on B.

In the next example, we illustrate how we calculate the positive region of a set of attributes ([15]).

Example 6.2.5. We now take the original data from Table 6.1 and we use Equation (6.3) to determine the indiscernibility relation. Again, let $B = \{a_4, a_5\}$. Let us take $\mathscr{I}_L$ as implicator and $\mathscr{T}_L$ as t-norm. Since d is qualitative, we can use the characterisation $\text{POS}_B(x) = (R_B{\downarrow_{\mathscr{I}}} R_d x)(x)$ for all $x \in U$. Let us take $x = y_3$. If $b = 1$, then $\mathscr{I}_L(a,b) = 1$ for all $a \in I$. With this in mind, we derive that

$$\begin{aligned}
\text{POS}_B(y_3) &= (R_B{\downarrow_{\mathscr{I}_L}} R_d y_3)(y_3) \\
&= \inf_{z \in U} \mathscr{I}_L(R_B(z,y_3), R_d(z,y_3)) \\
&= \min\{1, 1 - R_B(y_2,y_3), 1, 1 - R_B(y_4,y_3), 1, 1, 1\} \\
&= \min\Big\{1 - \mathscr{T}_L(R_{a_4}(y_2,y_3), R_{a_5}(y_2,y_3)), \\
&\qquad 1 - \mathscr{T}_L(R_{a_4}(y_4,y_3), R_{a_5}(y_4,y_3))\Big\}.
\end{aligned}$$

We first determine that $\overline{a_4} = \frac{244}{7}$ and $\sigma_{a_4} = 16.385$ and that $\overline{a_5} = \frac{1272}{7}$ and $\sigma_{a_5} = 131.176$. With this, we obtain that

$$\begin{aligned}
R_{a_4}(y_2,y_3) = 0.512 \text{ and } R_{a_5}(y_2,y_3) = 0.680, \\
R_{a_4}(y_4,y_3) = 0.817 \text{ and } R_{a_5}(y_4,y_3) = 0.710.
\end{aligned}$$

We continue our computation of the positive region:

$$\begin{aligned}
\text{POS}_B(y_3) &= \min\{1 - 0.192, 1 - 0.527\}. \\
&= 0.473.
\end{aligned}$$

We can do this for the other elements of U. The result is:

$$\text{POS}_B = \{(y_1,1),(y_2,0.808),(y_3,0.473),(y_4,0.473),(y_5,1),(y_6,1),(y_7,1)\}$$

where $(x,a) \in \text{POS}_B$ means that $\text{POS}_B(x) = a$. Note that in this case $\text{POS}_B(x) = 1$ if x is y_1, y_5, y_6 or y_7, just as in the crisp case.

Once we have fixed the fuzzy positive region, we can define measures that can act as stopping criteria for algorithms. Such a measure is an increasing $\mathcal{P}(\mathcal{A}) \to I$ mapping. An example of such a measure is a normalised extension of the degree of dependency ([15]): for $B \subseteq \mathcal{A}$, define γ_B and γ'_B by

$$\gamma_B = \frac{|\mathrm{POS}_B|}{|\mathrm{POS}_{\mathcal{A}}|} \text{ and } \gamma'_B = \frac{|\mathrm{POS}'_B|}{|\mathrm{POS}'_{\mathcal{A}}|}.$$

We assume that the denominators are not zero, but this would only be the case when the positive region of $\mathcal{A}$ would be empty and then every positive region would be empty. We do not consider these cases.

Jensen and Shen used $|U|$ as denominator instead of $|\mathrm{POS}_{\mathcal{A}}|$ and $|\mathrm{POS}'_{\mathcal{A}}|$ which will lead to smaller values for the dependency degrees. When the decision system is consistent, the results will be the same.

Instead of taking the average of the membership degrees of the B-positive region, we can also consider the most problematic element ([15]):

$$\delta_B = \frac{\min\limits_{x\in U} \mathrm{POS}_B(x)}{\min\limits_{x\in U} \mathrm{POS}_{\mathcal{A}}(x)} \text{ and } \delta'_B = \frac{\min\limits_{x\in U} \mathrm{POS}'_B(x)}{\min\limits_{x\in U} \mathrm{POS}'_{\mathcal{A}}(x)}.$$

Again we assume that the denominators are not zero. The four measures are clearly increasing functions and we have that

$$\gamma_{\mathcal{A}} = \gamma'_{\mathcal{A}} = \delta_{\mathcal{A}} = \delta'_{\mathcal{A}} = 1.$$

This means that these four measures fulfil the conditions of the function $\mathcal{M}$ from Definition 6.2.2 and we can use them to construct a modification of the QuickReduct algorithm (see Algorithm 3).

Algorithm 3 Modified QuickReduct to obtain a fuzzy $\mathcal{M}$-decision superreduct to degree α

$B \leftarrow \{\}$
do
$T \leftarrow B, \beta \leftarrow -1$
for each $a \in (\mathcal{A} \setminus B)$
if $\mathcal{M}(\gamma_{B\cup\{a\}}) > \beta$
$T \leftarrow B \cup \{a\}, \beta \leftarrow \mathcal{M}(\gamma_{B\cup\{a\}})$
$B \leftarrow T$
until$\mathcal{M}(\gamma_{B\cup\{a\}}) \geq \alpha$
return B

Dependency degrees are not only measures for subsets of $\mathcal{A}$, but there are also measures for attributes. For example, the *significance* of an attribute $a \in B$ ([37]):

$$\sigma_B(a) = \gamma_B - \gamma_{B\setminus\{a\}}.$$

If the significance of an attribute is 0, then we call the attribute *dispensable*. This means we can delete the attribute from our subset B without loss of dependency degree. If we look at the *indispensable* attributes, we obtain the *core* of $\mathscr{A}$: these are the attributes $a \in \mathscr{A}$ such that the dependency degree of $\mathscr{A}$ changes if the attribute is removed ([37]):

$$\begin{aligned}\text{Core}(\mathscr{A}) &= \{a \in \mathscr{A} \mid \gamma_{\mathscr{A}\setminus\{a\}} < \gamma_{\mathscr{A}}\} \\ &= \{a \in \mathscr{A} \mid \sigma_{\mathscr{A}}(a) > 0\}.\end{aligned}$$

The core of $\mathscr{A}$ contains the relevant attributes. We can also determine the core of $\mathscr{A}$ with other choices of dependency degree.

We continue with determining decision reducts based on fuzzy boundary regions.

6.2.2 Feature selection based on fuzzy boundary regions

The second technique is based on fuzzy boundary regions ([37]). We do not only take the lower, but also the upper approximation into account. Let B be a subset of $\mathscr{A}$ and let $\mathscr{I}$ and $\mathscr{T}$ be an implicator and a t-norm, respectively. The *fuzzy B-boundary region* in $x \in U$ is given by

$$\text{BNR}_B(x) = \sup_{y\in U}(R_B\!\uparrow_{\mathscr{T}} R_d y)(x) - \sup_{y\in U}(R_B\!\downarrow_{\mathscr{I}} R_d y)(x).$$

Again we want to find a decision (super)reduct B. Since we work with fuzzy sets, we need to take additional uncertainty into account. The *uncertainty degree* of a subset B of $\mathscr{A}$ is given by

$$\mu_B = \frac{1 - |\,\text{BNR}_B\,|}{1 - |\,\text{BNR}_{\mathscr{A}}\,|}.$$

We can again construct an algorithm similar to QuickReduct, where we want to maxime the uncertainty degree. Note that if the denominator of μ is zero, then $\text{BNR}_{\mathscr{A}} = U$ and then we have again that the positive region of $\mathscr{A}$ is empty. We do not consider these cases in applications.

The third technique uses fuzzy discernibility functions and determines all decision reducts.

6.2.3 Feature selection based on fuzzy discernibility functions.

Besides fuzzy positive and boundary regions, we can use fuzzy tolerance relations to define fuzzy discernibility functions. Recall that a^* is the Boolean variable associated with attribute a. If the decision system $(U, \mathscr{A} \cup \{d\})$ is consistent and $\mathscr{A} = \{a_1, \ldots, a_m\}$, we can rewrite Equation (6.2) as follows

$$\begin{aligned}&f(a_1^*, \ldots, a_m^*)\\ &\quad= \wedge\left\{\bigvee_{k=1}^{m} a_k^* \left[d(y_i) \neq d(y_j) \Rightarrow a_k(y_i) \neq a_k(y_j)\right] \mid 1 \leq i < j \leq n\right\}\\ &\quad= \wedge\left\{\bigvee_{k=1}^{m} a_k^* \left[a_k(y_i) = a_k(y_j) \Rightarrow d(y_i) = d(y_j)\right] \mid 1 \leq i < j \leq n\right\}\\ &\quad= \wedge\left\{\left[\bigwedge_{a_k^*=1}(a_k(y_i) = a_k(y_j)\right] \Rightarrow d(y_i) = d(y_j) \mid 1 \leq i < j \leq n\right\}\end{aligned}$$

with $y_i \in U$ for all $i \in \{1, \dots, n\}$.

We can generalise this by using t-norms, implicators and fuzzy indiscernibility relations to obtain a fuzzy discernibility function ([15]).

Definition 6.2.6. Let $\mathcal{T}$ be a t-norm and $\mathcal{I}$ an implicator. We define the *fuzzy discernibility function* $f: \{0,1\}^m \to I$ as

$$f(a_1^*, \dots, a_m^*) = \mathcal{T}(O_{ij}(a_1^*, \dots, a_m^*)) \text{ with } 1 \leq i < j \leq n$$

and with

$$O_{ij}(a_1^*, \dots, a_m^*) = \mathcal{I}(\mathcal{T}(R_{a_k}(y_i, y_j)), R_d(y_i, y_j))$$

where we take the t-norm over all a_k such that $a_k^* = 1$.

If $R_{a_k}(y_i, y_j)$ w.r.t. a_k decreases or $R_d(y_i, y_j)$ w.r.t. d increases, then O_{ij} increases. If R_{a_k} and R_d are crisp, we obtain the discernibility function from Equation (6.2).

We discuss the fuzzy discernibility function derived by Jensen and Shen ([37]). Let $\mathcal{N}$ be a negator. They define a *fuzzy clause* O'_{ij} as

$$O'_{ij}(a) = \mathcal{N}(R_a(y_i, y_j))$$

for $a \in A$ and $1 \leq i, j \leq n$. A fuzzy clause measures the fuzzy discernibility between two objects. If $O'_{ij}(a) = 1$, then y_i and y_j are distinct for a. If it is 0, then y_i and y_j are identical for a. When $O'_{ij}(a)$ is in $]0,1[$, we call the objects y_i and y_j *partly discernible*.

Definition 6.2.7. Let $\mathcal{N}$ be a negator and $\mathcal{I}$ an implicator. We define the *fuzzy discernibility function* $f': \{0,1\}^m \to I$ by

$$f'(a_1^*, \dots, a_m^*) = \min\left\{\mathcal{I}(O'_{ij}(d), \max\{O'_{ij}(a_k)\}) \mid 1 \leq i < j \leq n, a_k^* = 1\right\}$$

with $O'_{ij}(a) = \mathcal{N}(R_a(y_i, y_j))$.

When we take $\mathcal{T}$ the minimum t-norm and $\mathcal{I}$ a contrapositive implicator with respect to the negator $\mathcal{N}$, the fuzzy discernibility functions of Cornelis et al. and Jensen and Shen coincide. We derive:

$$f(a_1^*, \dots, a_m^*) = \min\{O_{ij}(a_1^*, \dots, a_m^*) \mid 1 \leq i < j \leq n\}$$

with

$$\begin{aligned} O_{ij}(a_1^*, \dots, a_m^*) &= \mathcal{I}(\min\{R_{a_k}(y_i, y_j)\}, R_d(y_i, y_j)) \\ &= \mathcal{I}(\mathcal{N}(R_d(y_i, y_j)), \mathcal{N}(\min\{R_{a_k}(y_i, y_j)\})) \\ &= \mathcal{I}(O'_{ij}(d), \max\{O'_{ij}(a_k)\}) \end{aligned}$$

where we every time take the a_k's into account for which $a_k^* = 1$. Hence,

$$\begin{aligned} f(a_1^*, \dots, a_m^*) &= \min\{\mathcal{I}(O'_{ij}(d), \max\{O'_{ij}(a_k)\}) \mid 1 \leq i < j \leq n, a_k^* = 1\} \\ &= f'(a_1^*, \dots, a_m^*). \end{aligned}$$

Again we need measures to construct stopping criteria ([15]). Let us look at the valuation $\mathscr{V}_B$ associated with $B \subseteq \mathscr{A}$: the value of $\mathscr{V}_B$ in $f(a_1^*, \ldots, a_m^*)$ is now in I and not in $\{0,1\}$. Recall that $\mathscr{V}_B(f(a_1^*, \ldots, a_m^*)) = f(\mathscr{V}_B(a_1^*), \ldots, \mathscr{V}_B(a_m^*))$ with $\mathscr{V}_B(a_k^*) = 1$ if $a_k \in B$ and $\mathscr{V}_B(a_k^*) = 0$ otherwise. Based on this, we introduce a normalised subset evaluation measure f_B ([15]):

$$f_B = \frac{\mathscr{V}_B(f(a_1^*, \ldots, a_m^*))}{\mathscr{V}_{\mathscr{A}}(f(a_1^*, \ldots, a_m^*))}.$$

We can also generalise Equation (6.2) by taking the average instead of the minimum ([15]). We obtain the function

$$g(a_1^*, \ldots, a_m^*) = \frac{2 \cdot \sum\limits_{1 \le i < j \le n} O_{ij}(a_1^*, \ldots, a_m^*)}{n(n-1)}.$$

This can be useful, since the function f is 0 as soon as one O_{ij} is 0. The associated measure with g is

$$g_B = \frac{\mathscr{V}_B(g(a_1^*, \ldots, a_m^*))}{\mathscr{V}_{\mathscr{A}}(g(a_1^*, \ldots, a_m^*))}.$$

Both measures f_B and g_B are increasing and it holds that $f_{\mathscr{A}} = g_{\mathscr{A}} = 1$. Let us illustrate these two measures ([15]).

Example 6.2.8. We take the data from Table 6.1 and we use Equation (6.3) to determine the indiscernibility relation. Let us again take $B = \{a_4, a_5\}$, $\mathscr{I} = \mathscr{I}_L$ and $\mathscr{T} = \mathscr{T}_L$.

We have that

$$\begin{aligned} f_B &= \frac{\mathscr{V}_B(f(a_1^*, \ldots, a_m^*))}{\mathscr{V}_{\mathscr{A}}(f(a_1^*, \ldots, a_m^*))} \\ &= \frac{f(0,0,0,1,1,0,0,0)}{f(1,1,1,1,1,1,1,1)} \\ &= \frac{\mathscr{T}_L(O_{ij}(0,0,0,1,1,0,0,0))}{1} \end{aligned}$$

with $1 \le i < j \le 7$. We calculate for example $O_{12}(0,0,0,1,1,0,0,0)$:

$$\begin{aligned} O_{12}(0,0,0,1,1,0,0,0) &= \mathscr{I}_L(\mathscr{T}_L(R_{a_4}(y_1,y_2), R_{a_5}(y_1,y_2)), R_d(y_1,y_2)) \\ &= \mathscr{I}_L(\mathscr{T}_L(0,0),0) \\ &= \mathscr{I}_L(0,0) \\ &= 1. \end{aligned}$$

With the results from Example 6.2.5 we can see that

$$O_{23}(0,0,0,1,1,0,0,0) = \mathscr{I}_L(\mathscr{T}_L(0.518, 0.680), 0) = \mathscr{I}_L(0.192, 0) = 0.808.$$

We obtain

$$\begin{aligned} f_B &= \mathscr{T}_L(1,1,1,1,1,1,0.808,1,1,1,1,0.473,1,1,1,1,1,1,1,1,1) \\ &= \mathscr{T}_L(0.808, 0.473) \\ &= 0.281. \end{aligned}$$

We can also do this for g_B, then we have that

$$\begin{aligned}
g_B &= \frac{\mathscr{V}_B(g(a_1^*, \ldots, a_m^*))}{\mathscr{V}_{\mathscr{A}}(g(a_1^*, \ldots, a_m^*))} \\
&= \frac{g(0,0,0,1,1,0,0,0)}{g(1,1,1,1,1,1,1,1)} \\
&= \frac{\sum\limits_{1 \leq i < j \leq 7} O_{ij}(0,0,0,1,1,0,0,0)}{\sum\limits_{1 \leq i < j \leq 7} 1} \\
&= \frac{20.281}{21} \\
&= 0.966.
\end{aligned}$$

There are some relations between the six measures we saw ([15]). For example,

$$\delta'_B \leq \gamma'_B \leq \gamma_B \text{ and } \delta'_B \leq \delta_B \leq \gamma_B$$

always holds for $B \subseteq \mathscr{A}$. If d is qualitative, then $\gamma_B = \gamma'_B$ and $\delta_B = \delta'_B$. These inequalities hold, because $\mathrm{POS}'_B \leq \mathrm{POS}_B$ and we have equalities when d is qualitative. We also have the following lemma ([15]).

Lemma 6.2.9. Let us assume that we use the same $\mathscr{I}$ and $\mathscr{T}$ to define the model, the indiscernibility relation and the discernibility functions. Let $U = \{y_1, \ldots, y_n\}$. For every $B \subseteq \mathscr{A}$ it holds that

1. if $\mathrm{POS}'_{\mathscr{A}} = U$, then $f_B \leq \delta'_B$ and $\gamma'_B \leq g_B$,
2. if $\mathscr{T} = \mathscr{T}_M$, then $f_B = \delta'_B$,
3. if $\mathrm{POS}'_{\mathscr{A}} = U$ and $g_B = 1$, then $\gamma'_B = \gamma_B = 1$.

Proof. 1. Assume $\mathrm{POS}'_{\mathscr{A}} = U$, then $\min\limits_{y \in U} \mathrm{POS}'_{\mathscr{A}}(y) = 1$. We also have that

$$\begin{aligned}
\mathrm{POS}'_{\mathscr{A}} = U &\Leftrightarrow \forall x \in U : (R_{\mathscr{A}} \downarrow_{\mathscr{I}} R_d x)(x) = 1 \\
&\Leftrightarrow \forall x \in U : \inf_{y \in U} \mathscr{I}(R_{\mathscr{A}}(y,x), R_d(y,x)) = 1 \\
&\Leftrightarrow \forall x, y \in U : \mathscr{I}(R_{\mathscr{A}}(y,x), R_d(y,x)) = 1 \\
&\Leftrightarrow \forall x, y \in U : \mathscr{I}(\mathscr{T}(R_a(y,x)), R_d(y,x)) = 1 \\
&\Leftrightarrow \forall i, j \in \{1, \ldots, n\} : \mathscr{I}(\mathscr{T}(R_a(y_i,y_j)), R_d(y_i,y_j)) = 1
\end{aligned}$$

where we take the t-norm over all attributes a. We obtain that

$$\begin{aligned}
\mathscr{V}_{\mathscr{A}}(f(a_1^*,\ldots,a_m^*)) &= f(1,1,\ldots,1)\\
&= \mathscr{T}(O_{ij}(1,1,\ldots,1)) \quad (\text{with } 1\le i<j\le n)\\
&= \mathscr{T}(\mathscr{I}(\mathscr{T}(R_a(y_i,y_j)),R_d(y_i,y_j))) \quad (\text{with } 1\le i<j\le n, a\in\mathscr{A})\\
&= \mathscr{T}(1,1,\ldots,1)\\
&= 1.
\end{aligned}$$

We derive that

$$\begin{aligned}
f_B &= \frac{\mathscr{V}_B(f(a_1^*,\ldots,a_m^*))}{\mathscr{V}_{\mathscr{A}}(f(a_1^*,\ldots,a_m^*))}\\
&= \mathscr{V}_B(f(a_1^*,\ldots,a_m^*))\\
&= \mathscr{T}(\mathscr{I}(R_B(y_i,y_j),R_d(y_i,y_j))) \quad (\text{with } 1\le i<j\le n)\\
&\le \min_{1\le i<j\le n} \mathscr{I}(R_B(y_i,y_j),R_d(y_i,y_j))\\
&= \min_{x,y\in U} \mathscr{I}(R_B(x,y),R_d(x,y))\\
&= \min_{y\in U}(R_B{\downarrow_{\mathscr{I}}}R_d y)(y)\\
&= \min_{y\in U}\mathrm{POS}'_B(y)\\
&= \delta'_B.
\end{aligned}$$

We obtain for γ'_B and g_B that

$$\begin{aligned}
\gamma'_B &= \frac{\sum\limits_{y\in U}(R_B{\downarrow_{\mathscr{I}}}R_d y)(y)}{n}\\
&= \frac{\sum\limits_{y\in U}\inf\limits_{x\in U}\mathscr{I}(R_B(x,y),R_d(x,y))}{n}\\
&= \frac{\sum\limits_{1\le j\le n}\inf\limits_{x\in U}\mathscr{I}(R_B(x,y_j),R_d(x,y_j))}{n}\\
&\le \frac{2\cdot\sum\limits_{1\le i<j\le n}\mathscr{I}(R_B(y_i,y_j),R_d(y_i,y_j))}{n(n-1)}\\
&= g_B.
\end{aligned}$$

2. If we take $\mathscr{T}=\mathscr{T}_M$, then for all $B\subseteq\mathscr{A}$ we have that

$$\mathscr{V}_B(f(a_1^*,\ldots,a_m^*)) = \min_{y\in U}\mathrm{POS}'_B(y)$$

and thus in particular for $B=\mathscr{A}$. Hence, $f_B=\delta'_B$.

3. Assume $g_B = 1$, then
$$\frac{2 \cdot \sum\limits_{1 \leq i < j \leq n} \mathscr{I}(R_B(y_i, y_j), R_d(y_i, y_j))}{n(n-1)} = 1.$$
This means that for all $i, j \in \{1, \ldots, n\}$ with $i < j$ we have
$$\mathscr{I}(R_B(y_i, y_j), R_d(y_i, y_j)) = 1.$$
Since R_B and R_d are reflexive and symmetric, we have for $j \leq i$ that
$$\mathscr{I}(R_B(y_i, y_j), R_d(y_i, y_j)) = 1,$$
or in other words
$$\forall j \in \{1, \ldots, n\} : \inf_{x \in U} \mathscr{I}(R_B(x, y_j), R_d(x, y_j)) = 1.$$
We conclude that $\gamma'_B = 1$. Since $\gamma'_B \leq \gamma_B$, γ_B is also 1.

□

This lemma shows that f and δ are essentially built upon the same idea, but there is a difference between g and γ: g evaluates all pairwise evalutions of $\mathscr{I}(R_B(x, y), R_d(x, y))$, while γ looks at the lowest value of $\mathscr{I}(R_B(x, y), R_d(x, y))$ for each $y \in U$ and then averages over these values. The last property tells us that for consistent data, a crisp g-decision reduct is always a crisp γ- and γ'-decision reduct.

Jensen and Shen ([37]) propose another measure. Let $\mathscr{S}$ be a t-conorm. We define the *satisfaction* of a fuzzy clause O'_{ij} for a subset B of $\mathscr{A}$ by
$$\mathrm{SAT}_B(O'_{ij}) = \mathscr{S}(O'_{ij}(a)) = \mathscr{S}(\mathscr{N}(R_a(y_i, y_j)))$$
where we take the t-conorm over all $a \in B$. If we take the decision attribute d into account, we define
$$\mathrm{SAT}_{B,d}(O'_{ij}) = \mathscr{I}(O'_{ij}(d), \mathrm{SAT}_B(O'_{ij})).$$
for a certain implicator $\mathscr{I}$. For a subset B of $\mathscr{A}$, we can also define the *total satisfiability* of all clauses:
$$\mathrm{SAT}(B) = \frac{\sum\limits_{1 \leq i < j \leq n} \mathrm{SAT}_{B,d}(O'_{ij})}{\sum\limits_{1 \leq i < j \leq n} \mathrm{SAT}_{\mathscr{A},d}(O'_{ij})}.$$
If $\mathrm{SAT}(B) = 1$, then we have found a decision superreduct. Note that SAT is monotone in B and can be used as a stopping criterium for a modified QuickReduct algorithm similar to Algorithm 3. We start with $B = \emptyset$ and we add the attribute a to B if
$$\mathrm{SAT}(B \cup \{a\}) \geq \mathrm{SAT}(B \cup \{c\})$$

for all $c \in \mathscr{A} \setminus B$. The algorithm stops when we have found a subset B such that $\mathrm{SAT}(B) = \mathrm{SAT}(\mathscr{A}) = 1$.

If we take the standard negator $\mathscr{N}_S$, an implicator $\mathscr{I}$ contrapositive w.r.t. $\mathscr{N}_S$, $\mathscr{T} = \min$ and $\mathscr{S} = \max$, we have a connection between SAT and g_B.

Proposition 6.2.10. Let $\mathscr{N}_S$ be the standard negator, $\mathscr{I}$ an implicator contrapositive w.r.t. $\mathscr{N}_S$, $\mathscr{T} = \min$ and $\mathscr{S} = \max$. Then

$$\forall B \subseteq \mathscr{A} : g_B = \mathrm{SAT}(B).$$

Proof. Fix $B \subseteq \mathscr{A}$ and $i < j \in \{1, \ldots, n\}$. We obtain that

$$\begin{aligned}
\mathrm{SAT}_{B,d}(O'_{ij}) &= \mathscr{I}(O'_{ij}(d), \mathrm{SAT}_B(O'_{ij})) \\
&= \mathscr{I}(\mathscr{N}(R_d(y_i, y_j)), \max\{\mathscr{N}(R_a(y_i, y_j))\}) \\
&= \mathscr{I}(\mathscr{N}(R_d(y_i, y_j)), \mathscr{N}(\min\{R_a(y_i, y_j)\})) \\
&= \mathscr{I}(\min\{R_a(y_i, y_j)\}, R_d(y_i, y_j)) \\
&= O_{ij}(a_1^*, \ldots, a_m^*)
\end{aligned}$$

where we take the maximum and the minimum over all $a \in B$, i.e., we only take into account those a_k's such that $\mathscr{V}_B(a_k^*) = 1$. So, we can write

$$\mathrm{SAT}_{B,d}(O'_{ij}) = O_{ij}(\mathscr{V}_B(a_1^*), \ldots, \mathscr{V}_B(a_m^*)).$$

We derive

$$\begin{aligned}
g_B &= \frac{\mathscr{V}_B(g(a_1^*, \ldots, a_m^*))}{\mathscr{V}_{\mathscr{A}}(g(a_1^*, \ldots, a_m^*))} \\
&= \frac{\sum\limits_{1 \leq i < j \leq n} O_{ij}(\mathscr{V}_B(a_1^*), \ldots, \mathscr{V}_B(a_m^*))}{\sum\limits_{1 \leq i < j \leq n} O_{ij}(\mathscr{V}_{\mathscr{A}}(a_1^*), \ldots, \mathscr{V}_{\mathscr{A}}(a_m^*))} \\
&= \frac{\sum\limits_{1 \leq i < j \leq n} \mathrm{SAT}_{B,d}(O'_{ij})}{\sum\limits_{1 \leq i < j \leq n} \mathrm{SAT}_{\mathscr{A},d}(O'_{ij})} \\
&= \mathrm{SAT}(B).
\end{aligned}$$

□

In the next section, we study some results of introducing fuzzy rough set models into feature selection.

6.3 Feature selection with fuzzy rough set models

In this section, we use fuzzy rough set theory to find all decision reducts. Again, we will build a discernibility function to do this. We start with a detailed overview of the approach of Tsang et al. ([60]), who used the model designed by Dubois and Prade.

Next, we give an overview of the approach of Chen et al. ([7]), who used the general fuzzy rough set model with a left-continuous t-norm $\mathscr{T}$ and its R-implicator $\mathscr{I}$ as fuzzy rough set model (see Definition 3.2.1). In [6], Chen et al. used the Łukasiewicz t-norm and implicator.

To end, we discuss some relations between different reducts. This was studied by Zhao and Tsang ([69]).

Throughout this section we work in the decision system $(U, \mathscr{A} \cup \{d\})$ with $U = \{y_1, \ldots, y_n\}$, $\mathscr{A} = \{a_1, \ldots, a_m\}$ and d the decision attribute.

6.3.1 Feature selection based on the general fuzzy rough set model

There are two key problems we should keep in mind when dealing with attribute selection with fuzzy rough sets. The first question is what we should keep invariant after reduction. In feature selection with rough set analysis, we keep the positive region of the decision attribute d invariant. Here we will see how we can change this condition to something we can use in an algorithm.

The second question is the selection of aggregation operator for several fuzzy similarity relations. We want that a smaller fuzzy similarity relation can provide a more precise lower approximation. As seen in Proposition 4.1.7, the general fuzzy rough set model is monotone with respect to fuzzy relations. Further, we know that with a reflexive fuzzy relation and a border implicator $\mathscr{I}$, the lower approximation of a fuzzy set A is contained in A. This shall fulfil our second question. That is why both the model designed by Dubois and Prade and the general fuzzy rough set model with an R-implicator $\mathscr{I}$ are good models to use in feature selection.

Dubois and Prade's model

We start by discussing the approach of Tsang et al. ([60]). We only need the lower approximation operator $R\downarrow$.

As seen in Section 6.2, we can associate each attribute $a \in \mathscr{A} \cup \{d\}$ with a fuzzy similarity relation R_a. This can be done in different ways, as illustrated in Example 6.2.1. Let $\mathscr{R}$ be the family of assiociated fuzzy similarity relations, i.e.,

$$\mathscr{R} = \{R_a \mid a \in \mathscr{A}\}.$$

Again, we call $\mathscr{R}$ the *conditional attributes set* and $(U, \mathscr{R} \cup R_d)$ a *fuzzy decision system*.

We take the minimum operator as aggregation operator and define the following relation on $U \times U$:

$$\mathrm{Sim}(\mathscr{R}) = \cap\{R \mid R \in \mathscr{R}\}.$$

This is again a fuzzy similarity relation. As before, we define the positive region as the union of lower approximations:

$$\forall x \in U : (\mathrm{POS}_{\mathrm{Sim}(\mathscr{R})} R_d)(x) = \sup_{y \in U}((\mathrm{Sim}(\mathscr{R}))\downarrow[y]_{R_d})(x).$$

A subset $\mathscr{P} \subseteq \mathscr{R}$ is a decision reduct if $\mathrm{POS}_{\mathrm{Sim}(\mathscr{P})} R_d = \mathrm{POS}_{\mathrm{Sim}(\mathscr{R})} R_d$ and if for all $R \in \mathscr{P}$ it holds that $\mathrm{POS}_{\mathrm{Sim}(\mathscr{P} \setminus \{R\})} R_d < \mathrm{POS}_{\mathrm{Sim}(\mathscr{P})} R_d$.

The collection of all the indispensable elements is again called the core:

$$\mathrm{Core}(\mathscr{R}) = \{R \in \mathscr{R} \mid \mathrm{POS}_{\mathrm{Sim}(\mathscr{R})} R_d > \mathrm{POS}_{\mathrm{Sim}(\mathscr{R} \setminus \{R\})} R_d\}$$

We will show that

$$\mathrm{Core}(\mathscr{R}) = \cap \mathrm{Red}(\mathscr{R})$$

where $\mathrm{Red}(\mathscr{R})$ is the collection of all decision reducts of the decision system.

We want to know under which conditions $\mathscr{P}$ could be a decision reduct of $\mathscr{R}$. To do that, we recall some properties about the structure of $R{\downarrow}A$ for A a fuzzy set and R a fuzzy similarity relation. We want to describe the lower approximation with fuzzy granules.

First we define a *fuzzy point*: let $\lambda \in]0,1]$, then the fuzzy point x_λ is defined by

$$\forall z \in U : (x_\lambda)(z) = \begin{cases} \lambda & z = x \\ 0 & z \neq x. \end{cases}$$

Note that $x_0 = \emptyset$. A *basic granule* $(x_\lambda)_R$ is a similarity class w.r.t. R, for such a fuzzy point. Let R be a fuzzy similarity relation and x_λ a fuzzy point. For every $z \in U$, we define

$$(x_\lambda)_R(z) = \begin{cases} \lambda & 1 - R(z,x) < \lambda \\ 0 & 1 - R(z,x) \geq \lambda. \end{cases}$$

Since $R(x,x) = 1$, we have that $x_\lambda \subseteq (x_\lambda)_R$. We have the following lemma that characterises $R{\downarrow}A$ ([60]).

Lemma 6.3.1. Let R be a fuzzy similarity relation and A a fuzzy set in U, then

$$R{\downarrow}A = \cup\{(x_\lambda)_R \mid (x_\lambda)_R \subseteq A, \lambda \in]0,1]\}.$$

Proof. We prove that

$$(x_\lambda)_R \subseteq R{\downarrow}A \Leftrightarrow (x_\lambda)_R \subseteq A.$$

Fix $x \in U$ and $\lambda \in]0,1]$. Assume $(x_\lambda)_R \subseteq A$ and take $y \in U$. If $1 - R(y,x) < \lambda$, then

$$(x_\lambda)_R(y) = \lambda \leq A(y) \leq \max\{1 - R(y,x), A(y)\}.$$

If $1 - R(y,x) \geq \lambda$, then also $\lambda \leq \max\{1 - R(y,x), A(y)\}$. So, we have for $z \in U$

$$(R{\downarrow}A)(z) = \inf_{y \in U} \max\{1 - R(y,z), A(y)\} \geq \lambda \geq (x_\lambda)_R(z).$$

This means that if $(x_\lambda)_R \subseteq A$, we have that $(x_\lambda)_R \subseteq R{\downarrow}A$.

On the other hand, suppose that $(x_\lambda)_R \subseteq R{\downarrow}A$. This means that $(R{\downarrow}A)(x) \geq \lambda$. Take $y \in U$. If $1 - R(y,x) \geq \lambda$, then $(x_\lambda)_R(y) = 0 \leq A(y)$. But if $1 - R(y,x) < \lambda$, then $A(y)$ has to be greater or equal than λ since

$$(R{\downarrow}A)(x) = \inf_{y \in U} \max\{1 - R(y,x), A(y)\} \geq \lambda$$

and thus $(x_\lambda)_R(y) = \lambda \leq A(y)$. This means that $(x_\lambda)_R \subseteq A$. □

It is also easy to see that

$$R{\downarrow}(x_\lambda)_R = \cup\{(x_\beta)_R \mid (x_\beta)_R \subseteq (x_\lambda)_R, \beta \in]0,1]\} = (x_\lambda)_R.$$

This holds, because the general fuzzy rough set model fulfils the inclusion property for a reflexive fuzzy relation and a border implicator and

$$(x_\lambda)_R \subseteq \cup\{(x_\beta)_R \mid (x_\beta)_R \subseteq (x_\lambda)_R, \beta \in]0,1]\}.$$

Also note that for all $x, y \in U$, $\lambda \in]0,1]$ we have either $(x_\lambda)_R = (y_\lambda)_R$ or $(x_\lambda)_R \cap (y_\lambda)_R = \emptyset$. Let us prove this. Assume that $(x_\lambda)_R \cap (y_\lambda)_R \neq \emptyset$, then there is a $z \in U$ such that $(x_\lambda)_R(z) \neq 0$ and $(y_\lambda)_R(z) \neq 0$, but then $(x_\lambda)_R(z) = (y_\lambda)_R(z) = \lambda$. This implies that $1 - R(z,x) < \lambda$ and $1 - R(z,y) < \lambda$ and thus $1 - R(x,y) < \lambda$ by min-transitivity. This means that $x_\lambda \subseteq (y_\lambda)_R$ and $y_\lambda \subseteq (x_\lambda)_R$, hence $(x_\lambda)_R = (y_\lambda)_R$.

Let us look again at the relation $\mathrm{Sim}(\mathscr{R})$. We have the following statements ([60]):

Lemma 6.3.2. Let $x, y \in U$ and $\lambda \in]0,1]$. It holds that

1. $(x_\lambda)_{\mathrm{Sim}(\mathscr{R})} = \bigcap_{R \in \mathscr{R}} (x_\lambda)_R$,
2. $(x_\lambda)_{\mathrm{Sim}(\mathscr{R})} = (y_\lambda)_{\mathrm{Sim}(\mathscr{R})}$ if and only if $(x_\lambda)_R = (y_\lambda)_R$ for every $R \in \mathscr{R}$.

Proof. 1. Take $z \in U$, then we have that:

$$\begin{aligned}(x_\lambda)_{\mathrm{Sim}(\mathscr{R})}(z) = \lambda &\Leftrightarrow 1 - (\mathrm{Sim}(\mathscr{R}))(z,x) < \lambda \\ &\Leftrightarrow \forall R \in \mathscr{R} : 1 - R(z,x) < \lambda \\ &\Leftrightarrow \forall R \in \mathscr{R} : (x_\lambda)_R(z) = \lambda \\ &\Leftrightarrow \bigcap_{R \in \mathscr{R}} (x_\lambda)_R(z) = \lambda.\end{aligned}$$

2. Assume there is an $R \in \mathscr{R}$ such that $(x_\lambda)_R \neq (y_\lambda)_R$, then $(x_\lambda)_R \cap (y_\lambda)_R = \emptyset$. Without loss of generality, this means that there is a $z \in U$ such that $(x_\lambda)_R(z) = \lambda$ and $(y_\lambda)_R(z) = 0$. By the first statement we obtain $(x_\lambda)_{\mathrm{Sim}(\mathscr{R})} \neq (y_\lambda)_{\mathrm{Sim}(\mathscr{R})}$.

   On the other hand, if for all $R \in \mathscr{R}$ hold that $(x_\lambda)_R = (y_\lambda)_R$, then

$$\bigcap_{R \in \mathscr{R}} (x_\lambda)_R = \bigcap_{R \in \mathscr{R}} (y_\lambda)_R,$$

hence $(x_\lambda)_{\mathrm{Sim}(\mathscr{R})} = (y_\lambda)_{\mathrm{Sim}(\mathscr{R})}$.

□

Since U is finite and

$$\left(\mathrm{POS}_{\mathrm{Sim}(\mathscr{R})} R_d\right)(x) = \sup_{z\in U}((\mathrm{Sim}(\mathscr{R}))\downarrow[z]_{R_d})(x),$$

we know that $\left(\mathrm{POS}_{\mathrm{Sim}(\mathscr{R})} R_d\right)(x)$ has to reach its maximum value for some z. This will be reached in x itself ([60]).

Lemma 6.3.3. Take $x, z \in U$ and $\lambda \in]0,1]$. If $(x_\lambda)_{\mathrm{Sim}(\mathscr{R})} \subseteq [z]_{R_d}$, then $(x_\lambda)_{\mathrm{Sim}(\mathscr{R})} \subseteq [x]_{R_d}$.

Proof. Take $x, z \in U$ and $\lambda \in]0,1]$ and assume $(x_\lambda)_{\mathrm{Sim}(\mathscr{R})} \subseteq [z]_{R_d}$. Then for every $y \in U$ we have that

$$(x_\lambda)_{\mathrm{Sim}(\mathscr{R})}(y) \leq R_d(y,z).$$

So, if we take $y = x$, then $\lambda \leq R_d(x,z)$. Because R_d is min-transitive, we obtain

$$\begin{aligned}(x_\lambda)_{\mathrm{Sim}(\mathscr{R})}(y) &\leq \min\{(x_\lambda)_{\mathrm{Sim}(\mathscr{R})}(y), \lambda\}\\ &\leq \min\{R_d(y,z), R_d(x,z)\}\\ &\leq R_d(x,y)\\ &= [x]_{R_d}(y)\end{aligned}$$

which implies that $(x_\lambda)_{\mathrm{Sim}(\mathscr{R})} \subseteq [x]_{R_d}$. □

If we fix λ such that $\lambda = \left(\mathrm{POS}_{\mathrm{Sim}(\mathscr{R})} R_d\right)(x)$, then there exists a $z \in U$ such that

$$\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow[z]_{R_d})(x).$$

Since we have by Lemma 6.3.1 that

$$\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow[z]_{R_d})(x) = \sup\{((x_\beta)_{\mathrm{Sim}(\mathscr{R})})(x) \mid (x_\beta)_{\mathrm{Sim}(\mathscr{R})} \subseteq [z]_{R_d}, \beta \in I\},$$

we have that $(x_\lambda)_{\mathrm{Sim}(\mathscr{R})} \subseteq [z]_{R_d}$, and thus $(x_\lambda)_{\mathrm{Sim}(\mathscr{R})} \subseteq [x]_{R_d}$. Hence, $\lambda \leq ((\mathrm{Sim}(\mathscr{R}))\downarrow[x]_{R_d})(x)$ and thus $((\mathrm{Sim}(\mathscr{R}))\downarrow[x]_{R_d})(x) = \lambda$.

From this we can conclude that keeping the positive region invariant after deleting attributes from $\mathscr{R}$ is the same as keeping $((\mathrm{Sim}(\mathscr{R}))\downarrow[x]_{R_d})(x)$ invariant for every $x \in U$. With this in mind, we can characterise a decision reduct of $\mathscr{R}$ ([60]).

Lemma 6.3.4. Suppose $\mathscr{P} \subset \mathscr{R}$, then $\mathscr{P}$ contains a decision reduct of $\mathscr{R}$ if and only if $\mathscr{P}$ satisfies $(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \subseteq [x]_{R_d}$ for $\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow[x]_{R_d})(x)$ and for all $x \in U$.

Proof. Suppose $\mathscr{P}$ contains a decision reduct of $\mathscr{R}$, then

$$\mathrm{POS}_{\mathrm{Sim}(\mathscr{R})} R_d = \mathrm{POS}_{\mathrm{Sim}(\mathscr{P})} R_d.$$

By Lemma 6.3.3, we have for $x \in U$ that

$$\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow [x]_{R_d})(x) = ((\mathrm{Sim}(\mathscr{P}))\downarrow [x]_{R_d})(x),$$

thus, $(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \subseteq (\mathrm{Sim}(\mathscr{P}))\downarrow [x]_{R_d}$ and hence by Lemma 6.3.1 $(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \subseteq [x]_{R_d}$.

On the other hand, we always have

$$\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow [x]_{R_d})(x) \geq ((\mathrm{Sim}(\mathscr{P}))\downarrow [x]_{R_d})(x).$$

Now, if $(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \subseteq [x]_{R_d}$, then by Lemma 6.3.1, $(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \subseteq (\mathrm{Sim}(\mathscr{P}))\downarrow [x]_{R_d}$. This implies that

$$\lambda \leq ((\mathrm{Sim}(\mathscr{P}))\downarrow [x]_{R_d})(x).$$

By Lemma 6.3.3, we have $\mathrm{POS}_{\mathrm{Sim}(\mathscr{R})} R_d = \mathrm{POS}_{\mathrm{Sim}(\mathscr{P})} R_d$ and hence, $\mathscr{P}$ contains a decision reduct of $\mathscr{R}$. □

Note that $\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow [x]_{R_d})(x)$ depends on x. Since $\mathscr{P} \subset \mathscr{R}$, we have

$$(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \supseteq (x_\lambda)_{\mathrm{Sim}(\mathscr{R})}.$$

Keeping the positive region invariant can be reduced to keep the inclusion

$$(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \subseteq [x]_{R_d}$$

for every x in U and $\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow [x]_{R_d})(x)$. We can characterise a decision reduct of $\mathscr{R}$ in another way ([60]):

Lemma 6.3.5. Suppose $\mathscr{P} \subset \mathscr{R}$, then $\mathscr{P}$ contains a decision reduct of $\mathscr{R}$ if and only if for every $x \in U$ and $\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow [x]_{R_d})(x)$, it holds that if $(y_\lambda)_{\mathrm{Sim}(\mathscr{R})} \nsubseteq [x]_{R_d}$ for $y \in U$, then

$$\mathrm{Sim}(\mathscr{P})(y,x) \leq 1 - \lambda.$$

Proof. Fix $x \in U$ and $\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow [x]_{R_d})(x)$. Assume that $\mathscr{P}$ contains a decision reduct of $\mathscr{R}$, then $(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \subseteq [x]_{R_d}$. Take $y \in U$. If $(y_\lambda)_{\mathrm{Sim}(\mathscr{R})} \nsubseteq [x]_{R_d}$, then $(y_\lambda)_{\mathrm{Sim}(\mathscr{P})} \nsubseteq [x]_{R_d}$. Since $(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \neq (y_\lambda)_{\mathrm{Sim}(\mathscr{P})}$, we have that $(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \cap (y_\lambda)_{\mathrm{Sim}(\mathscr{P})} = \emptyset$. We conclude that $\mathrm{Sim}(\mathscr{P})(y,x) \leq 1 - \lambda$.

On the other hand, if $(y_\lambda)_{\mathrm{Sim}(\mathscr{R})} \nsubseteq [x]_{R_d}$, then $\mathrm{Sim}(\mathscr{P})(y,x) \leq 1 - \lambda$ implies that

$$(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \cap (y_\lambda)_{\mathrm{Sim}(\mathscr{P})} = \emptyset.$$

This means that $(x_\lambda)_{\mathrm{Sim}(\mathscr{P})} \subset [x]_{R_d}$ and by Lemma 6.3.4, we conclude that $\mathscr{P}$ contains a decision reduct of $\mathscr{R}$. □

By Lemma 6.3.1, we have

$$(y_\lambda)_{\mathrm{Sim}(\mathscr{R})} \nsubseteq [x]_{R_d} \Leftrightarrow (\mathrm{Sim}(\mathscr{R}){\downarrow}[x]_{R_d})(y) < \lambda.$$

By Lemma 6.3.5, we have that keeping the positive region of the decision attribute invariant is equivalent to keeping

$$\mathrm{Sim}(\mathscr{P})(y,x) \leq 1-\lambda$$

invariant for $(y_\lambda)_{\mathrm{Sim}(\mathscr{R})} \nsubseteq [x]_{R_d}$ and $\lambda = ((\mathrm{Sim}(\mathscr{R})){\downarrow}[x]_{R_d})(x)$. This can easily be applied as stopping criteria in an algorithm to compute decision reducts. So, $\mathscr{P}$ is a decision reduct of $\mathscr{R}$ if and only if $\mathscr{P}$ is the minimal subset of $\mathscr{R}$ satisfying the conditions of Lemma 6.3.4 and 6.3.5.

We are going to use the above discussion to develop a reduction algorithm. We do this by constructing a discernibility matrix and discernibility function. The discernibility matrix O of $(U, \mathscr{R} \cup \{R_d\})$ is an $n \times n$-matrix with $\forall i, j \in \{1,\ldots,n\}$ and for all $y_i, y_j \in U$:

$$O_{ij} = \begin{cases} \{R \mid 1-R(y_i,y_j) \geq \lambda_i\} & \lambda_i > \lambda_j \\ \emptyset & \text{otherwise} \end{cases}$$

with $\lambda_i = ((\mathrm{Sim}(\mathscr{R})){\downarrow}[y_i]_{R_d})(y_i)$, $\lambda_j = ((\mathrm{Sim}(\mathscr{R})){\downarrow}[y_i]_{R_d})(y_j)$. Note that O does not have to be symmetric and that O_{ii} is empty. We study what $R \in O_{ij}$ means:

$$R \in O_{ij} \Rightarrow \left[((y_i)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R})} \cap ((y_j)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R})} = \emptyset \Rightarrow ((y_i)_{\lambda_i})_R \cap ((y_j)_{\lambda_i})_R = \emptyset\right].$$

We check this equation. Assume that $R \in O_{ij}$ and that $((y_i)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R})} \cap ((y_j)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R})} = \emptyset$, i.e., $\mathscr{R}$ distinguishes y_i and y_j. Now assume that $((y_i)_{\lambda_i})_R \cap ((y_j)_{\lambda_i})_R \neq \emptyset$, then there is an element $x \in U$ such that

$$((y_i)_{\lambda_i})_R(x) = ((y_j)_{\lambda_i})_R(x) = \lambda_i.$$

This means that $1-R(x,y_i) < \lambda_i$ and $1-R(x,y_j) < \lambda_i$ and since R is a fuzzy similarity relation we have that $1-R(y_i,y_j) < \lambda_i$, i.e., $R \notin O_{ij}$. This is a contradiction. So, if y_i and y_j are distinguishable by all the attributes and $R \in O_{ij}$, then y_i and y_j are distinguishable by R.

Now, if $\lambda_i = ((\mathrm{Sim}(\mathscr{R})){\downarrow}[y_i]_{R_d})(y_i) = ((\mathrm{Sim}(\mathscr{P})){\downarrow}[y_i]_{R_d})(y_i)$ for $\mathscr{P} \subset \mathscr{R}$, then

$$((y_i)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R})} \cap ((y_j)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R})} = \emptyset \Rightarrow ((y_i)_{\lambda_i})_{\mathrm{Sim}(\mathscr{P})} \cap ((y_j)_{\lambda_i})_{\mathrm{Sim}(\mathscr{P})} = \emptyset,$$

which is equivalent to saying that $\mathscr{P}$ contains an element in O_{ij}. So, O_{ij} is the collection of conditional attributes that can keep

$$((y_i)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R})} \cap ((y_j)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R})} = \emptyset$$

for $\lambda_j < \lambda_i$.

We denote the Boolean variable associated with R_i by R_i^*, $i \in \{1,\dots,m\}$. We define the discernibility function f of $(U,\mathscr{R}\cup\{R_d\})$ by

$$f(R_1^*,\dots R_m^*) = \wedge\left\{\vee O_{ij}^* \mid O_{ij}\neq\emptyset, 1\le i,j\le n\right\}$$

with $O_{ij}^* = \{R_k^* \mid R_k \in O_{ij}, 1\le k\le m\}$. Note that f is a mapping from $\{0,1\}^m$ to I.

We discuss that f represents all decision reducts of $\mathscr{R}$. First, we characterise the core of $\mathscr{R}$ ([60]).

Lemma 6.3.6. We have

$$\mathrm{Core}(\mathscr{R}) = \{R \mid \exists i,j\in\{1,\dots,n\} : O_{ij} = \{R\}\}.$$

Proof. We have

$$\begin{aligned} R\in\mathrm{Core}(\mathscr{R}) &\Leftrightarrow \mathrm{POS}_{\mathrm{Sim}(\mathscr{R})}R_d \neq \mathrm{POS}_{\mathrm{Sim}(\mathscr{R}\setminus\{R\})}R_d \\ &\Leftrightarrow \exists y_i\in U : ((y_i)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R}\setminus\{R\})} \nsubseteq [y_i]_{R_d} \\ &\quad\text{and } \exists y_j\in U : ((y_j)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R})} \nsubseteq [y_i]_{R_d} \\ &\quad\text{and } ((y_i)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R}\setminus\{R\})} = ((y_j)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R}\setminus\{R\})} \\ &\Leftrightarrow 1-R'(y_i,y_j)<\lambda, \forall R'\neq R, \text{ and } ((y_i)_{\lambda_i})_R \neq ((y_j)_{\lambda_i})_R \\ &\Leftrightarrow O_{ij} = \{R\} \end{aligned}$$

with $\lambda_i = ((\mathrm{Sim}(\mathscr{R}))\downarrow[y_i]_{R_d})(y_i)$. □

The statement $O_{ij}=\{R\}$ implies that R is the unique attribute to ensure

$$((y_i)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R})} \cap ((y_j)_{\lambda_i})_{\mathrm{Sim}(\mathscr{R})} = \emptyset$$

for $\lambda_j < \lambda_i$.

This means that $\mathscr{P}\subset\mathscr{R}$ contains a decision reduct of $\mathscr{R}$ if and only if

$$\forall O_{ij}\neq\emptyset : \mathscr{P}\cap O_{ij}\neq\emptyset, \tag{6.5}$$

or, $\mathscr{P}$ is a decision reduct of $\mathscr{R}$ if and only if $\mathscr{P}$ is minimal for Equation (6.5).

Now let F be the disjunctive normal form of the discernibility function f, i.e., there is an $l\in\mathbb{N}$ and there are $\mathscr{R}_k\subseteq\mathscr{R}$, $1\le k\le l$ such that

$$F = (\wedge\mathscr{R}_1^*)\vee\dots\vee(\wedge\mathscr{R}_l^*)$$

where every element in $\mathscr{R}_k$ only appears one time. We have the following theorem ([60]).

Theorem 6.3.7.

$$\mathrm{Red}(\mathscr{R}) = \{\mathscr{R}_1,\dots,\mathscr{R}_l\}.$$

Proof. We first prove that every $\mathscr{R}_k$ is a reduct of $\mathscr{R}$. For every $k \in \{1,\dots,l\}$ and for every $O_{ij} \neq \emptyset$, $i,j \in \{1,\dots,n\}$, we have that $\wedge\mathscr{R}_k^* \leq \vee O_{ij}^*$ since

$$\bigvee_{r=1}^{l}(\wedge\mathscr{R}_k^*) = \wedge\{\vee O_{ij}^* \mid O_{ij} \neq \emptyset\}$$

and thus, $\mathscr{R}_k \cap O_{ij} \neq \emptyset$ for every $O_{ij} \neq \emptyset$. Let $\mathscr{R}_k' = \mathscr{R}_k \setminus \{R\}$, then

$$F < \left(\bigvee_{r=1}^{k-1}(\wedge\mathscr{R}_r^*)\right) \vee (\wedge\mathscr{R}_k'^*) \vee \left(\bigvee_{r=k+1}^{l}(\wedge\mathscr{R}_r^*)\right).$$

If for every $O_{ij} \neq \emptyset$ we have that $\mathscr{R}_k' \cap O_{ij} \neq \emptyset$ and thus $\wedge\mathscr{R}_k'^* \leq \vee O_{ij}^*$ then

$$F \geq \left(\bigvee_{r=1}^{k-1}(\wedge\mathscr{R}_r^*)\right) \vee (\wedge\mathscr{R}_k'^*) \vee \left(\bigvee_{r=k+1}^{l}(\wedge\mathscr{R}_r^*)\right)$$

which is a contradiction. Hence, there is an $O_{i_0j_0} \neq \emptyset$ such that $\mathscr{R}_k' \cap O_{i_0j_0} = \emptyset$. This means that $\mathscr{R}_k$ is indeed a decision reduct of $\mathscr{R}$.

Now take $\mathscr{X} \in \text{Red}(\mathscr{R})$. For every $O_{ij} \neq \emptyset$, $i,j \in \{1,\dots,n\}$, we have that $\mathscr{X} \cap O_{ij} \neq \emptyset$, so

$$f \wedge (\wedge\mathscr{X}^*) = \wedge(\vee O_{ij}^*) \wedge (\wedge\mathscr{X}^*) = \wedge\mathscr{X}^*.$$

This implies that $\wedge\mathscr{X}^* \leq f = F$. Suppose for every k that $\mathscr{R}_k \setminus \mathscr{X} \neq \emptyset$, then take for every k an $R_k \in \mathscr{R}_k \setminus \mathscr{X}$. We rewrite F such that

$$F = \left(\bigvee_{r=1}^{l} R_k^*\right) \wedge \dots$$

and thus $\wedge\mathscr{X}^* \leq \bigvee_{r=1}^{l} R_k^*$. So, there is an R_{k_0} such that $\wedge\mathscr{X}^* \leq R_{k_0}^*$, which implies that $R_{k_0} \in \mathscr{X}$. This is a contradiction. There has to be a $k_1 \in \{1,\dots,l\}$ such that $\mathscr{R}_{k_1} \cap \mathscr{X} = \emptyset$, which implies $\mathscr{R}_{k_1} \subseteq \mathscr{X}$, but since they are both decision reducts, we have $\mathscr{X} = \mathscr{R}_{k_1} \in \{\mathscr{R}_1,\dots,\mathscr{R}_l\}$. □

From this, we obtain that $\text{Core}(\mathscr{R}) = \cap\text{Red}(\mathscr{R})$. Assume $R \in \text{Core}(\mathscr{R})$, then there is an O_{ij} such that $O_{ij} = \{R\}$. Then for every reduct $\mathscr{R}_k$, $1 \leq k \leq l$, we have that $\mathscr{R}_k \cap O_{ij} \neq \emptyset$ and so, $R \in \mathscr{R}_k$ for $1 \leq k \leq l$. This means that $R \in \cap\text{Red}(\mathscr{R})$. Now take $R \in \cap\text{Red}(\mathscr{R})$, then for every decision reduct $\mathscr{R}_k$ we have that $R \in \mathscr{R}_k$. This means that

$$\text{POS}_{\text{Sim}(\mathscr{R}_k)} R_d = \text{POS}_{\text{Sim}(\mathscr{R})} R_d \text{ and } \text{POS}_{\text{Sim}(\mathscr{R}_k \setminus \{R\})} R_d < \text{POS}_{\text{Sim}(\mathscr{R}_k)} R_d,$$

hence,

$$\text{POS}_{\text{Sim}(\mathscr{R} \setminus \{R\})} R_d < \text{POS}_{\text{Sim}(\mathscr{R})} R_d.$$

By definition, we have that $R \in \text{Core}(\mathscr{R})$.

Before we give the algorithm, we note that if $O_{ij} \cap \text{Core}(\mathscr{R}) \neq \emptyset$, then $\{R^*\} \wedge (\vee O_{ij}^*) = \{R^*\}$ for $R \in O_{ij} \cap \text{Core}(\mathscr{R})$. So, when we compute F from f, we should only consider the elements in $\text{Core}(\mathscr{R})$ and O_{ij} satisfying $O_{ij} \cap \text{Core}(\mathscr{R}) = \emptyset$ to reduce the computations.

Algorithm 4 Reduction algorithm based on fuzzy rough sets

1. Compute $\mathrm{Sim}(\mathscr{R})$.
2. Compute $(\mathrm{Sim}(\mathscr{R}))\downarrow[x]_{R_d}$ for every $x \in U$.
3. Compute O_{ij}: if $\lambda_j < \lambda_i$, then $O_{ij} = \{R \mid 1 - R(y_i, y_j) \geq \lambda_i\}$, otherwise, $O_{ij} = \emptyset$.
4. Compute the core as a collection of those O_{ij} with single element.
5. Delete those $O_{ij} = \emptyset$ or O_{ij} with non-empty overlap with the core.
6. Define $f = \wedge\{\vee O^*_{ij}\}$ with the O_{ij} left after the previous step.
7. Compute $F = (\wedge\mathscr{R}^*_1) \vee \ldots \vee (\wedge\mathscr{R}^*_l)$ from f.
8. Return all decision reducts $\mathscr{R}_1, \ldots, \mathscr{R}_l$.

Let U be a universe and d the decision attribute. Let $\lambda_i = ((\mathrm{Sim}(\mathscr{R}))\downarrow[y_i]_{R_d})(y_i)$ and $\lambda_j = ((\mathrm{Sim}(\mathscr{R}))\downarrow[y_i]_{R_d})(y_j)$ for $y_i, y_j \in U$, then we can construct Algorithm 4.

We study now what happens if we use the general fuzzy rough set model with a left-continuous t-norm and its R-implicator.

Using a left-continuous t-norm and its R-implicator

Chen et al. did something similar, but now they used the general fuzzy rough set model with a left-continuous t-norm $\mathscr{T}$ and its R-implicator $\mathscr{I}$ ([6, 7]). We have the same concepts as in the setting where we used Dubois and Prade's model, only the positive region of R_d relative to the fuzzy similarity $\mathrm{Sim}(\mathscr{R})$ is now defined by

$$\mathrm{POS}_{\mathrm{Sim}(\mathscr{R})} R_d = \bigcup_{x \in U} (\mathrm{Sim}(\mathscr{R}))\downarrow_{\mathscr{I}} [x]_{R_d}.$$

Note that in this setting, we can work with fuzzy $\mathscr{T}$-similarity relations instead of fuzzy min-similarity relations. We again want to know when $\mathscr{P} \subset \mathscr{R}$ contains a decision reduct of $\mathscr{R}$.

We first describe the basic granules ([7]). If $\lambda \in]0,1]$, then x_λ is a fuzzy point.

Lemma 6.3.8. Let R be a fuzzy $\mathscr{T}$-similarity relation and A a fuzzy setting in U, then

$$\begin{aligned} R\downarrow_{\mathscr{I}} A &= \cup\{R\uparrow_{\mathscr{T}}(x_\lambda) \mid R\uparrow_{\mathscr{T}}(x_\lambda) \subseteq A\}, \\ R\uparrow_{\mathscr{T}} A &= \cup\{R\uparrow_{\mathscr{T}}(x_\lambda) \mid x_\lambda \subseteq A\}. \end{aligned} \tag{6.6}$$

Proof. Recall that $\mathscr{T}$ is left-continuous. Fix $x \in U$ and $\lambda \in]0,1]$. To prove the first equality, we prove that

$$R\uparrow_{\mathscr{T}}(x_\lambda) \subseteq R\downarrow_{\mathscr{I}} A \Leftrightarrow R\uparrow_{\mathscr{T}}(x_\lambda) \subseteq A.$$

Take $z \in U$, then

$$\begin{aligned}
&(R\uparrow_{\mathcal{T}}(x_\lambda))(z) \leq (R\downarrow_{\mathcal{I}}A)(z) \\
&\Leftrightarrow \sup_{y\in U} \mathcal{T}(R(y,z), x_\lambda(y)) \leq \inf_{y\in U} \mathcal{I}(R(y,z), A(y)) \\
&\Leftrightarrow \mathcal{T}(R(x,z),\lambda) \leq \inf_{y\in U} \mathcal{I}(R(y,z), A(y)) \\
&\Leftrightarrow \forall y \in U : \mathcal{T}(R(x,z),\lambda) \leq \mathcal{I}(R(y,z), A(y)) \\
&\Leftrightarrow \forall y \in U : \mathcal{T}(\mathcal{T}(R(x,z),\lambda), R(y,z)) \leq A(y) \\
&\Leftrightarrow \forall y \in U : \mathcal{T}(\mathcal{T}(R(x,z), R(z,y)),\lambda) \leq A(y) \\
&\Leftrightarrow \forall y \in U : \mathcal{T}(R(x,y),\lambda) \leq A(y) \\
&\Leftrightarrow \forall y \in U : \sup_{u\in U} \mathcal{T}(R(u,y), x_\lambda(u)) \leq A(y) \\
&\Leftrightarrow R\downarrow_{\mathcal{T}}(x_\lambda) \subseteq A
\end{aligned}$$

where we used the the residual principle in the fourth step.

The second equality follows from the fact that

$$A = \bigcup\{x_\lambda \mid \lambda \in]0,1], \lambda \leq A(x)\} = \bigcup\{x_\lambda \mid \lambda \in]0,1], x_\lambda \subseteq A\}$$

and the fact that the upper approximation of a union is equal to the union of the upper approximations ([17]). The latter holds by Proposition 4.1.8 and by the fact that $\mathcal{T}$ is complete-distributive w.r.t the supremum. □

This means we can use the set $\{R\uparrow_{\mathcal{T}}(x_\lambda) \mid x \in U, \lambda \in]0,1]\}$ as basic granules. Now, take x and y in U. If $y \notin [x]_{R_d}$, then clearly

$$(R\downarrow_{\mathcal{I}}[x]_{R_d})(y) \leq \mathcal{I}(R(y,y), R_d(y,x)) = 0.$$

Now, for $y \in [x]_{R_d}$, we have the following lemma ([7]).

Lemma 6.3.9. Suppose $y \in [x]_{R_d}$, then we have that

$$R\uparrow_{\mathcal{T}}(y_\lambda) \subseteq R\downarrow_{\mathcal{I}}[x]_{R_d} \Leftrightarrow \forall z \notin [x]_{R_d} : (R\uparrow_{\mathcal{T}}(y_\lambda))(z) = 0.$$

Proof. Take $x, y \in U$ such that $y \in [x]_{R_d}$. If $R\uparrow_{\mathcal{T}}(y_\lambda) \subseteq R\downarrow_{\mathcal{I}}[x]_{R_d}$, then for $z \notin [x]_{R_d}$ we have $(R\downarrow_{\mathcal{I}}[x]_{R_d})(z) = 0$, hence $(R\uparrow_{\mathcal{T}}(y_\lambda))(z) = 0$.

On the other hand, suppose for all $z \notin [x]_{R_d}$ that $(R\uparrow_{\mathcal{T}}(y_\lambda))(z) = 0$. Since for all $u \in [x]_{R_d}$ it holds that $[x]_{R_d}(u) = 1$, we have

$$(R\uparrow_{\mathcal{T}}(y_\lambda))(u) \leq ([x]_{R_d})(u)$$

and thus $R\uparrow_{\mathcal{T}}(y_\lambda) \subseteq [x]_{R_d}$. By Equation (6.6) we have that $R\uparrow_{\mathcal{T}}(y_\lambda) \subseteq R\downarrow_{\mathcal{I}}[x]_{R_d}$. □

Note that since $y_\lambda \subseteq R\uparrow_{\mathcal{T}}(y_\lambda)$, we obtain the following equivalence from Lemma 6.3.9:

$$y_\lambda \subseteq R\downarrow_{\mathcal{I}}[x]_{R_d} \Leftrightarrow \forall z \notin [x]_{R_d} : (R\uparrow_{\mathcal{T}}(y_\lambda))(z) = 0.$$

We can now characterise decision reducts ([7]).

Lemma 6.3.10. Suppose $\mathscr{P} \subset \mathscr{R}$, then $\mathscr{P}$ contains a decision reduct of $\mathscr{R}$ if and only if for every $x \in U$:

$$(\mathrm{Sim}(\mathscr{P}))\uparrow_{\mathcal{T}} x_\lambda \subseteq [x]_{R_d},$$

with $\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow_{\mathcal{I}}[x]_{R_d})(x)$.

Proof. Take $x, y \in U$. We either have $[x]_{R_d} = [y]_{R_d}$ or $[x]_{R_d} \cap [y]_{R_d} = \emptyset$. So keeping

$$\mathrm{POS}_{\mathrm{Sim}(\mathscr{R})} R_d = \mathrm{POS}_{\mathrm{Sim}(\mathscr{P})} R_d$$

invariant is the same as keeping

$$(\mathrm{Sim}(\mathscr{R}))\downarrow_{\mathcal{I}}[x]_{R_d} = (\mathrm{Sim}(\mathscr{P}))\downarrow_{\mathcal{I}}[x]_{R_d}$$

invariant for every $x \in U$. By Equation (6.6) and Lemma 6.3.9, this latter statement is equivalent to

$$\forall y \in [x]_{R_d} : (\mathrm{Sim}(\mathscr{P}))\uparrow_{\mathcal{T}} y_\lambda \subseteq [x]_{R_d}$$

which is equivalent to

$$(\mathrm{Sim}(\mathscr{P}))\uparrow_{\mathcal{T}} x_\lambda \subseteq [x]_{R_d}$$

since $y \in [x]_{R_d}$ implies $[x]_{R_d} = [y]_{R_d}$. □

Note that λ depends on x. This lemma can be used to give us two other characterisations ([7]).

Lemma 6.3.11. Suppose $\mathscr{P} \subset \mathscr{R}$, then $\mathscr{P}$ contains a decision reduct of $\mathscr{R}$ if and only if for every $x, z \in U$:

$$\forall z \notin [x]_{R_d} : ((\mathrm{Sim}(\mathscr{P}))\uparrow_{\mathcal{T}} x_\lambda)(z) = 0,$$

with $\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow_{\mathcal{I}}[x]_{R_d})(x)$.

Proof. This follows from Equation 6.6, Lemma 6.3.9 and Lemma 6.3.10. □

Lemma 6.3.12. Suppose $\mathscr{P} \subset \mathscr{R}$, then $\mathscr{P}$ contains a decision reduct of $\mathscr{R}$ if and only if there exists a $P \in \mathscr{P}$ such that $\mathcal{T}(P(x,z), \lambda) = 0$ for every $x, z \in U$ and $z \notin [x]_{R_d}$ and $\lambda = ((\mathrm{Sim}(\mathscr{R}))\downarrow_{\mathcal{I}}[x]_{R_d})(x)$.

Proof. Take $x, z \in U$ such that $z \notin [x]_{R_d}$. We obtain

$$\begin{aligned}((\mathrm{Sim}(\mathscr{P}))\uparrow_{\mathscr{T}} x_\lambda)(z) &= \sup_{y\in U} \mathscr{T}(\mathrm{Sim}(\mathscr{P})(y,z), x_\lambda(y)) \\ &= \mathscr{T}(\mathrm{Sim}(\mathscr{P})(x,z),\lambda) \\ &= \min\{\mathscr{T}(P(x,z),\lambda) \mid P \in \mathscr{P}\}.\end{aligned}$$

The statement follows now from Lemma 6.3.11. □

Clearly $\mathscr{P}$ is a decision reduct of $\mathscr{R}$ if and only if $\mathscr{P}$ is minimal for the conditions in Lemma 6.3.11 and 6.3.12. This last characterisation can easily be used to design an algorithm to compute all decision reducts. We do this by constructing the discernibility matrix and function of the decision system $(U, \mathscr{R} \cup \{R_d\})$. We assume that $|U| = n$ and $|\mathscr{R}| = m$. The discernibility matrix O is an $n \times n$-matrix with the (i, j)-th entry defined by

$$O_{ij} = \begin{cases} \{R \in \mathscr{R} \mid \mathscr{T}(R(y_i, y_j), \lambda_i) = 0\} & y_j \notin [y_i]_{R_d} \\ \emptyset & \text{otherwise} \end{cases}$$

with $y_i, y_j \in U$, $1 \leq i, j \leq n$ and $\lambda_i = ((\mathrm{Sim}(\mathscr{R}))\downarrow_{\mathscr{I}}[y_i]_{R_d})(y_i)$. The matrix does not have to be symmetric and O_{ii} can be empty. O_{ij} is the collection of conditional attributes such that

$$(R\uparrow_{\mathscr{T}}(y_i)_{\lambda_i})(y_j) = 0$$

for $y_j \notin [y_i]_{R_d}$. The discernibility function f is constructed in the same way as before. If we denote the Boolean variable associated with R_i by R_i^*, $i \in \{1, \ldots, m\}$, then the discernibility function f of $(U, \mathscr{R} \cup \{R_d\})$ is the function

$$f(R_1^*, \ldots R_m^*) = \wedge \left\{\vee O_{ij}^* \mid O_{ij} \neq \emptyset, 1 \leq i, j \leq n\right\}$$

with $O_{ij}^* = \{R_k^* \mid R_k \in O_{ij}, 1 \leq k \leq m\}$. Again, f is a mapping from $\{0,1\}^m$ to I.

Now, f represents all decision reducts of $\mathscr{R}$. We can characterise the core of $\mathscr{R}$.

Lemma 6.3.13. We have

$$\mathrm{Core}(\mathscr{R}) = \{R \mid \exists i, j \in \{1, \ldots, n\} : O_{ij} = \{R\}\}.$$

Proof. We have

$$\begin{aligned}R \in \mathrm{Core}(\mathscr{R}) &\Leftrightarrow \mathrm{POS}_{\mathrm{Sim}(\mathscr{R})} R_d \neq \mathrm{POS}_{\mathrm{Sim}(\mathscr{R}\setminus\{R\})} R_d \\ &\Leftrightarrow \exists y_i, y_j \in U : \mathscr{T}(R(y_i, y_j), \lambda_i) = 0 \\ &\qquad \text{and } \forall R' \neq R : \mathscr{T}(R'(y_i, y_j), \lambda_i) > 0 \\ &\Leftrightarrow O_{ij} = \{R\}\end{aligned}$$

with $\lambda_i = ((\mathrm{Sim}(\mathscr{R}))\downarrow[y_i]_{R_d})(y_i)$. The statement $O_{ij} = \{R\}$ implies that R is the unique attribute to maintain $\mathscr{T}(R(y_i, y_j), \lambda_i) = 0$. □

This means that $\mathscr{P} \subset \mathscr{R}$ contains a decision reduct of $\mathscr{R}$ if and only if

$$\forall O_{ij} \neq \emptyset : \mathscr{P} \cap O_{ij} \neq \emptyset, \tag{6.7}$$

or, $\mathscr{P}$ is a desicion reduct of $\mathscr{R}$ if and only if $\mathscr{P}$ is minimal for Equation (6.7).

Now let F be the disjunctive normal form of the discernibility function f, i.e., there is an $l \in \mathbb{N}$ and there are $\mathscr{R}_k \subseteq \mathscr{R}$, $1 \leq k \leq l$ such that

$$F = (\wedge \mathscr{R}_1^*) \vee \ldots \vee (\wedge \mathscr{R}_l^*),$$

where every element in $\mathscr{R}_k$ only appears one time. We have the following theorem.

Theorem 6.3.14.

$$\text{Red}(\mathscr{R}) = \{\mathscr{R}_1, \ldots, \mathscr{R}_l\}.$$

Proof. The proof is the same as the proof of Theorem 6.3.7. □

As in the approach of Tsang et al. ([60]), we have that

$$\text{Core}(\mathscr{R}) = \cap \text{Red}(\mathscr{R}).$$

As before, we should only consider the elements in Core($\mathscr{R}$) and O_{ij} satisfying $O_{ij} \cap \text{Core}(\mathscr{R}) = \emptyset$ to reduce the computations.

Let U be a universe and d the decision attribute. With $\lambda_i = (\text{Sim}(\mathscr{R}))\downarrow[y_i]_{R_d})(y_i)$, we can construct algorithm 5 (see [6]). As we see, this is the same as Algorithm 4, only step 2 and 3 differ,

Algorithm 5 Reduction algorithm based on fuzzy rough sets 2

1. Compute Sim($\mathscr{R}$).
2. Compute $(\text{Sim}(\mathscr{R}))\downarrow_{\mathscr{I}}[x]_{R_d}$ for every $x \in U$.
3. Compute O_{ij}: if $y_j \notin [y_i]_{R_d}$, then $O_{ij} = \{R \mid \mathscr{T}(R(y_i, y_j), \lambda_i) = 0\}$, otherwise, $O_{ij} = \emptyset$.
4. Compute the core as a collection of those O_{ij} with single element.
5. Delete those $O_{ij} = \emptyset$ or O_{ij} with non-empty overlap with the core.
6. Define $f = \wedge\{\vee O_{ij}^*\}$ with the O_{ij} left after the previous step.
7. Compute $F = (\wedge \mathscr{R}_1^*) \vee \ldots \vee (\wedge \mathscr{R}_l^*)$ from f.
8. Return all decision reducts $\mathscr{R}_1, \ldots, \mathscr{R}_l$.

because we work with another fuzzy rough set model and we have found another criterium to define O.

We continue with discussing some relations between decision reducts.

6.3.2 Relations between decision reducts

We saw two approaches of how we can construct an algorithm to find all decision reducts. Zhao and Tsang ([69]) give us some relations between different decision reducts. We have the following set-up: a fuzzy decision system $(U, \mathscr{A} \cup \mathscr{D})$ with U the universe of the objects, $\mathscr{A}$ the set of conditional attributes and $\mathscr{D}$ the set of decision attributes, which in this case are all symbolic. Every subset $B \subseteq \mathscr{A}$ can be described by a fuzzy similarity relation R_B: for $x, y \in U$, $R_B(x, y)$ is given by

$$R_B(x, y) = \min\{R_a(x, y) \mid a \in B\},$$

as seen before. Let $\mathscr{I}$ be an implicator, then the positive region of B in x is given by

$$(\mathrm{POS}_B(C))(x) = \sup_{y \in U}(R_B \downarrow_{\mathscr{I}} [y]_{R_C})(x)$$

with $C \subseteq \mathscr{D}$ and $R_C(x, y) = \min\{R_d(x, y) \mid d \in C\}$. Since U is finite, the positive region of B reaches its maximum membership degree in a certain point $z \in U$ and as seen before, we have

$$(\mathrm{POS}_B(C))(x) = (R_B \downarrow_{\mathscr{I}} [x]_{R_C})(x).$$

We work again with the dependency degree of C on B:

$$\gamma_B(C) = \frac{|\,\mathrm{POS}_B(C)|}{|U|}.$$

Since the general fuzzy rough set model is monotone with respect to fuzzy sets, the positive region is also monotone with respect to fuzzy sets, i.e., if $B_1 \subseteq B_2 \subseteq \mathscr{A}$ and $C \subseteq \mathscr{D}$, then

$$\mathrm{POS}_{B_1}(C) \subseteq \mathrm{POS}_{B_2}(C).$$

Before we can study relations between decision reducts, we need the following two definitions. By Red_i, we denote the type (or set) of decision reducts obtained in model i.

Definition 6.3.15. Given two types of decision reducts, i.e., Red_1 and Red_2, that are obtained by two different fuzzy approximation operators. If

$$\forall B_1 \in \mathrm{Red}_1\, \exists B_2 \in \mathrm{Red}_2 \text{ such that } B_1 \subseteq B_2,$$
$$\forall B_3 \in \mathrm{Red}_2\, \exists B_4 \in \mathrm{Red}_1 \text{ such that } B_4 \subseteq B_3,$$

then we say that the type of decision reducts Red_1 *is included* by the type of decision reducts Red_2 or Red_2 *includes* Red_1.

We also want to know when two types of decision reducts are identical.

Definition 6.3.16. Given two types of decision reducts, i.e., Red_1 and Red_2, that are obtained by two different fuzzy approximation operators. If

$$\forall B_1 \in \text{Red}_1 \text{ it holds that } B_1 \in \text{Red}_2,$$
$$\forall B_2 \in \text{Red}_2 \text{ it holds that } B_2 \in \text{Red}_1,$$

then we say that the type of decision reducts Red_1 and the type of decision reducts Red_2 are *identical*. We denote this by $\text{Red}_1 = \text{Red}_2$.

We discuss some relations between different types of decision reducts. We will only give the results, the proofs can be found in [69]. The first two properties gives some information about decision reducts found by an S-implicator and decision reducts found by an R-implicator.

Proposition 6.3.17. Let $\mathscr{S}$ be a t-conorm and $\mathscr{I}_{\mathscr{S}}$ its S-implicator. Let $\mathscr{T}$ be a t-norm and $\mathscr{I}_{\mathscr{T}}$ its R-implicator. Let Red_1 be obtained by the fuzzy approximation operator $R{\downarrow}_{\mathscr{I}_{\mathscr{S}}}$ and let Red_2 be obtained by the fuzzy approximation operator $R{\downarrow}_{\mathscr{I}_{\mathscr{T}}}$. If $\mathscr{S}$ is the dual t-conorm of $\mathscr{T}$ w.r.t. the standard negator, then Red_2 includes Red_1.

If this t-norm is the Łukasiewicz t-norm, then both types are identical.

Proposition 6.3.18. Let $\mathscr{T}$ be the Łukasiewicz t-norm $\mathscr{T}_L$ and $\mathscr{S}$ its dual t-conorm w.r.t. the standard negator. Let Red_1 be obtained by the fuzzy approximation operator $R{\downarrow}_{\mathscr{I}_{\mathscr{S}_L}}$ and let Red_2 be obtained by the fuzzy approximation operator $R{\downarrow}_{\mathscr{I}_{\mathscr{T}_L}}$, then Red_1 and Red_2 are identical.

The following two theorems show how a t-norm can influence the attribute reductions. Let $x \in U$ and $C \subseteq \mathscr{D}$.

Proposition 6.3.19. Let $\mathscr{S}_1$ and $\mathscr{S}_2$ be two t-conorms. If Red_1 is obtained by the fuzzy approximation $(R{\downarrow}_{\mathscr{I}_{\mathscr{S}_1}}[x]_{R_c})(x)$ and Red_2 is obtained by the fuzzy approximation $(R{\downarrow}_{\mathscr{I}_{\mathscr{S}_2}}[x]_{R_c})(x)$, then Red_1 and Red_2 are identical.

Proposition 6.3.20. Let $\mathscr{T}_1$ and $\mathscr{T}_2$ be t-norms. If Red_1 is obtained by the fuzzy approximation $(R{\downarrow}_{\mathscr{I}_{\mathscr{T}_1}}[x]_{R_c})(x)$ and Red_2 is obtained by the fuzzy approximation $(R{\downarrow}_{\mathscr{I}_{\mathscr{T}_2}}[x]_{R_c})(x)$, and we have for all $a, b \in I$ that

$$\mathscr{I}_{\mathscr{T}_1}(a,0) = \mathscr{I}_{\mathscr{T}_1}(b,0) \Rightarrow a = b,$$
$$\mathscr{I}_{\mathscr{T}_2}(a,0) = \mathscr{I}_{\mathscr{T}_2}(b,0) \Rightarrow a = b,$$

then Red_1 and Red_2 are identical.

If $\mathscr{I}_{\mathscr{T}_2}$ does not fulfil the condition, but the other conditions are fulfilled, then Red_2 includes Red_1.

We end with a chronological overview of authors that use fuzzy rough sets for feature selection.

6.4 A chronological overview of fuzzy rough feature selection

The first to apply fuzzy rough sets to feature selection was Kuncheva ([39], 1992). However, her work is largely disconnected from the mainstream literature on the subject, both because of the rough set model used and the assumptions that are made about the data. She assumes that the data is characterised by a weak fuzzy partition[3] of U, i.e., a family $\mathscr{P} = \{P_1, \ldots, P_k\}$ of fuzzy sets in U such that $\bigcup_{i=1}^{k} \mathrm{supp}(P_i) = U$. This is called the *a priori classification* of the data.

Each subset B of the set of attributes $\mathscr{A}$ is assumed to induce a weak fuzzy partition $\mathscr{P}_B = \{B_1, \ldots, B_l\}$ of U, with l not necessarily equal to $|\mathscr{P}|$.

The fuzzy rough set model used by Kuncheva uses an inclusion measure, i.e., a mapping

$$\mathrm{Inc}\colon \mathscr{F}(U)^2 \to I$$

that evaluates the degree to which one fuzzy set is included into another one, as well as two thresholds λ_1 and λ_2 in I such that $\lambda_1 > \lambda_2$. Some examples of inclusion measures were discussed in Section 3.4.2.

Given a weak fuzzy partition $\mathscr{P} = \{P_1, \ldots, P_k\}$ of U, Kuncheva defined the lower approximation of a fuzzy set A in U by

$$R{\downarrow}_{\mathscr{P},\lambda_1} A = \bigcup_{\mathrm{Inc}(P_i,A) \geq \lambda_1} P_i.$$

The boundary region is given by

$$\mathrm{BNR}_{\mathscr{P},\lambda_1,\lambda_2} A = \bigcup_{\lambda_2 < \mathrm{Inc}(P_i,A) < \lambda_1} P_i.$$

To measure the quality of the approximation of the a priori classification by means of the attribute subset B, Kuncheva used the measure

$$\sum_{i=1}^{n} w_i \nu_{B,\lambda_1,\lambda_2}(P_i)$$

with $W = \langle w_1, \ldots, w_n \rangle$ a weight vector and

$$\nu_{B,\lambda_1,\lambda_2}(P_i) = \frac{1}{2}\left(\mathrm{SIM}(R{\downarrow}_{\mathscr{P}_B,\lambda_1} P_i, P_i) + 1 - \mathrm{SIM}\left(\mathrm{BNR}_{\mathscr{P}_B,\lambda_1,\lambda_2} P_i, P_i\right)\right)$$

where SIM is a similarity measure, i.e., a $\mathscr{F}(U)^2 \to I$ mapping that evaluates to what extent two fuzzy sets are similar.

A lot of pioneering work on fuzzy rough feature selection in the first half of the 2000's was done by Jensen and Shen. In [34] (and [35, 36, 58]) they proposed a reduction method based on fuzzy extensions of the positive region and the dependency measure based on fuzzy lower

[3] This is not the same as a $\mathscr{T}$-semipartition defined in Chapter 2.

approximations. However, in [60] it was noticed that there are problems with Jensen and Shen's approach. Before that, Bhatt and Gopal already had stated some problems with the approach of Jensen and Shen ([3, 4, 5]).

In [32], Hu et al. assumed that for every subset B of attributes, there exists a fuzzy similarity relation R_B. The fuzzy rough set model they use is the one designed by Dubois and Prade. They base the definition of a decision reduct on the positive region POS_B and the degree of dependency γ_B. They also introduce the *conditional entropy* $H(d|B)$ of the decision attribute d relative to B:

$$H(d|B) = -\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{|R_d x_i \cap R_B x_i|}{|R_B x_i|}\right).$$

They prove that B is a decision reduct if $H(d|B) = H(d|\mathscr{A})$ and

$$H(d|B \setminus \{a\}) > H(d|\mathscr{A})$$

for all a in B.

In a second approach, Hu et al. ([31]) assumed that each conditional attribute a generates a fuzzy similarity relation R_a in U and that $R_B = \bigcap_{a\in B} R_a$ for $B \subseteq \mathscr{A}$. Furthermore, they assumed the decision attribute d categorical, thus it induces a crisp equivalence relation in U. This leads to a partition of U. Given a fuzzy set A in U, a fuzzy similarity relation R in U, $0 \leq l < 0.5 < u \leq 1$, the approximations of A by R are given by the VQFRS model[4] with the couple of fuzzy quantifiers $(Q_{\geq u}, Q_{>l})$.

A very important paper from theoretical point of view, is by Tsang et al. ([60]). The approaches of Chen et al. ([6, 7]) and Zhao and Tsang ([69]) are also based on the general fuzzy rough set model. We studied these three approaches in Section 6.3.

Cornelis and Jensen ([14]) applied the VQFRS model to feature selection, but since the approximation operators defined by this model are not monotone w.r.t. the fuzzy relation, adding more attributes does not necessarily increase the positive region. This can give problems when applying the QuickReduct algorithm (see Algorithm 1 and 3).

In Jensen and Shen's second approach ([37]) three subset quality measures are presented. We discussed these measures in Section 6.2, just like the approach of Cornelis et al. ([15]) that defines an alternative definition for the positive region of a attribute subset B and an alternative measure for the degree of dependency γ_B.

Chen and Zhao ([10]) focused on a specific subset of decision classes (local reduction), instead of keeping the full positive region invariant (global reduction).

Chen et al. ([9]) used the definition of a decision reduct for fuzzy rough sets from [60]. They provided a fast algorithm to obtain one decision reduct, based on a procedure to find the minimal elements of the fuzzy discernibility matrix. The execution time is a lot faster then the proposals in [37] and [60].

[4]They did not make the link with the VQFRS model, since that model did not exist at the moment.

Currently, they are some recent papers about the subject: e.g., Derrac et al. ([18]) combined fuzzy rough feature selection with evolutionary instance selection, Chen et al. ([8]) considered feature selection with kernelised fuzzy rough sets and He and Wu ([25]) developed a new method to compute membership for fuzzy support vector machines by using Gaussian kernel-based fuzzy rough sets.

Chapter 7

Conclusion

In this thesis, we have seen that fuzzy rough set theory provides us with good techniques to construct algorithms for feature selection. We have introduced a general fuzzy rough set model with an implicator $\mathscr{I}$ and a conjunctor $\mathscr{C}$, that covers a lot of fuzzy rough set models in the literature. With the right choices for $\mathscr{I}$ and $\mathscr{C}$ and the fuzzy relation R, this model fulfils all the properties of the original rough set model of Pawlak. We can refine this model in a natural way, by using tight and loose approximation operators. We have also shown that it is very useful in applications such as feature selection.

Furthermore, we have studied some robust models. The soft fuzzy rough set model turns out to be ill-defined. Studying the properties of the variable precision fuzzy rough set model is very difficult, due to the complex definition of the model. Further study is required. We have shown that the OWA-based fuzzy rough set model is related to the vaguely quantified fuzzy rough set model (VQFRS) by using quantifiers to determine the weight vectors. The main advantage of the OWA-based fuzzy rough set model is that it is monotone with respect to fuzzy relations, a property that is not fulfilled by the VQFRS model. The OWA-based fuzzy rough set model also covers fuzzy rough set models based on robust nearest neighbour. Further work will be to study more properties of fuzzy rough set models and find connections between them. Defining new robust models is also a big challenge.

In Chapter 5, we saw that the properties of approximation operators and the properties of fuzzy relations are strongly related. This can help us to define new fuzzy rough set models. Another open problem is to develop axiomatic approaches for robust fuzzy rough set models.

Another important challenge is to find good approaches to use robust models in feature selection. Developing new algorithms will also be a subject of future research. For example, we want to construct an algorithm to determine all decision reducts for a fuzzy tolerance relation instead of a fuzzy similarity relation.

Bibliography

[1] B. De Baets and J. Fodor. Residual operators of uninorms. In Antonio Di Nola, editor, *Soft Computing 3*, pages 89–100. Springer-Verlag, 1999.

[2] B. De Baets and R. Mesiar. T-Partitions. *Fuzzy Sets and Systems*, 97:211–223, 1998.

[3] R.B. Bhatt and M.Gopal. Improved feature selection algorithm with fuzzy-rough sets on compact computational domain. *International Journal of General Systems*, 34(4):485–505, 2005.

[4] R.B. Bhatt and M.Gopal. On fuzzy-rough sets approach to feature selection. *Pattern Recognition Letters*, 26(7):965–975, 2005.

[5] R.B. Bhatt and M.Gopal. On the compact computational domain of fuzzy-rough sets. *Pattern Recognition Letters*, 26(11):1632–1640, 2005.

[6] D. Chen, E. Tsang, and S. Zhao. An approach of attributes reduction based on fuzzy T_L-rough sets. In *Proceedings IEEE International Conference on Systems, Man and Cybernites*, pages 486–491, 2007.

[7] D. Chen, E. Tsang, and S. Zhao. Attribute reduction based on fuzzy rough sets. In *Proceedings International Conference on Rough Sets and Intelligent Systems Paradigms, Lecture Notes in Computer Science 4585*, pages 73–89, 2007.

[8] D.G. Chen, Q.H. Hu, and Y.P. Yang. Parameterized attribute reduction with Gaussian kernel based fuzzy rough sets. *Information Sciences*, 181(23):5169–5179, 2011.

[9] D.G. Chen, L. Zhang, S.Y. Zhao, Q.H. Hu, and P.F. Zhu. A novel algorithm for finding reducts with fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, 20(2):385–389, 2012.

[10] D.G. Chen and S.Y. Zhao. Local reduction of decision system with fuzzy rough sets. *Fuzzy Sets and Systems*, 161(13):1871–1883, 2010.

[11] M. De Cock, C. Cornelis, and E.E. Kerre. Fuzzy rough sets: the forgotten step. *IEEE Transactions on Fuzzy Systems*, 15(1):121–130, 2007.

[12] C. Cornelis, M. De Cock, and A.M. Radzikowska. Vaguely quantified rough sets. In *Proceedings of 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC2007)*, pages 87–94, 2007.

[13] C. Cornelis, M. De Cock, and A.M. Radzikowska. Fuzzy rough sets: from theory into practice. In W. Pedrycz, A. Skowron, and V. Kreinovich, editors, *Handbook of Granular Computing*, pages 533–552. John Wiley and Sons, 2008.

[14] C. Cornelis and R. Jensen. A noise-tolerant approach to fuzzy-rough feature selection. In *Proceedings of the 2008 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2008)*, pages 1598–1605, 2008.

[15] C. Cornelis, G. Hurtado Martín, R. Jensen, and D. Ślęzak. Attribute selection with fuzzy decision reducts. *Information Sciences*, 180(2):209–224, 2010.

[16] C. Cornelis, N. Verbiest, and R. Jensen. Ordered weighted average based fuzzy rough sets. In *Proceedings of the 5th International Conference on Rough Sets and Knowledge Technology (RSKT2010)*, pages 78–85, 2010.

[17] C. Degang, Z. Wenxiu, D.S. Yeung, and E.C.C. Tsang. Rough approximations on a complete completely distributive lattice with applications to generalized rough sets. *Information Sciences*, 176:1829–1848, 2006.

[18] J. Derrac, C. Cornelis, S. García, and F. Herrera. Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection. *Information Sciences*, 186(1):73–92, 2012.

[19] D. Dubois and H. Prade. Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*, 17:191–209, 1990.

[20] D. Dubois and H. Prade. Putting fuzzy sets and rough sets together. In R. Słowiński, editor, *Intelligent Decision Support - Handbook of Applications and Advances of the Rough Sets Theory*, pages 203–232. Kluwer Academic Publishers, 1992.

[21] F. Esteva and L. Godo. Monoidal t-norm based logic: towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124:271–288, 2001.

[22] T.F. Fan, C.J. Liau, and D.R. Liu. Variable precision fuzzy rough set based on relative cardinality. In *Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS2012)*, pages 43–47, 2012.

[23] J. Fodor. Left-continuous t-norms in fuzzy logic: an overview. *Journal of applied sciences at Budapest Tech Hungary*, 1(2), 2004.

[24] S. Gottwald and S. Jenei. A new axiomatization for involutive monoidal t-norm-based logic. *Fuzzy Sets and Systems*, 124:303–307, 2001.

[25] Q. He and C.X. Wu. Membership evalution and feature selection for fuzzy support vector machine based on fuzzy rough sets. *Soft Computing*, 15(6):1105–1114, 2011.

[26] Q. Hu, S. An, and D. Yu. Soft fuzzy rough sets for robust feature evaluation and selection. *Information Sciences*, 180:4384–4400, 2010.

[27] Q. Hu, S. An, X. Yu, and D. Yu. Robust fuzzy rough classifiers. *Fuzzy Sets and Systems*, 183:26–43, 2011.

[28] Q. Hu, D. Yu, W. Pedrycz, and D. Chen. Kernelized fuzzy rough sets and their applications. *IEEE Transactions on Knowledge and Data Engineering*, 23(11):1649–1667, 2011.

[29] Q. Hu, L. Zhang, S. An, D. Zhang, and D. Yu. On robust fuzzy rough set models. *IEEE Transactions on Fuzzy Systems*, 20(4):636 – 651, 2012.

[30] Q. Hu, L. Zhang, D. Chen, W. Pedrycz, and D. Yu. Gaussian kernel based fuzzy rough sets: model, uncertainty measures and applications. *International Journal of Approximate Reasoning*, 51:453–471, 2010.

[31] Q.H. Hu, X.Z. Xie, and D.R. Yu. Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. *Pattern Recognition Letter*, 40(12):3509–3521, 2007.

[32] Q.H. Hu, D.R. Yu, and X.Z. Xie. Information-preserving hybrid data reduction based on fuzzy-rough techniques. *Pattern Recognition Letter*, 27(5):414–423, 2006.

[33] S. Jenei. New family of triangular norms via contrapositive symmetrization of residuated implicators. *Fuzzy Sets and Systems*, 110:157–174, 2000.

[34] R. Jensen and Q. Shen. Fuzzy-rough attribute reduction with application to web categorization. *Fuzzy Sets and Systems*, 141(3):469–485, 2004.

[35] R. Jensen and Q. Shen. Fuzzy-rough data reduction with ant colony optimization. *Fuzzy Sets and Systems*, 149(1):5–20, 2005.

[36] R. Jensen and Q. Shen. Fuzzy-rough sets assisted attribute selection. *IEEE Transactions on Fuzzy Systems*, 15(1):73–89, 2007.

[37] R. Jensen and Q. Shen. New approaches to fuzzy-rough feature selection. *IEEE Transactions on Fuzzy Systems*, 17(4):824–838, 2009.

[38] J.D. Katzberg and W. Ziarko. Variable precision extension of rough sets. *Fundamenta Informaticae*, 27(2-3):155–168, 1996.

[39] L. I. Kuncheva. Fuzzy rough sets: application to feature selection. *Fuzzy Sets and Systems*, 51:147–153, 1992.

[40] G. Liu. Axiomatic systems for rough sets and fuzzy rough sets. *International Journal of Approximate Reasoning*, 48:857–867, 2008.

[41] G. Liu. Using one axiom to characterize rough sets and fuzzy rough sets. *Information Sciences*, 223:285–296, 2013.

[42] W.N. Liu, J.T. Yao, and Y.Y. Yao. Rough approximations under level fuzzy sets. In *Rough Sets and Current Trends in Computing (RSCTC2004)*, pages 78–83, 2004.

[43] J.S. Mi, Y. Leung, H.Y. Zhao, and T. Feng. Generalized fuzzy rough sets determined by a triangular norm. *Information Sciences*, 178:3203–3213, 2008.

[44] J.S. Mi and W.X. Zhang. An axiomatic characterization of a fuzzy generalization of rough sets. *Information Sciences*, 160:235–249, 2004.

[45] M. Michał and B. Jayaram. *Fuzzy Implications*. Springer-Verlag Berlin, 2008.

[46] A. Mieszkowicz-Rolka and L. Rolka. Variable precision fuzzy rough sets. In J.F. Peters et al., editor, *Transactions on Rough Sets I*, pages 144–160. Springer-Verlag, 2004.

[47] A. Mieszkowicz-Rolka and L. Rolka. Fuzzy rough approximations of process data. *Internation Journal of Approximate Reasoning*, 49:301–315, 2008.

[48] N.N. Morsi and M.M. Yakout. Axiomatics for fuzzy rough set. *Fuzzy Sets Systems*, 100:327–342, 1998.

[49] C.V. Negoita and D.A. Ralescu. Representation theorems for fuzzy concepts. *Kybernetes*, 4(3):169–174, 1975.

[50] Z. Pawlak. Rough sets. *International journal of computer and information sciences*, 11(5):341–356, 1982.

[51] D. Pei. A generalized model of fuzzy rough sets. *International Journal of General Systems*, 34(5):603–613, 2005.

[52] J.A. Pomykala. Approximation operations in approximation space. *Bulletin of the Polish Academy of Sciences*, 35:653–662, 1987.

[53] A.M. Radzikowska and E.E. Kerre. A comparative study of fuzzy rough sets. *Fuzzy Sets and Systems*, 126:137–155, 2002.

[54] A.M. Radzikowska and E.E. Kerre. Characterisation of main classes of fuzzy relations using fuzzy modal operators. *Fuzzy Sets and Systems*, 152:223–247, 2005.

[55] D.A. Ralescu. A generalization of the representation theorem. *Fuzzy Sets and Systems*, 51:309–311, 1992.

[56] J. M. Fernández Salido and S. Murakami. On β-precision aggregation. *Fuzzy Sets and Systems*, 139:547 – 558, 2003.

[57] J. M. Fernández Salido and S. Murakami. Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations. *Fuzzy Sets and Systems*, 139:635–660, 2003.

[58] Q. Shen and R. Jensen. Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring. *Pattern Recognition*, 37(7):1351–1363, 2004.

[59] A. Skowron and C. Rauszer. The discernibility matrices and functions in information systems. In R. Słowiński, editor, *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, pages 331–362. Kluwer Academic Publishers, the Netherlands, 1992.

[60] E.C.C. Tsang, D.G. Chen, D.S. Yeung, X.Z. Wang, and J.W.T. Lee. Attributes reduction using fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, 16(5):1130–1141, 2008.

[61] W.Z. Wu, Y. Leung, and J.S. Mi. On characterizations of (I, T)-fuzzy rough approximation operators. *Fuzzy Sets and Systems*, 154:76–102, 2005.

[62] W.Z. Wu, J.S. Mi, and W.X. Zhang. Generalized fuzzy rough sets. *Information Sciences*, 151:263–282, 2003.

[63] W.Z. Wu and W.X. Zhang. Constructive and axiomatic approaches of fuzzy approximation operators. *Information Sciences*, 159:233–254, 2004.

[64] R. R. Yager. Families of OWA operators. *Fuzzy Sets and Systems*, 59:125–148, 1993.

[65] Y. Y. Yao. Combination of rough and fuzzy sets based on α-level sets. In T. Y. Lin and N. Cercone, editors, *Rough Sets and Data Mining: Analysis for Imprecise Data*, pages 301–321. Kluwer Academic Publishers, 1997.

[66] D.S. Yeung, D. Chen, E.C.C. Tsang, J.W.T. Lee, and W. Xizhao. On the generalization of fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, 13(3):343–361, 2005.

[67] L. A. Zadeh. Fuzzy sets. *Information and control*, 8:338–353, 1965.

[68] S. Zhao, E. C. C. Tsang, and D. Chen. The model of fuzzy variable precision rough sets. *IEEE Transactions on Fuzzy Systems*, 17(2):451–467, 2009.

[69] S. Zhao and E.C.C. Tsang. On fuzzy approximation operators in attribute reduction with fuzzy rough sets. *Information Sciences*, 178(16):3163–3176, 2008.

[70] X. Zhu and X. Wu. Class noise vs attribute noise: a quantitative study of their impacts. *Artificial Intelligence Review*, 22:177–210, 2004.

[71] W. Ziarko. Variable precision rough set model. *Journal of Computer and System Sciences*, 46:39–59, 1993.