INTERVAL-VALUED FUZZY INFERENCE SYSTEMS BASED ON THE BANDLER-KOHOUT SUBPRODUCT

LIM CHEE KAU

FACULTY OF COMPUTER SCIENCE
AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2015

INTERVAL-VALUED FUZZY INFERENCE SYSTEMS BASED ON THE BANDLER-KOHOUT SUBPRODUCT

LIM CHEE KAU

THESIS SUBMITTED IN FULFILMENT
OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTER SCIENCE
AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2015

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: LIM CHEE KAU

Registration/Matrix No.: WHA 110004

Name of Degree: DOCTOR OF PHILOSOPHY

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Interval-Valued Fuzzy Inference Systems based on BK Subproduct.

Field of Study: FUZZY INFERENCES

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness's Signature Date

Name:
Designation:

ABSTRACT

The studies of fuzzy relations by Bandler and Kohout, which are also known as the BK products, are well known in the literature as tools to study the composition of relations. In the past, BK products, particularly the BK subproduct, gained remarkable success in developing inference engines for numerous applications.

Though successful, there are still some limitations. First of all, this research starts with a survey on a set of inference structures formed by the BK subproduct in previous researches. The survey finds shortcomings in some inference structures. With excluding these candidates, a set of robust inference structures are obtained from the analysis.

Secondly, with the understanding that the ordinary type-1 fuzzy sets have limited ability in modeling uncertainty, a more general fuzzy set framework is proposed to improve the performance of BK products. Thus, extending BK products to interval-valued fuzzy sets is another contribution of this thesis. Since the subsethood measure is fundamental to the BK products, two interval-valued fuzzy subsethood measures are also developed in this research.

Moreover, this research suggests that, among all the features involved in inferences, certain features should have higher influence compared to the others. Therefore, to distinguish the influence of features towards inference results, a weight parameter is added. The computation of this weighted inference engine is also discussed.

In order to test the proposed inference engine, this research also proposes a new method to define membership degrees from statistical data. With this method, the BK subproduct is tested with 3 publicly available data sets. The results are compared. Experimental results show that the extension to interval-valued fuzzy sets and the additional weight parameter improve the quality of inferences.

ABSTRAK

Kajian perhubungan kabur oleh Bandler dan Kohout, ataupun yang dikenali sebagai BK products merupakan peralatan yang terkenal dalam literasi untuk mengkaji komposisi hubungan. Pada masa lalu, BK products, terutamanya BK subproduct mendapat kejayaaan yang cemerlang dalam pembangunan enjin inferens kepada banyak aplikasi.

Walaupun mendapat kejayaan, beberapa batasan masih wujud. Pertama sekali, pemeriksaan terhadap satu set struktur inferens berasal dari BK subproduct pada penyelidikan yang dahulu. Beberapa kelemahan telah dijumpai dalam sebilangan struktur inferens. Dengan tidak memasukan calon-calon ini, satu set struktur inferens yang mantap diperolehi dari analisis ini.

Seterusnya, dengan pengetahuan bahawa set kabur jenis pertama mempunyai keupayaan yang terhad dalam mewakili ketidakpastian, satu rangkaian set kabur yang lebih umum dicadangkan untuk meningkatkan prestasi BK products. Maka, melanjutkan BK products ke set kabur bernilai selang merupakan satu lagi sumbangan tesis ini. Oleh kerana pengukuran keahlian set merupakan asas BK products, dua pengukuran keahlian set untuk set kabur bernilai selang juga dibangunkan dalam penyelidikan ini.

Sebagai tambahan, penyelidikan ini mencadangkan, dalam semua ciri-ciri yang terlibat dalam inferens, sesetengah ciri-ciri seharusnya mempunyai pengaruhan yang lebih tinggi baerbanding dengan yang lain. Oleh itu, untuk membezakan pengaruhan ciri-ciri terhadap keputusan inferencs, satu parameter berat telah ditambah. Penyelesaian kepada enjin inferens berberat ini juga dibincangkan.

Untuk menguji enjin inferens ini, satu cara baru menentukan darjah keahlian dari data statistik untuk BK subproduct dicadangan. Dengan cara ini, BK subproduct telah diuji dengan 3 set data awam. Keputusan eksperimen menunjukan pelanjutan ke set kabur bernilai selang dan penambahan parameter berat telah meningkatkan kualiti inferens.

ACKNOWLEDGEMENTS

This could be the hardest chapter to write, because there are so many people I have to thank. Although listing out all their names is easy, but it is hard to fully express my gratitude to them with words.

First and foremost, I would like to express my appreciation to my supervisor, Dr Chan Chee Seng, for his motivation, patience and support. His guidance helped me in all the time of research and writing of this thesis. Also, with his support, I had opportunities to explore top conferences and meet great people. One simply could not wish for a better or friendlier supervisor.

Special thanks also goes to my co-supervisor, Prof. Bernard De Baets. He had been taking care of me while I was visiting University of Gent as exchange student. He also gives me constructive comments and the research would not be the same without his advises.

My deepest appreciation goes to my family. I owe a very important debt to my parents, who grown me up with their endless love, and continuously on my back for any important decisions I make today. My wife, Soo Yeong has been extraordinarily tolerant and supportive all the time, especially the tough moment I met in my life. There are no words to convey my gratitude for her understanding and support. Great appreciation also goes to Ruo Qing and Ruo Fan, my two lovely children who inspired and delighted me. They are wonderful. I also want to thank my brother for the encouragement and support.

Lastly, I would like to credit all my friends and colleagues who have been helping me in any form. There are too many of them for me to list. A great thank to all of you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

BK	Bandler and Kohout's.
COG	Center of Gravity.
COM	Center of Maxima.
CRI	Compositional Rule of Inference.
CWW	Computing With Words.
FOU	Footprint Of Uncertainty.
FWA	Fuzzy Weighted Average.
GMP	generalized *modus ponens*.
IVFS	Interval-Valued Fuzzy Set.
LMF	Lower Membership Function.
LWA	Linguistic Weighted Average.
MOM	Mean of Maxima.
T1FS	Type-1 Fuzzy Set.
T2FS	Type-2 Fuzzy Set.
UMF	Upper Membership Function.
WDBC	Wisconsin Diagnostic Breast Cancer.

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background

The *modus ponens* has been used in reasoning by many researches since antiquity. This rule of inference can be expressed as "P implies Q. P, therefore, Q", or:

$$\big(P \wedge (P \rightarrow Q)\big) \rightarrow Q. \tag{1.1}$$

The foundation of the *modus ponens* is classical two-valued logic. When fuzzy set theory was proposed, the *modus ponens* was extended to the generalized *modus ponens* (GMP) (Zadeh, 1973):

$$\big(P' \wedge (P \rightarrow Q)\big) \rightarrow Q' \tag{1.2}$$

where P, P', Q and Q' are fuzzy concepts.

Based on the GMP, the Compositional Rule of Inference (CRI) (Zadeh, 1973) is one of the most popular fuzzy inference schemes, where both the Mamdani (Mamdani & Assilian, 1975) and Sugeno (Takagi & Sugeno, 1985) inference engines are among the most well-known applications.

Then, in some later studies on deductive reasoning, such as De Baets and Kerre (1993b), Bodenhofer, Dankova, Stepnicka, and Novak (2007), Stepnicka and De Baets (2013), investigations on the CRI found that the Mamdani inference engine does not really make use of the concept of implication as suggested in Eq. (1.2). In another study, Stepnicka and Jayaram (2010) explicitly pointed out that another established fuzzy relational inference mechanism, the Bandler-Kohout relational products (Kohout & Bandler, 1980a, 1980b), particularly the Bandler-Kohout subproduct can form excellent inference schemes that model Eq. (1.2), which make use of implication operators. The Bandler-Kohout products, which are commonly abbreviated as the BK products in the literature (Kohout & Kim, 2002; Běhounek & Daňková, 2009; Kohout, 2009a; Mandal & Jayaram, 2012, 2013), is a study of composition of relations, or relations between two sets that are not directly related. In a review on fuzzy relational calculus (Kerre, 2007), the Bandler and Kohout's (BK) products were honoured as the the most important operation on relations.

Instead of investigating BK products as tool of rule-based reasoning, they can be studied as schemes of case-based reasoning. Case-based reasoning finds its advantages over rule-based reasoning in many ways. For example, case-based reasoning works successfully in domains that are not completely understood, where defining rules are not easy. Moreover, adding new cases to output is easy as it will not interfere with the existing cases. Kolodner (1992, pp. 28-30) holds a comprehensive discussion on the advantages of case-based reasoning. In the context of case-based reasoning, BK products have been implemented as inference schemes of medical expert systems (Yew & Kohout, 1996a, 1997; Lim, Yew, Ng, & Abdullah, 2002), information retrieval (Kohout & Bandler, 1985), path finding of autonomous underwater vehicles (Bui & Kim, 2006; Y.-i. Lee & Kim, 2008), land evaluation (Groenemans, Ranst, & Kerre, 1997) and etc.

1.2 Problem Formulation

Despite the success of the BK products in the past, there are still some limitations. For instance, De Baets and Kerre pointed out a lack of non-emptiness condition in the original definition of BK products (De Baets & Kerre, 1993a, 1994). An improvement was proposed in their paper in order to make BK products more robust. With this improvement, more logical connectives need to be instantiated to develop the fuzzy inference structures. Yew and Kohout (Yew & Kohout, 1996a, 1996b, 1997) showed a typical example of this work where a set of 23 inference structures based on the BK subproduct and its variants were developed. However, this research found that there are limitations in the formulation of these inference structures, which are not addressed. These limitations lead to the initialization of inappropriate logical connectives for the inference structures.

Secondly, similar to other fuzzy logic systems, a defuzzification module is used to prepare the results in the form that meets the output requirements (Figure 1.1). Meanwhile, one of the interesting features of the BK products is that it performs inferences using the fuzzy implication operators, which can be defined based on the needs of applications. Based on this special property, interval-valued inferencing was proposed in (Kohout & Bandler, 1992) to obtain more reliable results in inferences. However, despite of its great idea, this research found that the interval-valued inferencing is not only associates with low efficiency because of unnecessary computations, but also falls into the realm of dichotomy, i.e. an inference can only be 'accepted' or 'not accepted' (or, 'rejected' or 'not rejected' in the other way round). Certainly, an improvement of this defuzzification method is needed so that attention is given to both reliability and efficiency.

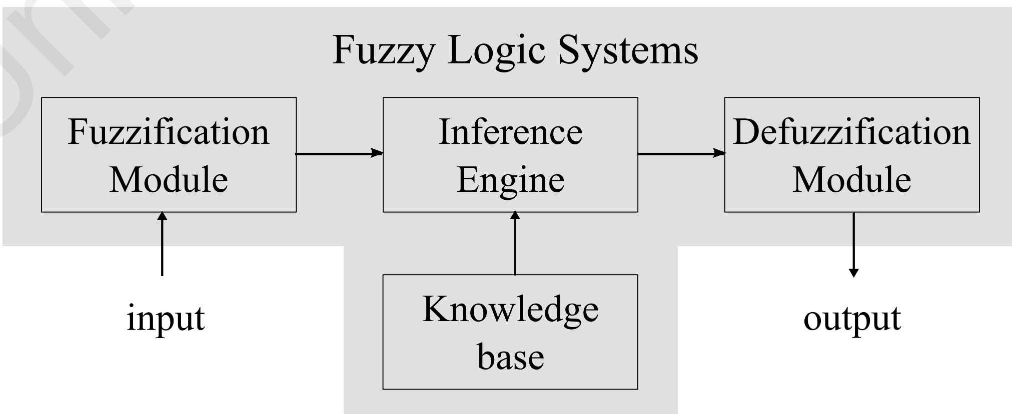


Figure 1.1: Typical Structure of a Fuzzy Logic Systems

More importantly, the implementations of the BK products in the literature are based on classical T1FS theory, which address uncertainty with single-point-values. Studies such as (Mendel, 2000, 2003) claimed that T1FSs have their limitation in addressing uncertainty with its crisp membership functions. While more complicated fuzzy set theories, such as the IVFSs and Type-2 Fuzzy Sets (T2FSs) (Zadeh, 1975b; Gorzalczany, 1987; Bustince, 2000; Mendel, John, & Liu, 2006) are being developed, extending the BK products to these fuzzy frameworks became a challenge that has not been attended so far.

Moreover, the BK products perform inferences by utilizing a set of common features that relate the inputs and outputs. In most cases, the BK products treat all the features equally, i.e. the importance of all features is similar. However, in practice, not all these features have the same influences towards expected inference results. This thesis argues that some features may have higher reliability or distinguishability than the others, and vice versa. In the literature of fuzzy logic researches, implementation of weight parameters is also not rare (Hoffmann, 2004; Ishibuchi & Yamamoto, 2005; Seki & Mizumoto, 2011; Xing & Ha, 2014). Therefore, adding a weight parameter to the BK products based inference engines is another challenge of this study.

Lastly, in previous implementations of the BK products, predefine rules (Bui & Kim, 2006) or experts knowledge (Groenemans et al., 1997) are required so that the knowledge based can be formed. In some cases, even the fuzzification modules are also predefined. However, an approach to train the BK products so that it can learn from examples cannot be found in the literature. The lack of this learning feature limits the application of the BK products in many fields as long as a predefined knowledge base is not available.

1.3 Thesis Objectives and Main Contributions

In Section 1.2, some research opportunities on the BK products have been pointed out. Generally, the aim of the research is to improve the reasoning performance of the BK products, particularly the most popular BK subproduct based inference engines. This aim is achieved by building the theoretical framework of a weighted inference engine which based on IVFS-based BK subproduct. More specifically, the objectives and contributions of this research are as follow:

(a) To improve the implementations of the BK products that can be found in the literature, particularly the implementations of BK subproduct.

This research discovers limitations presented in the list of inference structures that proposed by Yew and Kohout (1996a, 1996b, 1997). With an analysis on the properties of the logical connectives used in these inference structures, some of the inference structures with shortcomings are rejected. This bring the number of robust inference structures reduced from 19 to 8.

Besides, a defuzzification method was proposed as an improvement to interval-valued inferencing technique. For a given threshold, this improved defuzzification method not only proposes acceptable inferences, but also the reliabilities of these inferences.

(b) To improve the performance of the BK products with an extension to IVFSs, as well as additional weight parameter.

T2FS theory started to get attention in the late 1990s (John & Coupland, 2007) with the claim that they have a better capability in modeling uncertainties (Fazel Zarandi, Rezaee, Turksen, & Neshat, 2009; Choi & Rhee, 2009; Hwang, Yang, & Hung, 2011). However, due to high computational complexity and other reasons (Coupland, 2007; Mendel, 2004), T1FS theory is still dominant. In this research, the BK products are extended to be based on IVFSs, a special form of T2FSs.

Besides, researches reported that incorporating a weight parameter may improve the fuzzy systems by paying more attention to certain factors in reasoning (Hayashi, Otsubo, Murakami, & Maeda, 1999; Ishibuchi & Yamamoto, 2005; Y. Wang & Fan, 2007; Seki & Mizumoto, 2011). Thus, an IVFS-modeled weight parameter is added to the inference engines based on the BK products.

(c) To develop a learning mechanism for the BK subproduct based inference systems so that a system can be built without predefined data.

In the past, expert knowledge was needed in most of the implementations of the BK subproduct. In these cases, experts are expected to provide information that is necessary to construct the knowledge base, as well as in the fuzzification process. In this research, a method is proposed so that the knowledge base and fuzzification module (Figure 1.1) can be built from the learning process. With this mechanism, membership functions can be formed from a set of training data. Next, the distributions of data help to define the membership degrees for the knowledge base. This set of membership functions also serve to fuzzify the input data so that can be processed by the inference engines.

(d) To demonstrate the implementation of the BK subproduct based inference systems as a classifier.

Empirical study is not an objective of this research. However, to demonstrate the usefulness of the extension of BK subproduct, an experiment with limited data sets is conducted. In this experiment, three publicly available data sets are adopted so that inference engines derived from BK subproduct work as classifiers. Classification accuracies are compared among these BK subproduct based inference engines, as well as with other works in the literature.

1.4 Organization of the Thesis

This thesis is organized into 8 chapters.

Chapter 2 provides background knowledge required for this research. This chapter starts with a revision on the fundamental knowledge about fuzzy sets and fuzzy relations. A detailed review on the theory of BK products and improvement is provided in the subsequent section, followed by examples of applications in the literature. This chapter also discuss the concepts of IVFSs and T2FSs.

Chapter 3 discusses two shortcomings of the implementations of the BK subproduct in the past, namely the adoption of inappropriate logical connectives and the low performance of defuzzification modules. Improvements for both shortcomings are proposed here.

Chapter 4 starts with the introduction of two subsethood measures of IVFSs, followed by extending the BK products into the framework of IVFSs. Some interesting properties of the BK products in the IVFSs are also studied here. Lastly, the weight parameter is added to the BK products based inference templates.

Chapter 5 proposes a learning mechanism so that the BK subproduct based inference systems can be built from numerical data. Detailed algorithms are provided, followed by an example.

Chapter 6 demonstrates the application of the BK subproduct as a classifier. Three implementations of the BK subproduct are shown here, i.e. the original T1FSs based, improved IVFSs based and weighted.

Chapter 7 shows the results of the applications in Chapter 6, followed by a discussion and comparisons with other works in the literature.

Chapter 8 concludes the research. Last but not least, further research topics related to the work are proposed.

CHAPTER 2

BACKGROUND RESEARCH

2.1 Introduction

BK products concern the composition of relations between sets. To make the discussion on BK products efficient, this chapter starts with a brief discussion of the concepts of fuzzy sets, fuzzy relations and their characteristics. The BK products are discussed in detail in the second section of this chapter, followed by two extensions of the ordinary T1FS theory, namely IVFS theory and T2FS theory.

2.2 Fuzzy Sets and Fuzzy Relations

An element can be discriminated as member or non-member of a set in classical (crisp) set theory. However, due to some reasons, the boundary of a set may ambiguous, imprecise or uncertain. The reasons for this problem may come from:

- imperfect, or incomplete definitions;
- systemic or random errors in measurements;
- vagueness in natural languages, and etc.

Since crisp set theory is not capable of handling the aforementioned problems, fuzzy set theory was developed. If X is the universe of discourse and $x \in X$, a mapping $A : X \to [0,1]$ is a fuzzy set on X. Furthermore, $A(x)$ is the membership degree of x in A.

The same concept of membership degrees can be applied to the study of relations (Zadeh, 1965). Assume R is a relation between 2 universes X and Y, where $x \in X$ and $y \in Y$, then $R(x,y) \in [0,1]$ is the degree of relationship between x and y. There are at least 3 different characteristics of fuzzy relations that can be studied, namely, reflexivity, symmetry, and transitivity.

A relation R is called *reflexive* if $R(x,x) = 1$ for all $x \in A$. If R is not reflexive, it is called *irreflexive*. If the relation does not hold for all $x \in A$, then R is an *antireflexive* relation.

A relation R is called *symmetric* if and only if for $x,y \in A$, $R(x,y) = R(y,x)$. If R is not symmetric, it is called *irreflexive*. *asymmetric*. Subsequently, the relation is *antisymmetric* if $R(x,y) \neq 0$ and $R(y,x) \neq 0$ only when $x = y$.

For $x,y,z \in A$, a relation R is *transitive* if and only if:

$$R(x,z) \geq \max_{y \in A} \min\Big(R(x,y), R(y,z)\Big) \tag{2.1}$$

for all $(x,z) \in A \times A$. This relation is *non-transitive* if Eq. (2.1) does not hold for some $(x,z) \in A \times A$. Lastly, for all $(x,z) \in A \times A$, if:

$$R(x,z) < \max_{y \in A} \min\Big(R(x,y), R(y,z)\Big) \tag{2.2}$$

then R is *antitransitive*.

In some real-world applications, the fuzzy relation R may not satisfy the property of transitivity. Therefore, finding a modified fuzzy relation that is close to R, and contains R is a solution to this problem. This modified fuzzy relation $R^{\otimes}$ is called the *transitive closure* of R, and satisfy the following properties:

1. $R^{\otimes}$ is transitive;

2. R is a subset of $R^{\otimes}$;

3. the elements of the transitive closure have the smallest possible membership grades.

This transitive closure is given by:

$$R^{\otimes} = \bigcup_{k=1}^{\infty} R^k \tag{2.3}$$

2.3 BK Products

2.3.1 BK Products in Crisp Sets

The discussion of this section starts with a brief revisit of the fundamental definitions of BK (crisp) relational products. To make the discussion more concise, the following notations and definitions are used for the remaining of this thesis.

Let set $A = \{a_i \mid i = 1, \cdots, I\}$ and set $B = \{b_j \mid j = 1, \cdots, J\}$. R is defined as a relation from A to B such that $R \subseteq A \times B$. The abbreviation aRb shows that a is in relation R with b.

Definition 1 (Domain)**.** The domain of a relation R is the set of elements of A such that:

$$\text{dom}(R) = \{a \mid a \in A \text{ and } (\exists b \in B)(aRb)\} \tag{2.4}$$

Definition 2 (Range)**.** The range of a relation R is the set of elements in B such that:

$$\text{rng}(R) = \{b \mid b \in B \text{ and } (\exists a \in A)(aRb)\} \tag{2.5}$$

Definition 3 (Converse Relation)**.** The converse relation R^T is the reverse of relation R from B to A:

$$R^T = \{(b,a) \mid (b,a) \in B \times A \text{ and } aRb\} \tag{2.6}$$

Definition 4 (Afterset)**.** The afterset aR is the image of a in B under relation R in B:

$$aR = \{b \mid b \in B \text{ and } aRb\} \tag{2.7}$$

Definition 5 (Foreset)**.** The foreset Rb is the image of b in A under relation R^T:

$$Rb = \{a \mid a \in A \text{ and } aRb\} \tag{2.8}$$

Assume that there is another set $C = \{c_k \mid k = 1, \cdots, K\}$, and S is a crisp relation from set B to set C. The classical relational product, namely the circle product is defined as follows:

Definition 6 (Circle product)**.** The circle product gives all (a,c) couples for which there exist at least one b that is in relation R^T with $a \in A$ and relation S with $c \in C$.

$$R \circ S = \{(a,c) \mid (a,c) \in A \times C \text{ and } aR \cap Sc \neq \varnothing\} \tag{2.9}$$

Bandler and Kohout revised Definition (6) and proposed the crisp BK products (Kohout & Bandler, 1980b):

Definition 7 (BK Subproduct)**.** The BK subproduct gives all (a,c) couples for which the afterset aR is a subset of foreset Sc.

$$R \lhd_{\text{BK}} S = \{(a,c) \mid (a,c) \in A \times C \text{ and } aR \subseteq Sc\} \tag{2.10}$$

With this composition of relations, one can find the relation between an **object**, $a \in A$ and **target**, $c \in C$ if a set with common **features**, $B' \subseteq B$ appears in the middle (Figure 2.1).

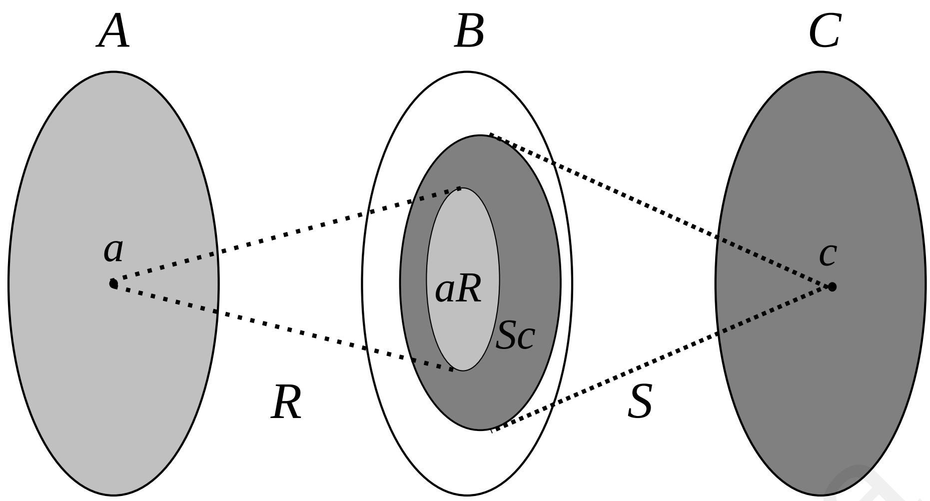


Figure 2.1: With BK subproduct, the relation between 2 sets which are not related directly can be retrieved if a set with common features exist.

Definition 8 (BK Superproduct)**.** The BK Superproduct gives all (a,c) couples for which the foreset *Sc* is a subset of afterset *aR*:

$$R \rhd_{\mathrm{BK}} S = \{(a,c) \mid (a,c) \in A \times C \text{ and } Sc \subseteq aR\} \tag{2.11}$$

Definition 9 (BK Square Product)**.** The BK square product gives all (a,c) couples for which the afterset *aR* is exactly equal to the foreset *Sc*:

$$R \diamond_{\mathrm{BK}} S = \{(a,c) \mid (a,c) \in A \times C \text{ and } aR = Sc\} \tag{2.12}$$

It is easy to see that $R \diamond_{\mathrm{BK}} S \equiv (R \lhd_{\mathrm{BK}} S) \bigcap (R \rhd_{\mathrm{BK}} S)$.

De Baets and Kerre (1993c, 1993a, 1994) found that there is a shortcoming in Definition (7), namely the lack of non-emptiness condition. They found that for an element *a* which finds no relation *R* with any elements in *B*, the image of *aR* is an empty set. Due to the empty set is a subset of all sets, this particular element *a* can have relation $R \lhd_{\mathrm{BK}} S$ with all the elements in *C* even if there is no image of *a* in *B* under relation *R*. For the BK superproduct and the square product, a similar imperfection holds. Thus, De Baets and Kerre concluded that a lot of unwanted couples are generated by the traditional BK relational products.

To resolve this shortcoming, De Baets and Kerre (1993c, 1993a, 1994) proposed that an additional term should be added to Definitions (7) - (9), so that the empty set is not counted:

$$R \triangleleft_{\mathrm{K}} S = \{(a,c) \mid (a,c) \in A \times C \text{ and } \varnothing \subset aR \subseteq Sc\} \tag{2.13}$$

$$R \triangleright_{\mathrm{K}} S = \{(a,c) \mid (a,c) \in A \times C \text{ and } \varnothing \subset Sc \subseteq aR\} \tag{2.14}$$

$$R \diamond_{\mathrm{K}} S = \{(a,c) \mid (a,c) \in A \times C \text{ and } \varnothing \subset aR = Sc\} \tag{2.15}$$

2.3.2 BK Products in Fuzzy Sets

As one can observe in Definitions (7) and (8), $aR \subseteq Sc$ and $Sc \subseteq aR$ are the keys in retrieving the relationship between a and c. Therefore, these crisp BK products can be extended to a fuzzy BK products easily by introducing a fuzzy subsethood measure - i.e. the possibility of a set is the subset of another given set. As proposed by Kohout and Bandler (1980b), for two fuzzy subsets P and Q, where both subsets are in the universe X and x is a general notation of elements in this universe, the possibility that P is a subset of Q is given as:

$$\pi(P \subseteq Q) = \bigwedge_{x \in X} (P(x) \to Q(x)) \tag{2.16}$$

where $\bigwedge$ is the infimum operator, $\to$ denotes a fuzzy implication operator, and $P(x)$ and $Q(x)$ represent the membership degrees of x in P and Q perspectively. The infimum operator is an aggregator that can be defined as min function in harsh criterion, or arithmetic mean in mean criterion. The fuzzy implication operator (Kohout & Bandler, 1980a; Willmott, 1980; Ruan, 1993) is a function such that: i) $[0,1]^2 \to [0,1]$; ii) decreasing in the first variable but increasing in the second variable; and iii) satisfying the following boundary conditions, namely $0 \to 0 = 1$, $0 \to 1 = 1$, $1 \to 1 = 1$ and $1 \to 0 = 0$.

There are a number of candidates of fuzzy implication operators in the literature (Table 2.1).

With the definition of fuzzy subsethood measure in Eq. (2.16), the *original fuzzy BK products* can be defined as follow:

$$\text{Fuzzy BK subproduct:} \qquad R \triangleleft_{\mathrm{BK}} S(a,c) = \bigwedge_{b \in B} (R(a,b) \to S(b,c)) \tag{2.17}$$

$$\text{Fuzzy BK superproduct:} \qquad R \triangleright_{\mathrm{BK}} S(a,c) = \bigwedge_{b \in B} (S(b,c) \to R(a,b)) \tag{2.18}$$

$$\text{Fuzzy BK square product:} \qquad R \diamond_{\mathrm{BK}} S(a,c) = \bigwedge_{b \in B} (R(a,b) \leftrightarrow S(b,c)) \tag{2.19}$$

where $a \leftrightarrow b = \min(a \to b, b \to a)$.

On the other hand, the fuzzy circle product is defined as follow:

$$\text{Fuzzy circle product:} \qquad R \circ S(a,c) = \bigvee_{b \in B} \tau(R(a,b), S(b,c)) \tag{2.20}$$

where $\bigvee$ is the supremum operator and τ is the t-norm.

With the consideration of non-emptiness condition, De Baets and Kerre (De Baets & Kerre, 1993a) proposed 2 sets of improvement. Each of the improvement requires an additional term to rectify the problem.

The first set of improvement is based on:

$$R \bowtie_{\mathrm{B}} S = (R \bowtie_{\mathrm{BK}} S) \cap (\mathrm{dom}(R) \times \mathrm{rng}(S))$$

where $\bowtie = \{\triangleleft, \triangleright, \diamond\}$. This expression leads to the first set of improved fuzzy BK products,

Table 2.1: Examples of the Fuzzy Implication Operators and Their Respective Definitions.

Name	Symbol	Definition
S# - Standard Sharp (Mizumoto & Zimmermann, 1982)	$r \to_{\text{S\#}} s$	$\begin{cases} 1 & \text{iff } r \neq 1 \text{ or } s = 1 \\ 0 & \text{otherwise} \end{cases}$
S - Standard Strict (Mizumoto & Zimmermann, 1982)	$r \to_{\text{S}} s$	$\begin{cases} 1 & \text{iff } r \leq 1 \\ 0 & \text{otherwise} \end{cases}$
S* - Standard Star (Mizumoto & Zimmermann, 1982)	$r \to_{\text{S}*} s$	$\begin{cases} 1 & \text{iff } r \leq s \\ s & \text{otherwise} \end{cases}$
G43 - Gaines 43 (Mizumoto & Zimmermann, 1982)	$r \to_{\text{G43}} s$	$\min(1, \frac{r}{s})$
G43' - Modified Gaines 43 (Mizumoto & Zimmermann, 1982)	$r \to_{\text{G43'}} s$	$\min(1, \frac{r}{s}, \frac{1-r}{1-s})$
KD - Kleene-Dienes (Kohout & Bandler, 1980a)	$r \to_{\text{KD}} s$	$\max(s, 1-r)$
R - Reichenbach (Kohout & Bandler, 1980a)	$r \to_{\text{R}} s$	$1 - r + rs$
L - Łukasiewicz (Zadeh, 1975a)	$r \to_{\text{Ł}} s$	$\min(1, 1 - r + s)$
W - Willmott (Willmott, 1980)	$r \to_{\text{W}} s$	$\min\Big(\max(1-r,s), \max(r,1-r), \max(s,1-s)\Big)$
Y - Yager (Yager, 1980)	$r \to_{\text{Y}} s$	s^r
EZ - Early Zadeh (Zadeh, 1975a)	$r \to_{\text{EZ}} s$	$(r \wedge s) \vee (1-r)$

namely the *fuzzy BK products (set B)*:

$$R \triangleleft_{\text{B}} S(a,c) = \min\Big(\bigwedge_{b \in B}(R(a,b) \to S(b,c)), \bigvee_{b \in B} R(a,b), \bigvee_{b \in B} S(b,c)\Big) \tag{2.21}$$

$$R \triangleright_{\mathrm{B}} S(a,c) = \min\Big(\bigwedge_{b\in B}(S(b,c) \to R(a,b)), \bigvee_{b\in B} R(a,b), \bigvee_{b\in B} S(b,c)\Big) \tag{2.22}$$

$$R \diamond_{\mathrm{B}} S(a,c) = \min\Big(\bigwedge_{b\in B}(R(a,b) \leftrightarrow S(b,c)), \bigvee_{b\in B} R(a,b), \bigvee_{b\in B} S(b,c)\Big) \tag{2.23}$$

The second set of improvements is based on:

$$R \bowtie_{\mathrm{K}} S = (R \bowtie_{\mathrm{BK}} S) \cap (R \circ S)$$

Thus, the second set of improved fuzzy BK products are *fuzzy BK products (set K)*:

$$R \triangleleft_{\mathrm{K}} S(a,c) = \min\Big(\bigwedge_{b\in B}\big((R(a,b) \to S(b,c)\big), \bigvee_{b\in B} \tau\big(R(a,b), S(b,c)\big)\Big) \tag{2.24}$$

$$R \triangleright_{\mathrm{K}} S(a,c) = \min\Big(\bigwedge_{b\in B}\big((S(b,c) \to R(a,b)\big), \bigvee_{b\in B} \tau\big(R(a,b), S(b,c)\big)\Big) \tag{2.25}$$

$$R \diamond_{\mathrm{K}} S(a,c) = \min\Big(\bigwedge_{b\in B}\big((R(a,b) \leftrightarrow S(b,c)\big), \bigvee_{b\in B} \tau\big(R(a,b), S(b,c)\big)\Big) \tag{2.26}$$

The initial term that originates from Bandler and Kohout is referred to as the implication term, whereas the term added by De Baets and Kerre is referred as the additional term.

2.3.3 Advantages and Disadvantages of the BK Products

The inference law of *modus ponens* and its generalization, the GMP are the fundamental mechanisms in developing inference schemes. The popular CRI scheme is claimed to be a model of inference scheme based on the *modus ponens* (Zadeh, 1973). However, studies (Dubois & Prade, 1996; Daňková, 2007; Novák & Lehmke, 2006; Stepnicka & De Baets, 2013) show that the CRI does not really forms fuzzy rules with implications as required by *modus ponens*. In contrast, the inference scheme formed by BK products do really make use of residual implication. This make BK products mathematically more appealing compare to the CRI. Therefore, the strong mathematical fundamental is one of the advantages of BK products.

Although BK products follow the inference law of *modus ponens*, the implementations of BK products do not require to define rules explicitly. This brings an advantage to BK products to fill the void where developing rule based systems is problematic. Defining rules may be insufficient in some cases where one has no enough domain knowledge. In some other cases, the increasing of antecedents may cause the rule sets lengthy and difficult to handle. On the other hand, construction of inference systems based on BK products is easier as long as the fuzzy relations between objects-features and features-targets can be defined. Also, the increasing of attributes in BK products does not increase the complexity of the system because each attribute works independently from another.

Even though BK products work well with numerical data, but they find limitation in dealing with categorical data, especially when order or rank does not exist in the data. For example, categorical attribute “shape” with data “square”, “triangle”, “octagon” and etc is hard to quantify so that membership degrees can be retrieve for BK products to compute.

2.3.4 Applications of BK Products

The BK products have been widely applied in various field of soft computing (Kohout & Kim, 2002; Kohout, 2009b). Among all, the BK subproduct is the most popular and we can find its application in medical expert systems (Yew & Kohout, 1996a, 1996b, 1997), information retrieval (Kohout & Bandler, 1985), path finding of autonomous underwater vehicles (Bui & Kim, 2006; Y.-i. Lee & Kim, 2008), land evaluation (Groenemans et al., 1997), scene classification (Vats, Lim, & Chan, 2012) and etc, whereas the BK square product is also used in applications such as medical diagnosis (Davis IV & Kohout, 2006) and pattern recognition (Davis IV, 2006).

However, one shortcoming is that all the above applications did not consider the influential difference of each individual feature, except the study of lands evaluation (Groenemans et al., 1997). In this lands evaluation study, lands are evaluated so that the most suitable land unit is selected for specific utilization. The selection of lands is based on a list of land qualities and each land quality carries its own weight (influence).

Yet, the implementation of the weight in this work required to fulfill a condition: $\sum_{n=1}^{N} w_n = 1$ where N is the number of features (land qualities) and w is the weight of feature n. This condition is too restrictive for a good implementation of weights because: *(i)* adding or decreasing features into consideration list will cause recalculation of all the weights. For instance, adding a new feature with weight $w_{N+1} \neq 0$ to the existing feature list will cause the total weight become $\sum_{n=1}^{N+1} w_n$. It is easy to verify that $\sum_{n=1}^{N+1} w_n = 1 + w_{N+1} > 1$ and the condition of total weight equal to 1 is no longer valid. Thus, a normalization is required so that the $\sum_{n=1}^{N+1} w_n = 1$ is fulfilled. *(ii)* importance or influence of a feature is not intuitive - i.e. comparing a system with such condition to an implementation of weights where $w_n \in [0,1]$ for all n, the weights of the later are much more intuitive as the weights close to 0 means less influence, while close to 1 means high influence. In (Groenemans et al., 1997), the weights can be small numbers close to 0 even if they have high influence in case the number of features N is large. Furthermore, this problem becomes much more complicated if new features are going to be added into consideration as one may not know what are the appropriate values that represent high (or low) influence.

On the other hand, Yew and Kohout (1996a, 1996b, 1997) built a medical application based on the BK products, where both the original and improved BK subproducts were developed into medical inference engines. In this application, the composition of relations between the patients and illnesses were studied through a set of features, namely signs and symptoms. Based on the the original fuzzy BK subproduct, the improved fuzzy BK subproduct (set B) and the improved fuzzy BK subproduct (set K), three fuzzy inference templates were built. These inference templates are named as Sub-BK inference template, Sub-B inference template and Sub-K inference template, respectively:

Sub-BK inference template

$$R \lhd_{\mathrm{BK}} S(a,c) = \curlywedge_2\big(R(a,b) \to S(b,c)\big) \tag{2.27}$$

Sub-B inference template

$$R \lhd_{\mathrm{B}} S(a,c) = \curlywedge_1\Big(\curlywedge_2\,(R(a,b) \to S(b,c)), \curlyvee_3\,R(a,b), \curlyvee_4\,S(b,c)\Big) \tag{2.28}$$

Sub-K inference template

$$R \lhd_{\mathrm{K}} S(a,c) = \curlywedge_1\Big(\curlywedge_2\,(R(a,b) \to S(b,c)), \curlyvee_3\big(\curlywedge_4\,(R(a,b), S(b,c))\big)\Big) \tag{2.29}$$

In these inference templates, the $\curlywedge_i$ ($i = \{1,2,4\}$) and $\curlyvee_j$ ($j = \{2,3\}$) are logical connectives instantiated with the checklist paradigm (Kohout & Bandler, 1980b, 1992) as follow:

$$
\begin{aligned}
&\curlywedge_1 = \{\min, \max\} \\
&\curlywedge_2 = \{\text{Arithmetic mean}, \text{AndTop}, \text{AndBot}\} \\
&\curlyvee_3 = \{\text{Arithmetic mean}, \text{OrTop}, \text{OrBot}\} \\
&\curlyvee_4 = \{\text{OrTop}\} \\
&\curlywedge_4 = \{\text{AndTop}, \text{AndBot}\}
\end{aligned} \tag{2.30}
$$

and AndTop, AndBot, OrTop and OrBot are defined as follow:

$$\text{AndTop}(p,q) = \min(p,q) \tag{2.31}$$

$$\text{AndBot}(p,q) = \max(0, p+q-1) \tag{2.32}$$

$$\text{OrTop}(p,q) = \max(p,q) \tag{2.33}$$

$$\text{OrBot}(p,q) = \min(1, p+q) \tag{2.34}$$

With these inference templates, a set of 23 inference structures were formed and tested in a medical expert system. Among all, 3 Sub-BK inference structures were instantiated from Eq. (2.27), and they are referred as the BK1, BK2 and BK3 respectively:

$$\text{BK1} : \frac{1}{|B|}\sum_{b \in B}(R(a,b) \to S(b,c))$$

$$\text{BK2} : \text{AndTop}(R(a,b) \to S(b,c))$$

$$\text{BK3} : \text{AndBot}(R(a,b) \to S(b,c))$$

There is only 1 instantiation for the Sub-B inference structure:

$$\text{B1}: \min\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \max\ R(a,b), \max\ S(b,c)\Big)$$

For the Sub-K inference template, a list of 19 inference structures were instantiated, namely K1 - K19:

$$\text{K1}: \min\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \frac{1}{|B|}\sum_{b\in B}\big(\text{AndTop}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K2}: \min\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \frac{1}{|B|}\sum_{b\in B}\big(\text{AndBot}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K3}: \max\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \frac{1}{|B|}\sum_{b\in B}\big(\text{AndTop}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K4}: \max\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \frac{1}{|B|}\sum_{b\in B}\big(\text{AndBot}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K5}: \max\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \text{OrBot}\big(\text{AndTop}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K6}: \max\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \text{OrBot}\big(\text{AndBot}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K7}: \min\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \text{OrBot}\big(\text{AndBot}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K8}: \max\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \text{OrTop}\big(\text{AndTop}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K9}: \min\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \text{OrBot}\big(\text{AndTop}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K10}: \min\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \text{OrTop}\big(\text{AndBot}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K11}: \max\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \text{OrBot}\big(\text{AndBot}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K12} : \min\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\to S(b,c)), \text{OrTop}\big(\text{AndTop}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K13} : \min\Big(\text{AndTop}(R(a,b)\to S(b,c)), \text{OrTop}\big(\text{AndTop}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K14} : \min\Big(\text{AndTop}(R(a,b)\to S(b,c)), \text{OrTop}\big(\text{AndBot}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K15} : \min\Big(\text{AndBot}(R(a,b)\to S(b,c)), \text{OrBot}\big(\text{AndBot}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K16} : \min\Big(\text{AndBot}(R(a,b)\to S(b,c)), \text{OrBot}\big(\text{AndTop}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K17} : \max\Big(\text{AndTop}(R(a,b)\to S(b,c)), \text{OrBot}\big(\text{AndTop}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K18} : \min\Big(\text{AndTop}(R(a,b)\to S(b,c)), \text{OrBot}\big(\text{AndTop}(R(a,b),S(b,c))\big)\Big)$$

$$\text{K19} : \min\Big(\text{AndTop}(R(a,b)\to S(b,c)), \text{OrBot}\big(\text{AndBot}(R(a,b),S(b,c))\big)\Big)$$

In this application, the feature set is a set of signs and symptoms. $R(a,b)$ denotes the membership degree of the fuzzy relation R between a patient a and a sign or symptom b, whereas $S(b,c)$ is the membership degree of the fuzzy relation S between a sign or symptom b and a disease c. By studying the fuzzy relations R and S, the inference engine of this medical expert system is able to infer the relation between patients a and diseases c. The higher the membership degree of composition of relation between a patient and a disease, the higher the possibility the patient is suffering from that disease.

2.4 Interval-Valued Fuzzy Sets

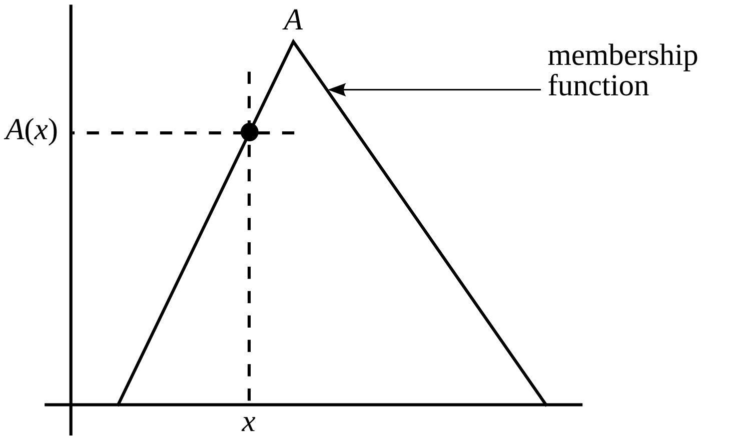
An interval-valued fuzzy set (IVFS) is a general form of T1FSs. According to Bustince (2000), it was developed by Sambuc (1975). In the past, it has been applied in various researches such as image processing (Bustince, Barrenechea, Pagola, & Fernandez, 2009; Fisher, 2007), forecasting (M & Mendez, 2007; Fazel Zarandi et al., 2009), reasoning (Xu, Kerre, Ruan, & Song, 2001; Turksen, 2002) and etc. In some literature, it is known as “interval type-2 fuzzy sets” (Aisbett, Rickard, & Morgenthaler, 2010). Over the years, T1FSs have been critisized for its limitation in representing uncertainty because T1FSs model uncertainties with point values. In many cases, it is too restrictive because point values do not able to capture uncertainties from multiple sources. Thus, the emerge of IVFSs relaxed the restriction by allowing uncertainty to be represented with an interval. The difference of IVFSs and T1FSs is discussed in the following.

Let A be a T1FS and $\tilde{A}$ be an IVFS, both in the universe X. It is common to define A as $(x, A(x))$ where $x \in X$ and $A(x) \in [0,1]$ is a point-valued membership function. Whereas for the IVFS, $\tilde{A} = \{(x, \tilde{A}(x))\}$ and $\tilde{A}(x) = [\underline{A}(x), \overline{A}(x)]$, where both $\underline{A}(x)$, $\overline{A}(x) \in [0,1]$ and $\underline{A}(x) \leq \overline{A}(x)$. Compared to T1FSs, whose membership functions are point-valued, the membership functions of $\tilde{A}$ are interval-values (IV) $[\underline{A}(x), \overline{A}(x)]$. $\underline{A}(x)$ is the Lower Membership Function (LMF) and $\overline{A}(x)$ is the Upper Membership Function (UMF). An area surrounded by a LMF and a UMF is defined as the Footprint Of Uncertainty (FOU) (Figure 2.2).

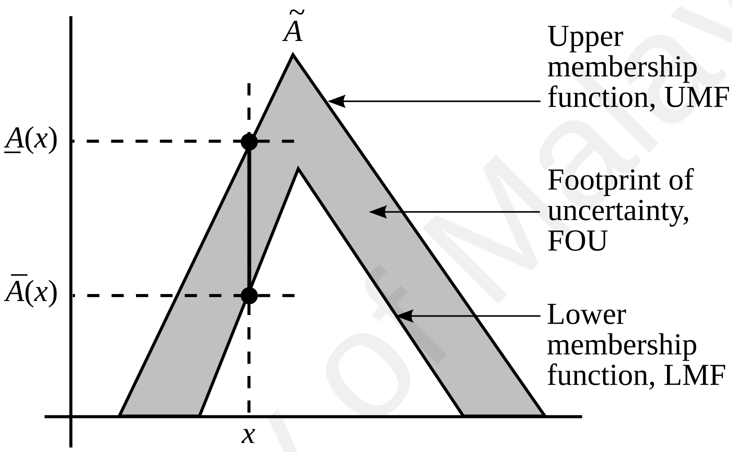
According to the Representation Theorem (Mendel & John, 2002), an IVFS can be considered as a collection of T1FSs:

$$\tilde{A} \rightarrow (\{A_1, A_2, \ldots, A_\eta\}) \tag{2.35}$$

where η is the total number of T1FSs “embedded” in the IVFS. In this sense, the computing of an IVFS can be reduced to the computing of multiple T1FSs. Assume that $\tilde{A} = \{x_i | i = \{1, 2, \cdots, I\}\}$ where $I \in \mathbb{N}$. If the membership degree corresponding to an x_i, $[\underline{A}(x_i), \overline{A}(x_i)]$ can be discretized into $J_i \in \mathbb{N}$ points, the total number of T1FSs that form



(a) Membership function of a T1FS.



(b) Membership function of an IVFS.

Figure 2.2: Comparison of membership functions of a T1FS and an IVFS.

the IVFS is given by η:

$$\eta = \prod_{i=1}^{I} J_i \tag{2.36}$$

2.5 Type-2 Fuzzy Sets

The theory of T2FSs first emerged in 1975 (Zadeh, 1975b, 1975d, 1975c). However, it was not widely adopted during the first few decades, due to the computational complexity (John & Coupland, 2007). With the development of tools and theories associated with T2FSs in recent years (N. N. Karnik, Mendel, & Liang, 1999; N. Karnik & Mendel, 2001; N. N. Karnik & Mendel, 2001; C.-H. Wang, Cheng, & Lee, 2004; Mitchell, 2006;

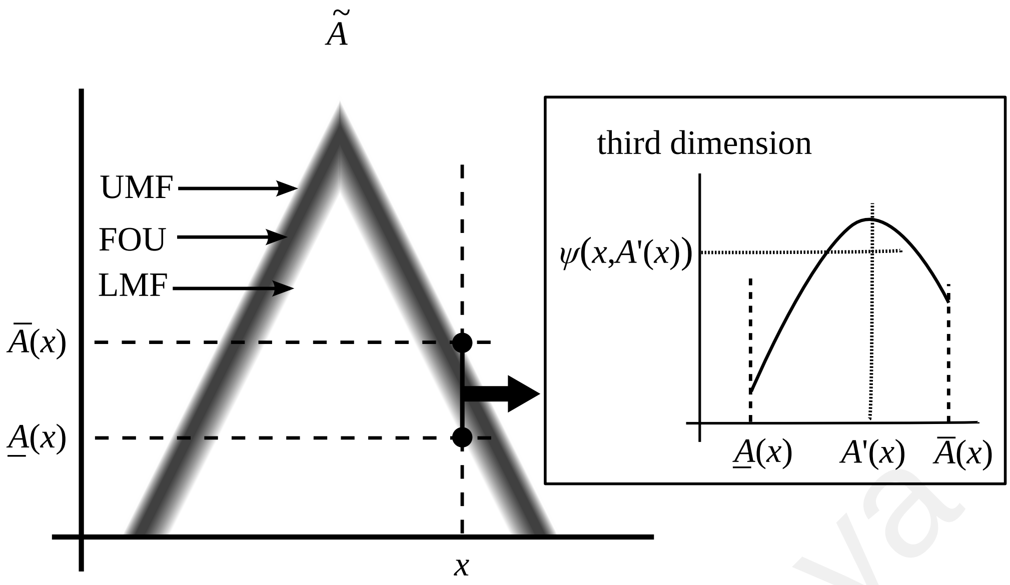


Figure 2.3: A T2FS and its membership function. The small box on the right is a vertical slice of element x.

Castro, Castillo, & Melin, 2007; Greenfield, Chiclana, Coupland, & John, 2009), the T2FSs have become more popular. However, a special type of T2FSs, i.e. the IVFSs is the main focus in this development (N. N. Karnik & Mendel, 2001). In general, T2FSs were claimed to have advantages in modeling problems where membership functions are ill-defined (Mizumoto & Tanaka, 1976). Thus, Computing With Words (CWW) (Zadeh, 1996, 1999, 2011) is one of the areas where T2FSs have found its applications (Mendel, 2007b, 2007a; D. Wu & Mendel, 2010).

Generally, a T2FS is a fuzzy set whose membership functions are T1FSs. If $\tilde{A}$ is a T2FS in universe X:

$$\tilde{A} = \left\{ \left(x, A(x), \psi(x, A(x)) \right) \right\} \tag{2.37}$$

where $x \in X$, $A(x) = [\underline{A}(x), \overline{A}(x)] \subseteq [0,1]$ and $\psi(x, A(x))$ is the vertical slice (Figure 2.3) that representing the membership degree of $A(x)$. For an $A'(x)$ where $\underline{A}(x) \leq A'(x) \leq \overline{A}(x)$, the secondary membership of $A'(x)$ is given by $\psi(x, A'(x)) \in (0,1]$. For $A'(x) < \underline{A}(x)$ and $A'(x) > \overline{A}(x)$, $\psi(x, A'(x)) = 0$.

An IVFS is a special case of T2FSs. In an IVFS, as long as $\underline{A}(x) \leq A'(x) \leq \overline{A}(x)$, the secondary memberships $\psi(x, A(x)) = 1$, otherwise $\psi(x, A(x)) = 0$. Representation Theorem (Mendel & John, 2002) is still applicable for T2FSs. In this case, a T2FS $\tilde{A}$ is a collection of T1FSs which associated with secondary membership degrees $\psi(x, A(x)) \in [0, 1]$.

2.6 Chapter Conclusion

In this chapter, the fundamentals of fuzzy relations were introduced. The BK products, which actually are the composition of relations are discussed in the subsequent section. Among all the three BK products, the BK subproduct is highlighted due to its popularity in the past. Then, two of its applications are examined in detail to facilitate the discussion in the following chapters. Some recent advances of fuzzy set theory are also discussed in this chapter, namely the IVFSs and T2FSs. Both of these T1FS extensions are getting more and more popular in the last decade.

CHAPTER 3

THE BK SUBPRODUCT IN FUZZY INFERENCE SYSTEMS

3.1 Introduction

In order to improve the implementation of the BK subproduct, a typical application as described in Section 2.3.4 is analysed in this chapter. Weaknesses in the inference engines are pointed out and suggestions are given in Section 3.2. Besides, an improvement of interval-valued inferencing (Yew & Kohout, 1996a), a defuzzification method for inference engines utilizing implication operators, is also discussed in Section 3.3.

3.2 Justification of Logical Connectives in Inference Structures

As discussed in Section 2.3.4, a list of 19 inference structures based on the Sub-K inference template were proposed by Yew and Kohout (1996a, 1996b, 1997). This section is dedicated to study the justification of adopting the logical connectives proposed in the research and highlights the improvements.

3.2.1 Shortcomings of Inference Structures

3.2.1 (a) Using max as Outer Logical Connective

Among these 19 inference structures (Yew & Kohout, 1996a, 1996b, 1997), 7 of them employed max as the outer logical connective, $\curlywedge_1$. These inference structures are K3, K4, K5, K6, K8, K11 and K17.

Please note that the purpose of $\curlywedge_1$ in the Sub-K inference template is to choose a candidate from two to be the result of a particular inference - the first candidate is the implication term that proposed by the original BK subproduct $\curlywedge_2(R(a,b) \rightarrow S(b,c))$, whereas the second candidate is the additional term $\curlyvee_3(\curlywedge_4(R(a,b),S(b,c))$, which was proposed by De Baets and Kerre (De Baets & Kerre, 1993a) to avoid the emptiness as discussed in Section 2.3.1 (i.e. object a that has no relations with any feature in set B eventually shows strong relations with all the elements in C). In case of the presence of emptiness, operations on the implication term yield 1.0 due to the nature of implication operators $r \rightarrow s = 1$ if $r = 0$, $\forall s \in [0,1]$.

Obviously, replacing $\curlywedge_1$ as min can cause the results of the particular inference come from the additional term. Though the problem of emptiness could be solved with the solution proposed by De Baets and Kerre (De Baets & Kerre, 1993a), max as $\curlywedge_1$ will yield the implication term as inference output, which implies that the non-emptiness condition is ignored again.

So, it is clear that max is not a valid outer connective in solving the Sub-K inference template due to the non-emptiness condition. The same conclusion holds if the Sub-B Inference Template is studied and the reason is trivial.

3.2.1 (b) Using AndBot to Aggregate the Results of Implications

In K15 and K16, AndBot was used as $\curlywedge_2$, the aggregation operator for implication results in the implication term. However, this logical connective may not work as expected in practice. We explain this in the following paragraphs.

In a J-variable environment, AndBot in Eq. (2.32) can be generalized as follow:

$$\text{AndBot}(\varphi) = \max\left(0, \sum_{j=1}^{J} \varphi_j - (J-1)\right) \tag{3.1}$$

where $j \in \{1, \cdots, J\}$. In K15 and K16, the result of each implication is aggregated with

Eq. (3.1), where φ_j should be substituted with $R(a,b_j) \to S(b_j,c)$.

The problem associated with AndBot as the aggregation operator is not prominent when J is small and results of most $R(a,b_j) \to S(b_j,c)$ are big. However, it is easy to verify that when J increases, the result of the aggregation decreases as long as the outcome of an implication is not 1.0. Some numerical examples that may happen in real applications are as follows: 1) If all the 10 implications yield 0.9, the result of the aggregation is 0.0; 2) Out of J implications, if an implication yields 0.0, the result of the whole aggregation is also 0.0, even though all the other implications yield 1.0.

In this sense, the demand for high $R(a,b_j)$ and $S(b_j,c)$ values simultaneously of this logical connective is not a desirable property of an aggregation operator. So, we conclude that K15 and K16, as well as any inference structures that adopt AndBot to aggregate the results of implications are not good inference structures practically.

3.2.1 (c) The Influence of De Baets and Kerre's Improvement on BK Sub-triangle Product

The validity of the BK subproduct comes from the subsethood measurement of one set in another, which is provided by the implication term. Apparently, the additional term added by De Baets and Kerre (1993a) to rectify the problem of emptiness has some degrees of influence in the K-series of inference structures, despite it is the solution to the non-emptiness condition.

Also, we learned in Section 3.2.1 (a) that choosing min as $\curlywedge_1$ is to avoid the emptiness in $R(a,b)$ and affects the inference result. However, the additional term that was proposed by De Baets and Kerre (1993a) generates smaller values in most cases, not only when the $R(a,b)$ is an empty set. And so, with the min outer connective, the real subsethood measurements are ignored occasionally.

To study this influence in more detail, let us consider the simplest instance where only one pair of fuzzy relations (r,s) is involved. Two possible candidates of implication operators, $\rightarrow_{\text{Ł}}$ and $\rightarrow_{\mathrm{KD}}$ are listed in Table 3.1a and Table 3.1b respectively. AndTop and AndBot, logical connectives that correspond to the additional term, are listed in Table 3.2a and Table 3.2b respectively. Table 3.3 shows the difference between these two terms when the values of implications were override by additional term. Those cells without any values represent the combinations of r and s for which the values of implications are smaller than or equal to the additional terms.

Table 3.1: Values Generated By Implication Operators

(a) $r \rightarrow_{\mathrm{L}} s$

r \ s	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
0.2	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
0.3	0.8	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
0.4	0.7	0.8	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0
0.5	0.6	0.7	0.8	0.9	1.0	1.0	1.0	1.0	1.0	1.0
0.6	0.5	0.6	0.7	0.8	0.9	1.0	1.0	1.0	1.0	1.0
0.7	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.0	1.0	1.0
0.8	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.0	1.0
0.9	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.0
1.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

(b) $r \rightarrow_{\mathrm{KD}} s$

r \ s	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	1.0
0.2	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.9	1.0
0.3	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.9	1.0
0.4	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.8	0.9	1.0
0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.7	0.8	0.9	1.0
0.6	0.4	0.4	0.4	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.7	0.3	0.3	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.8	0.2	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.9	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

From Table 3.3, it is clear that the solution proposed by De Baets and Kerre (1993a) has a certain influence on the inference structures in the case that $\curlywedge_1 = \min$ and there is only one pair of (r,s) relations. This causes a dilemma in choosing an appropriate outer

Table 3.2: Values Generated By AndTop And AndBot

(a) AndTop(r,s)

r \ s	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
0.3	0.1	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
0.4	0.1	0.2	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.4
0.5	0.1	0.2	0.3	0.4	0.5	0.5	0.5	0.5	0.5	0.5
0.6	0.1	0.2	0.3	0.4	0.5	0.6	0.6	0.6	0.6	0.6
0.7	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.7	0.7	0.7
0.8	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.8	0.8
0.9	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.9
1.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

(b) AndBot(r,s)

r \ s	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2
0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.3
0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.3	0.4
0.5	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.3	0.4	0.5
0.6	0.0	0.0	0.0	0.0	0.1	0.2	0.3	0.4	0.5	0.6
0.7	0.0	0.0	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.8	0.0	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0.9	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

connective since max is not a good connective either.

3.2.2 Candidates of Logical Connectives

With the argument that the influence of the additional term will be minor and tolerable once the total number of features, J increases, it is fair to solve the problems by reconstructing the set of influence structures with reasonable logical connectives.

Firstly, there is no reason that AndBot must be kept as a candidate of $⅄_2$. Subsequently, max should also be removed from the list of $⅄_1$ due to the shortcomings highlighted in Section 3.2.1 (a).

Moreover, in order to minimize the influence of the additional term, the set of logical

Table 3.3: Comparing Values Generated By Implication Operators and AND Operators

(a) $r \to_{Ł} s - \text{AndTop}(r,s)$ if $r \to_{Ł} s > \text{AndTop}(r,s)$

r \ s	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
0.2	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
0.3	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
0.4	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
0.6	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
1.0										

(b) $r \to_{Ł} s - \text{AndBot}(r,s)$ if $r \to_{Ł} s > \text{AndBot}(r,s)$

r \ s	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9
0.2	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.8
0.3	0.8	0.9	1.0	1.0	1.0	1.0	1.0	0.9	0.8	0.7
0.4	0.7	0.8	0.9	1.0	1.0	1.0	0.9	0.8	0.7	0.6
0.5	0.6	0.7	0.8	0.9	1.0	0.9	0.8	0.7	0.6	0.5
0.6	0.5	0.6	0.7	0.8	0.8	0.8	0.7	0.6	0.5	0.4
0.7	0.4	0.5	0.6	0.6	0.6	0.6	0.6	0.5	0.4	0.3
0.8	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.3	0.2
0.9	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1
1.0										

(c) $r \to_{\text{KD}} r - \text{AndTop}(r,s)$ if $s \to_{\text{KD}} s > \text{AndTop}(r,s)$

r \ s	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.9
0.2	0.7	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.8
0.3	0.6	0.5	0.4	0.4	0.4	0.4	0.4	0.5	0.6	0.7
0.4	0.5	0.4	0.3	0.2	0.2	0.2	0.3	0.4	0.5	0.6
0.5	0.4	0.3	0.2	0.1		0.1	0.2	0.3	0.4	0.5
0.6	0.3	0.2	0.1				0.1	0.2	0.3	0.4
0.7	0.2	0.1						0.1	0.2	0.3
0.8	0.1								0.1	0.2
0.9										0.1
1.0										

(d) $r \to_{\text{KD}} s - \text{AndBot}(r,s)$ if $r \to_{\text{KD}} s > \text{AndBot}(r,s)$

r \ s	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
0.2	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
0.3	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
0.4	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
0.6	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
0.7	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
0.9	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
1.0										

connectives can be further reduced, especially $\curlyvee_3$, which is the main determiner of the additional term. To reduce the influence of the additional term, a larger possible value should be generated by $\curlyvee_3$. Since both OrTop and OrBot are greater than the arithmetic mean, the list of logical connectives can be further reduced to :

$$\begin{aligned} \curlywedge_1 &= \{\text{min}\} \\ \curlywedge_2 &= \{\text{Arithmetic mean}, \ \text{AndTop}\} \\ \curlyvee_3 &= \{\text{OrTop}, \ \text{OrBot}\} \\ \curlywedge_4 &= \{\text{AndTop}, \ \text{AndBot}\} \end{aligned} \tag{3.2}$$

These logical connectives generate a list of 8 sub-K inference structures, namely K7, K9, K18, K19, K10, K12, K13 and K14. One can compare to the experiment results conducted by Yew and Kohout (1997) and find out that these are the top ranked (high Mean True Acceptance rate) inference structures in performance, especially K7, K9, K18 and K19, along with BK2 and BK3. On the other hand, Sub-K inference structures that use max as outer connective, such as K3, K4, K5, K6, K8, K11 and K17 have the highest Mean False Acceptance rate. All these inference structures have Mean False Acceptance rate from 0.56 to 0.96, due to the influence of the additional term and ignorance of the non-emptiness condition. The consistency of the experiment results (Yew & Kohout, 1997) shows that the theoretical discussion in this section is supported by empirical work.

3.3 Reliability Measure in Defuzzification with Interval-Valued Reasoning

Although the mechanism of inference engines developed by the BK subproduct work is different compared to CRI-based inference engine, the flow of information processing in both models is somewhat similar (Figure 1.1). For both models, the process of an inference starts with interpreting fuzzified input signals based on the predefined information stored in a knowledge base by an inference engine. Once all the input signals are interpreted, an aggregator is used to aggregate the results. To produce meaningful results in a system, a defuzzification module is used to prepare the results in the forms that meet the output requirements.

Some popular defuzzification methods for CRI-based inference engines include Center of Gravity (COG), Center of Maxima (COM), Mean of Maxima (MOM) and etc (Filev & Yager, 1991; Klir & Yuan, 1995; Fortemps & Roubens, 1996; Patel & Mohan, 2002). In these methods, a value is computed to represent an output set.

For the BK products-based inference engines, one of the interesting properties is that it performs inferences relying on fuzzy implication operators, which can be tailored based on the needs of applications. Based on this special property, interval-valued inferencing was proposed by Kohout and Bandler (1992) as a defuzzification method to obtain more reliable results in inferences. Instead of providing point-values as outputs, intervals are given. The intervals can be obtained with the implementation of the Kleene-Dienes and Łukasiewicz implementation operators in the inferencing. For example, for the original fuzzy BK subproduct (Eq. (2.17)):

$$R \triangleright_{BK} S(a,c) = \Big[\bigwedge_{b \in B} \big(R(a,b) \to_{\mathrm{KD}} S(b,c)\big), \bigwedge_{b \in B} \big(R(a,b) \to_{\mathrm{L}} S(b,c)\big)\Big] \tag{3.3}$$

An inference is considered as accepted if and only if the whole result interval is in the acceptance band $[\beta, 1]$, where β is the predefined acceptance threshold value. On the other hand, an inference is rejected if the whole result interval is in the rejection band $[0, \alpha]$, where α is the predefined rejection threshold value. The intervals that are partly in the acceptance band or rejection band are considered as falling into grey area.

However, despite its great idea, this interval-valued inferencing technique suffers from low efficiency because of unnecessary computation, as well as falling into the realm of dichotomy, i.e. an inference can only be 'accepted' or 'not accepted' (or 'rejected' or 'not rejected' in the other way round). In this section, an improvement to the interval-valued inferencing technique is explained. For a given threshold, this improved defuzzification method not only proposes acceptable inferences, but also the reliabilities of these inferences.

3.3.1 Limitations of Interval-Valued Inferencing

From the definition of Kleene-Dienes and Łukasiewicz fuzzy implication operators in Table 2.1, one can easily prove that:

$$\forall r, s \in [0,1] \qquad r \rightarrow_{\mathrm{KD}} s \;\; \leq \;\; r \rightarrow_{\text{Ł}} s \tag{3.4}$$

One can also refer to Table 3.1 for examples of calculations of $\rightarrow_{\mathrm{KD}}$ and $\rightarrow_{\text{Ł}}$.

With Eq. (3.4), it is clear that for an inference to obtain an interval in the acceptance band, computation using $\rightarrow_{\mathrm{KD}}$ is sufficient. This is because whenever a computation with $\rightarrow_{\mathrm{KD}}$ is in the acceptance band, the computation with $\rightarrow_{\text{Ł}}$ which yields a result greater or equal to $\rightarrow_{\mathrm{KD}}$ is always in the acceptance band too:

$$\forall r, s \in [0,1] \qquad r \rightarrow_{\mathrm{KD}} s \;\; \geq \beta \qquad \Rightarrow \qquad r \rightarrow_{\text{Ł}} s \;\; \geq \beta \tag{3.5}$$

On the other hand, if the computation using $\rightarrow_{\mathrm{KD}}$ is not in the acceptance band, the result of computation using $\rightarrow_{\text{Ł}}$ is not relevant any more because the interval is not entirely in the acceptance band and the inference is not going to be accepted. To check whether an inference is rejected, a similar argument holds, $\rightarrow_{\text{Ł}}$ is the only needed fuzzy implication for this purpose.

Furthermore, this defuzzification method may suffer from the same problem as in crisp systems - i.e. details are oversimplified by applying threshold values. For example, if $\beta = 0.80$, then the interval $\mathscr{I}_1 = [0.80, 0.90]$ will be accepted but $\mathscr{I}_2 = [0.79, 0.90]$ and $\mathscr{I}_3 = [0.79, 1.00]$ will be rejected, although the differences between $\mathscr{I}_1$ and $\mathscr{I}_2$ are very small, and $\mathscr{I}_3$ has a higher upper bound compared to $\mathscr{I}_1$.

As a summary from our findings, we can conclude that (Kohout, Stabile, Kalantar, & San-Andres, 1995; Kohout & Bandler, 1992; Yew & Kohout, 1996a) which perform interval-valued inferencing do not gain any significant advantages though (Yew & Kohout, 1996a) claimed that the intervals have better accuracy over point values.

3.3.2 Improving Interval-Valued Inferencing

Bandler and Kohout (Kohout & Bandler, 1980a) revised the fuzzy subset theory proposed by Zadeh (Zadeh, 1965) stating that a fuzzy set A is a subset of another fuzzy set B if and only if:

$$\forall x \in X \qquad A(x) \leq B(x) \tag{3.6}$$

Apparently, Zadeh's fuzzy subset theory is "an unconscious step backward to the realm of dichotomy" (Kohout & Bandler, 1980a). With this theory, a fuzzy set is either utterly a subset or not a subset of the other fuzzy set. To rectify this problem, a subsethood theory that based on implication operator was proposed (discussed in Section 2.3).

Based on the same justification, this research proposes a new interval-valued inferencing scheme that can rectify the limitations as discussed in Section 3.3.1. This new scheme not only provides information about whether an inference result is in the predefined band (either acceptant or rejection), but also measures the reliability of an inference based on a threshold value.

Basically, this improved method measures the reliability of an inference result. It

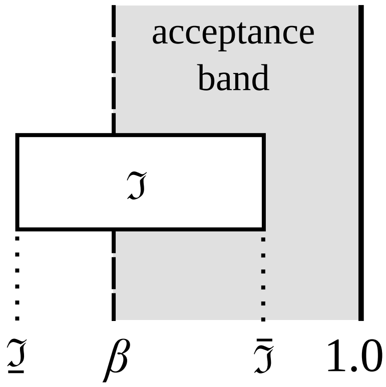


Figure 3.1: Interval $\mathfrak{I}$ which is partially covered in acceptance band $[\beta, 1.0]$. The reliability of this interval being accepted is given by Eq. (3.7).

evaluates the ratio of the output interval in the acceptance or rejection band. If the whole output interval is in the acceptance or rejection band, it is reasonable to assume that the inference is completely accepted or rejected and the reliability is 1.0. Otherwise, the coverage of the interval in the corresponding evaluation band gives the reliability of this inference.

Assume that the result of an inference is an interval $\mathfrak{I} = [\underline{\mathfrak{I}}, \overline{\mathfrak{I}}]$, where $\underline{\mathfrak{I}}$ and $\overline{\mathfrak{I}}$ correspond to the results of Kleene-Dienes and Łukasiewicz fuzzy implication operators respectively, and $\underline{\mathfrak{I}} < \overline{\mathfrak{I}}$. If $\overline{\mathfrak{I}} > \beta$, it implies that at least a portion of $[\underline{\mathfrak{I}}, \overline{\mathfrak{I}}]$ is in the acceptance band (Figure 3.1). Therefore, we can measure the reliability of this inference with result $\mathfrak{I}$ being accepted at threshold value β, $\Theta_{\mathfrak{I}}^{\beta}$:

$$\Theta_{\mathfrak{I}}^{\beta} = \begin{cases} 1 & \text{if } \underline{\mathfrak{I}} \geq \beta \\ \dfrac{\overline{\mathfrak{I}} - \beta}{\overline{\mathfrak{I}} - \underline{\mathfrak{I}}} & \text{if } \underline{\mathfrak{I}} < \beta < \overline{\mathfrak{I}} \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

and $\Theta \in [0, 1]$.

Similarly, the measurement of the reliability of an inference with interval $\mathfrak{I}$ in rejection band at threshold value α is given by Υ:

$$\Upsilon_{\mathfrak{I}}^{\alpha} = \begin{cases} 1 & \text{if } \overline{\mathfrak{I}} \leq \alpha \\ \dfrac{\alpha - \underline{\mathfrak{I}}}{\overline{\mathfrak{I}} - \underline{\mathfrak{I}}} & \text{if } \underline{\mathfrak{I}} < \alpha < \overline{\mathfrak{I}} \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

and $\Upsilon \in [0, 1]$.

The magnitude of Θ and Υ shows the reliability of an inference in the acceptance band or rejection band, respectively. Instead of utterly accepting or rejecting an inference, this new scheme provides a better tolerance towards uncertainty in inputs and the choice of acceptance and rejection threshold values.

In case of $\underline{\mathfrak{I}} = \overline{\mathfrak{I}}$, the result is a point value and the reliability measure is not applicable.

3.3.3 Consistency of Reliability Measure

Consistency, which means absence of contradiction in interpreting inference results, is an important property of a defuzzification module. A high consistency defuzzification module always shows better performance compared to the others.

One maybe concerned about the consistency of this improvement on the interval-valued inferencing, i.e. whether the reliabilities of 2 different intervals increase/decrease consistently when the threshold values (α and β) vary. More precisely, in acceptance bands, when β increases, will the reliability of an interval decrease to zero faster than the other intervals which previously have lower reliability? Similarly, in rejection bands, when α decreases, will the reliability of an interval decrease to zero faster than the other intervals which previously have lower reliability?

Consistency in acceptance band:

$$\Theta_{\mathfrak{I}_1}^{\beta_1} \leq \Theta_{\mathfrak{I}_2}^{\beta_1} \Rightarrow \Theta_{\mathfrak{I}_1}^{\beta_2} \leq \Theta_{\mathfrak{I}_2}^{\beta_2} \qquad \text{where } \beta_1 \leq \beta_2 \tag{3.9}$$

Consistency in rejection band:

$$\Upsilon_{\mathfrak{I}_1}^{\alpha_1} \geq \Upsilon_{\mathfrak{I}_2}^{\alpha_1} \Rightarrow \Upsilon_{\mathfrak{I}_1}^{\alpha_2} \geq \Upsilon_{\mathfrak{I}_2}^{\alpha_2} \qquad \text{where } \alpha_1 \leq \alpha_2 \tag{3.10}$$

2 intervals may form 4 possible combinations of arrangement, in the following, we examine these 4 cases when the threshold values of acceptance bands varies. A similar discussion applies to the study of consistency in the rejection band.

(a) Case 1 (Same $\underline{\mathfrak{I}}$)

Assume that the 2 intervals are $\mathfrak{I}_1$ and $\mathfrak{I}_2$. Both of them have the same lower bounds $\underline{\mathfrak{I}}$, and the upper bounds are $\overline{\mathfrak{I}}_1$ and $\overline{\mathfrak{I}}_2$ respectively and $\overline{\mathfrak{I}}_1 < \overline{\mathfrak{I}}_2$ (Figure 3.2).

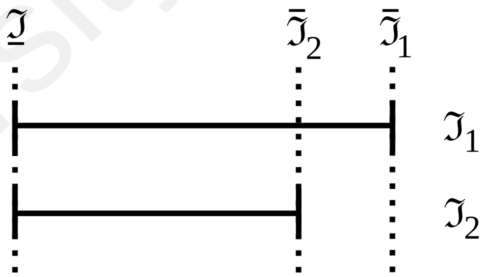


Figure 3.2: Case 1: Both intervals have the same lower bound but different upper bounds.

Both the reliabilities $\Theta_{\mathfrak{I}_1}$ and $\Theta_{\mathfrak{I}_2}$ are 1 when the acceptance threshold value, $\beta \leq \underline{\mathfrak{I}}$. When $\underline{\mathfrak{I}} < \beta < \overline{\mathfrak{I}}_2$, both $\Theta_{\mathfrak{I}_2}$ and $\Theta_{\mathfrak{I}_1}$ are decreasing. Decrease rate of $\Theta_{\mathfrak{I}_2}$ is higher compared to $\Theta_{\mathfrak{I}_1}$ until it reaches 0 at $\beta = \overline{\mathfrak{I}}_2$. This situation is considered as consistent because $\Theta_{\mathfrak{I}_2} \leq \Theta_{\mathfrak{I}_1}$ for any value of β.

(b) Case 2 (Same $\overline{\mathfrak{I}}$)

Both intervals $\mathfrak{I}_1$ and $\mathfrak{I}_2$ have the same upper bound $\overline{\mathfrak{I}}$, but the lower bounds are $\underline{\mathfrak{I}_1}$ and $\underline{\mathfrak{I}_2}$ respectively and $\underline{\mathfrak{I}_1} > \underline{\mathfrak{I}_2}$ (Figure 3.3).

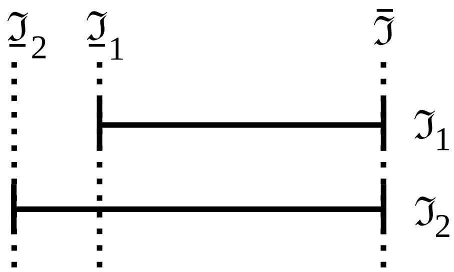


Figure 3.3: Case 2: Both intervals have the same upper bound but different lower bounds.

Both the reliabilities $\Theta_{\mathfrak{I}_1}$ and $\Theta_{\mathfrak{I}_2}$ are 1 when the acceptance threshold value $\beta \leq \underline{\mathfrak{I}_2}$. When β increases until $\underline{\mathfrak{I}_2} < \beta \leq \underline{\mathfrak{I}_1}$, $\Theta_{\mathfrak{I}_1} = 1$ but $\Theta_{\mathfrak{I}_2} \in (0,1)$. This is because the whole $\mathfrak{I}_1$ is in the acceptance band but only a portion of $\mathfrak{I}_2$ is accepted. When $\underline{\mathfrak{I}_1} < \beta \leq \overline{\mathfrak{I}}$, both $\Theta_{\mathfrak{I}_1}$ and $\Theta_{\mathfrak{I}_2}$ decrease with consistent rates to 0 until $\beta = \overline{\mathfrak{I}}$. Although $\Theta_{\mathfrak{I}_1}$ has higher decreasing rate compared to $\Theta_{\mathfrak{I}_2}$, but this does not affect the consistency because the condition $\Theta_{\mathfrak{I}_2} \leq \Theta_{\mathfrak{I}_1}$ always true until both reach 0 at the same time.

(c) Case 3 (Sequence)

In this case, both the lower and upper bounds of interval $\mathfrak{I}_1$ are higher than their counterparts of interval $\mathfrak{I}_2$ (Figure 3.4). Compared to case 2, the only major difference is $\overline{\mathfrak{I}}_2 < \overline{\mathfrak{I}}_1$ and this caused $\Theta_{\mathfrak{I}_1} > 0$ when $\Theta_{\mathfrak{I}_2}$ reach 0 at $\beta = \overline{\mathfrak{I}}_2$. Hence, the proposed defuzzification method is still showing its consistency in this case.

(d) Case 4 (Bounded)

In this case, interval $\mathfrak{I}_1$ is bounded in interval $\mathfrak{I}_2$ (Figure 3.5) so that $\underline{\mathfrak{I}_2} < \underline{\mathfrak{I}_1} < \overline{\mathfrak{I}}_1 < \overline{\mathfrak{I}}_2$.

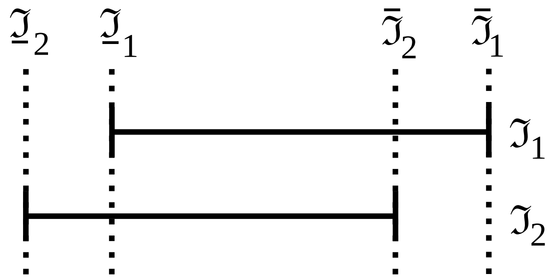


Figure 3.4: Case 3: Both upper and lower bounds of interval $\mathfrak{I}_1$ are greater than their counterparts of interval $\mathfrak{I}_2$

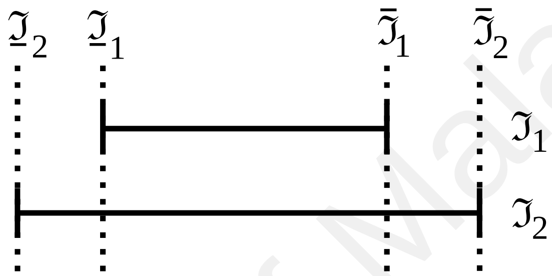


Figure 3.5: Case 4: Interval $\mathfrak{I}_1$ is bounded in interval $\mathfrak{I}_2$

It is obvious that this is the only case where the consistency of this defuzzification method does not hold. $\Theta_{\mathfrak{I}_1} = \Theta_{\mathfrak{I}_2} = 1$ when $\beta \leq \underline{\mathfrak{I}}_2$ at the initial. If β increases to $\underline{\mathfrak{I}}_2 < \beta < \underline{\mathfrak{I}}_1$, $\Theta_{\mathfrak{I}_2} < \Theta_{\mathfrak{I}_1} = 1$. In the range of $\underline{\mathfrak{I}}_1 < \beta < \overline{\mathfrak{I}}_1$, $\Theta_{\mathfrak{I}_1}$ will decrease at a higher rate compared to $\Theta_{\mathfrak{I}_2}$ if β increases, until $\Theta_{\mathfrak{I}_1} = 0$ when $\beta = \overline{\mathfrak{I}}_1$. But in the range of $\overline{\mathfrak{I}}_1 \leq \beta < \overline{\mathfrak{I}}_2$, $\Theta_{\mathfrak{I}_2} > 0$.

Anyway, the inconsistency of this defuzzification method does not totally denied its usefulness. In fact, the inconsistency only occurs in the case $\underline{\mathfrak{I}}_2 < \underline{\mathfrak{I}}_1$ and $\overline{\mathfrak{I}}_1 < \overline{\mathfrak{I}}_2$. i.e. if $\mathscr{R}_1 \to \mathscr{S}_2 = [\underline{\mathfrak{I}}_1, \overline{\mathfrak{I}}_1]$ and $\mathscr{R}_2 \to \mathscr{S}_2 = [\underline{\mathfrak{I}}_2, \overline{\mathfrak{I}}_2]$, the following set of 4 constraints must be fulfilled to cause inconsistency among 2 sets of inputs:

Constraint 1:

$$\mathscr{R}_1 \neq 1; \qquad \mathscr{S}_1 \neq 0; \qquad \mathscr{R}_2 \neq 0; \qquad \mathscr{S}_2 \neq 1 \tag{3.11}$$

Constraint 2:

$$\mathscr{S}_1 < \mathscr{R}_1 \tag{3.12}$$

Constraint 3:

$$\mathscr{R}_1 - \mathscr{S}_1 > \mathscr{R}_2 - \mathscr{S}_2 \tag{3.13}$$

Constraint 4:

either

$$\mathscr{S}_1 > \mathscr{S}_2 \qquad \text{and} \qquad \mathscr{S}_1 > 1 - \mathscr{R}_2 \tag{3.14}$$

or

$$\mathscr{R}_1 < \mathscr{R}_2 \qquad \text{and} \qquad \mathscr{R}_1 < 1 - \mathscr{S}_2 \tag{3.15}$$

Proof : Case 4 representing a scenario where the following two conditions fulfilled simultaneously:

$$\underline{\mathfrak{I}}_2 < \underline{\mathfrak{I}}_1 \qquad \text{and} \qquad \overline{\mathfrak{I}}_1 < \underline{\mathfrak{I}}_2$$

This can be illustrated as:

$$\max(\mathscr{S}_1, 1 - \mathscr{R}_1) > \max(\mathscr{S}_2, 1 - \mathscr{S}_2) \tag{3.16}$$

and

$$\min(1, 1 - \mathscr{R}_1 + \mathscr{S}_1) < \min(1, 1 - \mathscr{R}_2 + \mathscr{S}_2) \tag{3.17}$$

From Eq. (3.16), by finding the limits of each variables, we get Constraint 1:

$$\mathscr{S}_2 \neq 1; \qquad \mathscr{R}_2 \neq 0; \qquad \mathscr{S}_1 \neq 0; \qquad \mathscr{R}_1 \neq 1$$

Also from Eq. (3.16), in case of $\mathscr{S}_1 > (1-\mathscr{R}_1)$, we get:

$$\mathscr{S}_1 > \mathscr{S}_2; \qquad \mathscr{S}_1 > (1-\mathscr{R}_2) \tag{3.18}$$

or, alternatively, if $\mathscr{S}_1 \leq (1-\mathscr{R}_1)$:

$$\mathscr{R}_1 < \mathscr{R}_2; \qquad \mathscr{R}_1 < (1-\mathscr{S}_2) \tag{3.19}$$

Combining Eq. (3.18) and Eq. (3.19) form Constraint 4.

On the other hand, from Eq. (3.17), we can find 2 inequalities:

$$1-\mathscr{R}_1+\mathscr{S}_1 < 1 \tag{3.20}$$

and

$$1-\mathscr{R}_1+\mathscr{S}_1 < 1-\mathscr{R}_2+\mathscr{S}_2 \tag{3.21}$$

From Eq. (3.20), we can get Constraint 2:

$$\mathscr{S}_1 < \mathscr{R}_1$$

and from Eq. (3.21), we can get Constraint 3:

$$\mathscr{R}_1-\mathscr{S}_1 > \mathscr{R}_2-\mathscr{S}_2$$

□

This set of constraints highly limits the chance of getting pairs of $(\mathscr{R}_1,\mathscr{S}_1)$ and $(\mathscr{R}_2,\mathscr{S}_2)$ that will cause inconsistency. Thus, it is rather safe to conclude that this defuzzification method is consistent most of the time.

3.4 Chapter Conclusion

In this chapter, the implementations of the BK products, particularly the BK subproduct was reviewed. Two improvements were suggested so that better inference systems can be formed from the BK products.

The first revision is on the inference structures developed from the improved BK subproduct. Through the theoretical analysis, limitations were found on previous proposed inference structures. This result is also supported by the empirical studies in the past. Therefore, if an improved version of the BK subproduct is going to be considered in inference engines, only K7, K9, K18, K19, K10, K12, K13 and K14 are suggested.

The defuzzification method for the BK subproduct, i.e. interval-valued inferencing proposed by Kohout and Bandler (1992) has also been revised. Although the idea behind the interval-valued inferencing is great, it suffers from low efficiency and dichotomy. With the introduction of the reliability measure in Section 3.3, the shortcoming was resolved. Anyway, the proposed solution may suffer from inconsistency in specific cases. However, this inconsistency is not an issue if fixed thresholds are applied during the defuzzifications.

CHAPTER 4

DEVELOPING WEIGHTED IVFSS BASED BK PRODUCTS

4.1 Introduction

In this chapter, the extensions of BK products, particularly the BK subproduct will be discussed. Due to the fuzzy subsethood measure is the foundation of BK products, this chapter starts with a discussion on subsethood measures for IVFSs. With a proposed subsethood measure, the BK products are extended to IVFS-based. The incorporation of the weight parameter with BK products, as well as its computations are detailed in the later part of this chapter.

4.2 Subsethood Measure of Interval Valued Fuzzy Sets

BK products are based on the fuzzy subsethood measure with the fuzzy implication operators. However, all the fuzzy implication operators are only defined for point-valued membership degrees, but the membership degrees of IVFSs are intervals. To solve this problem, two IVFS subsethood measures based on the fuzzy implication operators are proposed in this research, namely the Complete Derivation Method and Border Evaluation Method.

4.2.1 Complete Derivation Method

Let $\tilde{P}$ and $\tilde{Q}$ be two IVFSs in universe X as shown in Figure 4.1. For an element x_i, the interval membership degrees for x_i in $\tilde{P}$ and $\tilde{Q}$ are $[\underline{P}(x_i), \overline{P}(x_i)]$ and $[\underline{Q}(x_i), \overline{Q}(x_i)]$ respectively.

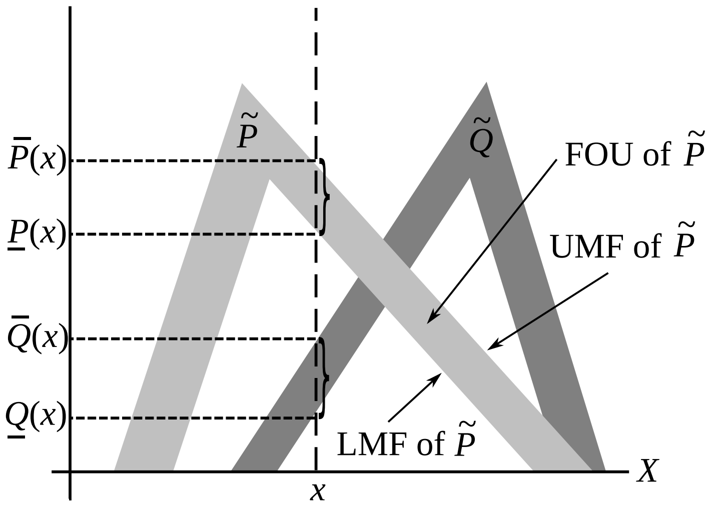


Figure 4.1: Two IVFSs $\tilde{P}$ and $\tilde{Q}$ in the same universe X.

Assume that both axes of element and membership degree are discrete or can be discretized. Representation Theorem (Mendel & John, 2002) suggests that, if the total number of T1FSs for $\tilde{P}$ is given by Eq. (2.36), then the number of T1FSs that pass through a discrete point $P(x_i)_j$ is given by $\frac{\eta_{\tilde{P}}}{J_i}$. For $\tilde{Q}$, if K_i is the number of discrete membership degrees in $[\underline{Q}(x_i), \overline{Q}(x_i)]$ and $k = \{1, 2, \cdots, K_i\}$, the number of T1FSs that pass through a discrete point $Q(x_i)_k$ is given by $\frac{\eta_{\tilde{Q}}}{K_i}$.

To formulate the fuzzy subsethood measure for IVFSs, we start with evaluating an arbitrary pair of point membership degrees $P(x_i)_j$ and $Q(x_i)_k$ in $\tilde{P}$ and $\tilde{Q}$ respectively, on a same element x_i. If these are the only points on x_i for $\tilde{P}$ and $\tilde{Q}$, following Eq. (2.16), the subsethood measure on this element can be written as $\pi(\tilde{P} \subseteq \tilde{Q})(x_i) = P(x_i)_j \to Q(x_i)_k$.

However, since there are $\frac{\eta_{\tilde{P}}}{J_i}$ and $\frac{\eta_{\tilde{Q}}}{K_i}$ of T1FSs on $P(x_i)_j$ and $Q(x_i)_k$ respectively, this implication involves a number of $\frac{\eta_{\tilde{P}}}{J_i} \times \frac{\eta_{\tilde{Q}}}{K_i}$ pairs of T1FSs. Therefore, this implication should be represented as:

$$\frac{\eta_{\tilde{P}}\eta_{\tilde{Q}}}{J_i K_i}\Big(P(x_i)_j \to Q(x_i)_k\Big) \tag{4.1}$$

If $P(x_i)_j$ is the only discrete point for $\tilde{P}(x_i)$, the subsethood measure can be evaluate by summing up all the implications of this point membership degree to all the $Q(x_i)_k$ (Fig

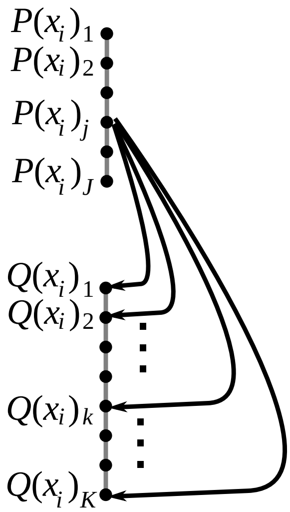


Figure 4.2: Implication of a single membership degree in $\tilde{P}(x_i)$ to all the discretized membership degrees of $\tilde{Q}(x_i)$

.

4.2):

$$\frac{\eta_{\tilde{P}}\eta_{\tilde{Q}}}{J_i K_i}\sum_{k=1}^{K_i}\left(P(x_i)_j \to Q(x_i)_k\right) \tag{4.2}$$

Generally, $[\underline{P}(x_i), \overline{P}(x_i)]$ are intervals with more than one discrete points. Therefore, if we generalized Eq. (4.2) to all the $P(x_i)_j$ in the element x_i and normalized it with the total number of T1FSs , we get the fuzzy subsethood measure for this element:

$$\pi(\tilde{P} \subseteq \tilde{Q})(x_i) = \frac{1}{J_i K_i}\sum_{j=1}^{J_i}\sum_{k=1}^{K_i}\left(P(x_i)_j \to Q(x_i)_k\right) \tag{4.3}$$

If we substitute Eq. (4.3) to Eq. (2.16), the IVFS subsethood measure extended from the original BK subproduct is:

$$\pi(\tilde{P} \subseteq \tilde{Q}) = \bigwedge_{i \in I}\frac{1}{J_i K_i}\sum_{j=1}^{J_i}\sum_{k=1}^{K_i}\left(P(x_i)_j \to Q(x_i)_k\right) \tag{4.4}$$

Employing a fuzzy implication operator, Eq. (4.4) will produce a subsethood measurement of $\tilde{P} \subseteq \tilde{Q}$ in the interval $[0,1]$. However, this research noticed that the subsethood measurements for crisp sets are Boolean (yes / no) and for the T1FSs are point-values. Therefore, it is reasonable to deduce that the subsethood measurements for IVFSs should be intervals instead of point-values as proposed by other subsethood measures, such as Nguyen and Kreinovich (2008), Yang and Lin (2009), Zheng, Xiao, Zhang, and Shi (2010), Rickard, Aisbett, and Greb (2009). Studies of Kohout and Bandler (1980b), Kohout and Bandler (1992), Lim and Chan (2012) suggested that Kleene-Dienes and Łukasiewicz implication operators are among the two most suitable candidates for lower and upper bounds BK subsethood measurements. Thus, the subsethood measure using this Complete Derivation Method is given by:

$$\pi(\tilde{P} \subseteq \tilde{Q}) = \Big[\bigwedge_{i \in I} \frac{1}{J_i K_i} \sum_{j=1}^{J_i} \sum_{k=1}^{K_i} \Big(P(x_i)_j \rightarrow_{\mathrm{BK}} Q(x_i)_k \Big) \, , \bigwedge_{i \in I} \frac{1}{J_i K_i} \sum_{j=1}^{J_i} \sum_{k=1}^{K_i} \Big(P(x_i)_j \rightarrow_{\text{Ł}} Q(x_i)_k \Big) \Big] \tag{4.5}$$

This method needs $2 \sum_{i}^{I} J_i K_i$ computations with implication operators to measure subsethood.

4.2.2 Border Evaluation Method

Analysis of intervals (Sengupta & Pal, 2000; Moore & Lodwick, 2003) shows that, with basic arithmetic operations, the bounds of an interval can be computed with the bounds of operands. For the case of implication operators, similar property holds for some implication operators that satisfy the hybrid monotonicity property (Ruan, 1993; Baczyński & Jayaram, 2008), i.e., for $r_1, r_2, s_1, s_2 \in [0, 1]$:

$$r_1 \leq r_2 \qquad \text{implies} \qquad r_1 \rightarrow s_1 \ \geq \ r_2 \rightarrow s_1$$

$$s_1 \leq s_2 \qquad \text{implies} \qquad r_1 \rightarrow s_1 \ \leq \ r_1 \rightarrow s_2$$

In Table 2.1, implication operators that pose this hybrid monotonicity property include $\rightarrow_{S\#}$, $\rightarrow_{S}$, $\rightarrow_{S*}$, $\rightarrow_{G43}$, $\rightarrow_{G43'}$, $\rightarrow_{KD}$, $\rightarrow_{R}$, $\rightarrow_{Ł}$ and $\rightarrow_{Y}$. With this property, one can easily deduce that with lower bound of the first operand and upper bound of the second operand, we can obtain the upper bound of the implication. Instead, upper bound of the first operand and lower bound of the second operand give the lower bound of the implication.

Therefore, with these implication operators, intervals that represent subsethood measure can be obtained. Instead of Eq. (4.5), this border evaluation method gives:

$$\pi(\tilde{P} \subseteq \tilde{Q}) = \left[\bigwedge_{i \in I} \left(\overline{P}(x) \rightarrow \underline{Q}(x) \right) \, , \, \bigwedge_{i \in I} \left(\underline{P}(x) \rightarrow \overline{Q}(x) \right) \right] \tag{4.6}$$

This method needs $2I$ computations using an implication operator to measure subsethood.

4.3 BK Products With Interval Valued Fuzzy Sets

4.3.1 Derivation

With the results in Section 4.2, IVFS-based BK products can be derived. Comparing the 2 sets of subsethood measures that have been derived in this chapter, a subsethood measure using Complete Derivation Method needs $2\sum_i^I J_i K_i$ computations using the implication operators (Eq. (4.5)), whereas the Border Evaluation Method needs only $2I$ (Eq. 4.6). Therefore, the Border Evaluation Method has advantage of lower computational cost. Moreover, if the implication operators that satisfy the hybrid monotonicity property are selected, this measure offers results that are more mathematically reliable because the whole intervals are used in computations, instead of the mean values of discretized intervals. Hence, the following discussion is only focus on the subsethood measure with Border Evaluation Method.

Assume that $\Rightarrow$ represents implication operators that possess hybrid monotonicity property. With subsethood measure proposed in Eq. 4.6, the original BK products that based on IVFSs, can be defined as follow:

Interval-valued fuzzy BK subproduct:

$$\tilde{R} \triangleleft_{\text{BK}} \tilde{S}(a,c) = \Big[\bigwedge_{b \in B} (\overline{R}(a,b) \Rightarrow \underline{S}(b,c)), \bigwedge_{b \in B} (\underline{R}(a,b) \Rightarrow \overline{S}(b,c))\Big] \tag{4.7}$$

Interval-valued fuzzy BK superproduct:

$$\tilde{R} \triangleright_{\text{BK}} \tilde{S}(a,c) = \Big[\bigwedge_{b \in B} (\overline{S}(b,c) \Rightarrow \underline{R}(a,b)), \bigwedge_{b \in B} (\underline{S}(b,c) \Rightarrow \overline{R}(a,b))\Big] \tag{4.8}$$

Interval-valued fuzzy BK square product:

$$\begin{aligned} \tilde{R} \diamond_{\text{BK}} \tilde{S}(a,c) = \Big[&\min\Big(\bigwedge_{b \in B} (\overline{R}(a,b) \Rightarrow \underline{S}(b,c)) \,,\, \bigwedge_{b \in B} (\overline{S}(b,c) \Rightarrow \underline{R}(a,b))\Big) \,, \\ &\min\Big(\bigwedge_{b \in B} (\underline{R}(a,b) \Rightarrow \overline{S}(b,c)) \,,\, \bigwedge_{b \in B} (\underline{S}(b,c) \Rightarrow \overline{R}(a,b))\Big)\Big] \end{aligned} \tag{4.9}$$

Due to the monotonicity property of supremum operator and t-norm, the extension of fuzzy circle product to IVFSs is straightforward:

Fuzzy circle product: $$\tilde{R} \circ \tilde{S}(a,c) = [\bigvee_{b \in B} \tau(\underline{R}(a,b), \underline{S}(b,c)) \,,\, \bigvee_{b \in B} \tau(\overline{R}(a,b), \overline{S}(b,c))] \tag{4.10}$$

Two sets of improved BK products (De Baets & Kerre, 1993a, 1993c) based on the IVFSs can be obtained with adding additional terms. The set B of the improved BK products are as follow:

Interval-valued fuzzy BK subproduct (set B):

$$\tilde{R} \triangleleft_{\mathrm{B}} \tilde{S}(a,c) = \Big[\min\Big(\bigwedge_{b\in B}(\overline{R}(a,b) \Rightarrow \underline{S}(b,c)), \bigvee_{b\in B}\underline{R}(a,b), \bigvee_{b\in B}\underline{S}(b,c)\Big) \\ \min\Big(\bigwedge_{b\in B}(\underline{R}(a,b) \Rightarrow \overline{S}(b,c)), \bigvee_{b\in B}\overline{R}(a,b), \bigvee_{b\in B}\overline{S}(b,c)\Big)\Big] \quad (4.11)$$

Interval-valued fuzzy BK superproduct (set B):

$$\tilde{R} \triangleright_{\mathrm{B}} \tilde{S}(a,c) = \Big[\min\Big(\bigwedge_{b\in B}(\overline{S}(b,c) \Rightarrow \underline{R}(a,b)), \bigvee_{b\in B}\underline{R}(a,b), \bigvee_{b\in B}\underline{S}(b,c)\Big) \\ \min\Big(\bigwedge_{b\in B}(\underline{S}(b,c) \Rightarrow \overline{R}(a,b)), \bigvee_{b\in B}\overline{R}(a,b), \bigvee_{b\in B}\overline{S}(b,c)\Big)\Big] \quad (4.12)$$

Interval-valued fuzzy BK square product (set B):

$$\tilde{R} \diamond_{\mathrm{B}} \tilde{S}(a,c) = \Big[\min\Big(\bigwedge_{b\in B}(\overline{R}(a,b) \Rightarrow \underline{S}(b,c)), \bigwedge_{b\in B}(\overline{S}(b,c) \Rightarrow \underline{R}(a,b)), \bigvee_{b\in B}\underline{R}(a,b), \bigvee_{b\in B}\underline{S}(b,c)\Big), \\ \min\Big(\bigwedge_{b\in B}(\underline{R}(a,b) \Rightarrow \overline{S}(b,c)), \bigwedge_{b\in B}(\underline{S}(b,c) \Rightarrow \overline{R}(a,b)), \bigvee_{b\in B}\overline{R}(a,b), \bigvee_{b\in B}\overline{S}(b,c)\Big)\Big] \quad (4.13)$$

Whereas the set K of improved BK products are:

Interval-valued fuzzy BK subproduct (set K):

$$\tilde{R} \triangleleft_{\mathrm{K}} \tilde{S}(a,c) = \Big[\min\Big(\bigwedge_{b\in B}(\overline{R}(a,b) \Rightarrow \underline{S}(b,c)), \bigvee_{b\in B}\tau(\underline{R}(a,b),\underline{S}(b,c))\Big) \\ \min\Big(\bigwedge_{b\in B}(\underline{R}(a,b) \Rightarrow \overline{S}(b,c)), \bigvee_{b\in B}\tau(\overline{R}(a,b),\overline{S}(b,c))\Big)\Big] \quad (4.14)$$

Interval-valued fuzzy BK superproduct (set K):

$$\tilde{R} \triangleright_{\mathrm{K}} \tilde{S}(a,c) = \Big[\min\Big(\bigwedge_{b\in B}(\overline{S}(b,c) \Rightarrow \underline{R}(a,b)), \bigvee_{b\in B}\tau(\underline{R}(a,b),\underline{S}(b,c))\Big)$$

$$\min\Big(\bigwedge_{b\in B}(\underline{S}(b,c) \Rightarrow \overline{R}(a,b)), \bigvee_{b\in B}\tau(\overline{R}(a,b),\overline{S}(b,c))\Big)\Big] \qquad (4.15)$$

Interval-valued fuzzy BK square product (set K):

$$\tilde{R} \diamond_{\mathrm{K}} \tilde{S}(a,c) = \Big[\min\Big(\bigwedge_{b\in B}(\overline{R}(a,b) \Rightarrow \underline{S}(b,c)), \bigwedge_{b\in B}(\overline{S}(b,c) \Rightarrow \underline{R}(a,b)), \bigvee_{b\in B}\tau(\underline{R}(a,b),\underline{S}(b,c))\Big),$$

$$\min\Big(\bigwedge_{b\in B}(\underline{R}(a,b) \Rightarrow \overline{S}(b,c)), \bigwedge_{b\in B}(\underline{S}(b,c) \Rightarrow \overline{R}(a,b)), \bigvee_{b\in B}\tau(\overline{R}(a,b),\overline{S}(b,c))\Big)\Big] \qquad (4.16)$$

In term of computation complexity, number of computations needed by all the Eq. (4.7) - (4.16) are directly proportional to the number of element in B. In another word, the computational time of these equations increase linearly with the increment of number of element in B. For instance, the number of implications that require by computing Eq. (4.7) is $2|B|$, where $|B|$ is the number of elements in B.

4.3.2 Properties

This research also studies and proves some properties possess by these IVFS-based relational compositions. These properties include: containment property, convertibility property, monotonicity property, property of interaction with union, property of interaction with intersection and property of non-propagation of error.

Containment Property

Set K of BK subproduct and superproduct show the following properties:

$$\tilde{R} \triangleleft_{\mathrm{K}} \tilde{S} \;\subseteq\; \tilde{R} \circ \tilde{S} \qquad (4.17)$$

$$\tilde{R} \triangleright_{\mathrm{K}} \tilde{S} \subseteq \tilde{R} \circ \tilde{S} \tag{4.18}$$

Assume that $\tilde{R}_i$ is a family of I interval-valued fuzzy relations from A to B, $i = \{1,2,\cdots,I\}$, both the original and improved interval-valued fuzzy BK products possess the following properties. Thus, in the following equations, $\triangleleft = \{\triangleleft_{\mathrm{BK}}, \triangleleft_{\mathrm{B}}, \triangleleft_{\mathrm{K}}\}$ and vice-versa for super and square products.

Convertibility Property

$$(\tilde{R} \circ \tilde{S})^T = \tilde{S}^T \circ \tilde{R}^T \tag{4.19}$$

$$(\tilde{R} \triangleleft \tilde{S})^T = \tilde{S}^T \triangleright \tilde{R}^T \tag{4.20}$$

$$(\tilde{R} \triangleright \tilde{S})^T = \tilde{S}^T \triangleleft \tilde{R}^T \tag{4.21}$$

$$(\tilde{R} \diamond \tilde{S})^T = \tilde{S}^T \diamond \tilde{R}^T \tag{4.22}$$

Eq. (4.21) is proved in Appendix A.

Monotonicity Property

$$\tilde{R}_1 \subseteq \tilde{R}_2 \quad \text{implies} \quad \tilde{R}_1 \circ \tilde{S} \subseteq \tilde{R}_2 \circ \tilde{S} \tag{4.23}$$

$$\tilde{R}_1 \subseteq \tilde{R}_2 \quad \text{implies} \quad \tilde{R}_1 \triangleright \tilde{S} \subseteq \tilde{R}_2 \triangleright \tilde{S} \tag{4.24}$$

Interaction With Union

$$\left(\bigcup_{i=1}^{n} \tilde{R}_i\right) \circ \tilde{S} = \bigcup_{i=1}^{n} \left(\tilde{R}_i \circ \tilde{S}\right) \tag{4.25}$$

$$\bigcap_{i=1}^{n}\left(\tilde{R}_i \triangleleft \tilde{S}\right) \subseteq \left(\bigcup_{i=1}^{n}\tilde{R}_i\right) \triangleleft \tilde{S} \subseteq \bigcup_{i=1}^{n}\left(\tilde{R}_i \triangleleft \tilde{S}\right) \quad (4.26)$$

$$\left(\bigcup_{i=1}^{n}\tilde{R}_i\right) \triangleright \tilde{S} \supseteq \bigcup_{i=1}^{n}\left(\tilde{R}_i \triangleright \tilde{S}\right) \quad (4.27)$$

Interaction With Intersection

$$\left(\bigcap_{i=1}^{n}\tilde{R}_i\right) \circ \tilde{S} \subseteq \bigcap_{i=1}^{n}\left(\tilde{R}_i \circ \tilde{S}\right) \quad (4.28)$$

Non-Propagation of Error

Compare to the T1FSs, the IVFSs are bipolar representations of uncertain information (Dubois & Prade, 2008). In another word, an IVFS is a T1FS that associates with certain degree of error. Thus, instead of representing membership functions of $\tilde{R}$ as $[\underline{R}(a,b),\overline{R}(a,b)]$, we can assume that a T1FS R is bounded with an error ε, so that $\tilde{R}(a,b) = [R(a,b)-\varepsilon(a,b), R(a,b)+\varepsilon(a,b)]$, where $R(a,b)-\varepsilon(a,b) = \underline{R}(a,b)$ and $R(a,b)+\varepsilon(a,b) = \overline{R}(a,b)$.

In such case, the error that associates does not propagate with the transitive closure of R. The proof of this property is presented in Appendix B.

4.4 Inference With Weighted BK Product

4.4.1 Weight and BK Products

Most of the time, we can group criteria that we need to consider into a few criteria sets during the reasoning processes. Among these criteria sets, some of them might have higher influence over the others in a decision making process. For instance, let us look at an example of medical diagnosis. A physician may consider the following 4 criteria sets during a medical diagnosis, namely symptoms, patient personal history, family history and environmental issue. However, not all the criteria sets have the same influence in the medical diagnosis. In the diagnosis of diseases such as breast cancer, physicians may take more consideration on symptoms (higher influence) found on patients compare to environmental issues (low influence). Hence, weights are useful in representing the influence of the criteria sets.

However, one should note that the weights should not be confused with the strength of criteria in the criteria sets. Similar case is applied to inference engines based on the BK products. Instead of criteria, features are considered here. Membership degrees of the object-feature relations ($\tilde{R}$) and feature-target relations ($\tilde{S}$) are the "strength of criteria" that determine the results of inferences. While features form the feature sets, influence of each feature sets are represented as weights. That is, weights are applied to each feature set and are modeled with the IVFSs in this research.

As mentioned earlier, the subsethood measure is the foundation of fuzzy BK products. Thus, one might argue that it is inappropriate to implement weights in the BK products-based inference engines because there is no well defined weighted subsethood measure in the literature. In fact, the weights are applied to the feature sets rather than the subsethood measurements. This argument can be explained with a multiple feature sets model as below.

Assume that the features in set B can be grouped into multiple feature sets B_m, $m = \{1, 2, \cdots, M\}$. Each feature set has a number of features. The relation between A and B_m is $\tilde{R}_m$, whereas $\tilde{S}_m$ is the relation between B_m and C. In this case, the images of $a\tilde{R}_m$ and $\tilde{S}_m c$ are $\tilde{P}_m$ and $\tilde{Q}_m$ respectively. Studying the subsethood measure of $\tilde{P}_m \subseteq \tilde{Q}_m$, one can get the BK products $\tilde{R}_m \bigstar \tilde{S}_m(a,c)$, where $\bigstar = \{\triangleleft, \triangleright, \diamond\}$(Figure 4.3).

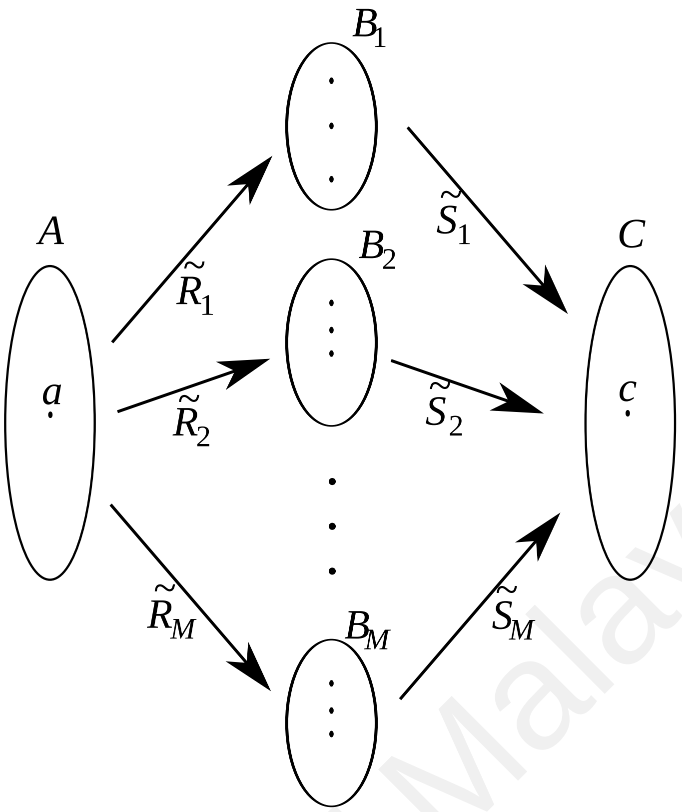


Figure 4.3: Dividing set B into multiple feature sets to form weighted BK products.

Each feature set carries different weights. Assume that the weight of $\tilde{R}_m \bigstar \tilde{S}_m(a,c)$ is $\tilde{W}_m$, the normalized aggregation of all the composition of relations is given as:

$$\tilde{R} \bigstar \tilde{S}(a,c) = \frac{\sum_{m=1}^{M} \tilde{W}_m \left(\tilde{R}_m \bigstar \tilde{S}_m(a,c) \right)}{\sum_{m=1}^{M} \tilde{W}_m} \tag{4.29}$$

Eq. (4.29) gives the weighted measure of the BK products. Here, since all $\tilde{R}_m \bigstar \tilde{S}_m(a,c)$ are intervals that only exist as numerators, whereas $\tilde{W}_m$ are IVFSs, the results of computations based on Eq. (4.29) are always IVFSs. The details of computing Eq. (4.29) are shown in the following subsection.

4.4.2 Computing the Weighted Average

Computing Eq. (4.29) is easy if all the parameters are crisp numbers. However, these parameters are fuzzy, and so the computation become slightly complicated, especially with a term $1/\sum_{m=1}^{M} \tilde{W}_m$. One of the closest problem to us that solved in the literature is Fuzzy Weighted Average (FWA) (Dong & Wong, 1987; Liou & Wang, 1992; D. H. Lee & Park, 1997). FWA computed the problems in the form of:

$$f = \sum_{m=1}^{M} \left(\omega_m \chi_m\right) \bigg/ \sum_{m=1}^{M} \omega_m \tag{4.30}$$

where all χ_m and ω_m are T1FS. Wu and Mendel (D. Wu & Mendel, 2007, 2008a) extended FWA to form Linguistic Weighted Average (LWA), which compute the problem where all χ_m and ω_m are IVFSs. Both FWA and LWA use α-cut decomposition theorem (Klir & Yuan, 1995) in computing the problems. With α-cut decomposition theorem, instead of performing computations directly on the sets (χ_m and ω_m) as whole, a number of $(\delta - 1)$ α-cuts are taken to break the sets into δ intervals. For each interval $\mathscr{I}_\iota$, $1 \leq \iota \leq \delta$, perform computation on the intervals obtained after the α-cut, i.e. χ_m^ι and ω_m^ι to yield an interval $\mathfrak{I}_\iota$. The composition of all the $\mathfrak{I}_\iota$ with corresponding α-cuts form the corresponding set $\mathfrak{I}$.

With the method discussed above, Eq. (4.29) can be computed easily. This research adopted the computation of LWA by D. Wu and Mendel(2007, 2008a) by assuming that the intervals $\tilde{R}_m \bigstar \tilde{S}_m(a,c)$ in Eq. (4.29) are special cases of IVFSs, where these fuzzy sets have rectangle membership functions and the UMF($\tilde{R}_m \bigstar \tilde{S}_m(a,c)$) = LMF($\tilde{R}_m \bigstar \tilde{S}_m(a,c)$). From here onwards, we denote $\tilde{R}_m \bigstar \tilde{S}_m(a,c)$ as Z_m and the lower and upper bounds of Z_m are denoted as $\underline{Z}_m$ and $\overline{Z}_m$ respectively. Thus, follow this notation scheme, $\tilde{R} \bigstar \tilde{S}(a,c)$ is denoted as $\tilde{Z}$.

Since the FOU of $\tilde{Z}$ is determined by UMF($\tilde{Z}$) and LMF($\tilde{Z}$), we can find $\tilde{Z}$ by calcu-

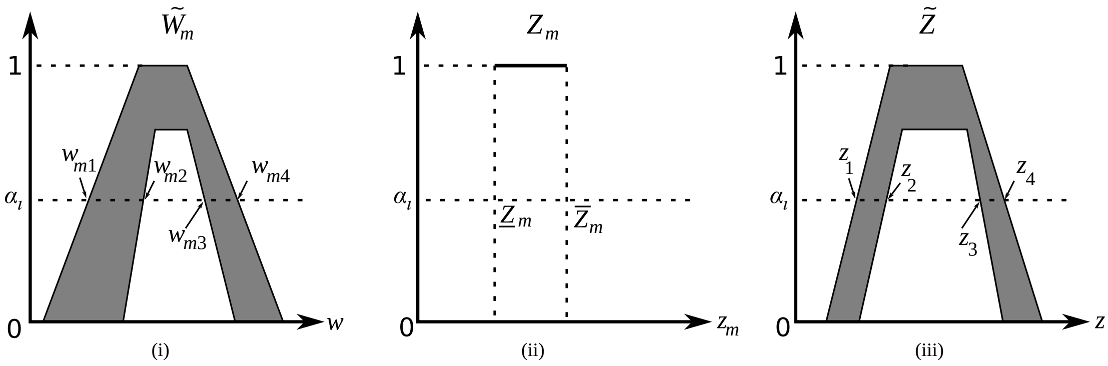


Figure 4.4: Notations used in finding weighted results of IVFS-based BK products.

lating these two boundaries only. In D. Wu and Mendel (2008a), the authors proved that the height of the output sets from LWA are equal to the minimum height of all Z_m and W_m. In the case of this research, since all UMF($\tilde{W}_m$) are normal and Z_m are intervals, the height of an UMF($\tilde{Z}$) is unity. On the other hand, the height of a LMF($\tilde{Z}$) is totally depends on LMF($\tilde{W}_m$). Assume that all the $\tilde{W}_m$ have trapezoidal (or triangular) shape FOU, the shape of $\tilde{Z}$ should be trapezoidal (or triangular) as well (Figure 4.4).

As described earlier in this section, a computation with Eq. (4.29) starts with taking $\delta - 1$ α-cuts to yield δ interval sets. For each interval set, find the interval that represent the FOU of $\tilde{Z}$ corresponding to each α-cut. For this purpose, notations that described in Figure 4.4 are used:

i) $\tilde{W}_m$: for an α-cut α_ι, $w_{\iota m1}$ and $w_{\iota m4}$ should be the leftmost and rightmost values of UMF($\tilde{W}_m$) respectively at α_ι. However, the variable ι is intentionally left out as an subscript of all variables here because it is independent from the calculation of each iteration, and to make the equations look more concise. Therefore, these variables become w_{m1} and w_{m4}. Similarly, w_{m2} and w_{m3} are the leftmost and rightmost values of LMF($\tilde{W}_m$) respectively.

ii) Z_m : $\underline{Z}_m$ is the lower bound of interval Z_m, whereas $\overline{Z}_m$ is the upper bound of this interval.

iii) $\tilde{Z}$: for an α-cut α_ι, z_1 and z_4 are the leftmost and rightmost values of UMF($\tilde{Z}$) respectively. Similarly, z_2 and z_3 are the leftmost and rightmost values of LMF($\tilde{Z}$) respectively.

By referring to the results of LWA (D. Wu & Mendel, 2007, 2008a), for each α-cut, the corresponding boundaries of UMF($\tilde{Z}$) and LMF($\tilde{Z}$) can be obtained by sorting $\overline{Z}_m$ and $\underline{Z}_m$ in ascending order first, then substituting the corresponding values into the following equations:

$$z_1 = \frac{\sum_{m=1}^{\beta_1} w_{m4}\underline{Z}_m + \sum_{m=\beta_1+1}^{M} w_{m1}\underline{Z}_m}{\sum_{m=1}^{\beta_1} w_{m4} + \sum_{m=\beta_1+1}^{M} w_{m1}} \tag{4.31}$$

$$z_2 = \frac{\sum_{m=1}^{\beta_2} w_{m3}\underline{Z}_m + \sum_{m=\beta_2+1}^{M} w_{m2}\underline{Z}_m}{\sum_{m=1}^{\beta_2} w_{m3} + \sum_{m=\beta_2+1}^{M} w_{m2}} \tag{4.32}$$

$$z_3 = \frac{\sum_{m=1}^{\beta_3} w_{m2}\overline{Z}_m + \sum_{m=\beta_3+1}^{M} w_{m3}\overline{Z}_m}{\sum_{m=1}^{\beta_3} w_{m2} + \sum_{m=\beta_3+1}^{M} w_{m3}} \tag{4.33}$$

$$z_4 = \frac{\sum_{m=1}^{\beta_4} w_{m1}\overline{Z}_m + \sum_{m=\beta_4+1}^{M} w_{m4}\overline{Z}_m}{\sum_{m=1}^{\beta_4} w_{m1} + \sum_{m=\beta_4+1}^{M} w_{m4}} \tag{4.34}$$

In these equations, β_1, β_2, β_3 and β_4 are the switching points in the range $[1, M]$ calculated with Karnik-Mendel algorithm (F. Liu & Mendel, 2008a; Mendel, 2009) such that:

$$\underline{Z}_{\beta_1} \leq z_1 \leq \underline{Z}_{\beta_1+1} \tag{4.35}$$

$$\underline{Z}_{\beta_2} \leq z_2 \leq \underline{Z}_{\beta_2+1} \tag{4.36}$$

$$\overline{Z}_{\beta_3} \leq z_3 \leq \overline{Z}_{\beta_3+1} \tag{4.37}$$

$$\overline{Z}_{\beta_4} \leq z_4 \leq \overline{Z}_{\beta_4+1} \tag{4.38}$$

4.4.3 Results Interpretation

The results obtained from Eq. (4.31)-(4.34) form a set of IVFSs. The meaning carried by this set of IVFSs is application dependent. In some applications, ranking algorithms suggested in Mitchell (2006) and D. Wu and Mendel (2009) are useful if comparisons between these IVFSs are needed to find their order, where control engineering applications may fall to this category. In some other cases, the results can be compared with a set of predefined IVFSs using the similarity measure (Nguyen & Kreinovich, 2008; D. Wu & Mendel, 2009).

4.5 Chapter Conclusion

In this chapter, two subsethood measures for IVFSs, namely the Complete Derivation Method and the Border Evaluation Method are proposed. The Complete Derivation Method, which requires much more computations, evaluates the whole interval membership degrees as discrete points. Taking the mean of implications of these discrete points, the subsethood of two sets are measured. The disadvantages of this method include the distortion during the discretization process, as well as higher computational cost. On the other hand, the Border Evaluation Method only considers the borders of interval membership degrees and evaluate the intervals as a whole to obtain the implication results. The second method has its advantage in computational efficiency but only applicable for implication operators that possess the hybrid monotonicity property.

Due to most of the common implication operators hold the hybrid monotonicity property, the subsethood measure with the Border Evaluation Method is used in developing the IVFSs-based BK products. 3 sets of the BK products were developed, namely original, improved set B and improved set K. Computation with these BK products return intervals.

Some of the important properties of IVFSs-based BK products are also examined in this chapter. Among all, one of the interesting property is related to the propagation of error associated with membership functions. If the interval membership functions of interval-valued fuzzy relations are assumed to be error carried by type-1 fuzzy relations, these errors do not propagate with the transitive closures of the relations.

Lastly, a weight parameter is added to the BK products, and the rational of it is discussed. While the weight parameter, $\tilde{W}$ is modeled with IVFSs, the computation of this fuzzy weighted equations is also discussed. With LWA algorithm, IVFSs are expected to be the output of the weighted BK products.

CHAPTER 5

LEARNING MECHANISM FOR BK PRODUCTS BASED INFERENCE ENGINES

5.1 Introduction

As like many other fuzzy logic systems, the BK products developed in Chapter 4 work only if one can determine the required membership degrees from the membership functions. In the case of the BK products, they are $\tilde{R}(a,b)$ (or $R(a,b)$ for the case of T1FS) - the membership degrees of relation between objects A and features B, and $\tilde{S}(b,c)$ (or $S(b,c)$ for the case of T1FS) - the membership degrees of relation between features B and targets C. In term of fuzzy logic systems (Figure 1.1), $\tilde{R}(a,b)$ is the result of the fuzzification module that gives the relation between set A and set B, whereas $S(b,c)$ forms the knowledge base that specify the relations between set B and set C.

In the past, many systems that based on BK products require expert knowledge to define the membership degrees, especially $S(b,c)$. The research on forming membership functions from numerical data set for BK products are hardly found in the literature. Therefore, the main contribution of this chapter is the development of a learning mechanism so that the BK products based inference systems can be formed if training examples are provided. This learning mechanism should be able to form membership functions for fuzzification purpose, as well as define membership degrees for the knowledge base.

5.2 Review Of Previous Works

Develop a learning fuzzy system requires the construction of membership functions for the involved parameters. This is a topic that has been studied extensively in the literature of the fuzzy rule based systems (Takagi & Sugeno, 1985; Nomura, Hayashi, & Wakami, 1992; T.-P. Hong & Lee, 1996; T. Hong & Chen, 1999). Some of these methods focus on constructing membership functions from numerical data. Among all, one of the widely used method is proposed by L. Wang and Mendel (1992). This method divides a data range into multiple areas, then a set of predefined membership functions are assigned to the data range. Thus, the membership degrees of an element can be obtained by mapping the data into the predefined membership functions. Instead of assigning a set of predefined membership functions into the range of data, T. P. Wu and Chen (1999) suggested that fuzzy compatible relations are required. With the transitive closures, fuzzy compatible relations are converted to fuzzy equivalence relations. Lastly, dividing the fuzzy equivalence relations with α-cuts brings a triangular membership function to each partition of the data range. There are also some methods which are based on probability density functions, such as the method proposed by Civanlar and Trussell (1986). However, all these methods focus only on type-1 fuzzy rule based systems. Hence, effort is needed so that these methods can be ported to work with inference engines that based on IVFS-based BK products.

Though there are few researches on constructing membership functions for the IVFSs or even T2FSs, these methods either require construction and analysis of questionaires (Mendel, 2007a; F. Liu & Mendel, 2008b), or focus only on image processing domain (Bustince et al., 2009; Choi & Rhee, 2009).

To make the numerical data interpretable by the BK products based inferences engines, a new method is needed. In the following, a method is proposed to construct the membership functions for the BK products. This method, which inspired by L. Wang and Mendel (1992) and Civanlar and Trussell (1986), not only serves for mapping the numerical data to membership degrees, but also helps in the training of inference engines to form knowledge bases.

5.3 Membership Functions For BK Subproduct Based Inference Structures

A typical fuzzy rule based system (C. C. Lee, 1990a; Lam & Seneviratne, 2008; Fazel Zarandi et al., 2009) has rules in the form:

$$\text{if } \mathsf{X}_1 \text{ is } U_1 \text{ and } \mathsf{X}_2 \text{ is } U_2 \text{ then } \mathsf{Y} \text{ is } V$$

Here, X_1 and X_2 are antecedents limited by linguistic terms U_1 and U_2 respectively, whereas Y is the consequent that related to another linguistic terms V. U_1 and U_2 can be in different universe, as well as the V.

In contrast, for the BK products based inference engines, there are no rules exist and inferences are only based on the relations of $\tilde{R}(a,b)$ and $\tilde{S}(b,c)$, which both come from the same domain for a feature b. Therefore, finding the values of $\tilde{R}(a,b)$ and $\tilde{S}(b,c)$ are interrelated and are in accordance with the domain of b. Moreover, the construction of membership is independence for each feature b.

Since $\tilde{R}(a,b)$ are mappings from a to b and $\tilde{S}(b,c)$ are reverse mappings from c to b, the domain of b should be studied. Assume that the number of features in set B is $J \in \mathbb{N}$ and $j = \{1, \cdots, J\}$. Each $a_i \in A$, where $i \in \{1, \cdots, I\}$, $I \in \mathbb{N}$, can be mapped to a feature b_j with a value L_{ji}. With all the L_{ji}, the domain of the feature b_j can be defined as an interval $L_j^{\mathtt{I}}$:

$$L_j^{\mathtt{I}} = \{L_{j1}, \cdots, L_{jI}\} = [\underline{L}_j, \overline{L}_j] \tag{5.1}$$

Divide this domain into multiple sections that each represent a linguistic term/concept. Overlapping between sections are allowed to reflect the nature of overlapping between linguistic terms. The number of section is application and feature dependent and subject to the number of linguistic terms one wants to define for this feature. At least two sections (e.g. “Low” and “High”) need to be defined but there is no rules on the maximum number of sections. However, the increment of sections will increase the computational cost of the inference systems.

For each section, form a interval-valued membership function. There is no rules on the shape of the membership functions, but for the sake of simplicity, the membership functions with trapezoidal UMF and triangular LMF are used in the following. Assume that all the UMFs are normal and the heights of LMFs are v.

Let $H_j \in \mathbb{N}$ be the number of membership functions defined in the domain of feature b_j, and $h_j = \{1, \cdots, H_j\}$. A membership function defined in this domain can be named as $\tilde{F}_{jh_j}$. Figure 5.1 shows an example where domain of b_j is divided into H_j membership functions that represent linguistic terms "very low", "low", "high", "medium low", "very high" and etc.

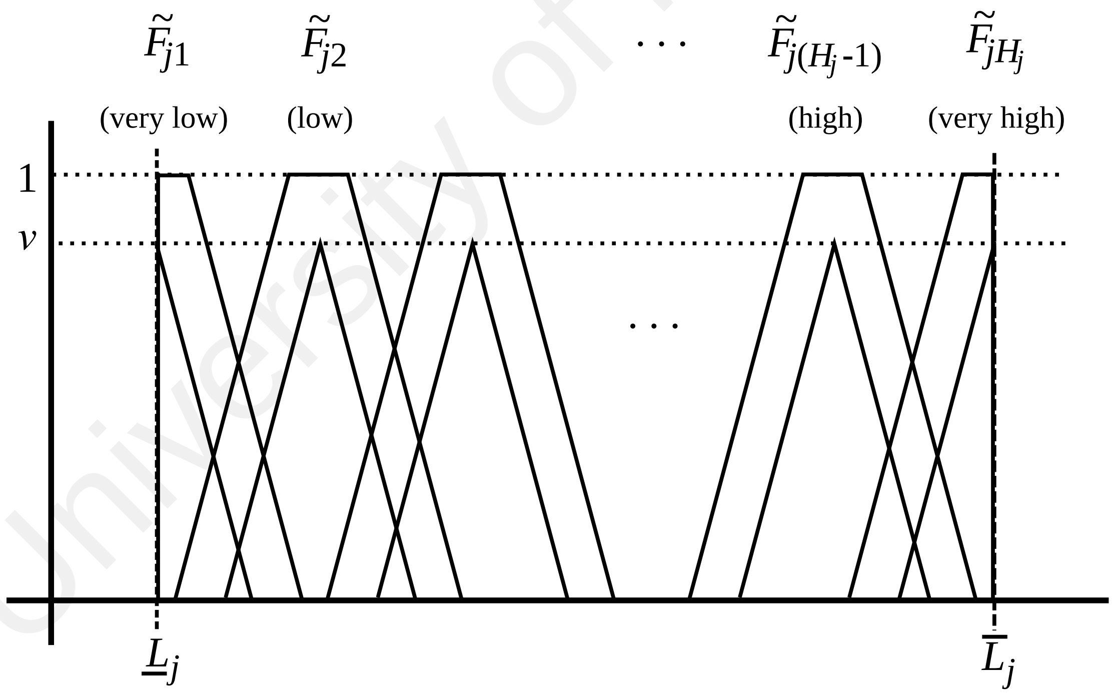


Figure 5.1: Define a number of H_j membership functions in the interval L_j^{I}.

For a membership function $\tilde{F}_{jh_j}$ that defined for a section, if the shape of this normal membership function is trapezoidal UMF and triangular LMF, one can define the parameters of this membership function as shown in Figure 5.2. Parameters 1-4 define the UMF, parameters 5-7 and ν define the LMF.

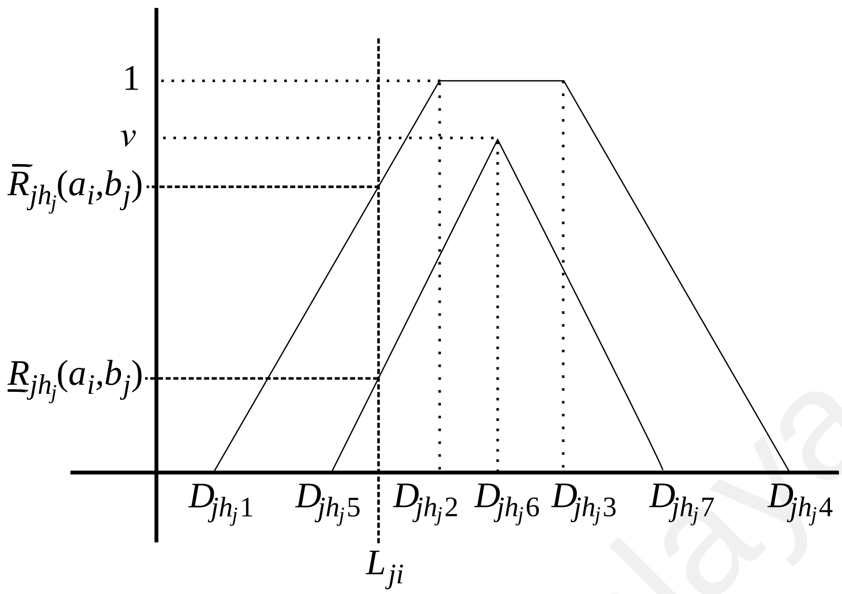


Figure 5.2: The 8 points definition of a membership function of IVFS and the mapping of membership degrees.

Assume that there are $K \in \mathbb{N}$ targets in C and $k = \{1, \cdots, K\}$. The composition of relations $\tilde{R} \triangleleft \tilde{S}(a_i, c_k)$ is meaningful only if a_i implies c_k. Therefore, to find $\tilde{S}(b, c_k)$, A is partitioned into K subsets according to c_k:

$$A = \{A_1, \cdots, A_k, \cdots, A_K\}, \tag{5.2}$$

$$a_i \in A_k \iff a_i \to c_k. \tag{5.3}$$

If a_i maps to b_j with a value L_{ji}, as proposed by Zadeh (1978) and Civanlar and Trussell (1986), the probability density function of L_{ji} in $[\underline{L}_j, \overline{L}_j]$ can be studied to find the membership degrees of $\tilde{S}(b, c_k)$. Therefore, let $|A_k|$ be the number of elements in A_k, and $|A^{(5,7)}_{kjh_j}|$ be the number of elements of a_i map to b_j with a value L_{ij} such that $D_{hj_h5} \leq L_{ij} \leq D_{hj_h7}$. With the information on the distribution of a_i in the lower membership functions range, the probability density functions can be plotted. Consequently, $\underline{S}_{jh_j}(b_j, c_k)$ can be find:

$$\underline{S}_{jh_j}(b_j, c_k) = \frac{|A^{(5,7)}_{kjh_j}|}{|A_k|} \tag{5.4}$$

Similarly, if $|A^{(1,4)}_{kjh_j}|$ is the number of elements of a_i map into the range $[D_{hj_h1}, D_{hj_h4}]$, or of UMF of b_j. The mapping values are denoted as L_{ij}. With the above information, the upper bound of $\tilde{S}_{jh_j}(b_j, c_k)$ can be find:

$$\overline{S}_{jh_j}(b_j, c_k) = \frac{|A^{(1,4)}_{kjh_j}|}{|A_k|} \tag{5.5}$$

The Eq. (5.4) and Eq. (5.5) find the relations between elements in set B and set C, which are required during the training process of a classifier. However, to prepare the inference engines for prediction, the test data should be fuzzified in the fuzzification module (Figure 1.1) to form $\tilde{R}(a, b)$.

With the membership functions defined, the finding of $\tilde{R}(a, b)$ are straight forward. Firstly, $\tilde{F}_{jh_j}$, the set of membership functions developed to find $\tilde{S}(b, c)$ must be adopted so that both $\tilde{S}(b, c)$ and $\tilde{R}(a, b)$ refer to the same set of membership functions. Subsequently, mapping of values described below gives the membership degrees of $\tilde{R}(a, b)$.

With relation $\tilde{R}(a_i, b_j)$, an a_i may maps to b_j with a value L_{ji} in the interval $L^{\mathtt{I}}_j$. If this L_{ji} falls into the section where $\tilde{F}_{jh_j}$ defines (i.e. in $[D_{jh_j1}, D_{jh_j4}]$), we can retrieve a membership degree for this membership function, $\tilde{R}_{jh_j}(a_i, b_j)$, otherwise $\tilde{R}_{jh_j}(a_i, b_j) = 0$ for this membership function (Figure 5.2). The upper and lower bounds of this interval membership degree, $[\underline{R}_{jh_j}(a_i,b_j),\overline{R}_{jh_j}(a_i,b_j)]$ are given by:

$$\underline{R}_{jh_j}(a_i,b_j) = \begin{cases} \dfrac{(\nu)(L_{ji}-D_{jh_j5})}{D_{jh_j6}-D_{jh_j5}} & \text{if } D_{jh_j5} < L_{ji} \leq D_{jh_j6}, \\ \dfrac{(\nu)(D_{jh_j7}-L_{ji})}{D_{jh_j7}-D_{jh_j6}} & \text{if } D_{jh_j6} < L_{ji} < D_{jh_j7}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.6}$$

$$\overline{R}_{jh_j}(a_i,b_j) = \begin{cases} \dfrac{L_{ji}-D_{jh_j1}}{D_{jh_j2}-D_{jh_j1}} & \text{if } D_{jh_j1} < L_{ji} < D_{jh_j2}, \\ 1 & \text{if } D_{jh_j2} \leq L_{ji} \leq D_{jh_j3}, \\ \dfrac{D_{jh_j4}-L_{ji}}{D_{jh_j4}-D_{jh_j3}} & \text{if } D_{jh_j3} < L_{ji} < D_{jh_j4}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.7}$$

It is possible that in the testing data set, there are some cases that a_i maps to L'_{ji} where $L'_{ji} < \underline{L}_j$. In such cases, it is wiser to reconsider both the membership degrees of $\underline{R}_{jh_j}(a_i,b_j)$ and $\overline{R}_{jh_j}(a_i,b_j)$ if $\tilde{F}_{jh_j}$ is a left-shoulder membership function. For the cases where $D_{jh_j1} = D_{jh_j2} = D_{jh_j5} = D_{jh_j6}$, one should set $\underline{R}_{jh_j}(a_i,b_j)$ and $\overline{R}_{jh_j}(a_i,b_j)$ to the heights of the corresponding membership functions, i.e. ν and 1 respectively. It is similar for the case when $L'_{ji} > \overline{L}_j$. For right-shoulder membership functions $\tilde{F}_{jh_j}$ such that $D_{jh_j3} = D_{jh_j4} = D_{jh_j6} = D_{jh_j7}$, one should also set the membership degrees of both $\underline{R}_{jh_j}(a_i,b_j)$ and $\overline{R}_{jh_j}(a_i,b_j)$ to ν and 1 respectively, if they are the heights of the corresponding membership functions.

As a summary of the section, this learning method forms H_j membership functions for a feature b_j, thus, for a c_k, it finds $\sum_{j=1}^{J} H_j$ membership degrees for both $\underline{S}(b,c_k)$ and $\overline{S}(b,c_k)$. The total number of membership degrees of both $\underline{S}(b,c)$ and $\overline{S}(b,c)$ is $K(\sum_{j=1}^{J} H_j)$. On the other hand, mapping of an object a_i also finds $\sum_{j=1}^{J} H_j$ membership degree for both $\underline{R}(a_i,b)$ and $\overline{R}(a_i,b)$. Therefore, with a data set with I objects, the total number of membership degrees of both $\underline{R}(a,b)$ and $\overline{R}(a,b)$ is $I\sum_{j=1}^{J} H_j$.

5.4 Algorithm

Based on the method described in Section 5.3, the algorithm to find the membership degrees of $\tilde{S}(b,c)$ is stated as follow:

Finding $\tilde{S}(b,c)$

For a feature b_j:

Step 1: With all the objects a_i, list all L_{ji}, the values that a_i map to the b_j;

Step 2: Find the interval that represents the domain of L_{ji}, $L_j^{\mathrm{I}} = [\underline{L}_j, \overline{L}_j]$;

Step 3: Divide the interval into H_j sections. Each section represents a linguistic term such as "high","low" and etc. The number of section/linguistic term is subject to the feature and application;

Step 4: For each section, construct a membership function. Name the constructed membership function in section h_j as $\tilde{F}_{jh_j}$;

Step 5: Divide the objects a_i into K subsets, each subset A_k consists only all the a_{ij} that bring the inference result c_k;

Step 6: For $k = 1$, count the number of elements in A_k, i.e. $|A_k|$;

Step 7: For $h = 1$, find $|A_{kjh_j}^{(1,4)}|$, the number of elements in A_k that map their values to the interval $[D_{hj_h1}, D_{hj_h4}]$;

Step 8: For $h = 1$, find $|A_{kjh_j}^{(5,7)}|$, the number of elements in A_k that map their values to the interval $[D_{hj_h5}, D_{hj_h7}]$;

Step 9: Compute the interval membership degrees $\underline{S}_{jh_j}(b_j, c_k)$ and $\overline{S}_{jh_j}(b_j, c_k)$ for $j = 1$ and $k = 1$ using Eq. (5.4) and Eq. (5.5);

Step 10: Repeat steps 7 - 9 for $h = \{2, \cdots, H_j\}$;

Step 11: Repeat steps 6 - 10 for $k = \{2, \cdots, K\}$.

The algorithm should be repeated J times for $j = \{1, \cdots, J\}$ to find all the membership degrees of relations between B and C.

The following algorithm finds $\tilde{R}(a,b)$. This algorithm can work independently from the finding $\tilde{S}(b,c)$ algorithm, but if both algorithms run on the same system, the same set of membership functions (developed in step 1 - 4 of finding $\tilde{S}(b,c)$ algorithm) must be use.

Finding $\tilde{R}(a,b)$

For a feature b_j:

Step 1: Adopt the membership functions developed in steps 1 - 4 of finding $\tilde{S}(b,c)$ algorithm;

Step 2: For $i = 1$, identify L_{ji}, the image of an a_i in b_j. Map this L_{ji} to all the membership functions $\tilde{F}_{jh_j}$ to find the $\underline{R}_{jh_j}(a_i, b_j)$ and $\overline{R}_{jh_j}(a_i, b_j)$ using Eq. (5.6) and Eq. (5.7) respectively;

However, If $\tilde{F}_{jh_j}$ is a left-shoulder membership function such that $D_{jh_j1} = D_{jh_j2} = D_{jh_j5} = D_{jh_j6}$, set $\underline{R}_{jh_j}(a_i, b_j) = \nu$ and $\overline{R}_{jh_j}(a_i, b_j) = 1$. If $\tilde{F}_{jh_j}$ is a right-shoulder membership function such that $D_{jh_j3} = D_{jh_j4} = D_{jh_j6} = D_{jh_j7}$, set $\underline{R}_{jh_j}(a_i, b_j) = \nu$ and $\overline{R}_{jh_j}(a_i, b_j) = 1$;

Step 3: Repeat step 2 for $i = \{2, \cdots, I\}$.

The algorithm should be repeated J times for $j = \{1, \cdots, J\}$ to find all the membership degrees of relations between A and B.

5.5 Example

Assume that there are $I = 12$ books arranged into $K = 2$ shelves, S1 and S2, depends on their features. These features are weight (b_1), thickness (b_2) and height (b_3). Table 5.1 shows the details of these books, namely a_1 to a_{12}.

Table 5.1: Data for example

book	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
b_1 (g)	150	450	750	250	150	250	550	600	400	350	550	450
b_2 (cm)	0.5	2.5	4.5	2.0	1.5	1.0	3.5	4.0	3.0	2.5	3.5	3.0
b_3 (cm)	20.5	22.0	27.0	27.0	17.0	23.0	23.0	18.0	21.0	22.5	22.0	23.0
Shelf	S1	S2	S2	S2	S1	S1	S2	S2	S1	S1	S2	S2

Finding $\tilde{S}(b,c)$

Firstly, the following steps compute $\tilde{S}(b_1,c)$:

Step 1: List L_{1i}:

L_{1i} = { 150 , 450 , 750 , 250 , 150 , 250 , 550 , 600 , 400 , 350 , 550 , 450 } .

Step 2: Find the domain of the feature b_1, i.e. L_1^I:

L_1^I = [150 , 750].

Step 3: Divide the domain L_1^I into sections:

Assume that dividing L_1^I into 3 sections is reasonable, which represent "Light"([150,350]), "Medium"([250,650]) and "Heavy"([550,750]).

Step 4: Construct a membership function $\tilde{F}_{1h}$ in each section.

Assume that 1) all the UMFs are trapezoidal and LMFs are triangular, as shown in Figure 5.2; 2) the heights of all the LMFs are 0.7; 3) the supports of the LMFs of $\tilde{F}_{1h}$ are 2cm lesser then their UMF counterparts on each side; 4) $D_{113} - D_{112} = 50$, $D_{123} - D_{122} = 100$ and $D_{133} - D_{132} = 50$. Therefore, for feature b_1:

$$UMF(\tilde{F}_{11}) = \begin{cases} 1 & \text{if } 150 \leq L_{1i} \leq 200, \\ \dfrac{350 - L_{1i}}{350 - 200} & \text{if } 200 < L_{1i} < 350, \\ 0 & \text{otherwise.} \end{cases} \tag{5.8}$$

$$LMF(\tilde{F}_{11}) = \begin{cases} \dfrac{0.7(300 - L_{1i})}{300 - 150} & \text{if } 150 \leq L_{1i} < 300, \\ 0 & \text{otherwise.} \end{cases} \tag{5.9}$$

$$UMF(\tilde{F}_{12}) = \begin{cases} \dfrac{L_{1i} - 250}{400 - 250} & \text{if } 250 < L_{1i} < 400, \\ 1 & \text{if } 400 \leq L_{1i} \leq 500, \\ \dfrac{650 - L_{1i}}{650 - 500} & \text{if } 500 < L_{1i} < 650, \\ 0 & \text{otherwise.} \end{cases} \tag{5.10}$$

$$LMF(\tilde{F}_{12}) = \begin{cases} \dfrac{0.7(L_{1i} - 300)}{450 - 300} & \text{if } 300 < L_{1i} \leq 450, \\ \dfrac{0.7(600 - L_{1i})}{600 - 450} & \text{if } 450 < L_{1i} < 600, \\ 0 & \text{otherwise.} \end{cases} \tag{5.11}$$

$$UMF(\tilde{F}_{13}) = \begin{cases} \dfrac{L_{1i} - 550}{700 - 550} & \text{if } 550 < L_{1i} < 700, \\ 1 & \text{if } 700 \leq L_{1i} \leq 750, \\ 0 & \text{otherwise.} \end{cases} \tag{5.12}$$

$$LMF(\tilde{F}_{13}) = \begin{cases} \dfrac{0.7(L_{1i} - 600)}{750 - 600} & \text{if } 600 < L_{1i} \leq 750, \\ 0 & \text{otherwise.} \end{cases} \tag{5.13}$$

Step 5: Divide a_i into $K = 2$ subsets:

For shelf S1, $A_1 = \{a_1, a_5, a_6, a_9, a_{10}\}$;

For shelf S2, $A_2 = \{a_2, a_3, a_4, a_7, a_8, a_{11}, a_{12}\}$.

Step 6: Count the number of elements in A_1:

From the results of the previous step, $|A_1| = 5$.

Step 7: Find the number of elements in the interval $[D_{111}, D_{114}]$:

$[D_{111}, D_{114}] = [150, 350]$. The books from A_1 with weight in this range are a_1, a_5, a_6 and a_{10}. Therefore, $|A_{111}^{(1,4)}| = 4$.

Step 8: Find the number of elements in the interval $[D_{115}, D_{117}]$:

$[D_{115}, D_{117}] = [150, 300]$. The books from A_1 with weight in this range are a_1, a_5 and a_6. Therefore, $|A_{111}^{(5,7)}| = 3$.

Step 9: Compute the membership degrees $\overline{S}_{11}(b_1, c_1)$ and $\underline{S}_{11}(b_1, c_1)$:

With Eq. 5.5, $\overline{S}_{11}(b_1, c_1) = \frac{4}{5} = 0.8$;

With Eq. 5.4, $\underline{S}_{11}(b_1, c_1) = \frac{3}{5} = 0.6$.

Step 10: Repeat steps 7-9 for $h = \{2, 3\}$:

$[D_{211}, D_{214}] = [250, 650]$. $|A_{112}^{(1,4)}| = 3$

$[D_{215}, D_{217}] = [300, 600]$. $|A_{112}^{(5,7)}| = 2$

Therefore, $\overline{S}_{12}(b_1, c_1) = \frac{3}{5} = 0.6$ and $\underline{S}_{12}(b_1, c_1) = \frac{2}{5} = 0.4$.

$[D_{211}, D_{314}] = [550, 750]$. $|A_{113}^{(1,4)}| = 0$

$[D_{215}, D_{317}] = [600, 750]$. $|A_{113}^{(5,7)}| = 0$

Therefore, $\overline{S}_{13}(b_1, c_1) = \frac{0}{5} = 0$ and $\underline{S}_{13}(b_1, c_1) = \frac{0}{5} = 0$.

Step 11: Repeat steps 6-10 for $h = \{2, 3\}$:

From step 5, $|A_2| = 7$;

$|A_{211}^{(1,4)}| = 1$ and $|A_{211}^{(5,7)}| = 1$, therefore $\overline{S}_{11}(b_1, c_2) = 0.14$ and $\underline{S}_{11}(b_1, c_2) = 0.14$;

$|A_{212}^{(1,4)}| = 6$ and $|A_{212}^{(5,7)}| = 5$, therefore $\overline{S}_{12}(b_1, c_2) = 0.86$ and $\underline{S}_{12}(b_1, c_2) = 0.71$;

$|A_{213}^{(1,4)}| = 4$ and $|A_{213}^{(5,7)}| = 2$, therefore $\overline{S}_{13}(b_1, c_2) = 0.57$ and $\underline{S}_{13}(b_1, c_2) = 0.29$.

Repeat the same procedure for features b_2 and b_3. Assume that only two membership functions are assigned to both b_2 and b_3, which are defined as:

$$UMF(\tilde{F}_{21}) = \begin{cases} 1 & \text{if } 0.5 \leq L_{2i} \leq 1.0, \\ \dfrac{3.0 - L_{2i}}{3.0 - 1.0} & \text{if } 1.0 < L_{2i} < 3.0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.14}$$

$$LMF(\tilde{F}_{21}) = \begin{cases} \dfrac{0.7(2.5 - L_{2i})}{2.5 - 0.5} & \text{if } 0.5 \leq L_{2i} < 2.5, \\ 0 & \text{otherwise.} \end{cases} \tag{5.15}$$

$$UMF(\tilde{F}_{22}) = \begin{cases} \dfrac{L_{2i} - 2.0}{4.0 - 2.0} & \text{if } 2.0 < L_{2i} < 4.0, \\ 1 & \text{if } 4.0 \leq L_{2i} \leq 4.5, \\ 0 & \text{otherwise.} \end{cases} \tag{5.16}$$

$$LMF(\tilde{F}_{22}) = \begin{cases} \dfrac{0.7(L_{2i} - 2.5)}{4.5 - 2.5} & \text{if } 2.5 < L_{2i} \leq 4.5, \\ 0 & \text{otherwise.} \end{cases} \tag{5.17}$$

$$UMF(\tilde{F}_{31}) = \begin{cases} 1 & \text{if } 17 \leq L_{3i} \leq 19, \\ \dfrac{23 - L_{3i}}{23 - 19} & \text{if } 19 < L_{3i} < 23, \\ 0 & \text{otherwise.} \end{cases} \tag{5.18}$$

$$LMF(\tilde{F}_{31}) = \begin{cases} \dfrac{0.7(21 - L_{3i})}{21 - 17} & \text{if } 17 \leq L_{3i} < 21, \\ 0 & \text{otherwise.} \end{cases} \tag{5.19}$$

$$UMF(\tilde{F}_{32}) = \begin{cases} \dfrac{L_{3i} - 21}{25 - 21} & \text{if } 21 < L_{3i} < 25, \\ 1 & \text{if } 25 \leq L_{3i} \leq 27, \\ 0 & \text{otherwise.} \end{cases} \tag{5.20}$$

$$LMF(\tilde{F}_{32}) = \begin{cases} \dfrac{0.7(L_{3i} - 23)}{27 - 23} & \text{if } 23 < L_{3i} \leq 27, \\ 0 & \text{otherwise.} \end{cases} \tag{5.21}$$

With these membership functions, one should get the following results:

$\overline{S}_{21}(b_2, c_1) = 1.0$ and $\underline{S}_{21}(b_2, c_1) = 0.8$;

$\overline{S}_{22}(b_2, c_1) = 0.4$ and $\underline{S}_{22}(b_2, c_1) = 0.4$;

$\overline{S}_{21}(b_2, c_2) = 0.29$ and $\underline{S}_{21}(b_2, c_2) = 0.14$;

$\overline{S}_{22}(b_2, c_2) = 1.0$ and $\underline{S}_{22}(b_2, c_2) = 0.86$;

$\overline{S}_{31}(b_3, c_1) = 1.0$ and $\underline{S}_{31}(b_3, c_1) = 0.6$;

$\overline{S}_{32}(b_3, c_1) = 0.6$ and $\underline{S}_{32}(b_3, c_1) = 0.2$;

$\overline{S}_{31}(b_3, c_2) = 0.71$ and $\underline{S}_{31}(b_3, c_2) = 0.14$;

$\overline{S}_{32}(b_3, c_2) = 0.86$ and $\underline{S}_{32}(b_3, c_2) = 0.57$.

Finding $\tilde{R}(a,b)$

Assume that a_{20} and a_{21} are two books with the features shown in Table 5.2. The following steps fuzzify these inputs to find $\tilde{R}(a,b)$.

Table 5.2: Example data for testing

book	a_{20}	a_{21}
b_1 (g)	500	280
b_2 (cm)	3.5	2.3
b_3 (cm)	28.0	22.0

Step 1: Membership functions adoption:

To find the membership degrees of $\tilde{R}_{1h_1}(a_{20}, b_1)$, Equations (5.8) - (5.13) are employed.

Step 2: Mapping to functions to find membership degrees:

With mapping $L_{1,20} = 500$ to Equations (5.8) - (5.13), which are special cases of Equations (5.6) and Eq. (5.7), one should able to get:

$\underline{R}_{11}(a_{20}, b_1) = 0$ and $\overline{R}_{11}(a_{20}, b_1) = 0$

$\underline{R}_{12}(a_{20}, b_1) = 0.47$ and $\overline{R}_{12}(a_{20}, b_1) = 1.0$

$\underline{R}_{13}(a_{20}, b_1) = 0$ and $\overline{R}_{13}(a_{20}, b_1) = 0$

Step 3: Repeat the process for the other objects:

One should find the following results for a_{21} with $L_{1,21} = 280$:

$\underline{R}_{11}(a_{21}, b_1) = 0.09$ and $\overline{R}_{11}(a_{21}, b_1) = 0.47$

$\underline{R}_{12}(a_{21}, b_1) = 0$ and $\overline{R}_{12}(a_{21}, b_1) = 0.2$

$\underline{R}_{13}(a_{21}, b_1) = 0$ and $\overline{R}_{13}(a_{21}, b_1) = 0$

For b_2 and b_3, repeat the process with Equations (5.14) - (5.21), the results are:

$\underline{R}_{21}(a_{20}, b_2) = 0$ and $\underline{R}_{21}(a_{20}, b_2) = 0$

$\underline{R}_{21}(a_{21}, b_2) = 0.07$ and $\underline{R}_{21}(a_{21}, b_2) = 0.35$

$\underline{R}_{22}(a_{20}, b_2) = 0.35$ and $\underline{R}_{22}(a_{20}, b_2) = 0.75$

$\underline{R}_{22}(a_{21}, b_2) = 0$ and $\underline{R}_{22}(a_{21}, b_2) = 0.15$

$\underline{R}_{31}(a_{20}, b_3) = 0$ and $\underline{R}_{31}(a_{20}, b_3) = 0$

$\underline{R}_{31}(a_{21}, b_3) = 0$ and $\underline{R}_{31}(a_{21}, b_3) = 0.25$

$\underline{R}_{32}(a_{20}, b_3) = 1$ and $\underline{R}_{32}(a_{20}, b_3) = 1$

$\underline{R}_{32}(a_{21}, b_3) = 0$ and $\underline{R}_{32}(a_{21}, b_3) = 0.25$

5.6 Chapter Conclusion

The performance of a fuzzy logic system is highly dependent on the quality of the fuzzification module and the knowledge base. Practically, both of these modules provide membership degrees to process by the inference engines. An outstanding learning mechanism should be able to generate expected membership degrees for both modules.

In this chapter, a learning mechanism for BK products is proposed. From a set of training data, this mechanism constructs a set of membership functions. With this set of membership functions, it eventually defines membership degrees for the knowledge base and fuzzification module. In the following chapter, this method will be implemented together with the BK subproduct to prove its usefulness.

CHAPTER 6

CLASSIFICATIONS WITH THE BK SUBPRODUCT

6.1 Introduction

In Chapter 4, the theoretical framework of weighted IVFS-based BK products are developed. In this chapter, an experiment is proposed to demonstrate the application of this framework, as well as to verify its advantages.

Among all, the BK subproduct has shown many good performance in the past (Yew & Kohout, 1997; Groenemans et al., 1997; Y.-i. Lee & Kim, 2008). Therefore, in this chapter, the discussion focuses on this relation composition. Three sets of BK subproduct based inference structures are going to test as classifiers. These inference structures are derived from the improved fuzzy (type-1) BK subproduct, IVFS-based improved BK subproduct and weighted IVFS-based improved BK subproduct respectively. Publicly available data sets are used for testing so that the comparison not only can be made between the inference structures, but also with other methods in the literature.

The details of the experiment settings are discussed in the following sections, including the data used in the tests, the inference structures, the training procedure, definition of weights, defuzzification process and etc.

6.2 Data Sets

Three publicly available data sets (Bache & Lichman, 2013) are used in the experiment, namely Statlog Heart (Statlog) (Michie, Spiegelhalter, & Taylor, 1994), Pima Indians Diabetes (Pima) (Smith, Everhart, Dickson, Knowler, & Johannes, 1988) and Wisconsin Diagnostic Breast Cancer (WDBC) (Street, Wolberg, & Mangasarian, 1993).

Table 6.1: A summary of data sets used in the experiment.

Name	Abbreviation	Instances	Attributes	Classes
Statlog Heart	Statlog	270	13	2
Pima Indians Diabetes*	Pima	768	8	2
Wisconsin Diagnostic Breast Cancer	WDBC	569	30	2

* This data set comes with some missing values.

A summary of these data sets are shown in Table 6.1.

For each data set, all the instances are divided into 2 groups: training group for learning and testing group for prediction. To minimize the bias in random sampling of training data (Delen, Walker, & Kadam, 2005), *k*-fold cross validation (Kohavi, 1995) is used in the experiment. With the *k*-fold cross validation, each data set is randomly divided into *k* mutually exclusive groups. The number of instances in each group is approximately equal. While some of these groups are used for training, the remaining groups are used to examine the classification accuracy of the BK subproduct. To learn the characteristics of the BK subproduct in different training environment, four training-testing data ratios are adopted, namely 1:4, 1:1 (2-fold cross validation), 4:1 (5-fold cross validation) and 9:1 (10-fold cross validation). For each train-test ratio, 30 tests are conducted. In each of these 30 train-test iteration, an independent set of random training and testing data are generated. The average accuracy of classification for each data set is computed at the end of the experiment to learn the classification ability of the BK subproduct.

$$\text{average accuracy} = \frac{1}{N}\sum_{n=1}^{N}\frac{(\text{Total number of instances that predicted correctly})_n}{(\text{Total number of instances})_n} \tag{6.1}$$

where $n = \{1, \cdots, N\}$ and N is the total tests run for each training-testing data ratio and equals to 30 in this case.

6.2.1 Statlog Heart Data Set

The Statlog data set consists of 270 records that can be classified into 2 classes, namely presence and absence (150 instances) of heart-disease (120 instances). This data set is originated from the Cleveland Clinic Foundation to classify those instances based on 13 attributes, including (1) age, (2) sex, (3) chest pain, (4) resting blood pressure, (5) serum cholesterol, (6) fasting blood sugar, (7) resting electrocardiographic results, (8) maximum heart rate achieved, (9) exercise induced angina, (10) oldpeak, (11) the slope of the peak exercise ST segment, (12) number of major vessels colored by flourosopy and (13) thal. No missing value is found in this data set.

6.2.2 Pima Indians Diabetes Data Set

Pima is a data set originated from the National Institute of Diabetes and Digestive and Kidney Diseases. This 768 instances data set concern with the presence (268 instances) or absence (500 instances) of diabetes among Pima-Indian heritage females with age is at least 21 years old. A total of 8 attributes can be found on these patients, i.e. (1) number of times pregnant, (2) plasma glucose concentration a 2 hours in an oral glucose tolerance test, (3) diastolic blood pressure, (4) triceps skin fold thickness, (5) 2 hours serum insulin, (6) body mass index, (7) diabetes pedigree function and (8) age.

Some missing values are reported in the data set. This can be verified easily as in some attributes such as the diastolic blood pressure and body mass index, the values of some instances are zero, which is biologically impossible. However, in this experiment, no special treatment is taken to this problem due to the lack of information. Therefore, these values are treated as it.

6.2.3 Wisconsin Diagnostic Breast Cancer Data Set

This data set is donated by the University of Wisconsin. With fine needle aspirate (FNA) of a breast mass, digitized image is taken. For each, 30 attributes that describe characteristics of the cell nuclei are recorded. These attributes are derived from: (1) radius, (2) texture, (3) perimeter , (4) area, (5) smoothness, (6) compactness, (7) concavity, (8) concave points, (9) symmetry and (10) fractal dimension. A total of 569 instances can be found in this data set, classified into 2 classes, namely malignant (212 instances) and benign (357 instances). No missing value is found.

6.3 Design Of The Inference Engines

In the discussion on the improved fuzzy (type-1) BK subproduct in Section 3.2, a few outstanding inference structures have been discussed and evaluated. Among all, K9 is one of the top ranked inference structures (Yew & Kohout, 1997) :

$$\text{K9}: \quad \min\Big(\frac{1}{|B|}\sum_{b\in B}(R(a,b)\rightarrow S(b,c)), \text{OrBot}\big(\text{AndTop}(R(a,b),S(b,c))\big)\Big) \tag{6.2}$$

where $|B| \in \mathbb{N}$ is the number of elements in B, $\text{AndTop}(p,q) = \min(p,q)$ and $\text{OrBot}(p,q) = \min(1, p+q)$ respectively, $\forall p,q \in [0,1]$.

To implement this inference structure, the $R(a,b)$ and $S(b,c)$ have to be defined. This research design the inference engine as follow: A is a set of all the instances (Table 6.1), whereas B is a set includes all the features and C is the classes. Therefore, for a defined membership function, $R(a,b)$ is the membership degree of relation between instance a and feature b, whereas $S(b,c)$ is the membership degree of relation between feature b and class c. With this design, K9 computes the membership degree of the relation between instances and classes. Comparison between these output membership degrees give results about which class an instance belongs to.

For the inferences with IVFS-based BK subproduct, Eq. (4.14) is initialized with logical connectives. Comparisons are only meaningful if the logical connectives that used in K9 are adopted. With this, V9, an inference structure based on IVFSs is obtained. The intervals that this inference structure computes are:

$$\Big[\min\Big(\frac{1}{|B|}\sum_{b\in B}(\overline{R}(a,b)\Rightarrow \underline{S}(b,c)),\mathrm{OrBot}\big(\mathrm{AndTop}(\underline{R}(a,b),\underline{S}(b,c))\big)\Big),$$
$$\min\Big(\frac{1}{|B|}\sum_{b\in B}(\underline{R}(a,b)\Rightarrow \overline{S}(b,c)),\mathrm{OrBot}\big(\mathrm{AndTop}(\overline{R}(a,b),\overline{S}(b,c))\big)\Big),\Big] \tag{6.3}$$

where $\underline{R}$ and $\overline{R}$ are the lower bound and upper bound of membership degrees of relations between instance a and feature b respectively, and the lower bound and upper bound of membership degrees of relation between b and class c are given by $\underline{S}$ and $\overline{S}$.

The last inference structure, W9, is a weighted inference structure based on V9. With features in set B be divided into M feature subsets according to their weight in inferences, an interval can be computed for each feature subset with Eq. (6.3). Define $\tilde{W}_m$, weight of feature subset $m = \{1,\cdots,M\}$ and inference structure W9 can be formed with replacing $\tilde{R}_m \bigstar \tilde{S}_m(a,c)$ in Eq. (4.29) with intervals compute from Eq. (6.3). Because of $\tilde{W}_m$ are IVFSs, the outputs of this inference structure are IVFSs as well.

6.4 Construction of Membership Functions

In Section 6.3, all the three constructed inference structures require membership degrees $\tilde{R}(a,b)$ and $\tilde{S}(b,c)$ (or $R(a,b)$ and $S(b,c)$) for computations. In term of fuzzy logic systems (Figure 1.1), $\tilde{R}(a,b)$ are the results of the fuzzification module that give the relations between set A and set B, whereas $S(b,c)$ form the knowledge base that specifies the relations between set B and set C. Therefore, retrieving these membership degrees enable inference engines operate according to expectation.

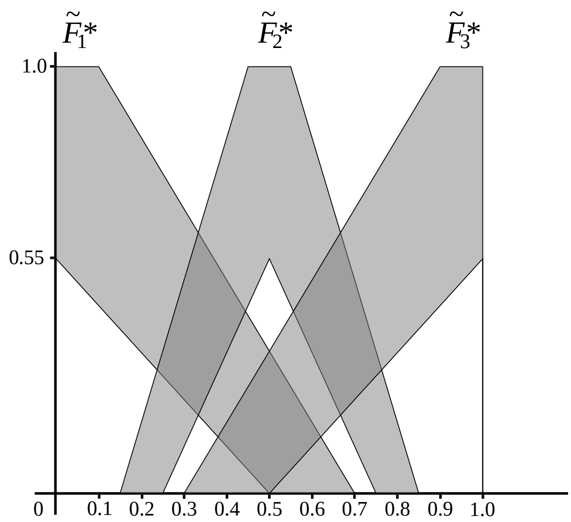


Figure 6.1: Definition of 3 standard membership functions, $\tilde{F}_1^*$, $\tilde{F}_2^*$ and $\tilde{F}_3^*$.

This section adopts the learning mechanism that defined in Chapter 5 to form the set of membership functions, follow by training of the knowledge base and convert test data to membership degrees.

Instead of define H_j membership functions for each b_i, this experiment fix $H = 3$ for all the features. A set of three standard membership functions (Figure 6.1) are defined, namely "Low"($\tilde{F}_1^*$), "Medium" ($\tilde{F}_2^*$) and "High" ($\tilde{F}_3^*$). For each b_j, find the corresponding limits of train data, $\underline{L}_j$ and $\overline{L}_j$. Next, scale the standard membership functions to this interval $[\underline{L}_j, \overline{L}_j]$. Let $L^{\#} = \overline{L}_j - \underline{L}_j$, Table 6.2 shows the scaling equations for feature b_j, which follow the 8 points definition of interval valued membership functions in Figure 5.2.

With this, the Statlog, Pima and WDBC data sets form 13×3, 8×3 and 30×3 membership functions respectively, for each of the inference system. Follow the algorithm of building knowledge base in Section 5.4, all the $\tilde{S}(b,c)$ can be computed.

Since T1FSs are special cases of IVFSs, for the experiment with T1FS-based BK inference structure, the similar approach can be used if the interval membership functions $\tilde{F}_{j\gamma}$ that used to generate membership degrees are replaced with type-1 membership func-

Table 6.2: Scaling from standard membership functions $\tilde{F}_1^*$, $\tilde{F}_2^*$ and $\tilde{F}_3^*$ to membership functions for b_j, namely $\tilde{F}_{j1}$, $\tilde{F}_{j2}$ and $\tilde{F}_{j3}$.

	$\tilde{F}_{j\gamma}$		
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$
$D_{j\gamma 1}$	$\underline{L}_j$	$\underline{L}_j + 0.15L^\#$	$\underline{L}_j + 0.30L^\#$
$D_{j\gamma 2}$	$\underline{L}_j$	$\underline{L}_j + 0.45L^\#$	$\underline{L}_j + 0.90L^\#$
$D_{j\gamma 3}$	$\underline{L}_j + 0.10L^\#$	$\underline{L}_j + 0.55L^\#$	$\overline{L}_j$
$D_{j\gamma 4}$	$\underline{L}_j + 0.70L^\#$	$\underline{L}_j + 0.85L^\#$	$\overline{L}_j$
$D_{j\gamma 5}$	$\underline{L}_j$	$\underline{L}_j + 0.25L^\#$	$\underline{L}_j + 0.50L^\#$
$D_{j\gamma 6}$	$\underline{L}_j$	$\underline{L}_j + 0.50L^\#$	$\overline{L}_j$
$D_{j\gamma 7}$	$\underline{L}_j + 0.70L^\#$	$\underline{L}_j + 0.75L^\#$	$\overline{L}_j$
ν	0.55	0.55	0.55

tions, $F_{j\gamma}$. To make the comparisons between T1FS- and IVFS-based BK subproduct fair, the following three criteria should be considered for the replacement type-1 membership functions:

1. The membership functions should be normal;
2. Each membership function that replace an interval-valued membership function should be in the FOU of the original IVFS;
3. A membership function set for a feature should form symmetry pattern.

Five sets of T1FSs that meet the above criteria are formed, they are: 1) trapezoids that equal to the UMF of the IVFSs, denoted as Tra-out; 2) Trapezoids that the points where membership degrees 0 are the LMF of the IVFSs, but the points where membership degrees 1 are the UMF of IVFSs, denoted Tra-in; 3) Equilateral triangles that the points where membership degrees 0 are the UMF of the IVFSs, denoted as Tri-out; 4) Equilateral

Table 6.3: Coordinates of standard membership functions for T1FS.

Tra-out	F_1^*	$\{(0.00,0.00),(0.00,1.00),(0.10,1.00),(0.70,0.00)\}$
	F_2^*	$\{(0.15,0.00),(0.45,1.00),(0.55,1.00),(0.85,0.00)\}$
	F_3^*	$\{(0.30,0.00),(0.90,1.00),(1.00,1.00),(1.00,0.00)\}$
Tra-in	F_1^*	$\{(0.00,0.00),(0.00,1.00),(0.10,1.00),(0.50,0.00)\}$
	F_2^*	$\{(0.25,0.00),(0.45,1.00),(0.55,1.00),(0.75,0.00)\}$
	F_3^*	$\{(0.50,0.00),(0.90,1.00),(1.00,1.00),(1.00,0.00)\}$
Tri-out	F_1^*	$\{(0.00,0.00),(0.00,1.00),(0.70,0.00)\}$
	F_2^*	$\{(0.15,0.00),(0.50,1.00),(0.85,0.00)\}$
	F_3^*	$\{(0.30,0.00),(1.00,1.00),(1.00,0.00)\}$
Tri-mid	F_1^*	$\{(0.00,0.00),(0.00,1.00),(0.60,0.00)\}$
	F_2^*	$\{(0.20,0.00),(0.50,1.00),(0.80,0.00)\}$
	F_3^*	$\{(0.40,0.00),(1.00,1.00),(1.00,0.00)\}$
Tri-in	F_1^*	$\{(0.00,0.00),(0.00,1.00),(0.50,0.00)\}$
	F_2^*	$\{(0.25,0.00),(0.50,1.00),(0.75,0.00)\}$
	F_3^*	$\{(0.50,0.00),(1.00,1.00),(1.00,0.00)\}$

triangles that the points where membership degrees 0 are the middle points between UMF and LMF of the IVFSs, denoted as Tri-mid; 5) Equilateral triangles that the points where membership degrees 0 are the LMF of the IVFSs, denoted as Tri-in. Table 6.3 shows the coordinates of their standard membership functions F_1^*, F_2^* and F_3^*. With these standard membership functions, $S(b,c)$ can be found.

For the fuzzification module, the similar membership function sets, $\tilde{F}_{j1}$, $\tilde{F}_{j2}$ and $\tilde{F}_{j3}$ (or F_{j1}, F_{j2} and F_{j3} for the case of T1FSs) are used again. Map the test data to the corresponding membership functions as described in the fuzzification algorithm in Section 5.4, all the $\tilde{R}(a,b)$ (or $R(a,b)$) can be computed.

6.5 Inference and Defuzzification

Using the settings described in the previous sections, the classifications can be done with each of the defined inference structure. The Kleene-Dienes implication operator is adopted for the experiment.

An instance in a data set is assume to be an object to infer, with the attributes are the features used for inferences. The classes of the instances are the objects that an inference targeted for. A complete inference for an instance a_i involves computing the relationships $\tilde{R} \triangleleft \tilde{S}(a_i, c)$ (or $R \triangleleft S(a_i, c)$) of the instance with all the c in the object set C. Therefore, with the definition of 3 standard membership functions for each feature (refer to Section 6.4), the total number of computations $R(a,b) \to S(b,c)$ involve in a complete inference for an instance is $3JK$ if inference structure K9 (based on T1FS) is used, where J is the total number of features and K is the total number of objects. For the case where inference structure V9 or W9, which is based on IVFS is used, the number will be $2 \times 3JK$, if the weight computation is not considered. The weights computation is discussed in the next section.

After the inferences, the results from the inference engines are defuzzified to produce meaningful information. One should able to find a lot of discussions in the literature on the defuzzification of inference results of rule-based systems based on T1FSs (C. C. Lee, 1990b; Filev & Yager, 1991; Patel & Mohan, 2002; Dubois, 2011). For the case of IVFSs and T2FSs, an extra procedure of type reduction (N. Karnik & Mendel, 1998; Greenfield, Chiclana, & John, 2009) involve, which reduces the IVFSs or T2FSs to T1FSs.

Despite there is a broad range of defuzzification methods in the literature, to interpret the results of a classifier based on the BK subproduct, a ranking method is sufficient. Generally, ranking methods compare the results of $\tilde{R} \triangleleft \tilde{S}(a_i, c_k)$ for all $k \in K$. Find $c_{k'}$ where $\tilde{R} \triangleleft \tilde{S}(a_i, c_{k'})$ is greater than $\tilde{R} \triangleleft \tilde{S}(a_i, c_k)$ for $k' \in K$ but $k \neq k'$. The ranking shows that a_i has strongest relation to $c_{k'}$. Therefore, a_i belongs to this class.

For the results of the K9 inference structure, the ranking process is easy because the outputs of this inference structure are point values in the range $[0, 1]$ and can be compared directly. For V9 inference structure, the outputs are intervals. To compare these intervals, the method described in Section 3.3 is applied. If $\Im_{ik} = [\underline{\Im}_{ik}, \overline{\Im}_{ik}]$ is the interval generated for target c_k by object a_i, the procedure to compare $\Im_{ik}$ across all intervals where $k \in \{1, \cdots, K\}$ is as follow: Firstly, a dynamic threshold β_i is generated with the mean of the alternate sequence of boundary values of $\Im_{ik}$ (for example, if K is an even number, $\beta_i = (\underline{\Im}_{i1} + \overline{\Im}_{i2} + \cdots + \overline{\Im}_{iK})/K$). Secondly, find Θ^{β}_{ik}, the reliability of each interval $\Im_{ik}$ using Eq. (3.7). Compare these reliability measures, which are point-valued. The target k' that with highest reliability interval is the one that have the strongest relation to the instance a_i.

For the case of the W9 inference structures, since the weights are modelled with IVFSs, the outputs are also IVFSs. There are some ranking algorithms for IVFSs in the literature (Mitchell, 2006; D. Wu & Mendel, 2009; Zhang, Joshua, & Lim, 2011). In this experiment, the study of D. Wu and Mendel (2009) is adopted to reduce the IVFS outputs to intervals. Lastly, method that use for rank the output of the V9 (described in the previous paragraph) is used to rank the results.

6.6 Determination Of Weight

In the process of weight determination, features/attributes are grouped according to their influence towards the classification results. In this experiment, features are divided into two groups, namely “low” and “high”, according to the influence. The membership functions of these groups are shown in Figure 6.2.

An algorithm is needed to distinguish whether a feature should belongs to high or low weight group. This weighing algorithm examines each attribute iteratively, using the training data.

Firstly, the data used during the training stage are employed again. Fuzzify the training data by mapping it to the membership function sets $\tilde{F}_{j1}$, $\tilde{F}_{j2}$ and $\tilde{F}_{j3}$. This form

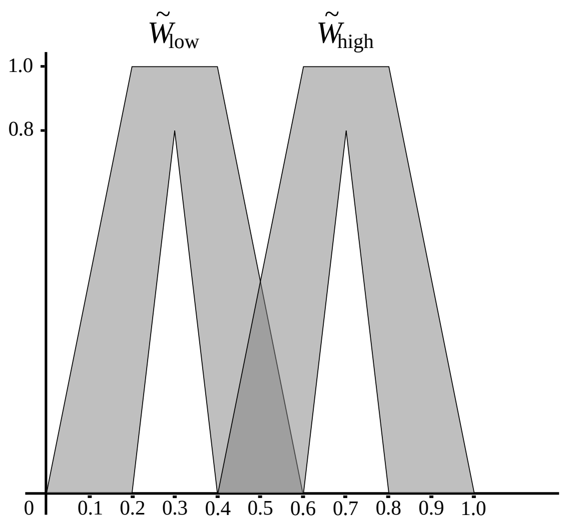


Figure 6.2: Definition of the low and high weight functions.

fuzzy relation $\tilde{R}(a^*, b)$ where $a^* \in A^*$ is the training data.

For each feature $b_j \in B$, find separately $\tilde{R} \triangleleft \tilde{S}(a^*, c)$ with V9 inference structure, as what described in Section 6.5, but on training data. Assume I^* is the total number of instances in training data. With this step, for a feature b_j, we find I^*K intervals that measuring the relations of objects and targets through this feature.

Convert these interval results to point-values by finding their arithmetic mean. For an $a^*_{i^*}$, $i^* \in I^*$, we get K point results. If the point value related to c'_k is the highest compare to the other, we can conclude that, for this particular $a^*_{i^*}$, b_j leads to infer for c'_k. Please note that for the computation on single feature b_j, it is common that multiple objects find the same highest value, or even all objects find the same value. Anyway, the same conclusion still can be made.

Compare the defuzzified inference results with the ground truth and count G_j, the number of instances that lead to correct inferences by b_j. The higher the G_j, the higher the influence feature b_j lead to correct inference. Therefore, the weight of b_j is given by w_j:

$$w_j = \frac{G_j}{I^*} \tag{6.4}$$

Find the median of the weight across all features, namely w^{M}. If the weight of a feature is smaller than w^{M} and a define value, 0.5, the feature falls to the group with low weight, otherwise it is in the high weight group. With this, all the features are divided into 2 weight groups.

6.7 Chapter Conclusion

This chapter discusses the settings of experiment so that the IVFS-based BK subproduct can be examined. The aim of this experiment is to demonstrate the implemention BK subproduct based inference structure as classifier for three publicly available data sets.

K9, the inference structure with outstanding performance is adopted in the experiment, and going to compare with another two newly developed inference structures, namely V9 and W9. Method described in Chapter 5 plays a main role in preparing data for the inferences. However, instead of defining membership functions for each feature separately, three standard membership functions are defined and scaled to the domain of each feature. With the train data, the weight group of each feature is determined. Lastly, the inference results are compared to find the accuracy of classifications.

CHAPTER 7

RESULTS AND DISCUSSION

7.1 Introduction

In previous chapter, an experiment is set up to demonstrate the application of BK subproduct as classifier, as well as to examine the two improvements over the classical fuzzy BK products, namely extension to IVFSs and the additional weight parameter. The experiment involves 3 publicly available data sets, each data set is tested with 4 different train-test ratios.

In this chapter, a common measuring model that based on accuracy, sensitivity and specificity is adopted. Through the experiment results, the improvements from the extension to IVFSs and the weight parameter are examine separately. Apart from this, the classification results are also compared with a few state-of-the-art classifiers in the literature. The training mechanism that introduced in the Chapter 5 are also discussed at the end of this chapter.

7.2 Measuring Model

Firstly, the measuring model is described briefly. The inference results and the ground truth can be compared with a matrix similar to Figure 7.1. In term of medical diagnosis, true positive (TP) is the case where a patient (instance) is classified as sick; true negative (TN) is the case where a healthy person is inferred as healthy; false positive (FP), or false alarm is the case where a healthy person is wrongly classified as sick and false negative (FN) is the case where a patient is wrongly inferred as healthy. A good classifier should aim for high TP and TN, but low FP and FN.

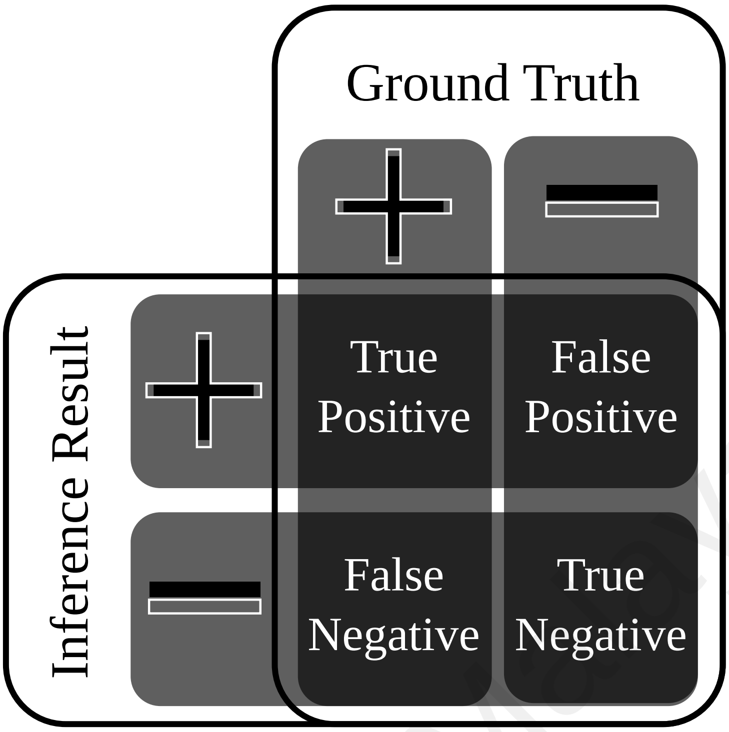


Figure 7.1: The confusion matrix that provides the fundamental of evaluation.

Among all, the most popular measurement in evaluating the performance of a classifier is accuracy. If $\mathfrak{T}^+$ denotes TP, $\mathfrak{T}^-$ denotes TN, $\mathfrak{F}^+$ denotes FP and $\mathfrak{F}^-$ denotes FN, the accuracy of a classifier is defined as:

$$\text{accuracy} = \frac{\mathfrak{T}^+ + \mathfrak{T}^-}{\mathfrak{T}^+ + \mathfrak{T}^- + \mathfrak{F}^+ + \mathfrak{F}^-} \tag{7.1}$$

Besides the accuracy, the sensitivity and specificity are another two popular indexes that measure the capability of a classifier. With the knowledge of TP and FN, sensitivity is given by:

$$\text{sensitivity} = \frac{\mathfrak{T}^+}{\mathfrak{T}^+ + \mathfrak{F}^-} \tag{7.2}$$

It is also equal to the number of $\mathfrak{T}^+$ divides by the population of patients. Therefore, one can explain the sensitivity as the probability of a patient to be inferred correctly in the patients population.

Specificity is the defined as:

$$\text{specificity} = \frac{\mathfrak{T}^-}{\mathfrak{T}^- + \mathfrak{F}^+} \tag{7.3}$$

It is also equal to the number of $\mathfrak{T}^-$ divides by the population of healthy people. Therefore, one can explain the specificity as the probability of a healthy people to be inferred correctly in the healthy people population.

A good classification algorithm should aim for high scores in accuracy, sensitivity and specificity.

7.3 Improvement With Interval Valued Fuzzy Sets

IVFSs have been adopted in many researches in the past decade (Mitchell, 2005; Melgarejo & Pena-Reyes, 2007; Poornaselvan, Kumar, & Vijayan, 2008; Zaher & Hagras, 2010). Although most of the researches claim that IVFSs help in the design of their applications, but not really all of them compared the results of their IVFS-based systems with T1FS-based systems. In this section, the discussion focuses on the results and comparisons of inference structures based on T1FSs and IVFSs. Firstly, the results of the experiment for WDBC, Statlog and Pima are presented in Table 7.1, 7.2 and 7.3 respectively.

7.3.1 Accuracy

In term of accuracy, it is clear that V9, the IVFS-based inference structure shows its advantages compare to all the implementations of K9s in most cases. Compare to the worst performing K9 implementations, V9 shows improvements from the range 0.571% to 32.690%, with the mean of improvement 9.758%. Compare to the best performing K9 implementations, V9 still performing well in most cases with maximum advantage of

Table 7.1: Averages (30 runs) of accuracy, sensitivity and specificity of classifications with BK subproduct for Wisconsin Diagnostic Breast Cancer.

Measurement	Train-Test Ratio	Inference Structures						
		K9 (Tra-out)	K9 (Tra-in)	K9 (Tri-out)	K9 (Tri-mid)	K9 (Tri-in)	V9 (IVFS)	W9 (Weighted)
Accuracy	1:4	78.750	92.449	81.564	92.683	91.601	92.763	95.526
	1:1	71.942	93.146	75.579	92.164	92.199	93.181	94.210
	4:1	62.738	93.828	66.775	92.483	93.272	94.150	95.730
	9:1	61.579	94.328	65.029	92.749	93.684	94.269	94.561
Sensitivity	1:4	67.422	98.988	72.166	94.876	99.302	84.019	86.011
	1:1	56.154	99.116	61.992	92.849	99.518	86.081	89.733
	4:1	40.974	98.838	47.498	90.992	99.406	89.770	93.725
	9:1	37.355	99.511	42.937	90.438	99.712	90.432	91.314
Specificity	1:4	98.167	81.368	100.000	96.395	83.943	97.932	98.364
	1:1	99.434	82.758	99.239	90.919	79.422	97.263	98.530
	4:1	99.773	85.407	99.601	94.946	82.959	96.720	97.746
	9:1	100.000	85.845	100.000	96.395	83.943	96.500	96.592

2.469%. The mean of improvements by V9 across all the best performing K9 implementations is 0.453%. In this class, only a few exceptions found in: i) WDBC test: it lost 0.059% to K9 (Tra-in) in train-test ratio 9:1; ii) Pima: it lost 0.163% and 0.243% respectively, to K9 (Tra-in) with train-test ratio 1:4 and 1:1. However, for the experiment with the Statlog heart disease, K9 with Tra-in as membership functions is not giving very good accuracy, even if compares to other K9 implementations. This represents that a T1FS membership functions adopted by the K9 is able to capture uncertainty in some cases but not some other. On the other hand, the maximum improvement brings by the adoption of IVFSs in this experiment is 2.469%.

Although K9s, the implementations of BK subproduct with T1FSs show higher accuracies in some cases, it is not enough to justify that T1FSs implementations are better then IVFSs in general. One can observe that the accuracy of K9s varies according to the

Table 7.2: Averages (30 runs) of accuracy, sensitivity and specificity of classifications with BK subproduct for Statlog Heart Disease.

Measurement	Train-Test Ratio	Inference Structures						
		K9 (Tra-out)	K9 (Tra-in)	K9 (Tri-out)	K9 (Tri-mid)	K9 (Tri-in)	V9 (IVFS)	W9 (Weighted)
Accuracy	1:4	80.463	80.617	80.278	80.803	80.448	80.849	84.259
	1:1	80.123	79.630	80.123	80.593	79.951	80.914	84.395
	4:1	80.864	79.951	80.173	80.593	79.630	83.333	85.926
	9:1	80.247	79.753	80.000	81.111	79.506	82.099	84.444
Sensitivity	1:4	82.589	78.268	81.725	78.324	78.123	77.858	79.814
	1:1	80.662	77.544	81.910	78.562	77.631	77.420	80.961
	4:1	81.910	77.631	80.662	78.562	77.544	80.582	83.897
	9:1	84.297	76.418	83.152	82.533	77.261	79.833	86.125
Specificity	1:4	78.748	82.457	77.731	80.147	81.385	83.231	87.838
	1:1	79.934	81.468	78.793	82.379	81.958	83.976	88.496
	4:1	78.793	81.958	79.934	82.379	81.468	85.253	84.590
	9:1	77.059	82.498	77.731	80.147	81.385	84.089	83.459

membership functions adopted, especially for experiment on WDBC (Table 7.1). In this experiment, the accuracy of K9s varies a lot. For example, in the test with train-test ratio 9:1, the accuracy results with K9 (Tra-in) and K9 (Tra-out) are 94.328% and 61.579% respectively. The standard deviation of these accuracy results is 14.9. One may argue that Tra-in is a better membership function compare to Tra-out. But, as pointed out earlier, Tra-out shows better results then Tra-in in the experiment with Statlog.

It is trivial that the changes in accuracy is because of the selection of membership functions. Once the membership functions adopted are not able to model the uncertainty as expected, the accuracy results drop. The uncertainty, which is ill-defined most of the time, may not be sufficient to represent with crisp, point-valued membership functions effectively. To capture uncertainty, a range of membership functions, or the interval-valued membership functions show their advantages (Turksen, 1986). Therefore, V9, the

Table 7.3: Averages (30 runs) of accuracy, sensitivity and specificity of classifications with BK subproduct for Pima Indians Diabetes.

Measurement	Train-Test Ratio	Inference Structures						
		K9 (Tra-out)	K9 (Tra-in)	K9 (Tri-out)	K9 (Tri-mid)	K9 (Tri-in)	V9 (IVFS)	W9 (Weighted)
Accuracy	1:4	71.523	72.136	71.236	71.111	71.772	71.973	72.835
	1:1	71.076	73.325	71.068	71.910	72.804	73.082	74.002
	4:1	71.082	73.853	71.364	72.597	72.662	74.632	75.693
	9:1	73.463	75.065	72.770	74.589	74.069	75.931	77.229
Sensitivity	1:4	45.262	45.432	41.648	36.342	40.506	36.992	38.177
	1:1	41.318	49.096	37.565	43.364	43.138	41.191	42.900
	4:1	41.624	54.319	38.519	46.184	48.069	45.210	46.320
	9:1	40.588	56.827	36.629	47.461	50.687	47.212	50.150
Specificity	1:4	85.422	86.364	90.958	88.369	85.984	90.542	90.225
	1:1	87.382	86.627	89.418	87.581	89.076	90.593	91.102
	4:1	86.840	84.289	88.883	86.762	85.775	90.211	91.203
	9:1	89.954	84.312	90.958	88.369	85.984	90.591	90.989

BK subproduct inference structure that based on IVFSs capable to show better accuracy in most of the cases.

One may also argue that, rather than representing uncertainty with intervals, we should clarify the uncertainty involved and model it carefully with T1FSs. This recall us about the doubts on fuzzy logic theory in early days (McCloskey & Glucksberg, 1978; Osherson & Smith, 1981; Zadeh, 1999), where some held strong believe that clear boundaries exist in object classifications and fuzzy sets theory was negated. Researches including Zadeh (1982), Kosko (1990) and Belohlavek, Klir, Harold W, and Way (2009) have clear and complete response to those questions. In the same vein, for the questions on the need of IVFSs instead of T1FSs, one might have to accept that most classical T1FSs-based fuzzy systems, with the assumption that uncertainty are crisp and well defined are special cases. The fact is, this crisp criteria may not always fit to all cases. Forcing point-

Table 7.4: Standard Deviations of Accuracies of K9 and V9

Data sets	Train-Test Ratio	Inference Structures	
		K9	V9
WDBC	1:4	1.532	1.525
	1:1	0.9	1.816
	4:1	2.215	1.968
	9:1	3.225	3.103
Statlog	1:4	2.182	1.98
	1:1	2.45	2.432
	4:1	4.391	4.33
	9:1	7.178	6.562
Pima	1:4	1.889	2.145
	1:1	1.735	2.14
	4:1	3.521	4.02
	9:1	4.921	5.159

valued membership functions in a system is just neglecting the fact that uncertainty is ill-defined in some cases. Therefore, high accuracy results in some T1FS-based inference structures are not guaranteed, they are just some special cases. Furthermore, if membership functions of T1FSs can be optimized to achieve the best accuracy, then the same strategy also can be applied to membership functions of IVFSs as well (Castillo, Huesca, & Valdez, 2005).

Table 7.4 compares the standard deviations of accuracy between V9 and the best performing K9 (Tra-In for both WDBC and Pima, Tra-Out for Statlog).

The standard deviation is a measure used to quantify the amount of dispersion from the mean values. In the measurement of accuracy, the lower the standard deviations, the closer the accuracy of runs to the mean accuracy. From the Table 7.4, one can observe that the standard deviations of tests on V9 are close to K9.

7.3.2 Sensitivity and Specificity

The measurement of sensitivity shows the probability of a patient to be inferred correctly in the patients population, whereas specificity is the probability of a healthy people to be inferred correctly in the healthy people population.

If the sensitivity and specificity results are compare across all the implementations of K9s, one can get an interesting finding: if a K9 implementation is having relatively high sensitivity in an experiment, the specificity is low; on the contrary, if a K9 implementation is having relatively low sensitivity in an experiment, the specificity is high. For example, K9 (Tra-out) and K9 (Tri-out) have gain high sensitivity in experiment with Statlog, but the specificity is low relative to K9 (Tra-in) and K9 (Tri-in). In contrast, for the experiment with Pima and WDBC data sets, K9 (Tra-in) and K9 (Tri-in) have relatively high sensitivity, but the K9 (Tra-out) and K9 (Tri-out) implementations are the ones with high specificity. In other words, K9 (Tra-in) and K9 (Tri-in) have relatively low specificity in the experiment with both WDBC and Pima, and K9 (Tra-out) and K9 (Tri-out) are having low sensitivity in the same experiment.

This results show that both K9 (Tra-in) and K9 (Tri-in) have stronger tendency to classify an instance as patient in the experiment for Pima and WDBC, but in the experiment with Statlog, they incline to label an instance as healthy. The features of K9 (Tra-out) and K9 (Tri-out) are the other way round. On the other hand, V9 that based on IVFSs is more consistent compare to these K9 implementations.

7.4 Improvement With Weight Parameter

From Table 7.1, 7.2 and 7.3, it is obvious that the additional weight parameter improves all the accuracy and sensitivity results, with only minor declination on the specificity in a few cases. The improvement to accuracy is ranged from 0.292% to 3.481%. However, it should be stressed that the algorithm that discussed in Section 6.6 is just a simple implementation for the purpose to prove the ability of the weight parameter. For application that aim for higher accuracy, some feature selection algorithms (Jain & Zongker, 1997; Guyon & Elisseeff, 2003; Peng, Long, & Ding, 2005) can be acquired.

Table 7.5: Improvement of accuracy by tuning of control parameter.

Dataset	Train-test ratio	Control Parameter	Accuracy
Statlog	1:4	0.5	84.182%
	1:4	0.6	85.818%
	1:1	0.5	84.395%
	1:1	0.6	84.938%
	4:1	0.5	85.926%
	4:1	0.6	86.667%
	9:1	0.5	84.444%
	9:1	0.6	86.420%
WDBC	1:4	0.5	93.772%
	1:4	0.7	95.526%
	1:1	0.5	95.298%
	1:1	0.7	95.298%
	4:1	0.5	95.730%
	4:1	0.7	96.226%
	9:1	0.5	94.561%
	9:1	0.7	94.620%
Pima	1:4	0.5	72.835%
	1:4	0.6	72.271%
	1:1	0.5	74.002%
	1:1	0.6	74.019%
	4:1	0.5	74.784%
	4:1	0.6	75.714%
	9:1	0.5	77.229%
	9:1	0.6	77.316%

With these feature selection algorithms, important features can be selected and assigned with high weights.

Although the scope of the thesis limits to the demonstration of the advantage of weight parameter, but not the optimization of the weights, it is worth to discuss the possibility of further improving results with this parameter. Besides the incorporation of feature selection algorithms that discussed on the above, other methods of improving the accuracy results include: i) optimization of the weight functions (Figure 6.2), and ii) fine tune the weight group control parameter. The optimization of weight functions involve

Table 7.6: Standard Deviations of Accuracies of V9 and W9

Data sets	Train-Test Ratio	Inference Structures	
		V9	W9
WDBC	1:4	1.525	2.125
	1:1	1.816	1.816
	4:1	1.968	1.971
	9:1	3.103	2.913
Statlog	1:4	1.98	2.768
	1:1	2.432	4.374
	4:1	4.33	4.046
	9:1	6.562	8.521
Pima	1:4	2.145	3.096
	1:1	2.14	2.887
	4:1	4.02	3.67
	9:1	5.159	4.751

the study of the shape of functions and numbers of weight functions/groups. Notice that a set of standard weight functions are used across all the three data sets in the experiment. It is not surprising that the results can be improved with the implementation of dedicated weight function sets that specially designed for each data set.

The fine tuning of weight group control parameter involve the control of number of features in each weight group. Recall that in Section 6.6, weight of each feature w_j is computed. A feature b_j falls to the group with low weight if w_j smaller then w^M and a control parameter 0.5. With the adjustment of this control parameter, the number of features in a weight group increases or decreases and this directly affect the accuracy of classification. Table 7.5 shows the accuracy results of the experiment if this control parameter is changed.

Lastly, the standard deviations of the accuracies of W9 and V9 are presented in Table 7.6.

One can see that the standard deviations of accuracy by W9 do not show major differences compare to the V9's, except in 2 tests with Statlog. From this observation, one can conclude that the improvement brings by the additional weight parameter is rather uniform.

7.5 Compare To Other Classifiers

In this section, the accuracy of the proposed V9 and W9 inference structures are compared to some classifiers in the literature. Due to many of the state-of-the-art researches only report the results with 10 runs instead of 30, a set of 10-run-result from each data set is extracted for the comparisons in this section.

Tables 7.7, 7.8 and 7.9 compare the results of experiment with Statlog, WDBC and Pima respectively, with some state-of-the-art works. Among all, Mantas and Abellán (2014) work on all the three data sets that this thesis works on; Hu (2013) and Jiang and Li (2013) work on both Statlog and Pima; Pacheco et al. (2012) work on both Statlog and WDBC; Li and Liu (2010) work on both WDBC and Pima. From the comparisons, it is clear that the weighted IVFS-based BK subproduct has its advantages compare to other classifiers.

7.6 The Mechanism Of Generating Membership Functions

Chapter 5 proposed a membership function definition mechanism for BK products-based inference engines. This mechanism not only helps in training the inference engines, but also serve to fuzzify test data. In this section, some features of this mechanism are discussed.

Basically, this method trains a system that provides information on the distribution of data in specific ranges. Therefore, as to many other training methods, the prediction accuracy increases when the number of training data increases. Meanwhile, this method shows an advantage: it manage to train a system with limited data, as long as the pattern of data distribution is obtained. This conclusion can be further affirmed with the experiment

Table 7.7: Comparison between classifier on the accuracy of Statlog data set

Author	Method	Train-Test Ratio	Runs	Accuracy
Mantas and Abellán(2014)	Credal-C4.5 (no pruning)	9:1	10	80.04%
Mantas and Abellán(2014)	Credal-C4.5 (pruning)	9:1	10	80.33%
Chen et al.(2014)	ACO-S1	9:1	10	81.85%
Chen et al.(2014)	ACO-S2	9:1	10	75.93%
Chen et al.(2014)	ACO-S3	9:1	10	82.96%
Yeh, Su, and Lee(2013)	SPDI	9:1	10	83.333%
Hu(2013)	RSRC-P	4:1	5	84.0%
Jiang and Li(2013)	AVDM	4:1	5	83.33%
Pacheco et al.(2012)	GRASP	9:1	10	78.1%
This study	V9	1:4	10	80.787%
This study	V9	1:1	10	81.556%
This study	V9	4:1	10	82.407%
This study	V9	9:1	10	84.074%
This study	W9	1:4	10	84.182%
This study	W9	1:1	10	85.556%
This study	W9	4:1	10	86.296%
This study	W9	9:1	10	84.815%
This study	V9	1:4	30	80.849%
This study	V9	1:1	30	80.914%
This study	V9	4:1	30	83.333%
This study	V9	9:1	30	82.099%
This study	W9	1:4	30	84.182%
This study	W9	1:1	30	84.395%
This study	W9	4:1	30	85.926%
This study	W9	9:1	30	84.444%

results of V9. For example, in Table 7.7, V9 that trained with only 20% of the total data manage to achieve accuracy of 80.849%, which is higher than some other methods trained with 90% of data.

In term of training efficiency, this method also shows its advantage. Since data distribution can be compute easily, the training process for V9 is fast. This can be observed in Table 7.10, where the training time for an experiment with train-test ratio 9:1 is presented. The hardware platform that running this experiment is a laptop computer with Intel Core

Table 7.8: Comparison between classifier on the accuracy of WDBC data set

Author	Method	Train-Test Ratio	Runs	Accuracy
Mantas and Abellán(2014)	Credal-C4.5 (no pruning)	9:1	10	95.08%
Mantas and Abellán(2014)	Credal-C4.5 (pruning)	9:1	10	95.12%
Koloseni et al.(2013)	Generalize DE	1:1	30	93.64%
Pacheco et al.(2012)	GRASP	9:1	10	94.8%
Li and Liu(2010)	SVM Gaussian	1:4	30	83.14%
Li and Liu(2010)	SVM Polynoimal	1:4	30	58.58%
Li and Liu(2010)	SVM CPBK	1:4	30	93.26%
This study	V9	1:4	10	92.983%
This study	V9	1:1	10	93.825%
This study	V9	4:1	10	94.730%
This study	V9	9:1	10	94.737%
This study	W9	1:4	10	94.518%
This study	W9	1:1	10	93.825%
This study	W9	4:1	10	96.136%
This study	W9	9:1	10	95.263%
This study	V9	1:4	30	92.763%
This study	V9	1:1	30	94.210%
This study	V9	4:1	30	94.150%
This study	V9	9:1	30	94.269%
This study	W9	1:4	30	93.772%
This study	W9	1:1	30	95.298%
This study	W9	4:1	30	95.730%
This study	W9	9:1	30	94.561%

i7-2670QM CPU @ 2.20GHz and 8GB RAM, whereas the software platform is Octave 3.6.4 running on Linux with kernel 3.11.0 (32 bits) with Physical Address Extension enabled (so that able to utilize 8GB of RAM). From the Table 7.10 and 6.1, one should able to see the training time increases with the number of features.

Table 7.9: Comparison between classifier on the accuracy of Pima data set

Author	Method	Train-Test Ratio	Runs	Accuracy
Mantas and Abellán(2014)	Credal-C4.5 (no pruning)	9:1	10	73.19%
Mantas and Abellán(2014)	Credal-C4.5 (pruning)	9:1	10	74.15%
Hu(2013)	RSRC-P	4:1	5	74.6%
Jiang and Li(2013)	AVDM	4:1	5	76.23%
Li and Liu(2010)	SVM Gaussian	1:2	30	64.00%
Li and Liu(2010)	SVM Polynoimal	1:2	30	62.52%
Li and Liu(2010)	SVM CPBK	1:2	30	71.15%
This study	V9	1:4	10	72.520%
This study	V9	1:1	10	72.370%
This study	V9	4:1	10	75.390%
This study	V9	9:1	10	76.493%
This study	W9	1:4	10	74.553%
This study	W9	1:1	10	74.036%
This study	W9	4:1	10	74.675%
This study	W9	9:1	10	76.753%
This study	V9	1:4	30	71.973%
This study	V9	1:1	30	73.082%
This study	V9	4:1	30	74.632%
This study	V9	9:1	30	75.931%
This study	W9	1:4	30	72.271%
This study	W9	1:1	30	74.002%
This study	W9	4:1	30	74.784%
This study	W9	9:1	30	77.229%

Table 7.10: Training time for V9 with train-test ratio 9:1

Data Sets	Time for 30 sets data (s)	Average time for a set of data
WDBC	2.49214	0.08307
Statlog	0.84584	0.02819
Pima	0.71643	0.02388

CHAPTER 8

CONCLUSION AND FUTURE WORK

8.1 Conclusion

Human reasoning is a mysterious phenomenon that scientists are trying to simulate with machines in the past few decades. With the knowledge that "soft" boundaries exist in concepts formation of human beings (Zadeh, 1997), fuzzy set theory has emerged to become one of the most important methodology in capturing notions. Among all reasoning methods, the fuzzy BK products, which make use of the concept of implications as suggested by the *modus ponens* (Eq. 1.1) and GMP (Eq. 1.2) are proved to be excellent (Stepnicka & Jayaram, 2010).

The classical fuzzy BK products are based on T1FSs. With the understanding that IVFSs have the advantage in capturing uncertainty, fuzzy BK products are extended to IVFS-based BK products in this research. This is done by firstly define the subsethood measures of IVFSs. Two subsethood measures are developed in the research, namely Complete Derivation Method and Border Evaluation Method. With Border Evaluation Method, which is mathematically more reliable and lower computational cost, IVFS-based BK subproduct, superproduct and square product are developed. The properties of these IVFS-based composition of relations are also discussed.

Furthermore, with the consideration that each feature involves in reasoning may have different level of influence towards the results (S.-M. Chen, 1994; Xing & Ha, 2014), a weight parameter is added to form weighted IVFS-based BK products. The computation of this weighted relational products is also provided by adopting the algorithm of LWA (D. Wu & Mendel, 2007, 2008a).

Putting the BK subproduct into the context of fuzzy logic systems, this research studies the implementation of this composition of relations as an inference engine. A list of inference structures developed from the BK subproduct are examined. With theoretical analysis and evaluation, a few good performing inference structures are identified, including K9, K7 and etc. Based on K9, the IVFSs version of K9 is formed, namely V9. Furthermore, weight parameter is added to form a weighted inference structure, W9.

A novel method of training is also developed to construct the knowledge based for the BK subproduct based inference engines. This method, which based on the probability density functions serves as a fast but effective method to retrieve the membership degrees of relations between features and targets. Apart from this, the membership functions developed in this process also serve in fuzzification of the input data.

For the defuzzification module, a reliability measure is proposed so that the interval results from V9 can be compared. The outputs of W9, which are IVFSs, are also compared with this method after a type reduction procedure.

The developed inference systems are implemented as classifiers so that the performance can be compared. An experiment on three publicly available data sets is carried out, each with four train-test ratios. These data sets are Statlog Heart, Pima Indians Diabetes and Wisconsin Diagnostic Breast Cancer. The classification accuracies of K9, V9 and W9 are examined and compared. In comparing K9 and V9, the results show that V9 recorded higher accuracy in most of the cases, with highest improvement up to 32.690%. Another advantage shown by V9 is its ability in capturing uncertainty: with an interval-valued membership function, V9 able to produce best results most of the time compared to type-1 membership functions in the range of the interval, which the performance are not stable.

For the comparison between W9 and V9, one can find that the weight parameter improves the accuracy in all settings of the experiment. The range of the improvements is from 0.206% to 4.024%. This result is based on a simple algorithm that assign weights to features with multiple independent trainings. Although better algorithms can be adopt here, but the result is enough to demonstrate that the advantage of adding weight parameter to inference structures, which is also the aim of the research.

8.2 Future Work

As mentioned in Section 8.1, a good weight assignment mechanism is a key to boost the performance of the weighted BK subproduct based inference engines, in term of efficiency and accuracy. The advancement of feature selection algorithms (Dash & Liu, 1997; H. Liu & Yu, 2005) may provide clues in developing this mechanism. Feature selection is a process that selecting a subset of features that are useful in system construction (Guyon & Elisseeff, 2003). In this case, feature selection algorithms can be used to find out the features with high level of influence and assign them to high weight groups, and the others in medium or low weight groups. Besides, with the assumption that similar pattern may exist in a feature for a inference target, measure of statistical dispersion may also gives hints to weight assignment. However, in order to compare the dispersion across features, a dimensionless measures such as Coefficient Of Variation and Quartile Coefficient Of Dispersion (Bonett, 2006) is required.

This research developed an efficient learning algorithm. However, one can see that once a membership function is defined, the corresponding membership degrees $\overline{S}(b,c)$ and $\underline{S}(b,c)$ are only related to the supports of the membership functions, but not the shapes in overall, including the value of v. Currently, the shape of the membership functions only serve for the fuzzification of input data. It is reasonable to believe that the shape of the membership functions may provide useful information in the calibration of $\overline{S}(b,c)$ and $\underline{S}(b,c)$, but their relations are still unclear. This is one of the topic that provides research opportunity for the future. Besides, three standard membership functions are used across all the features in all the data sets. Is the distribution of data related to the optimal number and shape of membership functions? How can one calibrate the standard membership functions to reach better classification accuracy? Yet, these are the questions that have to be answered in the future.

Last but not least, the BK products are studies on the composition of relations between sets that are not directly related. Classifier is only one of the possible applications of BK products. Since relations provides important notions in human reasoning, it is possible to apply BK products in other problems such as control engineering and computing with words. In control engineering, sensors are used to measure the outputs of a system. The measurements are used as feedback to calibrate the input levels so that the system can achieve desire performance. In the past, fuzzy rule-based systems are widely studied as the core of controllers (C. C. Lee, 1990b, 1990a; Zaher & Hagras, 2010). With the advantages of the BK products, it is possible to replace the rule-based systems in the controllers with the weighted IVFS-based BK products to achieve similar, or even higher performance and efficiency. In the study of computing with words, researchers are aiming to replace numerical information in the input/output processes of information systems, with words that modeled by fuzzy sets (Zadeh, 1996, 1999). In this kind of information systems, fuzzy logic systems are the core of mapping these input fuzzy sets to the outputs. In the literature, two main research directions in computing with words are the modeling of words with fuzzy sets (Herrera & Martinez, 2000; F. Liu & Mendel, 2008b) and the core processing systems (D. Wu & Mendel, 2008b, 2010). In the later, BK products that extended in this thesis may show their advantages.

REFERENCES

Aisbett, J., Rickard, J. T., & Morgenthaler, D. G. (2010). Type-2 Fuzzy Sets as Functions on Spaces. *IEEE Transactions on Fuzzy Systems*, *18*(4), 841–844.

Bache, K., & Lichman, M. (2013). *UCI machine learning repository.* Retrieved from `http://archive.ics.uci.edu/ml`

Baczyński, M., & Jayaram, B. (2008). *Fuzzy Implications*. Berlin: Springer.

Belohlavek, R., Klir, G. J., Harold W, L. I., & Way, E. C. (2009). Concepts and fuzzy sets : Misunderstandings , misconceptions , and oversights. *International Journal of Approximate Reasoning*, *51*(1), 23–34.

Bodenhofer, U., Dankova, M., Stepnicka, M., & Novak, V. (2007, June). A Plea for the Usefulness of the Deductive Interpretation of Fuzzy Rules in Engineering Applications. *2007 IEEE International Fuzzy Systems Conference*, 1–6.

Bonett, D. G. (2006). Confidence interval for a coefficient of quartile variation. *Computational statistics & data analysis*, *50*(11), 2953–2957.

Bui, L.-d., & Kim, Y.-g. (2006). An Obstacle-avoidance Technique for Autonomous Underwater Vehicles Based On BK-products of Fuzzy Relation. *Fuzzy Sets and Systems*, *157*, 560 – 577.

Bustince, H. (2000, March). Indicator of inclusion grade for interval-valued fuzzy sets. Application to approximate reasoning based on interval-valued fuzzy sets. *International Journal of Approximate Reasoning*, *23*(3), 137–209.

Bustince, H., Barrenechea, E., Pagola, M., & Fernandez, J. (2009, July). Interval-valued fuzzy sets constructed from matrices: Application to edge detection. *Fuzzy Sets and Systems*, *160*(13), 1819–1840.

Běhounek, L., & Daňková, M. (2009, April). Relational compositions in Fuzzy Class Theory. *Fuzzy Sets and Systems*, *160*(8), 1005–1036.

Castillo, O., Huesca, G., & Valdez, F. (2005). Evolutionary computing for optimizing type-2 fuzzy systems in intelligent control of non-linear dynamic plants. In *Fuzzy information processing society, 2005. nafips 2005. annual meeting of the north american* (pp. 247–251).

Castro, J. R., Castillo, O., & Melin, P. (2007, June). An Interval Type-2 Fuzzy Logic Toolbox for Control Applications. *2007 IEEE International Fuzzy Systems Conference*(August), 1–6.

Chen, S.-M. (1994, January). A weighted fuzzy reasoning algorithm for medical diagnosis. *Decision Support Systems*, *11*(1), 37–43.

Chen, Y., Wong, M.-L., & Li, H. (2014, May). Applying Ant Colony Optimization to configuring stacking ensembles for data mining. *Expert Systems with Applications*, *41*(6), 2688–2702.

Choi, B.-I., & Rhee, F. C.-H. (2009, June). Interval type-2 fuzzy membership function generation methods for pattern recognition. *Information Sciences*, *179*(13), 2102–2122.

Civanlar, M. R., & Trussell, H. J. (1986). Constructing membership functions using statistical data. *Fuzzy Sets and Systems*, *18*(1), 1–13.

Coupland, S. (2007). Geometric type-1 and type-2 fuzzy logic systems. *Fuzzy Systems, IEEE Transactions on*, *15*(1), 3–15.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent data analysis*, *1*(3), 131–156.

Davis IV, W. L., & Kohout, L. (2006). Enhancing Performance of Relational Fuzzy Neural Networks with

Square BK-Products. *Proceedings of the 9th Joint Conference on Information Sciences (JCIS).*

Davis IV, W. L. (2006). *Enhancing pattern classification with relational fuzzy neural networks and square BK-products* (Unpublished doctoral dissertation). Florida State University.

Daňková, M. (2007). On approximate reasoning with graded rules. *Fuzzy Sets and Systems*, *158*, 652–673.

De Baets, B., & Kerre, E. E. (1993a, November). Fuzzy relational compositions. *Fuzzy Sets and Systems*, *60*(1), 109–120.

De Baets, B., & Kerre, E. E. (1993b, November). The generalized modus ponens and the triangular fuzzy data model. *Fuzzy Sets and Systems*, *59*(3), 305–317.

De Baets, B., & Kerre, E. E. (1993c). A revision of Bandler-Kohout compositions of relations. *Mathematica Pannonica*, *4*(1), 59–78.

De Baets, B., & Kerre, E. E. (1994). Fuzzy Relations and Applications. *Advance In Electronics And Electron Physics*, *89*(1), 255–324.

Delen, D., Walker, G., & Kadam, A. (2005, June). Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial intelligence in medicine*, *34*(2), 113–27.

Dong, W. M., & Wong, F. S. (1987). Fuzzy weighted averages and implementation of the extension principle. *Fuzzy Sets and Systems*, *21*, 183–199.

Dubois, D. (2011, July). The role of fuzzy sets in decision sciences: Old techniques and new directions. *Fuzzy Sets and Systems*, *184*(1), 3–28.

Dubois, D., & Prade, H. (1996). What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, *84*, 169–185.

Dubois, D., & Prade, H. (2008). An introduction to bipolar representations of information and preference. *International Journal of Intelligent Systems*, *23*, 866–877.

Fazel Zarandi, M., Rezaee, B., Turksen, I., & Neshat, E. (2009, January). A type-2 fuzzy rule-based expert system model for stock price analysis. *Expert Systems with Applications*, *36*(1), 139–154.

Filev, D. P., & Yager, R. R. (1991, December). A generalized defuzzification method via bad distributions. *International Journal of Intelligent Systems*, *6*(7), 687–697.

Fisher, P. (2007, February). What is Where? Type-2 Fuzzy Sets for Geographical Information. *Computational Intelligence Magazine, IEEE*, *2*(1), 9–14.

Fortemps, P., & Roubens, M. (1996, September). Ranking and defuzzification methods based on area compensation. *Fuzzy Sets and Systems*, *82*(3), 319–330.

Gorzalczany, M. B. (1987). A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy sets and systems*, *21*, 1–17.

Greenfield, S., Chiclana, F., Coupland, S., & John, R. (2009, June). The collapsing method of defuzzification for discretised interval type-2 fuzzy sets. *Information Sciences*, *179*(13), 2055–2069.

Greenfield, S., Chiclana, F., & John, R. (2009, August). Type-reduction of the discretised interval type-2 fuzzy set. *2009 IEEE International Conference on Fuzzy Systems*, 738–743.

Groenemans, R., Ranst, V. E., & Kerre, E. E. (1997). Fuzzy relational calculus in land evaluation. *Geoderma*, *77*, 283–298.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, *3*, 1157–1182.

Hayashi, K., Otsubo, A., Murakami, S., & Maeda, M. (1999). Realization of nonlinear and linear PID controls using simplified indirect fuzzy inference method. *Fuzzy Sets and Systems*, *105*, 409–414.

Herrera, F., & Martinez, L. (2000, October). A 2-tuple fuzzy linguistic representation model for computing with words. *Fuzzy Systems, IEEE Transactions on*, *8*(6), 746–752.

Hoffmann, F. (2004, January). Combining boosting and evolutionary algorithms for learning of fuzzy classification rules. *Fuzzy Sets and Systems*, *141*(1), 47–58.

Hong, T., & Chen, J. (1999). Finding relevant attributes and membership functions. *Fuzzy Sets and Systems*, *103*, 389–404.

Hong, T.-P., & Lee, C.-Y. (1996). Induction of fuzzy rules and membership functions from training examples. *Fuzzy Sets and Systems*, *84*, 33–47.

Hu, Y.-C. (2013, July). Rough sets for pattern classification using pairwise-comparison-based tables. *Applied Mathematical Modelling*, *37*(12-13), 7330–7337.

Hwang, C.-M., Yang, M.-S., & Hung, W.-L. (2011, March). New Similarity and Inclusion Measures between Type-2 Fuzzy Sets. In *Advances in type-2 fuzzy logic systems (t2fuzz), 2011 ieee symposium on* (Vol. 57, pp. 82–87).

Ishibuchi, H., & Yamamoto, T. (2005). Rule Weight Specification in Fuzzy Rule-Based Classifiation Systems. *IEEE Transactions on Fuzzy Systems*, *13*(4), 428–435.

Jain, A., & Zongker, D. (1997). Feature selection: evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(2), 153–158.

Jiang, L., & Li, C. (2013, July). An Augmented Value Difference Measure. *Pattern Recognition Letters*, *34*(10), 1169–1174.

John, R., & Coupland, S. (2007). Type-2 fuzzy logic: A historical view. *Computational Intelligence Magazine, IEEE*, *2*(1), 57–62.

Karnik, N., & Mendel, J. M. (1998). Type-2 fuzzy logic systems: type-reduction. *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218)*, 2046–2051.

Karnik, N., & Mendel, J. M. (2001, February). Centroid of a type-2 fuzzy set. *Information Sciences*, *132*(1-4), 195–220.

Karnik, N. N., & Mendel, J. M. (2001, September). Operations on type-2 fuzzy sets. *Fuzzy Sets and Systems*, *122*(2), 327–348.

Karnik, N. N., Mendel, J. M., & Liang, Q. (1999). Type-2 Fuzzy Logic Systems. *IEEE Transactions on Fuzzy Systems*, *7*(6), 643–658.

Kerre, E. E. (2007). An overview of fuzzy relational calculus and its applications. In *Mdai 07* (pp. 1–13). Springer.

Klir, G. J., & Yuan, B. (1995). *Fuzzy Sets And Fuzzy Logic: Theory And Applications*. New Jersey: Prentice Hall.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International joint conference on artificial intelligence* (pp. 1137—-1145).

Kohout, L. J. (2009a). Power Sets and Implication Operators Revisited: A Retrospective Look at the Foundational and Conceptual Issues in Bandler and Kohout's Paper After 29 Years. In (chap. 7).

Kohout, L. J. (2009b). Power Sets, Implications and Set Inclusions Revisited - Retrospect and Prospect: A Review of Bandler and Kohout's Paper and a Survey of 30 Years of Subsequent Developments. In

Views on fuzzy sets and systems from different perspectives (Vol. 243, pp. 129–159). Springer.

Kohout, L. J., & Bandler, W. (1980a). Fuzzy Power Sets And Fuzzy Implication Operators. *Fuzzy Sets and Systems*, *4*, 13–30.

Kohout, L. J., & Bandler, W. (1980b, January). Semantics of implication operators and fuzzy relational products. *International Journal of Man-Machine Studies*, *12*(1), 89–116.

Kohout, L. J., & Bandler, W. (1985). Relational-Product Architectures for Information Processing. *Information Sciences*, *37*, 25–37.

Kohout, L. J., & Bandler, W. (1992). How the checklist paradigm elucidates the semantics of fuzzy inference. *IEEE International Conference on Fuzzy Systems*, 571–578.

Kohout, L. J., & Kim, E. (2002, April). The role of BK-products of relations in soft computing. *Soft Computing*, *6*(2), 92–115.

Kohout, L. J., Stabile, I., Kalantar, H., & San-Andres, F. (1995). Parallel interval-based reasoning in medical knowledge-based system Clinaid. *Reliable Computing*, *1*(2), 109–140.

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review*, *6*(1), 3–34.

Koloseni, D., Lampinen, J., & Luukka, P. (2013, August). Differential evolution based nearest prototype classifier with optimized distance measures for the features in the data sets. *Expert Systems with Applications*, *40*(10), 4075–4082.

Kosko, B. (1990). Fuziness vs. Probability. *International Journal of General Systems*, *17*, 211–240.

Lam, H., & Seneviratne, L. D. (2008, June). Stability analysis of interval type-2 fuzzy-model-based control systems. *IEEE Systems, Man, and Cybernetics, Part B*, *38*(3), 617–28.

Lee, C. C. (1990a). Fuzzy logic in control systems: fuzzy logic controller. Part I. *IEEE transactions on systems, man, and cybernetics*, *20*(2), 404–418.

Lee, C. C. (1990b). Fuzzy logic in control systems: fuzzy logic controller, part II. *IEEE Transactions on Systems, Man, and Cybernetics*, *20*(2), 419–435.

Lee, D. H., & Park, D. (1997, April). An efficient algorithm for fuzzy weighted average. *Fuzzy Sets and Systems*, *87*(1), 39–45.

Lee, Y.-i., & Kim, Y.-G. (2008, May). Comparison of fuzzy implication operators by means of fuzzy relational products used for intelligent local path-planning of AUVs. *Soft Computing*, *13*(6), 535–549.

Li, D.-C., & Liu, C.-W. (2010, April). A class possibility based kernel to increase classification accuracy for small data sets using support vector machines. *Expert Systems with Applications*, *37*(4), 3104–3110.

Lim, C. K., & Chan, C. S. (2012, June). Fuzzy set and multi descriptions property. In *2012 ieee international conference on fuzzy systems* (pp. 1–8). Brisbane: Ieee.

Lim, C. K., Yew, K. M., Ng, K. H., & Abdullah, B. J. J. (2002, September). A proposed hierarchical fuzzy inference system for the diagnosis of arthritic diseases. *Australasian physical & engineering sciences in medicine*, *25*(3), 144–50.

Liou, T.-S., & Wang, M.-J. J. (1992, August). Fuzzy weighted average: An improved algorithm. *Fuzzy Sets and Systems*, *49*(3), 307–315.

Liu, F., & Mendel, J. M. (2008a). Aggregation Using the Fuzzy Weighted Average as Computed by the Karnik–Mendel Algorithms. *IEEE Transactions on Fuzzy Systems*, *16*(1), 1–12.

Liu, F., & Mendel, J. M. (2008b). Encoding words into interval type-2 fuzzy sets using an interval approach.

Fuzzy Systems, IEEE Transactions on, *16*(6), 1503–1521.

Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *Knowledge and Data Engineering, IEEE Transactions on*, *17*(4), 491–502.

M, M. d. l. A. H., & Mendez, G. M. (2007). Modelling and Prediction of the MXNUSD Exchange Rate Using Interval Singleton Type-2 Fuzzy Logic Systems. *Computational Intelligence Magazine, IEEE*(February), 38–41.

Mamdani, E. H., & Assilian, S. (1975). An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. *International Journal of Man-Machine Studies*, *7*, 1–13.

Mandal, S., & Jayaram, B. (2012). with Yager ' s Classes of Fuzzy Implications. (i), 385–394.

Mandal, S., & Jayaram, B. (2013). Monotonicity of SISO Fuzzy Relational Inference Mechanism with Yager's Class of Fuzzy Implications. *Pattern Recognition and Machine Intelligence*, 569–574.

Mantas, C. J., & Abellán, J. (2014, August). Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data. *Expert Systems with Applications*, *41*(10), 4625–4637.

McCloskey, M. E., & Glucksberg, S. (1978, July). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, *6*(4), 462–472.

Melgarejo, M., & Pena-Reyes, C. (2007, February). Implementing Interval Type-2 Fuzzy Processors. *Computational Intelligence Magazine, IEEE*, *2*(1), 63–71.

Mendel, J. M. (2000). Interval type-2 fuzzy logic systems: theory and design. *IEEE Transactions on Fuzzy Systems*, *8*(5), 535–550.

Mendel, J. M. (2003). Type-2 fuzzy sets: some questions and answers. *IEEE Neural Networks Society*(August), 10–13.

Mendel, J. M. (2004, February). Computing Derivatives in Interval Type-2 Fuzzy Logic Systems. *IEEE Transactions on Fuzzy Systems*, *12*(1), 84–98.

Mendel, J. M. (2007a, February). Computing with words and its relationships with fuzzistics. *Information Sciences*, *177*(4), 988–1006.

Mendel, J. M. (2007b, January). Computing with words: Zadeh, turing, popper and occam. *Computational Intelligence Magazine, IEEE*, *29*(4), 518–24.

Mendel, J. M. (2009, August). Enhanced Karnik–Mendel Algorithms. *IEEE Transactions on Fuzzy Systems*, *17*(4), 923–934.

Mendel, J. M., & John, R. (2002, April). Type-2 fuzzy sets made simple. *IEEE Transactions on Fuzzy Systems*, *10*(2), 117–127.

Mendel, J. M., John, R. I., & Liu, F. (2006). Interval Type-2 Fuzzy Logic Systems Made Simple. *IEEE Transactions on Fuzzy Systems*, *14*(6), 808–821.

Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). *Machine learning, neural and statistical classification*.

Mitchell, H. B. (2005, February). Pattern recognition using type-II fuzzy sets. *Information Sciences*, *170*(2-4), 409–418.

Mitchell, H. B. (2006). Ranking type-2 fuzzy numbers. *Fuzzy Systems, IEEE Transactions on*, *14*(2), 287–294.

Mizumoto, M., & Tanaka, K. (1976, August). Some properties of fuzzy sets of type 2. *Information and Control*, *31*(4), 312–340.

Mizumoto, M., & Zimmermann, H. (1982). Comparison of fuzzy reasoning methods. *Fuzzy sets and systems*, *8*(3), 253–283.

Moore, R., & Lodwick, W. (2003, April). Interval analysis and fuzzy set theory. *Fuzzy Sets and Systems*, *135*(1), 5–9.

Nguyen, H., & Kreinovich, V. (2008). Computing Degrees of Subsethood and Similarity for Interval-Valued Fuzzy Sets : Fast Algorithms. In *Proceedings of the 9th international conference on intelligent technologies* (pp. 47–55). Samui, Thailand: Citeseer.

Nomura, H., Hayashi, I., & Wakami, N. (1992). A learning method of fuzzy inference rules by descent method. In *Fuzzy systems, 1992., ieee international conference on* (pp. 203–210).

Novák, V., & Lehmke, S. (2006). Logical structure of fuzzy IF-THEN rules. *Fuzzy Sets and Systems*, *157*, 2003–2029.

Osherson, D. N., & Smith, E. E. (1981, February). On the adequacy of prototype theory as a theory of concepts. *Cognition*, *9*(1), 35–58.

Pacheco, J., Alfaro, E., Casado, S., Gámez, M., & García, N. (2012, February). A GRASP method for building classification trees. *Expert Systems with Applications*, *39*(3), 3241–3248.

Patel, a., & Mohan, B. (2002, December). Some numerical aspects of center of area defuzzification method. *Fuzzy Sets and Systems*, *132*(3), 401–409.

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238.

Poornaselvan, K., Kumar, T. G., & Vijayan, V. P. (2008). Agent Based Ground Flight Control Using Type-2 Fuzzy Logic and Hybrid Ant Colony Optimization to a Dynamic Environment. *2008 First International Conference on Emerging Trends in Engineering and Technology*, 343–348.

Rickard, J. T., Aisbett, J., & Greb, G. (2009, February). Fuzzy Subsethood for Fuzzy Sets of Type-2 and Generalized Type-n. *IEEE Transactions on Fuzzy Systems*, *17*(1), 50–60.

Ruan, D. (1993, February). Fuzzy implication operators and generalized fuzzy method of cases. *Fuzzy Sets and Systems*, *54*(1), 23–37.

Sambuc, R. (1975). *Fonctions and floues: application a l'aide au diagnostic en pathologie thyroidienne*. Faculté de Médecine de Marseille.

Seki, H., & Mizumoto, M. (2011, December). On the Equivalence Conditions of Fuzzy Inference Methods—Part 1: Basic Concept and Definition. *IEEE Transactions on Fuzzy Systems*, *19*(6), 1097–1106.

Sengupta, A., & Pal, T. K. (2000, November). On comparing interval numbers. *European Journal of Operational Research*, *127*(1), 28–43.

Smith, J. W., Everhart, J., Dickson, W., Knowler, W., & Johannes, R. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the symposium on computer applications and medical care* (pp. 261–265). IEEE Computer Society Press.

Stepnicka, M., & De Baets, B. (2013, December). Implication-based models of monotone fuzzy rule bases. *Fuzzy Sets and Systems*, *232*, 134–155.

Stepnicka, M., & Jayaram, B. (2010). On the Suitability of the Bandler–Kohout Subproduct as an Inference Mechanism. *Fuzzy Systems, IEEE Transactions on*, *18*(2), 285–298.

Street, W. N., Wolberg, W. H., & Mangasarian, O. L. (1993, July). Nuclear Feature Extraction For Breast Tumor Diagnosis. In *Is&t/spie 1993 international symposium on electronic imaging: Science and technology* (Vol. 1905, pp. 861–870).

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of system and its applications to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-15*(1), 116–132.

Turksen, I. B. (1986). Interval valued fuzzy sets based on normal forms. *Fuzzy Sets and Systems*, *20*, 191–210.

Turksen, I. B. (2002). Type 2 representation and reasoning for CWW. *Fuzzy Sets and Systems*, *127*, 17–36.

Vats, E., Lim, C. K., & Chan, C. S. (2012). A BK subproduct approach for scene classification. In *Proceedings of the iieej image electronics and visual computing workshop.* Kuching, Malaysia.

Wang, C.-H., Cheng, C.-S., & Lee, T.-T. (2004, June). Dynamical optimal training for interval type-2 fuzzy neural network (T2FNN). *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, *34*(3), 1462–77.

Wang, L., & Mendel, J. M. (1992). Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics*, *22*(6), 1414–1427.

Wang, Y., & Fan, Z. (2007, August). Fuzzy preference relations: Aggregation and weight determination. *Computers & Industrial Engineering*, *53*(1), 163–172.

Willmott, R. (1980). Two Fuzzier Implication Operators In The Theory Of Fuzzy Power Sets. *Fuzzy Sets and Systems*, *4*, 31–36.

Wu, D., & Mendel, J. M. (2007). Aggregation using the linguistic weighted average and interval type-2 fuzzy sets. *Fuzzy Systems, IEEE Transactions on*, *15*(6), 1145–1161.

Wu, D., & Mendel, J. M. (2008a). Corrections to “Aggregation Using the Linguistic Weighted Average and Interval Type-2 Fuzzy Sets”. *Fuzzy Systems, IEEE Transactions on*, *16*(6), 1664–1666.

Wu, D., & Mendel, J. M. (2008b, January). A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets. *Information Sciences*, *178*(2), 381–402.

Wu, D., & Mendel, J. M. (2009, March). A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets. *Information Sciences*, *179*(8), 1169–1192.

Wu, D., & Mendel, J. M. (2010, June). Computing with words for hierarchical decision making applied to evaluating a weapon system. *Fuzzy Systems, IEEE Transactions on*, *18*(3), 441–460.

Wu, T. P., & Chen, S. M. (1999, January). A new method for constructing membership functions and fuzzy rules from training examples. *IEEE transactions on systems, man, and cybernetics. Part B*, *29*(1), 25–40.

Xing, H.-J., & Ha, M.-H. (2014, May). Further improvements in Feature-Weighted Fuzzy C-Means. *Information Sciences*, *267*, 1–15.

Xu, Y., Kerre, E. E., Ruan, D., & Song, Z. (2001, April). Fuzzy reasoning based on the extension principle. *International Journal of Intelligent Systems*, *16*(4), 469–495.

Yager, R. R. (1980). An approach to inference in approximate reasoning. *International Journal of Man-Machine Studies*(July 1979), 323–338.

Yang, M.-S., & Lin, D.-C. (2009, March). On similarity and inclusion measures between type-2 fuzzy sets with an application to clustering. *Computers & Mathematics with Applications*, *57*(6), 896–907.

Yeh, C.-Y., Su, W.-P., & Lee, S.-J. (2013, July). An efficient multiple-kernel learning for pattern classification. *Expert Systems with Applications*, *40*(9), 3491–3499.

Yew, K. M., & Kohout, L. J. (1996a). Classes of Iinterval-valued Fuzzy Relational Inference Templates. *Malaysian Journal of Computer Sciences*, *9*(1), 42–48.

Yew, K. M., & Kohout, L. J. (1996b). Interval-valued Fuzzy Relational Inference Structures. In *Proceedings iasted/ismm international conference on intelligent information management systems* (pp. 173–175). Washington DC.

Yew, K. M., & Kohout, L. J. (1997). Interval-based reasoning in medical diagnosis. In *Proceedings intelligent information systems. iis'97* (pp. 32–36). IEEE Comput. Soc.

Zadeh, L. A. (1965, January). Fuzzy Sets. *Information and Control*, *8*(1), 338.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *Systems, Man and Cybernetics, IEEE Transactions on*(1), 28–44.

Zadeh, L. A. (1975a). *Calculus of fuzzy restrictions*. Electronics Research Laboratory, University of California.

Zadeh, L. A. (1975b). The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences*, *8*(3), 199–249.

Zadeh, L. A. (1975c). The concept of a linguistic variable and its application to approximate reasoning-III. *Information Sciences*, *9*(1), 43–80.

Zadeh, L. A. (1975d). The concept of a linguistic variable and its application to approximate reasoning–II. *Information sciences*, *8*(4), 301–357.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, *1*, 3–28.

Zadeh, L. A. (1982, July). A note on prototype theory and fuzzy sets. *Cognition*, *12*, 291–297.

Zadeh, L. A. (1996, May). Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems*, *4*(2), 103–111.

Zadeh, L. A. (1997, September). Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*, *90*(2), 111–127.

Zadeh, L. A. (1999). From Computing with Numbers to Computing with Words — From Manipulation of Measurements to Manipulation of Perceptions. *IEEE Transactions on Circuits and Systems*, *46*(1), 105–119.

Zadeh, L. A. (2011, July). A Note on Z-numbers. *Information Sciences*, *181*(14), 2923–2932.

Zaher, M., & Hagras, H. (2010, November). Data generated type-2 fuzzy logic model for control of wind turbines. *2010 10th International Conference on Intelligent Systems Design and Applications*, 80–85.

Zhang, F., Joshua, I., & Lim, C. (2011). Comparison between interval type-2 fuzzy sets based on a fuzzy preference relation with probability degree. In *Machine learning and cybernetics (icmlc), 2011 international conference on* (Vol. 1, pp. 215–219). IEEE.

Zheng, G., Xiao, J., Zhang, Y., & Shi, G. (2010). An inclusion measure between general type-2 fuzzy sets. In *Fuzzy systems and knowledge discovery (fskd), 2010 seventh international conference on* (Vol. 1, pp. 99–103). IEEE.