FORMULATIONS OF SHADOWED SETS AND THREE-WAY APPROXIMATIONS OF FUZZY SETS

A Thesis
Submitted to the Faculty of Graduate Studies and Research
In Partial Fulfilment of the Requirements
For the Degree of
Masters of Science
In
Computer Science
University of Regina

By
Shu Wang
Regina, Saskatchewan
January, 2016

Copyright 2016: Shu Wang
Shu Wang, candidate for the degree of Master of Science in Computer Science, has presented a thesis titled, *Formulations of Shadowed Sets and Three-Way Approximations of Fuzzy Sets*, in an oral examination held on December 17, 2015. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner: Dr. Taehan Bae, Department of Mathematics & Statistics

Supervisor: Dr. Yiyu Yao, Department of Computer Science

Committee Member: Dr. Lisa Fan, Department of Computer Science

Committee Member: L. Saxton, Department of Computer Science

Chair of Defense: Dr. Hairuo Qing, Department of Geology
ABSTRACT

A three-way, three-valued, or three-region approximation of a fuzzy set is constructed from a pair of thresholds \((\alpha, \beta)\) with \(0 \leq \beta < \alpha \leq 1\) on its membership function. An element whose membership grade equals to or is greater than \(\alpha\) is put into the positive region, an element whose membership grade equals to or is less than \(\beta\) is put into the negative region, and an element whose membership grade is between \(\beta\) and \(\alpha\) is put into the boundary region. A fundamental issue in constructing such three-way approximations is the interpretation and determination of a pair of thresholds on the unit interval \([0, 1]\). Shadowed sets, proposed by Pedrycz, are an example of three-way approximations of fuzzy sets, in which one obtains an analytic solution of the thresholds by searching for a balance of uncertainty introduced by the three regions.

In this thesis, we introduce a general framework for determining the thresholds by optimizing an objective function. Within the framework, we critically review existing studies and present new formulations according to three principles, i.e., a principle of uncertainty invariance, a principle of minimum distance, and a principle of least cost. When applying the principle of uncertainty invariance, we maintain the uncertainty of a fuzzy set in a three-valued set. When applying the principle of minimum distance, we compute thresholds by minimizing the distance between a fuzzy set and a three-valued set. When applying the principle of minimum cost, we compute the thresholds by minimizing the costs of three-way approximations of a fuzzy set.

In our model, membership grades are mapped to three points, denoted as \(n\),
m and p. Membership grades of objects in the positive region are mapped to p, membership grades of the objects in the boundary region are mapped to m, and membership grades of the objects in negative region are mapped to n. In our new formulation, \{n, m, p\} is a different system. The set \{n, m, p\} denotes points that do not necessarily belong to the unit interval [0, 1]. We introduce a distance function between elements of [0, 1] and elements of \{n, m, p\} and study properties that should be satisfied. The problem of finding a pair of thresholds is transformed into a problem of minimizing the cost of a three-valued set with respect to a distance function.
ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to my advisor Dr. Yiyu Yao for the continuous support of my study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of my master research and writing of this thesis. I could not have imagined having a better advisor and mentor for my master’s study.

I deeply appreciate the financial support from my supervisor Dr. Yiyu Yao the Faculty of Graduate Studies and Research, and the Department of Computer Science at the University of Regina, which helped to satisfy my daily expenses and kept me concentrated in my studies and research.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Lisa Fan and Dr. Lawrence Saxton, for their insightful comments and encouragement, but also for the hard question which help me to widen my research from various perspectives.

Special thanks go to all my colleagues and friends in my office for all the supports and constructive suggestions on my research, especially, Xiaofei Deng, Mengjun Hu, Cong Gao, Kuifei Liu who provided invaluable insights, constant encouragements and sincere friendship.

Last but not the least, I would like to thank my family: my parents, my aunts, and my brothers and sister for supporting me spiritually throughout writing this thesis and my life in general.
ACKNOWLEDGEMENT

I would like to give special thanks to my supervisor, Dr. Yiyu Yao, for his invaluable suggestions throughout all stages of my studies. I would like to thank my external examiner and my internal committee members, Dr. Taehan Bae, Dr. Lawrence Saxton, and Dr. Lisa Fan for their constructive suggestions and comments. I also would like to thank the defense chair, Dr. Hairuo Qing, for overseeing the defense agenda. Further appreciation is extended to the Faculty of Graduate Studies and Research (FGSR) for fundings and all kinds of opportunities.
# TABLE OF CONTENTS

ABSTRACT i

ACKNOWLEDGEMENT iii

TABLE OF CONTENTS v

LIST OF FIGURES viii

1 INTRODUCTION 1

1.1 An Overview of Three-Way Decisions .......................... 1

1.2 The Problem of Three-Way Approximations of Fuzzy Sets .... 3

1.3 Motivation and an Outline of the Proposed Framework ....... 5

1.4 Contributions of the Thesis .................................. 7

1.5 Thesis Structure ............................................. 8

2 A BRIEF INTRODUCTION TO FUZZY SETS AND THREE-WAY DECISIONS 9

2.1 Fuzzy Sets ................................................... 9

2.2 Three-Way Decisions ......................................... 11

2.2.1 Pawlak Rough Set Approximations .......................... 11

2.2.2 Decision-Theoretic Rough Sets ............................. 13

2.3 Three-Way Approximations of Fuzzy Sets ...................... 16
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.1 Qualitative Three-Way Characterization</td>
<td>17</td>
</tr>
<tr>
<td>2.3.2 Zadeh’s Quantitative Three-way Approximations</td>
<td>18</td>
</tr>
<tr>
<td>2.3.3 Three-Valued Sets as Approximations of Fuzzy Sets</td>
<td>19</td>
</tr>
<tr>
<td>2.3.4 Construction of Three-Way Approximations by Optimization</td>
<td>21</td>
</tr>
<tr>
<td>3 AN ADDITIVE MEASURE AND FORMULATIONS BASED ON THREE PRINCIPLES</td>
<td>25</td>
</tr>
<tr>
<td>3.1 A Decision-Theoretic Optimization Framework Based on Additive Measures</td>
<td>25</td>
</tr>
<tr>
<td>3.2 Shadowed Sets According to the Principle of Uncertainty Invariance</td>
<td>27</td>
</tr>
<tr>
<td>3.2.1 Principle of Uncertainty Invariance</td>
<td>27</td>
</tr>
<tr>
<td>3.2.2 Pedrycz’s Formulation</td>
<td>28</td>
</tr>
<tr>
<td>3.2.3 Tahayori, Sadeghian and Pedrycz’s Reformulation</td>
<td>31</td>
</tr>
<tr>
<td>3.3 Three-Way Approximations of Fuzzy Sets According to the Principle of Minimum Distance</td>
<td>34</td>
</tr>
<tr>
<td>3.3.1 Principle of Minimum Distance</td>
<td>34</td>
</tr>
<tr>
<td>3.3.2 Formulations Based on Minkowski Distance</td>
<td>36</td>
</tr>
<tr>
<td>3.4 Three-Way Approximations of Fuzzy Sets According to the Principle of Least Cost</td>
<td>42</td>
</tr>
<tr>
<td>3.4.1 Principle of Least Cost</td>
<td>42</td>
</tr>
<tr>
<td>3.4.2 Decision-Theoretic Three-Way Approximations of Fuzzy Sets</td>
<td>44</td>
</tr>
<tr>
<td>4 A MORE GENERAL DISTANCE-BASED FORMULATION</td>
<td>50</td>
</tr>
<tr>
<td>4.1 Semantic Distance</td>
<td>50</td>
</tr>
<tr>
<td>4.2 Construction of Three-Way Approximations</td>
<td>53</td>
</tr>
<tr>
<td>5 CONCLUSIONS AND FUTURE WORK</td>
<td>60</td>
</tr>
<tr>
<td>5.1 Summary</td>
<td>60</td>
</tr>
<tr>
<td>5.2 Future Work</td>
<td>61</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Summary of existing studies and main works of the thesis</td>
<td>5</td>
</tr>
<tr>
<td>3.1</td>
<td>A shadowed set approximation of a fuzzy set</td>
<td>29</td>
</tr>
<tr>
<td>3.2</td>
<td>Tahayori, Sadeghian and Pedrycz’s reformulation of shadowed sets</td>
<td>34</td>
</tr>
<tr>
<td>3.3</td>
<td>A three-way approximation of a fuzzy set</td>
<td>42</td>
</tr>
<tr>
<td>4.1</td>
<td>Three-way approximations of a fuzzy set by using three actions</td>
<td>51</td>
</tr>
<tr>
<td>4.2</td>
<td>Semantic distance between elements in $[0, 1]$ and ${n, m, p}$</td>
<td>55</td>
</tr>
</tbody>
</table>
Chapter 1

INTRODUCTION

In this chapter, we give an overview of three-way decisions and state the problem of three-way approximations of fuzzy sets. We discuss motivations of the thesis, and the main contributions of the thesis, and the structure of the thesis.

1.1 An Overview of Three-Way Decisions

Three-way decisions can be treated as ternary classifications based on whether an object satisfies a set of criterions [41]. Suppose $U$ is a finite nonempty set of objects and $C$ is a finite set of conditions. Each condition in $C$ may be a criterion, an objective, or a constraint. For simplicity, we refer to conditions in $C$ as criteria. Our decision task is to classify objects of $U$ according to whether they satisfy the set of criteria [41]. A theory of three-way decisions is constructed based on the notions of acceptance, rejection and non-commitment. If we can infer with high degree of confidence that an object satisfies the set of criteria, we make an acceptance decision; if we can infer with high confidence that an object does not satisfy the set of criteria, we make a rejection decision; if we cannot make a confident inference of either one, we make a non-commitment decision. A three-way decision model is an extension of the commonly used binary-decision model with an added third option, which defers a
definite decision of acceptance or rejection. For a non-commitment decision, we look for further information to help us make a definite decision.

We use the strategy of a deferred decision very often in our daily life. For example, if we want to learn new knowledge about machine learning, we check the library to find some books. By looking at the titles of books, we can make a three-way decision: borrow these books that seem to be very relevant, do not borrow books that seem to be irrelevant, and keep a waiting list of books that are somewhat relevant for the future use. In doing so, we can borrow books in the list at later time when we cannot find sufficient information in the borrowed books.

We describe the basic idea of three-way decision by using this example. With respect to a book, three-way decisions have the following three actions:

- borrow the books,
- do not borrow the books, and
- put it on the list, we could read it later

Considering possible states of the book regarding the content, we simplify the evaluation step by giving the following $3 \times 2$ cost matrix:

<table>
<thead>
<tr>
<th>Action</th>
<th>Machine learning</th>
<th>Not about machine learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Borrow</td>
<td>no cost</td>
<td>cost of time of reading</td>
</tr>
<tr>
<td>Do not borrow</td>
<td>cost of missing a book</td>
<td>no cost</td>
</tr>
<tr>
<td>Checklist</td>
<td>cost of time checking</td>
<td>cost of time checking and reading</td>
</tr>
</tbody>
</table>

Table 1.1: Example of three-way decisions

Whenever there is a high degree of uncertainty about the state of the book, we can call for a deferred decision. Suppose we think that the book might has several chapters talk about machine learning. We can delay making a definite decision, which
is borrowing or ignoring, and put it on a waiting list to read later. The example demonstrates the case where three-way decisions show more benefit than two-way decisions. However, it should be pointed out that three-way decisions may not work well in all cases. It will not perform well or may fail in some cases.

Three-way decisions are related to many other studies, such as interval sets, shadowed sets and fuzzy sets. This thesis mainly focuses on shadowed sets, fuzzy sets and reviews the existing studies about three-way approximations on fuzzy sets.

1.2 The Problem of Three-Way Approximations of Fuzzy Sets

A fuzzy set, proposed by Zadeh [43], assigns objects with membership grades in the unit interval $[0, 1]$ so that one can describe a concept with an unsharp, gradually changing boundary. There are advantages and shortcomings of using a membership grade in the unit interval $[0, 1]$. On one hand, one can distinguish objects in details by using many levels of membership grade. On the other hand, since there are an infinite number of values, it is difficult to interpret and understand a fuzzy membership in practice. There are studies showing that human can process 4 units of information [11]. In many applications, it may happen that we do not need to distinguish objects that are very similar to each other, and a very precise membership value may not offer much more significant information. Therefore, an approximation may be sufficient. When approximating a fuzzy set, estimating the uncertainty and representing it with numeric values may add errors into the approximation and costs associate with the errors. There is a trade-off between the accuracy and cost in approximating a fuzzy set, sometimes costs may be high despite the small error when obtaining a precise numeric value. Accordingly, approximating a fuzzy set by using a few levels is of great importance. In some situations, more often or not, it is sufficient
to approximate a fuzzy set within a few levels [14].

Zadeh [43] first briefly discussed three-way approximations of a fuzzy set through a pair of thresholds and their connection to three-valued logics. Many authors investigated similar three-way approximations. For example, Yao [38] used a pair of an $\alpha$-cut and a $\beta$-cut to approximate a fuzzy set in formulating interval sets, where an $\alpha$-cut of a fuzzy set is the set of elements whose membership grades are at or above $\alpha$. Based on the notion of rough fuzzy sets proposed by Dubois and Prade [15, 16], Banerjee and Pal [2] introduced a three-way approximation by first constructing a rough fuzzy set of a fuzzy set and then computing a pair of an $\alpha$-cut of the lower approximation and a $\beta$-cut of upper approximation of the rough fuzzy set. These studies take the pair of thresholds as a primitive notion and have not investigated the crucial issue regarding the interpretation and computation of the required thresholds.

Pedrycz [29, 30] proposed the notion of shadowed sets as three-way approximations of fuzzy sets. One of the main contributions is a conceptual framework interpreting the required pair of thresholds and a practical method for computing the thresholds through optimizing an objective function. Recently, Tahayori, Sadeghian and Pedrycz [36] provided an alternative formulation of shadowed sets based on a modified objective function that is semantically superior. In addition, they obtained an analytic solution. One can consider further improvements and extensions by paying attention to two restrictions introduced by shadowed sets. One restriction is the use of the same shadowed set of a fuzzy set independent of particular applications. The other restriction is the requirement that the pair of thresholds $(\alpha, \beta)$ must satisfy the conditions $\alpha + \beta = 1$ and $\beta < \alpha$, which implies $\alpha > 0.5$. In contrast to decision-theoretic rough sets in which the pair of thresholds can be interpreted by classification errors, there is a difficulty in interpreting thresholds in three-valued approximations of a fuzzy sets. The introduction of a shadowed set induced by a fuzzy set attempts to address this problem [29].
In attempt to avoid the two restrictions, Deng and Yao [14] introduced a decision-theoretic formulation of three-way approximations of fuzzy sets. Their formulation is able to produce different three-way approximations of the same fuzzy set in different applications. Although the formulation removes the condition $\alpha + \beta = 1$, it still requires that $\alpha > 0.5$. A new problem introduced in their formulation is assigning a grade value of 0.5 to elements in the shadowed region, a representation suggested by Cattaneo and Ciucci [4]. This representation is also implicitly used in the formulation by Tahayori, Sadeghian and Pedrycz [36].

In Figure 1.1, we summarize the existing studies of three-way approximations of fuzzy sets and introduce the main work of this thesis which are denoted by *.

![Figure 1.1: Summary of existing studies and main works of the thesis](image)

### 1.3 Motivation and an Outline of the Proposed Framework

In three-way approximations, with respect to two thresholds, we divide a universal set into three regions, which are the positive region, negative region and boundary
region. In existing studies, the objects’ membership grades greater than or equal to \(\alpha\) to 1 are elevated to 1, and the objects’ membership grades lower than or equal to \(\beta\) to 0 are reduced to 0. The examination of existing studies calls for new research questions. Instead of using 1 and 0 to represent two extreme membership grade of a fuzzy sets, we use \(p\) and \(n\) to represent 1 and 0, respectively. We use \(m\) to represent the membership grades of objects in the boundary region. We can immediately get a new three-way approximation based on the an understanding of \(\{n, m, p\}\) as follows:

\[
T_{(\mu_A)}(x) = \begin{cases} 
  p, & \mu_A(x) \geq \alpha, \\
  m, & \beta < \mu_A(x) < \alpha, \\
  n, & \mu_A(x) \leq \beta.
\end{cases}
\]

We can treat three-way approximations of fuzzy sets as a transformation from the system defined by unit interval \([0, 1]\) to a new system defined by \(\{n, m, p\}\). This transformation immediately provide us with an optimization-based framework to interpret and determine a pair of thresholds for constructing a three-way approximation.

We can interpret the thresholds based on three principles, i.e., a principle of uncertainty invariance [21,22], a principle of minimum distance [3,27], and a principle of least cost [14]. The three principles have been used implicitly in existing studies. We represent a decision-theoretic optimization framework based on additive measures by introducing a measure \(Q\) to quantity the degree of discrepancy between a three-way approximation and the original fuzzy set. In this thesis, we consider a component-wise defined additive discrepancy measure. The total discrepancy can be measured by take a sum of each component discrepancy. The problem of finding a pair of thresholds to get the minimum total discrepancy is transformed to a problem to find a pair of thresholds to minimize each component discrepancy. Three rules are introduced to make a decision, if there is a tie, a tie-breaking rule can be used. We examine the framework with three principles under three existing formulations.
An examination of questions raised by existing formulations enables us to introduce a new formulation. In existing studies, there is a restrictions that \( \{n, m, p\} \) denotes specific points in the unit interval \([0,1]\). They in fact consider a subsystem of the unit interval \([0,1]\). In attempt to avoid the restriction, we introduce a different system, that is, elements in \( \{n, m, p\} \) do not necessarily belong to the unit interval. A three-way decision can be constructed, and a distance function is needed to evaluate the quality of three-way approximations. Therefore, we introduce a distance function called the semantic distance and study properties of the distance function. A new formulation based on the semantic distance is proposed and a pair of thresholds is determined that leads to a three-way approximation of a fuzzy set with least cost.

### 1.4 Contributions of the Thesis

The three main contributions of this thesis can be summarized as follows:

**Introduction of an optimization-based framework.** In three-way approximations, the fundamental problem is to interpret and determine the pair of thresholds. In this thesis, we propose an optimization-based framework to solve this problem. We consider a class of component-wise defined additive discrepancy measures and find a pair of thresholds which can lead to a minimum value of each component. In order to get the minimum value of each component, we introduce three rules to help us make decisions. If there is a tie, we have a tie-breaking rule can be used to help us make a decision.

**Applications of three principles.** We have identified three principles that can be used to get the pair of thresholds, namely, a principle of uncertainty invariance, a principle of minimum distance and a principle of minimum costs. Three principles lead to the formulations of shadowed sets, distance-based three-way approximations of fuzzy sets, and cost-sensitive approach to three-way approximations of fuzzy sets.
Introduction of a more general formulation based on semantic distances.

From reviewing the existing studies, we find out that $n$, $m$ and $p$ are always treated as three specific values in unit interval $[0, 1]$. We want to treat $\{n, m, p\}$ as a different system from the unit interval $[0, 1]$. An object with the membership grade equals to or is greater than $\alpha$ is put in the positive region and we map its membership grade to $p$; an object with the membership grade equals to or is less than $\beta$ is put in the negative region and we map its membership grade to $n$; other objects are put in the boundary region and we map their membership grade to $m$. To evaluate the three-way approximation, we introduce a semantic distance function between the unit interval $[0, 1]$ and $\{n, m, p\}$. The properties of the semantic distance function are studies. We construct a three-way approximation by a pair of thresholds that minimize the cost of mapping each object to three-valued sets.

1.5 Thesis Structure

The remaining parts of this thesis are organized as follows: Chapter 2 reviews the basic concept about fuzzy sets and three-way decisions, as well as background knowledge about decision-theoretic rough sets. An optimization-based framework for constructing three-way approximations of fuzzy sets is also introduced in Chapter 2. In Chapter 3, a decision-theoretic optimization framework based on additive measures is introduced. Within the framework, we review the existing studies based on three principles and identify their possible improvements and extensions. In Chapter 4, we introduce a definition of semantic distance and propose a more general framework. We prove that there exist a pair of two thresholds that allows us to construct three-way approximations of fuzzy sets. Chapter 5 concludes the thesis and discusses possible future work.
Chapter 2

A BRIEF INTRODUCTION TO FUZZY SETS AND THREE-WAY DECISIONS

This chapter introduces the basic notions and concepts of fuzzy sets and three-way decisions. In addition, we introduce the construction of three-way approximations of fuzzy sets, which provides background of our research.

2.1 Fuzzy Sets

Fuzzy sets allow us to represent vague concepts expressed in natural language. The representation depends not only on the concept, but also on the context in which it is used. For example, high temperature in winter can be different concept from high temperature in summer.

A set is defined by a function, usually called a characteristic function, that declares which elements of the universal set $U$ are members of the set and which are not. In fuzzy sets, this function can be generalized such that the values assigned to the elements of the universal set fall within the unit interval $[0, 1]$, each member-
ship function maps elements of a given universal set $U$ into real numbers in $[0, 1]$. Larger values denote higher degrees of set membership. Such a function is called a membership function. We can have fuzzy sets denoted as following:

$$\mu_A : U \rightarrow [0, 1],$$  \hspace{1cm} (2.1)

where $\mu_A(x)$ denotes the membership grade assigned to each element.

One of the most important concepts of fuzzy sets is the concept of an $\alpha$-cut, which is denoted as $^\alpha \mu_A$. Given a fuzzy set $\mu_A$ defined on $U$ and any number $\alpha \in [0, 1]$, the $^\alpha \mu_A$ is the crisp set:

$$^\alpha \mu_A = \{ x | \mu_A(x) \geq \alpha \},$$

$$^{\alpha+} \mu_A = \{ x | \mu_A(x) > \alpha \}. \hspace{1cm} (2.2)$$

That is, the $^\alpha \mu_A$ of a fuzzy set $\mu_A$ is the crisp set that contains all the elements of the universal set $U$ whose membership grades in $\mu_A$ are greater than or equal to the specified value of $\alpha$, and $^{\alpha+} \mu_A$ is often called strong $\alpha$-cut.

In fuzzy sets, when $\mu_A$ is a fuzzy set and $x$ is a relevant object, the proposition "$x$ is a member of" is not necessarily either true or false, as required by two-valued logic, but is maybe true only to some degree, and it can be true only to other degree, the degree to which $x$ is actually a member of $\mu_A$. This gives us a good reason to apply three-way decision on fuzzy sets.

A three-way approximation of a fuzzy set is constructed by transforming fuzzy membership grades in the unit interval $[0, 1]$ into three values based on a pair of thresholds $(\alpha, \beta)$ with $0 \leq \beta < \alpha \leq 1$. The three values are obtained by dividing $[0, 1]$ into three parts, namely, grades at or above $\alpha$, grades at or below $\beta$ and grades between $\beta$ and $\alpha$. As a result, one divides the universal set into three pair-wise disjoint sets called three regions. This process can be interpreted in the framework
of three-way decisions proposed by Yao [41].

Suppose that a fuzzy set models a concept with unsharp boundary [15, 23]. One region represents the set of elements with membership grades close to 1 and these elements are accepted to be instances of the concept modeled by the fuzzy set; another region represents the set of elements with membership grades close to 0 and these elements are rejected to be instances of the concept modeled by the fuzzy set; the third region represents the set of elements that are neither close to 1 nor close to 0 and these elements are neither accepted nor rejected to be instances of the concept modeled by the fuzzy set. Therefore, we can get the support and core of a fuzzy set.

\[
\text{Support}(\mu_A) = 0^+\mu_A = \{x|\mu_A(x) > 0\}
\]
\[
\text{Core}(\mu_A) = 1\mu_A = \{x|\mu_A(x) \geq 1\}. \quad (2.3)
\]

The support of a fuzzy set \(\mu_A\) within a universal set \(U\) is the crisp set that contains all the elements of \(U\) that have positive membership grades in \(\mu_A\). The support can be treated as a strong \(\alpha\)-cut with \(\alpha = 0\). And the 1-cut, is often called the core of \(\mu_A\).

### 2.2 Three-Way Decisions

The essential ideas of three-way decisions are described in terms of a ternary classification according to evaluations of a set of criteria.

#### 2.2.1 Pawlak Rough Set Approximations

In Pawlak rough set theory, let \(U\) denotes a finite non-empty universal set of objects, \(E \subseteq U \times U\) is a binary relation defined on \(U\). \(E\) is reflexive (\(\forall x \in U, xEx\)), symmetric (\(\forall x, y \in U, xEy \rightarrow yEx\)) and transitive (\(\forall x, y, z \in U, xEy \wedge yEz \rightarrow xEz\)). The
object $x$ which satisfies $[x]_E = [x] = \{ y \in U | xEy \}$ belongs to the same equivalence class. Pawlak define rough set approximations by using a pair of lower and upper approximations:

$$\text{apr}(X) = \{ x \in U | [x] \subseteq U \},$$
$$\overline{\text{apr}}(X) = \{ x \in U | [x] \cap X \neq \emptyset \}$$

The lower approximation $\text{apr}(X)$ contains objects that must be an instance of the set, and the upper approximation $\overline{\text{apr}}(X)$ contains objects that may be an instance of the set.

By definition, it follows that $\text{apr}(X) \subseteq X \subseteq \overline{\text{apr}}(X)$, which contributes to the fact that $X$ lies between its lower and upper approximations. Therefore, a pair of lower and upper approximations of $X$ divides object in universal set $U$ into three region represented by follows:

$$\text{POS}(X) = \text{apr}(X),$$
$$\text{NEG}(X) = U - \overline{\text{apr}}(X),$$
$$\text{BND}(X) = \overline{\text{apr}}(X) - \text{apr}(X)$$

$$= (\text{POS}(X) \cup \text{NEG}(X))^c. \tag{2.5}$$

By definition, the three regions are pair-wise disjoint and their union is the universal set $U$. Since some of the regions may be empty, the three regions do not necessarily form a partition of the universe. The positive region $\text{POS}(X)$ is defined by the lower approximation, the negative region $\text{NEG}(X)$ is defined by complement of the upper approximation, and the set $\{ \overline{\text{apr}}(X) - \text{apr}(X) \}$ define the boundary region. There is another way to define three region instead of using lower and upper
approximations, which is shown as follows:

\[
\begin{align*}
\text{POS}(X) & = \{ x \in U | [x] \subseteq X \}, \\
\text{NEG}(X) & = \{ x \in U | [x] \subseteq X^c \} \\
& = \{ x \in U | [x] \cup X = \emptyset \} \\
\text{BND}(X) & = (\text{POS}(X) \cup \text{NEG}(X))^c \\
& = \{ x \in U | \neg([x] \subseteq X^c) \land \neg([x] \subseteq X) \} \\
& = \{ x \in U | [x] \cap X^c \neq \emptyset \land [x] \cap X \neq \emptyset \}. 
\end{align*}
\]

(2.6)

Equivalently,

\[
\begin{align*}
apr(X) & = \text{POS}(X), \\
\overline{apr}(X) & = \text{POS}(X) \cup \text{NEG}(X). 
\end{align*}
\]

(2.7)

We can get the two representations are equivalent and one can either use the pair of lower and upper approximations or the three regions to approximate a set.

2.2.2 Decision-Theoretic Rough Sets

Decision-theoretic rough sets (DTRS) is a probabilistic extension of rough set classification. First created in 1990 by Yao [42], the extension makes use of loss functions to derive \(\alpha\) and \(\beta\) region parameters. Like rough sets, the lower and upper approximations of a set are used.

Using the Bayesian decision procedure, the DTRS approach allows for minimum-risk decision making based on observed evidence. Let \(X \subseteq U\) denote the concept, \(\Omega = \{X, X^c\}\) denote the two states of an object \(x \in U\) and \(\text{Actions} = \{a_n, a_m, a_p\}\) denote the three actions which leads to the three decisions, that is, acceptance, rejection and non-commitment. In decision-theoretic aspect, we assume that each action associates
with a certain loss or cost. We have Table 2.1 represent a loss function of three-way decisions. In the table, $\lambda_{NP}$, $\lambda_{BP}$, and $\lambda_{PP}$ denote the loss of taking actions $a_n$, $a_m$, and $a_p$, if $x$ is an instance of $X$, respectively. Equivalently, $\lambda_{NN}$, $\lambda_{BN}$, and $\lambda_{PN}$, denote the loss of taking actions $a_n$, $a_m$, and $a_p$, if $x$ is an non-instance of $X$, respectively.

<table>
<thead>
<tr>
<th>Action</th>
<th>$x \in X$</th>
<th>$x \in X^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_p$</td>
<td>$\lambda_{PP} = \lambda(a_p</td>
<td>X)$</td>
</tr>
<tr>
<td>$a_n$</td>
<td>$\lambda_{NP} = \lambda(a_n</td>
<td>X)$</td>
</tr>
<tr>
<td>$a_m$</td>
<td>$\lambda_{BP} = \lambda(a_m</td>
<td>X)$</td>
</tr>
</tbody>
</table>

Table 2.1: Loss function of three-way decisions

Suppose $a_i \in Actions$ and $w_j \in \Omega$. For an equivalent class $[x]$, the expected loss of taking an action $a_i$ is defined by:

$$R(a_i|[x]) = \sum_{w_j \in \Omega} \lambda(a_i|w_j) Pr(w_j|[x]), \quad (2.8)$$

where $\lambda(a_i|w_j)$ denotes the loss of taking action $a_i$ given the state $w_j$, and $Pr(w_j|[x])$ is the probability of state $w_j$ given $[x]$. We can get the expected risk of three actions as follows:

$$R(a_p|[x]) = \lambda_{PP} Pr(X|[x]) + \lambda_{PN} Pr(X^c|[x]),$$
$$R(a_n|[x]) = \lambda_{NP} Pr(X|[x]) + \lambda_{NN} Pr(X^c|[x]),$$
$$R(a_m|[x]) = \lambda_{BP} Pr(X|[x]) + \lambda_{BN} Pr(X^c|[x]). \quad (2.9)$$

The overall cost is given by:

$$R = \sum_{x \in U} R(a_i|[x]). \quad (2.10)$$

To minimize the overall risk, we can choose an action that minimizes the conditional
risk in Equation 2.9. Therefore, we use the following rules:

(P) \text{If } R(a_p[x]) \leq R(a_n[x]) \text{ and } R(a_p[x]) \leq R(a_m[x]), \text{ then take action } a_p,

(M) \text{If } R(a_m[x]) < R(a_n[x]) \text{ and } R(a_m[x]) < R(a_p[x]), \text{ then take action } a_m,

(N) \text{If } R(a_n[x]) \leq R(a_m[x]) \text{ and } R(a_n[x]) < R(a_p[x]), \text{ then take action } a_n.

When there is a tie, a tie-breaking rule can be used to select one of the action. In this thesis, we break ties by using the order of $a_n$, $a_m$, and $a_p$. To simplify these rules, we need to consider some properties of costs. Actions $a_n$ and $a_p$ will remove all uncertainty in the original set and $a_m$ will still retain some uncertainty. Therefore, we assume:

(i) \quad \lambda_{PP} > 0, \quad \lambda_{BP} > 0, \quad \lambda_{NP} > 0;
\quad \lambda_{NN} > 0, \quad \lambda_{BN} > 0, \quad \lambda_{PN} > 0;

(ii) \quad \lambda_{PP} < \lambda_{BP} < \lambda_{NP};

(iii) \quad \lambda_{NN} < \lambda_{BN} < \lambda_{PN}.

Properties in (i) maybe interpreted as normalization conditions. They simply suggest that cost are positive. Properties (ii) states that for a positive object, the cost of making a boundary decision should be greater than making an acceptance decision but be smaller than a rejection decision. Property (iii) similarly states that for a negative object, the cost of making a boundary decision should be greater than making a rejection decision but be smaller than an acceptance decision. In addition, we assume there is no conflict between acceptance and rejection decisions. To do this, we make
further assumption:

\[
\frac{\lambda_{NP} - \lambda_{BP}}{\lambda_{BN} - \lambda_{NN}} > \frac{\lambda_{BP} - \lambda_{PP}}{\lambda_{PN} - \lambda_{BN}} \tag{2.11}
\]

We can simplify the rules into:

(P) \quad \text{If } Pr(X|\{x\}) \geq \alpha, \text{ \ then \ take \ action } a_p,

(M) \quad \text{If } \beta < Pr(X|\{x\}) < \alpha, \text{ \ then \ take \ action } a_m,

(N) \quad \text{If } Pr(X|\{x\}) \leq \beta, \text{ \ then \ take \ action } a_n,

where \(\alpha\) and \(\beta\) are given by:

\[
\alpha = \frac{\lambda_{PN} - \lambda_{BN}}{\lambda_{PN} - \lambda_{BN} + \lambda_{BP} - \lambda_{PP}} = (1 + \frac{\lambda_{BP} - \lambda_{PP}}{\lambda_{PN} - \lambda_{BN}})^{-1},
\]

\[
\beta = \frac{\lambda_{BN} - \lambda_{NN}}{\lambda_{BN} - \lambda_{NN} + \lambda_{NP} - \lambda_{BP}} = (1 + \frac{\lambda_{NP} - \lambda_{BP}}{\lambda_{BN} - \lambda_{NN}})^{-1}. \tag{2.12}
\]

From the assumption, it is not difficult to verify that \(0 < \beta < \alpha < 1\).

### 2.3 Three-Way Approximations of Fuzzy Sets

The notions of support and core provide a qualitative three way characterization of a fuzzy set by using the two extreme membership grades 1 and 0. By relaxing \((1, 0)\) into a pair of thresholds \((\alpha, \beta)\) with \(0 \leq \beta < \alpha \leq 1\), one obtains quantitative approximations of a fuzzy set. We give an optimization-based framework for interpreting and determining the required thresholds according to three principles.
2.3.1 Qualitative Three-Way Characterization

A fuzzy set represents a concept with an unsharp and gradually changing boundary. A fuzzy set over a universal set $U$ is defined by a membership function:

$$\mu_A : U \rightarrow [0, 1],$$

(2.13)

where $[0, 1]$ is the unit interval. The numeric value $\mu_A(x) \in [0, 1]$ is called the membership grade of $x \in U$ in the fuzzy set $A$. Intuitively, an object with a full membership grade of 1 is viewed as a typical instance of the concept, an object with a membership grade of 0 is viewed as a non-instance of the concept, and a membership grade between 0 and 1 denotes the degree to which an object belongs to the concept. An element with a larger grade indicates that it has a higher degree of being an instance of the concept.

Qualitatively speaking, with respect to a fuzzy set, we can classify elements in $U$ based on full, zero and partial (i.e., neither 1 nor 0) membership grades. This immediately leads to a three-way characterization of a fuzzy set [14], namely, the positive, negative and boundary regions:

$$\text{POS}_{(0,1)}(\mu_A) = \{ x \in U \mid \mu_A(x) = 1 \} = \{ x \in U \mid \mu_A(x) \geq 1 \},$$

$$\text{NEG}_{(0,1)}(\mu_A) = \{ x \in U \mid \mu_A(x) = 0 \} = \{ x \in U \mid \mu_A(x) \leq 0 \},$$

$$\text{BND}_{(0,1)}(\mu_A) = \{ x \in U \mid 0 < \mu_A(x) < 1 \}. \quad (2.14)$$

In the equations, we re-express conditions $\mu_A(x) = 1$ and $\mu_A(x) = 0$ equivalently as $\mu_A(x) \geq 1$ and $\mu_A(x) \leq 0$, respectively. The new form provides hints for generalizations of the qualitative three-way characterization by using other values in the unit
interval $[0, 1]$. The three regions are pair-wise disjoint and their union is the universal set. Some of them may be empty and, hence, they do not necessarily form a partition of $U$. In this thesis, we call the triplet $(\text{POS}_{(0,1)}(\mu_A), \text{BND}_{(0,1)}(\mu_A), \text{NEG}_{(0,1)}(\mu_A))$, a tripartition of $U$ with an understanding that some of them may in fact be empty. The positive region is commonly known as the core of a fuzzy set, and the union of positive and boundary regions $\text{POS}_{(0,1)}(\mu_A) \cup \text{BND}_{(0,1)}(\mu_A)$ is known as the support of a fuzzy set $[16, 23]$.

The division of the universe according to full membership grade, zero membership grade and partial membership grade captures the qualitative nature of a fuzzy set. Both the positive and negative regions contain elements without uncertainty or fuzziness. The boundary region actually contains elements with fuzziness. However, with qualitative characterization, all elements in the boundary region are considered to be the same without considering the actual magnitude of their membership grades.

### 2.3.2 Zadeh’s Quantitative Three-way Approximations

The construction of the qualitative approximation of a fuzzy set uses only the two end points of the unit interval $[0, 1]$, namely, the full membership grade 1 and the zero membership grade 0. An object is not put into the positive region even though its membership grade is very close to 1, and an object is not put into the negative region even though its membership grade is very close to 0. This may be very restrictive for practical uses. A proposal to address such an issue was in fact made by Zadeh [43] by replacing $(1, 0)$ with a pair of thresholds $(\alpha, \beta)$. He suggested the following procedure for constructing a quantitative three-way approximation of a fuzzy set, in which a fuzzy set is denoted by $f_A(x) : U \rightarrow [0, 1]$:

"... one can introduce two levels $\alpha$ and $\beta$ ($0 < \alpha < 1, 0 < \beta < 1, \beta < \alpha$) and agree to say that (1) ‘$x$ belongs to $\mathcal{A}$’ if $f_A(x) \geq \alpha$; (2) ‘$x$ does not belong to $\mathcal{A}$’ if $f_A(x) \leq \beta$; and (3) ‘$x$ has an indeterminate status relative
to $\mathcal{A}'$ if $\beta < f_A(x) < \alpha$. This leads to a three-valued logic (Kleene, 1952) with three truth values: $T$ ($f_A(x) \geq \alpha$), $F$ ($f_A(x) \leq \beta$), and $U$ ($\beta < f_A(x) < \alpha$).

The connection to Kleene’s three-valued logic [19] provides an interpretation of the three-way approximations.

Quantitative approximations of a fuzzy set by a pair of thresholds $(\alpha, \beta)$ with $0 \leq \beta < \alpha \leq 1$ can be viewed as a specific model of three-way decisions [39–41]. Three-way decisions concern division of a universal set into three pair-wise disjoint regions according to the actions of acceptance, rejection and non-commitment. In the context of approximating a fuzzy set, the actions of acceptance and rejection can be interpreted as treating a membership grade close to 1 as 1 and treating a membership grade close to 0 as 0, respectively. One takes a non-commitment action if the membership grade is neither close to 1 nor close to 0. The induced three regions are defined by:

$$\text{POS}_{(\alpha, \beta)}(\mu_A) = \{ x \in U \mid \mu_A(x) \geq \alpha \},$$
$$\text{BND}_{(\alpha, \beta)}(\mu_A) = \{ x \in U \mid \beta < \mu_A(x) < \alpha \},$$
$$\text{NEG}_{(\alpha, \beta)}(\mu_A) = \{ x \in U \mid \mu_A(x) \leq \beta \}. \quad (2.15)$$

When $\alpha = 1$ and $\beta = 0$, we obtain the qualitative three-way approximations of a fuzzy set. The use of other thresholds make the quantitative formulation practically more useful than the qualitative formulation in Equation (2.14). A crucial issue is the interpretation and determination of the required pair of thresholds.

### 2.3.3 Three-Valued Sets as Approximations of Fuzzy Sets

We introduce the concept of three-valued sets, based on three-valued logics, for studying three-way approximations of fuzzy sets. Suppose we use a set of three values
\{n, m, p\} to represent status or our knowledge about a set of objects \(U\). A three-
valued set \(T\) is defined as a mapping from \(U\) to \(\{n, m, p\}\) namely, \(T : U \rightarrow \{n, m, p\}\).
One may interpret a three-valued set based on an understanding of the set of three
values \(\{n, m, p\}\).

A three-way approximation of a fuzzy set suggested by Zadeh [43] can be repre-
sented as the following three-valued set \(T_{(\alpha, \beta)}(\mu_A)\):

\[
T_{(\alpha, \beta)}(\mu_A)(x) = \begin{cases} 
    p, & \mu_A(x) \geq \alpha, \\
    m, & \beta < \mu_A(x) < \alpha, \\
    n, & \mu_A(x) \leq \beta.
\end{cases}
\]  

(2.16)

In contrast to existing studies, we use \(p\) and \(n\), instead of 1 and 0, to differentiate
them from the original full and zero membership grades. In other words, \(p\) represents
a membership grade that may be only close to 1 rather than exactly 1; \(n\) represents
a membership grade that may be only close to 0 rather than exactly 0. The concept
of shadowed sets as an approximation of fuzzy sets, proposed by Pedrycz [29], is
also a special case of three-valued sets, in which the set of three values is given by
\(\{0, [0, 1], 1\}\).

To transform a fuzzy set \(\mu_A\) into a three-valued set \(T_{(\alpha, \beta)}(\mu_A)\), we consider a set
of three actions \(\text{Actions} = \{a_n, a_m, a_p\}\). The actions \(a_n, a_m\) and \(a_p\) change the
membership grade of an element to \(n\), \(m\) and \(p\), respectively. A transformation
function \(\tau : U \rightarrow \{a_n, a_m, a_p\}\) tells us to take an action \(\tau(x) \in \{a_n, a_m, a_p\}\) for \(x \in U\).
With respect to a transformation function, we arrive at a three-way approximation
of a fuzzy set \(\mu_A\):

\[
T_\tau(\mu_A)(x) = \begin{cases} 
    p, & \tau(x) = a_p, \\
    m, & \tau(x) = a_m, \\
    n, & \tau(x) = a_n.
\end{cases}
\]  

(2.17)

Let \(\mathcal{F}\) be the set of all fuzzy sets on \(U\) and let \(\mathcal{T}\) be the set of all three-valued sets on
It can be verified that $T$ can, in fact, be generated from all possible transformation functions for any fuzzy set in $\mathcal{F}$.

A pair of thresholds provides a special way to producing a three-way approximation. A transformation function defined by Equation (2.16) is given by:

$$
\tau_{(\alpha,\beta)}(x) = \begin{cases} 
a_p, & \mu_A(x) \geq \alpha, 
a_m, & \beta < \mu_A(x) < \alpha, 
a_n, & \mu_A(x) \leq \beta. \end{cases}
$$

(2.18)

The subscript $(\alpha,\beta)$ of $T_{(\alpha,\beta)}(\mu_A)$ in Equation (2.16), in fact, denotes that the three-way approximation of a fuzzy set $\mu_A$ is defined by the transformation function $\tau_{(\alpha,\beta)}$.

We may view a three-way approximation of a fuzzy set as a transformation or modification of the original membership grade, in the system defined by the unit interval $[0,1]$, into a new grade in a different or the same system. When the set of three values $\{n,m,p\}$ is used, we explicitly consider three-way approximations of fuzzy sets in a system different from the unit interval $[0,1]$. In a special case, we may choose a setting with $n = 0$, $m \in (0,1)$ and $p = 1$. Consequently, three-way approximations are in the same system defined by the unit interval $[0,1]$.

2.3.4 Construction of Three-Way Approximations by Optimization

This transformation view immediately gives rise to an optimization-based framework for interpreting and determining a transformation function in general and a pair of required thresholds in specific.

As the first step, we need to introduce a measure to quantify the degree of discrepancy between a three-way approximation induced by a transformation function $\tau$ and the original fuzzy set. Suppose $Q : \mathcal{F} \times T \rightarrow \mathbb{R}^+$ maps a pair of a fuzzy
set and a three-valued set to a non-negative real number, where $\mathbb{R}^+$ is the set of non-negative real numbers. The quantity $Q(\mu_A, T_\tau(\mu_A))$ measures the degree of discrepancy, in terms of error, distance, cost or change of uncertainty, between a fuzzy set $\mu_A$ and a three-valued set $T_\tau(\mu_A)$. Without loss of generality, we assume that a lower discrepancy value indicates a better approximation.

In general, a meaningful discrepancy measure must satisfy some conditions. As minimum requirements, we need to consider two orderings on a set of membership grades. Following studies in three-valued logics [7, 8], we consider the following two orderings [18, 37]:

\begin{align*}
\text{truth ordering} & : \ n \preceq_t m \preceq_t p, \\
\text{information ordering} & : \ m \preceq_i n, \ m \preceq_i p.
\end{align*}

According to the truth ordering, the values $n$, $m$ and $p$ represent, respectively, the low, middle and high grades of membership. The two opposite values $n$ and $p$ stand for the negative and positive, respectively, and $m$ stands for the middle of the positive and negative. An element with the grade $p$ belongs to a fuzzy set more than an element with the grade $m$, and both of them are more than an element with the grade $n$. The information ordering represents the knowledge or uncertainty about objects. In other words, it suggests that we have less information, and hence more uncertainty, about objects with the membership grade $m$, and more information, and hence less uncertainty, about objects with membership grades $n$ and $p$. Similarly, for the unit interval $[0, 1]$, the two orderings are given by:

\begin{align*}
\text{truth ordering} & : \ a \preceq_i b \iff a \leq b, \\
\text{information ordering} & : \ \text{for } a, b \leq \frac{1}{2}, a \preceq_i b \iff a \geq b, \\
& \text{for } a, b > \frac{1}{2}, a \preceq_i b \iff a \leq b.
\end{align*}
According to the truth ordering, 0 is lowest membership grade and 1 is the highest membership grade. According to the uncertainty order, membership grades 0 and 1 are most informative and membership grade 1/2 is least informative. Thus, one can draw a corresponding between \( n \) and 0, \( m \) and 1/2, and \( p \) and 1, respectively. In addition, a discrepancy measure must take into consideration of consistency of corresponding truth and information orderings in systems \( \{n, m, p\} \) and \([0, 1]\).

Different transformation functions induce different three-valued approximations. Once a measure of discrepancy is established, the second step is to search for a transformation function or a pair of thresholds by optimizing an objective function expressed through the measure of discrepancy. In general, the problem is to find a solution to the following optimization problem:

\[
\arg \min_{\tau} Q(\mu_A, T_{\tau}(\mu_A)),
\]

where \( \arg \) denotes the argument, namely, a pair of threshold, that minimizes the discrepancy \( Q \). For the class of special transformation functions defined by pairs of thresholds, the problem becomes:

\[
\arg \min_{(\alpha, \beta)} Q(\mu_A, T_{(\alpha, \beta)}(\mu_A)), \quad 0 \leq \beta < \alpha \leq 1.
\]

One may apply different optimization methods.

Different interpretations of discrepancy measures suggest different classes of three-way approximations. In this thesis, we consider three possible interpretations, namely, \( Q \) as a measure of difference of uncertainties of \( \mu_A \) and \( T_{\tau}(\mu_A) \), as a distance between \( \mu_A \) and \( T_{\tau}(\mu_A) \), and as a cost for changing \( \mu_A \) into \( T_{\tau}(\mu_A) \), respectively. They immediately offer three interpretations according to the following principles of optimization:

- a principle of uncertainty invariance,
• a principle of minimum distance, and

• a principle of least cost.

The three principles have been used implicitly in existing studies. We make explicit references to the three principles in order to show the differences and connections of different formulations. To apply the principle of uncertainty invariance, $Q$ measures the absolute difference between uncertainty of the original fuzzy set and uncertainty of its three-way approximations. To apply the principle of minimum distance, $Q$ measures the distance between the original fuzzy set and its three-way approximations. The first two principles normally consider only the changes of membership grades without considering the costs associated with changes. They produce cost-insensitive methods. Three-way approximations obtained depend on the shape of a fuzzy membership function, with respect to an uncertainty measure or a distance. A fuzzy set always has the same approximation independent of particular applications.

The principle of least cost considers the costs or risks associated with changes, which leads to cost-sensitive methods. By assigning different costs of actions to reflect truthfully the consequences of changes of membership grades, a fuzzy set may be approximated differently in different applications. The principle of least cost may be viewed as a generalization of the principle of minimum distance: the latter can be obtained from the former if we assume the same cost for all changes [14]. In other words, when applying the principle of least cost, one considers both the magnitude of changes and associated costs. Decision-theoretic three-way approximations of fuzzy sets [14] are an application of the principle of least cost.
Chapter 3

AN ADDITIVE MEASURE AND FORMULATIONS BASED ON THREE PRINCIPLES

This chapter first introduce a general framework which can be used to solve the fundamental issue of three-way approximations, that is, how to determine the value of the pair of the thresholds. We examine the framework with three principles under three frameworks.

3.1 A Decision-Theoretic Optimization Framework Based on Additive Measures

In general, finding a pair of thresholds depends on a discrepancy measure $Q$. Defining and interpreting discrepancy measures $Q$ is a fundamental issue that needs further study. There are many possible ways to define a discrepancy measure and each of them captures a different semantical aspect of three-way approximations.

We consider a class of component-wise defined additive discrepancy measures of
the following form:

$$Q(\mu_A, T_\tau(\mu_A)) = \sum_{x \in U} q(\mu_A(x), T_\tau(\mu_A)(x)),$$

where $q(\mu_A(x), T_\tau(\mu_A)(x)) \geq 0$ measures the discrepancy of approximations with respect to element $x$. The overall discrepancy of approximations is simply a summation of the discrepancy of approximations of individual elements. For an additive measure, the optimization problem can be stated as follows:

$$\arg \min_\tau \sum_{x \in U} q(\mu_A(x), T_\tau(\mu_A)(x)), \quad 0 \leq \beta < \alpha \leq 1.$$

Since $q(\mu_A(x), T_\tau(\mu_A)(x)) \geq 0$, the summation reaches the minimum value if and only if each term $q(\mu_A(x), T_\tau(\mu_A)(x))$ is minimum. This immediately leads to a simple decision-theoretic optimization formulation for finding a pair of thresholds.

Let $Actions = \{a_n, a_m, a_p\}$ denote a set of three actions. That is, for each element $x$, we take one of the three actions: changing $\mu_A(x)$ into $n$, $m$, and $p$, respectively. We should choose an action so that $q$ is minimum, that is,

(P) If $q(\mu_A(x), p) \leq q(\mu_A(x), n)$ and $q(\mu_A(x), p) \leq q(\mu_A(x), m)$,
then take action $a_p$,

(M) If $q(\mu_A(x), m) < q(\mu_A(x), n)$ and $q(\mu_A(x), m) < q(\mu_A(x), p)$,
then take action $a_m$,

(N) If $q(\mu_A(x), n) \leq q(\mu_A(x), m)$ and $q(\mu_A(x), n) < q(\mu_A(x), p)$,
then take action $a_n$.

The three rules enable us to determine a transformation function. In the next three sections, we review the main results of shadowed sets and three-way approximations.
of fuzzy sets within the optimization-based framework.

3.2 Shadowed Sets According to the Principle of Uncertainty Invariance

In this section, we revisit two formulations of shadowed sets in the light of a principle of uncertainty invariance.

3.2.1 Principle of Uncertainty Invariance

The principle of uncertainty invariance or information preservation is first formulated by Klir [21,22]. The principle “facilitates connections among representations of uncertainty and information in alternative mathematical theories. The principle requires that the amount of uncertainty (and information) be preserved when a representation of uncertainty in one mathematical theory is transformed into its counterpart in another theory” [22]. A fundamental task, when applying the principle, is to define a measure of uncertainty or information. One may derive many versions of the same principle by using different uncertainty or information measures.

If we interpret the original fuzzy set and its three-way approximations as different representations of the same concept, we can immediately apply the principle of uncertainty invariance. According to the principle, a three-way approximation $T_{(\alpha,\beta)}(\mu_A)$ must have the same amount of uncertainty as the original fuzzy set $\mu_A$. Since this may not always be an achievable goal, one may require that the absolute difference of the uncertainties of the two sets is minimum. Let $u : F \rightarrow \mathbb{R}$ denote an uncertainty measure of a fuzzy set and, for simplicity, let $u : T \rightarrow \mathbb{R}$ denote an uncertainty measure of a three-valued set. The principle of uncertainty invariance suggests the
following optimization problem for finding a pair of thresholds:

$$\arg \min_{(\alpha, \beta)} |u(\mu_A) - u(T_{(\alpha, \beta)}(\mu_A))|; \quad (3.3)$$

where $| \cdot |$ denotes the absolute value.

Pedrycz [29] proposes the notion of shadowed sets by implicitly applying the principle of uncertainty invariance. He gave a simple method for minimizing the objective function (3.3) based on the following observation. When changing a fuzzy set into a three-valued set, one decreases uncertainty for some elements and increases uncertainty for some other elements. If one can balance the amount of increased and the amount of decreased uncertainty, then the uncertainty of a three-way approximation has the same uncertainty as the original fuzzy set. By considering different definitions of uncertainty of a fuzzy sets, different formulations of shadowed sets can be derived.

### 3.2.2 Pedrycz’s Formulation

A shadowed set $S$, proposed by Pedrycz [29], is defined by a mapping from a universal set $U$ to the set $\{0, [0, 1], 1\}$, that is, $S : U \rightarrow \{0, [0, 1], 1\}$. Elements of $U$ with membership grade 1 constitute the core of $S$ and elements with membership grade $[0, 1]$ form the shadow of $S$. It is possible that either the core or the shadow of a shadowed set may be empty. It may be commented that this definition of a shadowed set is not necessarily related to a fuzzy set.

Pedrycz [29–34] provides a framework for constructing a shadowed set as a three-way approximation of a fuzzy set. Given a pair of thresholds $(\alpha, \beta)$ with $0 \leq \beta < \alpha$. \]
$\alpha \leq 1$, one can construct a shadowed set from a fuzzy set $\mu_A$ as follows [29]:

$$S(\alpha, \beta)(\mu_A)(x) = \begin{cases} 1, & \mu_A(x) \geq \alpha, \\ [0, 1], & \beta < \mu_A(x) < \alpha, \\ 0, & \mu_A(x) \leq \beta. \end{cases}$$

(3.4)

In terms of three-way decisions, a shadowed set can be conveniently interpreted as three regions: the positive region represented by membership grade 1, the negative region represented by membership grade 0, and the boundary region represented by membership grade $[0, 1]$. The boundary region is in fact the shadowed area.

![Figure 3.1: A shadowed set approximation of a fuzzy set](image)

As shown in Figure 3.1, the transformation of a fuzzy set into a shadowed set can be interpreted in terms of three actions. The elevation action lifts the membership grade $\mu_A(x)$ to 1, the reduction action suppresses the membership grade $\mu_A(x)$ to 0, and the extending action stretches the membership grade $\mu_A(x)$ into the unit interval $[0, 1]$. For a shadowed set, membership grades 0 and 1 have the minimum uncertainty 0 and the membership grade $[0, 1]$ has the maximum uncertainty 1. The elevation and reduction actions lead to a decrease of uncertainty and the extending operation leads to an increase of uncertainty. One may interprets the sum of the elevated area...
and the reduced area as the amount of decreased uncertainty and the shadow is the amount of increased uncertainty. According to the principle of uncertainty invariance, one immediately have the following objective function, as suggested by Pedrycz [29]:

$$\text{Elevated Area}(S_{(\alpha,\beta)}(\mu_A)) + \text{Reduced Area}(S_{(\alpha,\beta)}(\mu_A)) = \text{Shadow}(S_{(\alpha,\beta)}(\mu_A)).$$

(3.5)

It represents a balance of increased and decreased uncertainties.

In practical situations, it may be difficult to find a pair of thresholds that satisfies condition (3.5), particularly, when $U$ is a finite universe. Thus, one can minimize the following absolute difference:

$$V_{(\alpha,\beta)}(\mu_A) = |\text{Elevated Area}(S_{(\alpha,\beta)}(\mu_A)) + \text{Reduced Area}(S_{(\alpha,\beta)}(\mu_A)) - \text{Shadow}(S_{(\alpha,\beta)}(\mu_A))|,$$

(3.6)

where card(·) denotes the cardinality of a set. For simplicity, we assume that universal set $U$ is finite and hence, summation is used in Equation (3.6). An optimal pair of thresholds $(\alpha, \beta)$ can be obtained by taking the arguments that minimize the objective function $V_{(\alpha,\beta)}(\mu_A)$:

$$\arg \min_{(\alpha,\beta)} V_{(\alpha,\beta)}(\mu_A).$$

(3.7)

That is, determining a pair of thresholds becomes a problem of optimization.

A difficulty with the formulation is to find an analytical solution to the objective function in Equation (3.6). The optimal pair of thresholds $(\alpha, \beta)$ depends on specific membership functions. For easy calculation, Pedrycz [29] imposes an additional condition:

$$\alpha + \beta = 1.$$

(3.8)
Consequently, one only needs to compute a single threshold $\alpha$. By the condition $\beta < \alpha$, the additional condition introduces a symmetric model with $\alpha > 0.5$. The symmetric model may not be suitable for all situations. It may happen that some practical applications require $\alpha < 0.5$.

Another difficulty with shadowed sets is the semantical interpretation of the objective function $V_{(\alpha, \beta)}(\mu_A)$ for finding an optimal pair of thresholds. Pedrycz interpreted $V_{(\alpha, \beta)}(\mu_A)$ as a balance of a trade-off of vagueness of three regions. While elevation and reduction decrease uncertainty of some objects, the change of membership grades to the unit interval $[0, 1]$ increases the uncertainty of some other objects. In the formulation, the size of an area is implicitly used as a measure of uncertainty. While the evaluated and reduced areas may be interpreted as errors induced by changes of membership grades, the shadowed area cannot be interpreted in a similar way [14]. Justification of the meaningfulness of $V_{(\alpha, \beta)}(\mu_A)$ remains to be an issue [13, 14, 36].

### 3.2.3 Tahayori, Sadeghian and Pedrycz’s Reformulation

Tahayori, Sadeghian and Pedrycz [36] argue that the objective function $V_{(\alpha, \beta)}(\mu_A)$ does not fully comply with the intended semantics interpretation of a shadowed set. Specifically, the original fuzziness of the elements in the shadowed region is not properly considered. In other words, the cardinality $\text{card}(\{x \in U \mid \beta < \mu_A(x) < \alpha\})$ does not correctly represent the amount of the increased vagueness or uncertainty, as it also includes the original uncertainty given by the membership grade $\beta < \mu_A(x) < \alpha$.

To resolve this difficulty, they give a new formulation based on a measure of fuzziness of a fuzzy set. We review their method with some slight modifications by using our notational system.

A fuzziness measure of a fuzzy set quantifies the underlying vagueness or uncertainty of a fuzzy set. For simplicity, we consider a class of additive fuzziness measures
of the following form [20, 24, 26]:

$$\varphi(\mu_A) = h(\sum_{x \in U} f(\mu_A(x))),$$

(3.9)

where $h : R^+ \rightarrow R^+$ is a monotonically increasing function and $f$ is a measure of
the fuzziness of a membership grade in the unit interval and satisfies the following
properties [20]:

1. $f(0) = f(1) = 0$,

2. $f$ is monotonically increasing on $[0, 0.5]$, monotonically decreasing on $[0.5, 1]$, and reaches the maximum value at 0.5,

3. $f$ is symmetric with respect to $x = 0.5$, i.e., $f(a) = f(1 - a)$ for $a \in [0, 1]$.

The additivity enables us to compute the fuzziness of a fuzzy set easily by using the fuzziness of each element in $U$.

By the condition $f(0) = f(1) = 0$, the fuzziness of the elevated and reduced regions is zero in the shadowed set. The amount of the decrease of fuzziness introduced by a shadowed set is given by:

$$\sum_{\mu_A(x) \geq \alpha} f(\mu_A(x)) + \sum_{\mu_A(x) \leq \beta} f(\mu_A(x)).$$

(3.10)

Elements in the shadowed region have the maximum fuzziness $f(0.5)$, and the amount of the decrease of fuzziness introduced by a shadowed set is given by:

$$\sum_{\beta < \mu_A(x) < \alpha} (f(0.5) - f(\mu_A(x))).$$

(3.11)

In this way, the original fuzziness of an element in the shadowed region with respect to the fuzzy set, i.e., $f(\mu_A(x))$, is taken into consideration. In contrast, Pedrycz's original formulation uses $\text{card}(\{x \in U \mid \beta < \mu_A(x) < \alpha\})$, which can be interpreted
by a fuzziness variance with \( f(0.5) = 1 \) and does not exclude the amount of fuzziness \( f(\mu_A(x)) \).

To ensure that the constructed shadowed set has the same level of fuzziness as the original fuzzy set, it is required that

\[
\sum_{\mu_A(x) \geq \alpha} f(\mu_A(x)) + \sum_{\mu_A(x) \leq \beta} f(\mu_A(x)) = \sum_{\beta < \mu_A(x) < \alpha} (f(0.5) - f(\mu_A(x))).
\] (3.12)

Equivalently, one can minimizing the following objective function:

\[
\left| \sum_{\mu_A(x) \geq \alpha} f(\mu_A(x)) + \sum_{\mu_A(x) \leq \beta} f(\mu_A(x)) - \sum_{\beta < \mu_A(x) < \alpha} (f(0.5) - f(\mu_A(x))) \right|. \tag{3.13}
\]

In addition to its semantical advantages, the reformulation of Tahayori, Sadeghian and Pedrycz [36] explores the monotonicity and symmetry of a measure of fuzziness of a membership grade for computing the required thresholds. The details can be found in their paper. We only summarize the main results.

As shown by Figure 3.2(a), the increase and decrease of fuzziness can be controlled by a threshold \( \gamma \) on a measure of fuzziness \( f \). Due to the symmetry of \( f \), \( \gamma \) induces a pair of threshold \((1 - \alpha, \alpha)\) with \( \alpha > 0.5 \). Based on the pair of threshold \((1 - \alpha, \alpha)\), as shown in Figure 3.2(b), one can construct a shadowed set. In summary, they transform the problem of determining pair of thresholds \((1 - \alpha, \alpha)\) into the problem of determining a threshold \( \gamma \) on \( f(\mu_A(x)) \). Furthermore, analytic formulas can be obtained for \( \gamma \) and, hence, for \((1 - \alpha, \alpha)\), based on the notion of a gradual number of cardinality of a fuzzy set. The result is a symmetric shadowed set model with \( \alpha + \beta = 1 \).

With the reformulation, the conditions \( \alpha + \beta = 1 \) and \( \alpha > 0.5 \) are still imposed, which limits the possible applications of shadowed sets. In addition, it is implicitly assumed the membership grade of an element in the shadowed region is 0.5, i.e., the grade with maximum fuzziness.
Ben and Coroianu [1] gave another application of the principle of uncertainty invariance. They constructed nearest interval, triangular and trapezoidal approximations, respectively, by preserving the ambiguity of the original fuzzy set.

3.3 Three-Way Approximations of Fuzzy Sets According to the Principle of Minimum Distance

This section discusses constructions of three-way approximations by using the principle of minimum distance.

3.3.1 Principle of Minimum Distance

The distance between two points has an intuitively appealing geometric interpretation. If one travels from one point to another point, the distance between the two points may be related to the time required or the amount of efforts demanded. The principle of minimum distance may take one of two forms. If one considers multiple paths between two points, the principle requires one travels along the shortest path between the two points. The principle may be interpreted as a principle of minimum path or a principle of least effort [45]. If one considers many points that are adjacent to a given point, the principle suggest that one should choose a node with a minimum
distance to the given node. In this case the principle is in fact a nearest neighbor principle [12,28]. In this thesis, we take the nearest neighbor interpretation.

The principle of minimum distance has been widely used in fuzzy set theory [3,27]. For approximating a fuzzy set, the principle requires that a good approximation must be closest to the original fuzzy set with respect to a distance measure between fuzzy sets. Chakrabarty et al. [5] introduced the notion of the nearest ordinary set of a fuzzy set within the theory of rough sets. Nguyen, Pedrycz and Kreinovich [27] provided a detailed examination of various distance-based approximations of a fuzzy set. They gave a distance-based interpretation of shadowed sets. Grzegorzewski [17] examined nearest interval approximation of a fuzzy number based on a distance measure, where a fuzzy number is a special fuzzy set. Ciucci and Flaminio [9] introduced the notion of generalized inner and outer approximations of fuzzy sets, which can also be interpreted in terms of the principle of minimum distance.

To apply the principle of minimum distance for constructing three-way approximations of a fuzzy set, we assume that \( n, m, \) and \( p \) are three distinct values in the unit interval \([0,1]\). This ensures that a distance function can be used. Without loss of generality, we use \( 0, 0 < m < 1 \) and \( 1 \) to represent \( n, m, \) and \( p \), respectively. That is, if the membership grade of an element is equal to or greater than \( \alpha \), we elevate its membership grade to 1; if the membership grade of an element is equal to or lower than \( \beta \), we reduce its membership grade to 0; if the membership grade is between \( \beta \) and \( \alpha \), we change its membership grade to \( m \). Thus, a three-way approximation is in fact a three-valued fuzzy set.

Let \( D : \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{R}^+ \) denote a distance measure between two fuzzy sets. Many classes of distance measures have been proposed and studied [6,35]. For example,
Liu [25] suggested that a distance measure should satisfy the following conditions:

\[(D1) \quad D(\mu_A, \mu_B) = D(\mu_B, \mu_A), \text{ for all } \mu_A, \mu_B \in \mathbb{F},\]
\[(D2) \quad D(\mu_A, \mu_A) = 0, \text{ for all } \mu_A \in \mathbb{F},\]
\[(D3) \quad D(\mu_A, \mu_A^c) = \max_{\mu_A, \mu_B \in \mathbb{F}} D(\mu_A, \mu_B), \forall \mu_A \in \mathbb{F},\]
\[(D4) \quad \text{For all } \mu_A, \mu_B, \mu_C \in \mathbb{F}, \text{ if } \mu_A \subset \mu_B \subset \mu_C, \text{ then } D(\mu_A, \mu_B) \leq D(\mu_A, \mu_C) \text{ and } D(\mu_B, \mu_C) \leq D(\mu_A, \mu_C),\]

where \(\mu_A^c(x) = 1 - \mu_A(x)\) denotes the complement of a fuzzy set \(\mu_A\). Many studies used this class of distance measures [10, 44].

According to the principle of minimum distance, one searches for three-valued approximation that has the minimum distance from the original fuzzy set. That is, we have the following optimal problem:

\[
\arg \min_{\tau} D(\mu_A, T_\tau(\mu_A)).
\] (3.14)

In general, we may not have a method to solve this problem independent of distance measures. Thus, we focus on specific classes of distance measures.

\subsection{3.3.2 Formulations Based on Minkowski Distance}

Consider the class of distance measures between fuzzy sets defined by Minkowski r-metric [6, 46]: for two fuzzy sets \(\mu_A, \mu_B \in \mathbb{F},\)

\[
D_r(\mu_A, \mu_B) = \left( \sum_{x \in U} |\mu_A(x) - \mu_B(x)|^r \right)^{\frac{1}{r}}.
\] (3.15)

The class of Minkowski distance measure satisfies conditions (D1)–(D4). When \(r \geq 1\), for a finite universe, \(D_r\) satisfies the property of a distance function, name-
ly, the reflexivity (D1), symmetry (D2) and the triangle inequality $D_r(\mu_A, \mu_C) \leq D_r(\mu_A, \mu_B) + D_r(\mu_B, \mu_C)$.

When $r = 1$, $D_1(\mu_A, \mu_B)$ is the Hamming distance. When $r = 2$, $D_2(\mu_A, \mu_B)$ the familiar Euclidean distance. When $r$ approaches $\infty$, the distance between $\mu_A$ and $\mu_B$ is dominated by an element with the largest difference $|\mu_A(x) - \mu_B(x)|$. That is,

\[ D_\infty(\mu_A, \mu_B) = \max_{x \in U} |\mu_A(x) - \mu_B(x)|. \] (3.16)

In this thesis, we only consider a finite $r \geq 1$.

For a finite $r \geq 1$, we know that $D_r$ takes a minimum value if and only if $|\mu_A(x) - \mu_B(x)|$ is minimum for every $x \in U$. Accordingly, we can transform the optimization problem with respect to $r \geq 1$:

\[ \arg \min_{\tau} D_r(\mu_A, T_\tau(\mu_A)), \] (3.17)

to an optimization problem with respect to $r = 1$:

\[ \arg \min_{\tau} \sum_{x \in U} |\mu_A(x) - T_\tau(\mu_A)(x)|. \] (3.18)

Immediately, we can apply the results regarding an additive discrepancy measure given in Section 3.1.

By inserting $d(\mu_A(x), T_\tau(\mu_A)(x)) = |\mu_A(x) - T_\tau(\mu_A)(x)|$ as the discrepancy mea-
sure in rules (P), (M) and (N) of Section 3.1, we have:

(P) \hspace{1cm} \text{If } |\mu_A(x) - 1| \leq |\mu_A(x) - 0| \land |\mu_A(x) - 1| \leq |\mu_A(x) - m|, \\
then \text{ take action } a_p, \text{ that is, } T_{\tau}(\mu_A)(x) = 1,

(M) \hspace{1cm} \text{If } |\mu_A(x) - m| < |\mu_A(x) - 0| \land |\mu_A(x) - m| < |\mu_A(x) - 1|, \\
then \text{ take action } a_m, \text{ that is, } T_{\tau}(\mu_A)(x) = m,

(N) \hspace{1cm} \text{If } |\mu_A(x) - 0| \leq |\mu_A(x) - m| \land |\mu_A(x) - 0| < |\mu_A(x) - 1|, \\
then \text{ take action } a_n, \text{ that is, } T_{\tau}(\mu_A)(x) = 0.

We determine a transformation by simplifying rules (P), (M) and (N).

Rule (P): The first condition $|\mu_A(x) - 1| \leq |\mu_A(x) - 0|$ of rule (P) can be equivalently expressed as $1 - \mu_A(x) \leq \mu_A(x)$. That is, $\mu_A(x) \geq 0.5$. The second condition $|\mu_A(x) - 1| \leq |\mu_A(x) - m|$ of rule (P) can be expressed as:

$$
[(\mu_A(x) \geq m) \land (1 - \mu_A(x) \leq \mu_A(x) - m)] \lor [(\mu_A(x) < m) \land (1 - \mu_A(x) \leq m - \mu_A(x))],
$$

which is equivalent to

$$
[(\mu_A(x) \geq m) \land (\mu_A(x) \geq \frac{1 + m}{2})] \lor [(\mu_A(x) < m) \land (m \geq 1)].
$$

Since $\mu_A(x) \geq (1 + m)/2$ implies $\mu_A(x) \geq m$, we do not need to consider $\mu_A(x) \geq m$. By assumption, $m \geq 1$ is not possible. Therefore, the condition can be simplified in to $\mu_A(x) \geq (1 + m)/2$. Since $\mu_A(x) \geq (1 + m)/2 \Rightarrow \mu_A(x) \geq 1/2$, the entire condition of rule (P) can be expressed as $\mu_A(x) \geq (1 + m)/2$. Consequently, we have a simple rule:

(P) \hspace{1cm} \text{If } \mu_A(x) \geq \frac{1 + m}{2}, \text{ then take action } a_p.
In other words, rule (P) uses one threshold.

Rule (M): The first condition $|\mu_A(x) - m| < |\mu_A(x) - 0|$ of rule (M) can be expressed as:

$$[(\mu_A(x) \geq m) \land (\mu_A(x) - m < \mu_A(x))] \lor [(\mu_A(x) < m) \land (m - \mu_A(x) < \mu_A(x))],$$

which is equivalent to:

$$[(\mu_A(x) \geq m) \land (m > 0)] \lor [(\mu_A(x) < m) \land (\mu_A(x) > \frac{m}{2})].$$

Therefore,

$$[\mu_A(x) \geq m] \lor [(\mu_A(x) < m) \land (\mu_A(x) > \frac{m}{2})],$$

which is equivalent to

$$\mu_A(x) > \frac{m}{2}. \quad (3.19)$$

According to the discussion of the second condition of rule (P), the second condition of rule (M) is equivalent to:

$$\mu_A(x) < \frac{1 + m}{2}. \quad (3.20)$$

By combining the two inequalities (3.19) and (3.20), we have:

$$\frac{m}{2} < \mu_A(x) < \frac{1 + m}{2}. \quad (3.21)$$

Rule (M) can be expressed as:

(M) If $\frac{m}{2} < \mu_A(x) < \frac{1 + m}{2}$, then take action $a_m$. 

39
Rule (M) uses a pair of thresholds.

Rule (N): The first condition $|\mu_A(x) - 0| \leq |\mu_A(x) - m|$ of rule (N) is related to the first condition of rule (M). Thus, we have $\mu_A(x) \leq m/2$. The second condition of rule (M) is related to the first condition of rule (P). Thus, we have $\mu_A(x) \leq 1/2$. Since $m \leq 1$, we have $m/2 \leq 1/2$. The combination of two conditions of rule (N) can be simply expressed as $\mu_A(x) \leq m/2$. Consequently, rule (N) can be expressed as:

\[\text{(N) If } \mu_A(x) \leq \frac{m}{2}, \text{ take action } a_n.\]

Again, rule (N) only uses one threshold.

Our analysis shows that the transformation function according to the principle of minimum distance is, in fact, a transformation function defined by a pair of thresholds. By setting,

\[\alpha = \frac{1 + m}{2}, \quad \beta = \frac{m}{2}, \quad (3.22)\]

we immediately arrive at $(\alpha, \beta)$-transformation function defined by Equation (2.18). It can be shown that

\[0 < \beta < m < \alpha < 1. \quad (3.23)\]

As a final result, we produce the following three-way approximation:

\[T_{(\alpha, \beta)}\mu_A(x) = \begin{cases} 1, & \mu_A(x) \geq \alpha, \\ m, & \beta < \mu_A(x) < \alpha, \\ 0, & \mu_A(x) \leq \beta, \end{cases} \quad (3.24)\]

where the pair of thresholds depend on the choice of $m$.

Consider a special case in which $m = 0.5$. According to Equation (3.22), we have $\alpha = 0.75$ and $\beta = 0.25$. This produces the $(0.75, 0.25)$-three-way approximation discussed by Deng and Yao [14].
Figure 3.3 illustrates \((\alpha, \beta)\)-three-way approximation of a fuzzy set. There are connections and differences between shadowed sets and distance based three-way approximations of fuzzy sets. Both of them can be explained in terms of elevation and reduction operations, as shown by Figures 3.1 and 3.3. While the computation and interpretation of the elevated and reduced areas are the same, the computation and interpretation of shadows are different. With respect to \(r = 1\), we can express the distance between \(\mu_A\) and \(T_{(\alpha, \beta)}(\mu_A)\), that is, \(D_1(\mu_A, T_{(\alpha, \beta)}(\mu_A))\), as the summation of three types of areas:

\[
D_1(\mu_A, T_{(\alpha, \beta)}(\mu_A)) = \sum_{\mu_A(x) \geq \alpha} (1 - \mu_A(x)) + \sum_{\mu_A(x) \leq \beta} \mu_A(x) + \sum_{\beta < \mu_A(x) < m} |m - \mu_A(x)|
\]

\[
= \text{Elevated Area}(T_{(\alpha, \beta)}(\mu_A)) + \text{Reduced Area}(T_{(\alpha, \beta)}(\mu_A))
\]

\[
+ \text{Shadow}(T_{(\alpha, \beta)}(\mu_A)). \tag{3.25}
\]

The elevated area can be considered as errors of elevation to 1, and the reduced area as errors of reduction to 0 for both models. However, the interpretation of shadows are different. The shadow of distance-based approximation is the error for changing the membership grade to \(m\). The objective function (3.25) may be interpreted as total errors induced by an \((\alpha, \beta)\)-transformation function. Thus, a distance based three-way approximation is obtained by minimizing total error. In contrast, for a shadowed set, the shadow area is a summation of errors of elevating \(\mu_A(x)\) to 1 and reducing \(\mu_A(x)\) to 0. The meaning of objective function (3.6) as defined by a balance of errors of three areas is not entirely clear. Further investigations are needed.
3.4 Three-Way Approximations of Fuzzy Sets According to the Principle of Least Cost

This section presents a generalized version of decision-theoretic three-way approximations of fuzzy sets [14]. Instead of using 0.5, we use a number $0 < m < 1$, as given in Equation (3.24).

3.4.1 Principle of Least Cost

In a formulation with a principle of minimum distance, we consider three actions. An underlying assumption implicitly made is that all three actions are of the same cost. In practical situations, one may argue that different actions may have different costs. This leads to a more general cost-sensitive approach to three-way approximations of fuzzy sets [14]. The corresponding principle used is a principle of least cost. One searches for an approximation with the least cost. A distance-based formulation is a special case in which costs of all actions are the same, which may be considered as a cost-insensitive approach.

For cost-sensitive approaches, the costs of changing a membership grade to $m$
for $m$ may be different. Thus, we divide action $a_m$ into two different actions $a_{m^\uparrow}$ and $a_{m^\downarrow}$. Similar to a distance-based formulation as shown by Figure 3.3, we have a set of four actions $\text{Actions} = \{a_n, a_{m^\uparrow}, a_{m^\downarrow}, a_p\}$, where the two new actions $a_{m^\uparrow}$ and $a_{m^\downarrow}$ denote elevating $\mu_A(x)$ to $m$ and reducing $\mu_A(x)$ to $m$, respectively. A transformation function $\tau(x) \in \{a_n, a_{m^\uparrow}, a_{m^\downarrow}, a_p\}$ indicates an action we take for $x \in U$. We assume that each action $a \in \text{Actions}$ associates with certain cost. Table 3.1 summarizes costs of all actions, in which $\lambda$’s are interpreted as the unit costs of various actions.

<table>
<thead>
<tr>
<th>Action</th>
<th>Fuzzy set membership grade</th>
<th>Three-way membership grade</th>
<th>Error</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_n$</td>
<td>$\mu_A(x)$</td>
<td>0</td>
<td>$\mu_A(x) - 0$</td>
<td>$\lambda_n$</td>
</tr>
<tr>
<td>$a_{m^\uparrow}$</td>
<td>$\mu_A(x) &lt; m$</td>
<td>$m$</td>
<td>$m - \mu_A(x)$</td>
<td>$\lambda_{m^\uparrow}$</td>
</tr>
<tr>
<td>$a_{m^\downarrow}$</td>
<td>$\mu_A(x) \geq m$</td>
<td>$m$</td>
<td>$\mu_A(x) - m$</td>
<td>$\lambda_{m^\downarrow}$</td>
</tr>
<tr>
<td>$a_p$</td>
<td>$\mu_A(x)$</td>
<td>1</td>
<td>$1 - \mu_A(x)$</td>
<td>$\lambda_p$</td>
</tr>
</tbody>
</table>

Table 3.1: Errors and costs of various actions

For an object $x \in U$, the decision error $E(\mu_A(x), T_\tau(\mu_A)(x))$ can be defined as a change of membership grade after the approximation, that is,

$$E(\mu_A(x), T_\tau(\mu_A)(x)) = |\mu_A(x) - T_\tau(\mu_A)(x)|. \hspace{1cm} (3.26)$$

The cost of taking action $\tau(x) \in A$ for $x$ can be computed by:

$$C(\tau(x)|\mu_A(x)) = E(\mu_A(x), T_\tau(\mu_A)(x)) \lambda_{\tau(x)},$$

$$= |\mu_A(x) - T_\tau(\mu_A)(x)| \lambda_{\tau(x)}. \hspace{1cm} (3.27)$$

The overall cost is given by:

$$C(T_\tau(\mu_A)) = \sum_{x \in U} C(\tau(x)|\mu_A(x)). \hspace{1cm} (3.28)$$
According to the principle of least cost, by minimizing the objective function $C(\tau(x)|\mu_A(x))$, we obtain an optimal transformation function $\tau$:

$$\arg \min_{\tau} C(T_\tau(\mu_A)).$$  \hspace{1cm} (3.29)

Since the overall cost is a summation of costs for individual objects, we can immediately apply the general framework given in Section 3.1.

### 3.4.2 Decision-Theoretic Three-Way Approximations of Fuzzy Sets

According to the principle of least cost, we need to find a three-way approximation that minimizes the overall cost given by Equation (3.28). This can be achieved by taking an action $\tau(x) \in A$ for object $x$ that has minimum $C(\tau(x)|\mu_A(x))$ value. We consider two cases: $\mu_A(x) \geq m$ and $\mu_A(x) < m$.

**Case 1 ($\mu_A(x) \geq m$):** We can take one of three actions $a_n$, $a_{m\downarrow}$ and $a_p$. By using $C(\tau(x)|\mu_A(x))$ as a measure in rules (N), (M) and (P) in Section 3.1, we have:

- **(P1)** If $C(a_p|\mu(x)) \leq C(a_n|\mu(x)) \land C(a_p|\mu(x)) \leq C(a_{m\downarrow}|\mu(x))$,
  then take action $a_p$, that is, $T_\tau(\mu_A)(x) = 1$,

- **(M1)** If $C(a_{m\downarrow}|\mu(x)) < C(a_n|\mu(x)) \land C(a_{m\downarrow}|\mu(x)) < C(a_p|\mu(x))$,
  then take action $a_{m\downarrow}$, that is, $T_\tau(\mu_A)(x) = m$,

- **(N1)** If $C(a_n|\mu(x)) \leq C(a_{m\downarrow}|\mu(x)) \land C(a_n|\mu(x)) < C(a_p|\mu(x))$,
  then take action $a_n$, that is, $T_\tau(\mu_A)(x) = 0$.

To simplify these rules, we need to consider some properties of costs. Actions $\lambda_p$ and $\lambda_n$ will remove all uncertainty in the original membership grade $\mu_A(x)$ and $\lambda_{m\downarrow}$ and
\(\lambda_{m^\uparrow}\) will still retain some uncertainty. Therefore, we assume:

1. \(\lambda_p > 0, \quad \lambda_n > 0, \quad \lambda_{m^\uparrow} > 0, \quad \lambda_{m^\downarrow} > 0;\)
2. \(\lambda_{m^\downarrow} \leq \lambda_n;\)
3. \(\lambda_{m^\uparrow} \leq \lambda_p.\)

Properties in (i) may be interpreted as normalization conditions. They simply suggest that cost are positive. Properties (ii) and (iii) state that \(\lambda_{m^\downarrow}\) and \(\lambda_{m^\uparrow}\) have the same or lower cost than \(\lambda_n\) and \(\lambda_p\), respectively.

Rule (P1): The first condition can be expressed as \((1 - \mu_A(x))\lambda_p \leq \mu_A(x)\lambda_n\). Since properties in (i) imply that \(\lambda_p + \lambda_n > 0\), we have \(\mu_A(x) \geq \lambda_p/(\lambda_p + \lambda_n)\).

The second condition of rule (P1) can be equivalently expressed as \((1 - \mu_A(x))\lambda_p \leq (\mu_A(x) - m)\lambda_{m^\downarrow}\). By properties in (i), we have \(\lambda_p + m\lambda_{m^\downarrow} > 0\). Thus, the condition can be expressed as \(\mu_A(x) \geq (\lambda_p + m\lambda_{m^\downarrow})/(\lambda_p + \lambda_{m^\downarrow})\). By assumption (ii), we have \((\lambda_p + m\lambda_{m^\downarrow})/(\lambda_p + \lambda_{m^\downarrow}) > \lambda_p/(\lambda_p + \lambda_n)\) and \((\lambda_p + m\lambda_{m^\downarrow})/(\lambda_p + \lambda_{m^\downarrow}) > m.\) Therefore, the condition of rule (P1) can be simply written as:

(P1) **If** \(\mu_A(x) \geq \alpha, \text{ then take } a_p, \text{ that is, } T_r(\mu_A)(x) = 1,\)

where

\[
\alpha = \frac{\lambda_p + m\lambda_{m^\downarrow}}{\lambda_p + \lambda_{m^\downarrow}}. \tag{3.30}
\]

Since \(\alpha > m\), the condition of \(\mu_A(x) \geq m\) is implied by \(\mu_A(x) \geq \alpha\), which is not needed in the rule.

Rule (M1): The first condition of rule (M1) is given by \((\mu_A(x) - m)\lambda_{m^\downarrow} < \mu_A(x)\lambda_n\). By property (ii), \(\lambda_n - \lambda_{m^\downarrow} \geq 0.\) Assuming that \(\lambda_n - \lambda_{m^\downarrow} \neq 0\), we have \(\mu_A(x) > -m\lambda_{m^\downarrow}/(\lambda_n - \lambda_{m^\downarrow}).\) With the properties in (i) and (ii), we know that \(-m\lambda_{m^\downarrow}/(\lambda_n - \lambda_{m^\downarrow}) >
\( \lambda_{m_\downarrow} < 0 \). Hence, \( \mu_A(x) > -m\lambda_{m_\downarrow}/(\lambda_n - \lambda_{m_\downarrow}) \) always holds. When \( \lambda_n - \lambda_{m_\downarrow} = 0 \), the condition becomes \( -m\lambda_m \downarrow < 0 \), which is true. Therefore, the first condition of rule (M1) always holds and does not need to be considered. The second condition of (M1) can be equivalently expressed as:

\[
C(a_{m_\downarrow}|\mu_A(x)) < C(a_P|\mu_A(x)) \iff \mu_A(x) < \alpha. \tag{3.31}
\]

By combining conditions \( \mu_A(x) \geq m > 0 \) and \( \mu_A(x) < \alpha \), we can express rule (M1) as:

(M1) \textbf{If} \( \mu_A(x) \geq m \land \mu_A(x) < \alpha \), \textbf{then} take \( a_{m_\downarrow} \), that is, \( T_r(\mu_A)(x) = m \).

Since \( \alpha > m \), we need to explicitly state the condition \( \mu_A(x) \geq m \).

Rule (N1): Based on the analysis of rules (P1) and (M1), assuming \( \lambda_n \neq \lambda_{m_\downarrow} \), the conditions for rule (N1) can be expressed as:

\[
\mu_A(x) \geq m \land \mu_A(x) < \frac{-m\lambda_{m_\downarrow}}{(\lambda_n - \lambda_{m_\downarrow})} \land \mu_A(x) \leq \frac{(\lambda_p + m\lambda_{m_\downarrow})}{(\lambda_p + \lambda_{m_\downarrow})}. \tag{3.32}
\]

Recall that \( -m\lambda_{m_\downarrow}/(\lambda_n - \lambda_{m_\downarrow}) < 0 \), which requires \( \mu_A(x) < 0 \). Thus, it is impossible to apply rule (N1) for reducing the membership grade. When \( \lambda_n = \lambda_{m_\downarrow} \), the conditions of rule (N1) can be expressed as:

\[
\mu_A(x) \geq m \land -m\lambda_{m_\downarrow} > 0 \land \mu_A(x) \leq \frac{1 + m}{2}. \tag{3.33}
\]

Since \( -m\lambda_{m_\downarrow} > 0 \) contradicts the assumption of positive costs, it is also not possible to use rule (N1).
The two remaining rules for Case 1 can be simply expressed as:

(P1) If \( \mu_A(x) \geq \alpha \), then take \( a_p \), that is, \( T_\tau(\mu_A)(x) = 1 \),

(M1) If \( m \leq \mu_A(x) < \alpha \), then take \( a_{m\downarrow} \), that is, \( T_\tau(\mu_A)(x) = m \).

A single threshold \( m < \alpha \leq 1 \) is used to decide either elevating a membership grade that equals to or is above \( m \) to 1 or reducing it to \( m \).

**Case 2** (\( \mu_A(x) < m \)): With respect to the three actions \( a_p \), \( a_{m\uparrow} \) and \( a_n \), we have:

(P2) If \( C(a_p|\mu_A(x)) \leq C(a_n|\mu_A(x)) \land C(a_p|\mu_A(x)) \leq C(a_{m\uparrow}|\mu_A(x)) \),
then take action \( a_p \), that is, \( T_\tau(\mu_A)(x) = 1 \),

(M2) If \( C(a_{m\uparrow}|\mu_A(x)) < C(a_n|\mu_A(x)) \land C(a_{m\uparrow}|\mu_A(x)) < C(a_p|\mu_A(x)) \),
then take action \( a_{m\uparrow} \), that is, \( T_\tau(\mu_A)(x) = m \),

(N2) If \( C(a_n|\mu_A(x)) \leq C(a_{m\uparrow}|\mu_A(x)) \land C(a_n|\mu_A(x)) < C(a_p|\mu_A(x)) \),
then take action \( a_n \), that is, \( T_\tau(\mu_A)(x) = 0 \).

By using an analysis similar to Case 1, we can obtain the following two simplified rules:

(M2) If \( \beta < \mu_A(x) < m \), then take \( a_{m\uparrow} \), that is, \( T_\tau(\mu_A)(x) = m \),

(N2) If \( \mu_A(x) < \beta \), then take \( a_n \), that is, \( T_\tau(\mu_A)(x) = 0 \),

where

\[
\beta = \frac{m\lambda_{m\uparrow}}{\lambda_n + \lambda_{m\uparrow}}. \tag{3.34}
\]

That is, we use a threshold \( \beta \) to decide either reducing to membership grade that equals to or is below \( m \) to 0 or elevating to \( m \).
By combining rules (P1), (M1), (M2) and (N2), we eventually have four rules:

(P) \[ \text{If } \mu_A(x) \geq \alpha, \text{ then take action } a_p, \text{ that is, } T_\tau(\mu_A)(x) = 1, \]

(M↓) \[ \text{If } m \leq \mu_A(x) < \alpha, \text{ then take action } a_m↓, \text{ that is, } T_\tau(\mu_A)(x) = m, \]

(M↑) \[ \text{If } \beta < \mu_A(x) < m, \text{ then take action } a_m↑, \text{ that is, } T_\tau(\mu_A)(x) = m, \]

(N) \[ \text{If } \mu_A(x) \leq \beta, \text{ then take action } a_n, \text{ that is, } T_\tau(\mu_A)(x) = 0, \]

where \(0 \leq \beta < m < \alpha \leq 1\). The required pair of thresholds of an \((\alpha, \beta)\)-transformation function is computed systematically from the costs of various actions.

In the special case when \(m = 0.5\), from Equations (3.30) and (3.34), we obtain the result of the original decision-theoretic three-way approximations of fuzzy sets [14]:

\[
\alpha = \frac{2\lambda_p + \lambda_m↓}{2(\lambda_p + \lambda_m↓)}, \quad \beta = \frac{\lambda_m↑}{2(\lambda_n + \lambda_m↑)}.
\] (3.35)

In the mean-value-based formulation [13], the \(m\) is the mean grade of all objects with \(0 < \mu_A(x) < 1\).

To derive a distance-based approximation, we assume that \(\lambda_n = \lambda_m↑ = \lambda_m↓ = \lambda_p\). According to Equations (3.30) and (3.34), we have:

\[
\alpha = \frac{1 + m}{2}, \quad \beta = \frac{m}{2}.
\] (3.36)

That is, distance-based approximation is a special case. Relationships between cost-
sensitive and distance-based formulations can also be seen as follows:

\[
C(T_{r}(\mu_{A})) = \sum_{\mu_{A}(x) \geq \alpha} (1 - \mu_{A}(x)) \lambda_{p} + \sum_{\mu_{A}(x) \leq \beta} \mu_{A}(x) \lambda_{n} \\
+ \sum_{\beta < \mu_{A}(x) \leq m} (m - \mu_{A}(x)) \lambda_{m^\uparrow} + \sum_{m < \mu_{A}(x) < \alpha} (\mu_{A}(x) - m) \lambda_{m^\downarrow} \\
= \text{Elevated Area}(T_{r}(\mu_{A})) \lambda_{p} + \text{Reduced Area}(T_{r}(\mu_{A})) \lambda_{n} \\
+ \text{Elevated Shadow} \ (T_{r}(\mu_{A})) \lambda_{m^\uparrow} + \text{Reduced Shadow} \ (T_{r}(\mu_{A})) \lambda_{m^\downarrow}.
\]

(3.37)

When \(\lambda_{m^\uparrow} = \lambda_{m^\downarrow} = \lambda_{m}\), the summation of the last two terms is \(\text{Shadow}(S_{(\alpha,\beta)}(\mu_{A})) \lambda_{m}\).

Comparing with the error-based computation given by Equation (3.26), the costs \(\lambda\)'s in \(C(T_{r}(\mu_{A}))\) indicates the advantages of a cost-sensitive approach.
Chapter 4

A MORE GENERAL
DISTANCE-BASED
FORMULATION

In the previous chapters, by taking \( n, m \) and \( p \) as three specific values in the unit interval \([0, 1]\), we in fact considered a subsystem of the unit interval \([0, 1]\). Consequently, a distance function can be easily introduced and interpreted. In this chapter, we may treat \( \{n, m, p\} \) to be a different system in general. This requires a semantic-based definition of distance.

4.1 Semantic Distance

In Section 2.3.3, we have given a general form of a three-way approximation of a fuzzy set \( \mu_A(x) \) as a three-valued set:

\[
T_{(\alpha,\beta)}(\mu_A)(x) = \begin{cases} 
  p, & \mu_A(x) \geq \alpha, \\
  m, & \beta < \mu_A(x) < \alpha, \\
  n, & \mu_A(x) \leq \beta. 
\end{cases} \quad (4.1)
\]
Figure 4.1: Three-way approximations of a fuzzy set by using three actions

That is, the membership grade $\mu_A(x)$ is transformed into one of three values in \{n, m, p\}. As shown in Figure 4.1, we take one of three actions $a_n, a_m$ and $a_p$. Since a different system is used, actions are interpreted in terms of mappings, instead of elevation and reduction. If $\mu_A(x) \geq \alpha$, we map $\mu_A(x)$ to $p$ by taking action $a_p$; if $\mu_A(x) \leq \beta$, we map $\mu_A(x)$ to $n$ by taking action $a_n$; if $\beta < \mu_A(x) < \alpha$, we map $\mu_A(x)$ to $m$ by taking action $a_m$.

To evaluate the quality of different three-way approximations, we need to consider the distance between a three-valued set and a fuzzy set. This can be achieved by introducing a distance function $D_s : [0, 1] \times \{n, m, p\} \rightarrow R^+$ between elements in \{n, m, p\}. We assume that with respect to the first argument, $D_s$ is a continuous function.

According to the argument given in Section 3.1, a semantic distance must be, at least, consistent with the truth orderings and information orderings in the two systems of [0, 1] and \{n, m, p\}. Such a requirement on consistency may be stated in terms of monotonicity of distance function.

With respect to the first argument, $D_s$ must satisfy the following monotonicity
properties:

\[(\text{MP}_1) \quad \mu_A(x_1) < \mu_A(x_2) \implies D_s(\mu_A(x_2), p) < D_s(\mu_A(x_1), p),\]
\[(\text{MP}_2) \quad \mu_A(x_1) < \mu_A(x_2) \implies D_s(\mu_A(x_1), n) < D_s(\mu_A(x_2), n),\]
\[(\text{MP}_3) \quad m \leq \mu_A(x_1) < \mu_A(x_2) \implies D_s(\mu_A(x_1), m) < D_s(\mu_A(x_2), m),\]
\[(\text{MP}_4) \quad \mu_A(x_1) < \mu_A(x_2) < m \implies D_s(\mu_A(x_2), m) < D_s(\mu_A(x_1), m).\]

As shown by Figure 4.1(a), for \( p \), monotonicity (\( \text{MP}_1 \)) requires that \( D_s(\cdot, p) \) is strictly monotonically decreasing with respect to membership grades, with the maximum value \( D_s(0, p) \) and the minimum value \( D_s(1, p) \). As shown in Figure 4.1(c), for \( n \), monotonicity (\( \text{MP}_2 \)) states that \( D_s(\cdot, n) \) is strictly monotonically increasing with respect to membership grades, which the maximum value is \( D_s(1, n) \) and the minimum value is \( D_s(0, n) \). For \( m \), we introduce a point \( m \in [0, 1] \) and consider a V-shaped function as shown in Figure 4.1(b). The two monotonicity properties (\( \text{MP}_3 \)) and (\( \text{MP}_4 \)) require that \( D_s(\cdot, m) \) is strictly monotonically decreasing in \([0, 1]\) and strictly monotonically increasing in \([m, 1]\), with the minimum value \( D_s(m, m) \). Since less than relation \(<\) on \([0, 1]\) is related to the truth ordering and information ordering, these monotonicity properties reflect the consistency of \( D_s \) with respect to two truth orderings and information orderings. For monotonicity with respect to the second argument, we consider:

\[(\text{MP}_5) \quad \text{For } \mu_A(x) \leq m, \quad D_s(\mu_A(x), m) < D_s(\mu_A(x), p),\]
\[(\text{MP}_6) \quad \text{For } \mu_A(x) > m, \quad D_s(\mu_A(x), m) < D_s(\mu_A(x), n).\]

That is, when \( \mu_A(x) \leq m, \mu_A(x) \) is semantically closer to \( m \) than to \( p \); when \( \mu_A(x) > m, \mu_A(x) \) is semantically closer to \( m \) than to \( n \). Monotonicity properties (\( \text{MP}_5 \)) and (\( \text{MP}_6 \)) are related to both truth ordering and informative ordering on \( \{n, m, p\} \).
In order to ensure that distances between membership grades in [0, 1] and membership grades in \{n, m, p\} are measured uniformly, we consider the following standardization conditions:

\[
(S_1) \quad D_s(0, n) = D_s(1, p) = D_s(m, m),
\]

\[
(S_2) \quad D_s(0, p) = D_s(1, n) > \max(D_s(0, m), D_s(1, m)).
\]

To some extent, the two conditions are consistent with an interpretation of \(n\) and \(p\) as two opposition, and \(m\) as the middle. Condition \((S_1)\) states that the minimum distances from \(n, m\) and \(p\), respectively to elements in \([0, 1]\) are the same. Condition \((S_2)\) states the maximum distance from \(n\) and \(p\), respectively, to elements in \([0, 1]\) are the same, and it is strictly greater than both \(D_s(0, m)\) and \(D_s(1, m)\). The inequality in \((S_2)\) is infant implied by \(0 < m < 1\), \((MP_5)\) and \((MP_6)\).

### 4.2 Construction of Three-Way Approximations

For an object \(x \in U\), the cost associated with an action \(\tau(x) \in \{a_n, a_m, a_p\}\) is computed by:

\[
C(\tau(x)|\mu_A(x)) = D_s(\mu_A(x), T_{\tau(x)}(\mu_A)(x))\lambda_{\tau(x)}.
\]  

(4.2)

The overall cost is given by:

\[
C(T_\tau(\mu_A)) = \sum_{x \in U} C(\tau(x)|\mu_A(x)).
\]

(4.3)
According to the principle of least cost, by minimizing the objective function \( C_s(\tau(x)|\mu_A(x)) \), we obtain an optimal transformation function \( \tau \):

\[
\arg \min_{\tau} C(T_\tau(\mu_A)). \tag{4.4}
\]

Again, the overall cost is a summation of costs for individual objects. We can also apply the general framework given in Section 3.1 by following the formulation of Section 3.3.

By using \( C(\tau(x)|\mu_A(x)) \) as a measure in rules (P), (M) and (N) in Section 3.1, we have:

\begin{align*}
\text{(P) } \quad & \text{If } C(a_p|\mu(x)) \leq C(a_n|\mu(x)) \land C(a_p|\mu(x)) \leq C(a_m|\mu(x)), \\
& \text{then take action } a_p, \text{ that is, } T_\tau(\mu_A(x)) = p, \\
\text{(M) } \quad & \text{If } C(a_m|\mu(x)) < C(a_n|\mu(x)) \land C(a_m|\mu(x)) < C(a_p|\mu(x)), \\
& \text{then take action } a_m, \text{ that is, } T_\tau(\mu_A(x)) = m, \\
\text{(N) } \quad & \text{If } C(a_n|\mu(x)) \leq C(a_m|\mu(x)) \land C(a_n|\mu(x)) < C(a_p|\mu(x)), \\
& \text{then take action } a_n, \text{ that is, } T_\tau(\mu_A(x)) = n.
\end{align*}

Let \( \lambda_n, \lambda_m \) and \( \lambda_p \) denote, respectively, the cost of actions \( a_n, a_m \) and \( a_p \). We can make assumptions similar to those in Section 3.4.2:

\begin{align*}
\text{(I) } & \lambda_p > 0, \lambda_m > 0, \lambda_n > 0, \\
\text{(II) } & \lambda_m \leq \lambda_p, \\
\text{(III) } & \lambda_m \leq \lambda_n.
\end{align*}

Properties in (I) may be interpreted as normalization conditions. They simply suggest that cost are positive. Properties (II) and (III) states that \( \lambda_m \) have the same or lower
Figure 4.2: Semantic distance between elements in $[0, 1]$ and $\{n, m, p\}$
cost than $\lambda_n$ and $\lambda_p$, respectively. Based on assumptions (I), (II) and (III), we simplify the three rules.

Rule (P): The two conditions can be expressed as $D_s(\mu_A(x), p)\lambda_p \leq D_s(\mu_A(x), n)\lambda_n$ and $D_s(\mu_A(x), p)\lambda_p \leq D_s(\mu_A(x), m)\lambda_m$, respectively. To further simplify the two conditions, we need to consider $D_s(\mu_A(x), p) = 0$ and $D_s(\mu_A(x), p) \neq 0$. For the case of $D_s(\mu_A(x), p) = 0$, by monotonicity (MP$_1$), we must have $\mu_A(x) = 1$. Since all cost must be positive, the two conditions are satisfied when $\mu_A(x) = 1$. If $D_s(\mu_A(x), p) \neq 0$, we have:

$$\frac{D_s(\mu_A(x), n)}{D_s(\mu_A(x), p)} \geq \frac{\lambda_p}{\lambda_n} \land \frac{D_s(\mu_A(x), m)}{D_s(\mu_A(x), p)} \geq \frac{\lambda_p}{\lambda_m}. \quad (4.5)$$

For the first condition, according to monotonicity properties (MP$_1$) and (MP$_2$), the assumption of positive costs, the normalization conditions (S$_1$) and (S$_2$), and $D_s(\cdot, T_\tau(\mu_A)(x))$ is continuous, there must exist a unique number $\gamma \in [0, 1]$, such that $D_s(\mu_A(x), p)\lambda_p = D_s(\mu_A(x), n)\lambda_n$ and

$$\frac{D_s(\mu_A(x), n)}{D_s(\mu_A(x), p)} \geq \frac{\lambda_p}{\lambda_n} \Leftrightarrow \mu_A(x) \geq \gamma. \quad (4.6)$$

For the second condition, we need to consider $\mu_A(x) > m$ and $\mu_A(x) \leq m$, separately. Assume that $\mu_A(x) > m$, according to monotonicity properties (MP$_3$) and (MP$_6$), normalization conditions (S$_1$) and (S$_2$), and $D_s(\cdot, T_\tau(\mu_A)(x))$ is continuous, there must exist a unique number $m < \alpha < 1$, such that $D_s(\mu_A(x), p)\lambda_p = D_s(\mu_A(x), m)\lambda_m$ and

$$\frac{D_s(\mu_A(x), m)}{D_s(\mu_A(x), p)} \geq \frac{\lambda_p}{\lambda_m} \Leftrightarrow \mu_A(x) \geq \alpha. \quad (4.7)$$

When $\mu_A(x) \leq m$, monotonicity (MP$_5$) suggests $D_s(\mu_A(x), m) < D_s(\mu_A(x), p)$. On the other hand, assumption (II) suggests $\lambda_m \leq \lambda_p$. It follows that $D_s(\mu_A(x), m)\lambda_m < D_s(\mu_A(x), p)\lambda_p$, which implies the first condition cannot hold. Therefore, for $\mu_A(x) \leq m$, the first condition is satisfied for all $x \in (4.6)$. On the other hand, assumption (II) suggests $\lambda_m \leq \lambda_p$. It follows that $D_s(\mu_A(x), m)\lambda_m < D_s(\mu_A(x), p)\lambda_p$, which implies the first condition cannot hold. Therefore, for $\mu_A(x) \leq m$, the first condition is satisfied for all $x$.

56
Consider now the relationship between $\gamma$ and $\alpha$. If $\gamma \leq m$, by the fact that $\alpha > m$, we can conclude that $\alpha > \gamma$. Suppose $\gamma > m$, we want to show that $\alpha \geq \gamma$. From the earlier discussion, we know that $\alpha$ is the solution to the equation $D_s(\mu_A(x), p)\lambda_p = D_s(\mu_A(x), m)\lambda_m$ and for $\mu_A(x) \geq \alpha$, $D_s(\mu_A(x), m)\lambda_m \geq D_s(\mu_A(x), p)\lambda_p$ holds. According to monotonicity (MP$_6$) and assumption (III), for $\mu_A(x) > m$, $D_s(\mu_A(x), n)\lambda_n > D_s(\mu_A(x), m)\lambda_m$. It follows that for $\mu_A(x) > \alpha$, $D_s(\mu_A(x), n)\lambda_n > D_s(\mu_A(x), p)\lambda_p$. One can conclude that the solution to equation $D_s(\mu_A(x), n)\lambda_n = D_s(\mu_A(x), p)\lambda_p$ cannot be greater than $\alpha$. That is, $\alpha \geq \gamma$. In summary, the condition $\mu_A(x) \geq \gamma \land \mu_A(x) \geq \alpha$ can be simplified into $\mu_A(x) \geq \alpha$. Hence, we have the following simplified rule:

(P) \textbf{If } \mu_A(x) \geq \alpha, \textbf{ then } \text{take action } a_p, \text{ that is, } T_\tau(\mu_A)(x) = p,

where the threshold $\alpha > m$ is the solution to equation $D_s(\mu_A(x), p)\lambda_p = D_s(\mu_A(x), m)\lambda_m$.

Given a semantic distance satisfies the required monotonicity and standardization conditions, its value can be easily computed.

\textbf{Rule (M):} The two conditions are $D_s(\mu_A(x), m)\lambda_m < D_s(\mu_A(x), n)\lambda_n$ and $D_s(\mu_A(x), m)\lambda_m < D_s(\mu_A(x), p)\lambda_p$, respectively. If $D_s(\mu_A(x), m) = 0$, by (MP$_3$) and (MP$_4$), we must have $\mu_A(x) = m$. According to the assumption of non-negative costs, the two conditions must hold. That is, we take action $a_m$ for $\mu_A(x) = m$. If $D_s(\mu_A(x), m) \neq 0$, we can simplify the conditions into:

$$\frac{D_s(\mu_A(x), n)}{D_s(\mu_A(x), m)} > \frac{\lambda_m}{\lambda_n} \land \frac{D_s(\mu_A(x), p)}{D_s(\mu_A(x), m)} > \frac{\lambda_m}{\lambda_p}. \tag{4.8}$$

For the first condition, we need to consider $\mu_A(x) > m$ and $\mu_A(x) \leq m$, separately. For $\mu_A(x) > m$, monotonicity (MP$_6$) suggests $D_s(\mu_A(x), m) < D_s(\mu_A(x), n)$. On the other hand, assumption (III) suggests $\lambda_m \leq \lambda_n$. It follows that $D_s(\mu_A(x), m)\lambda_m <
\( D_s(\mu_A(x), n) \lambda_n \), which implies the first condition holds. When \( \mu_A(x) \leq m \), according to the monotonicity properties (MP\(_2\)), (MP\(_4\)) and (MP\(_6\)), normalization conditions (S\(_1\)) and (S\(_2\)) and \( D_s(\cdot, T_\tau(\mu_A)(x)) \) is continuous, there must exist a unique number \( 0 < \beta < m \), such that \( D_s(\mu_A(x), n) \lambda_n = D_s(\mu_A(x), m) \lambda_m \) and

\[
\frac{D_s(\mu_A(x), n)}{D_s(\mu_A(x), m)} > \frac{\lambda_m}{\lambda_n} \iff \mu_A(x) > \beta.
\]

(4.9)

By combining the case of \( \mu_A(x) > m \) and \( \mu_A(x) \leq m \), the first condition can be simply expressed as \( \mu_A(x) > \beta, 0 < \beta < m \).

For the second condition, based on the analysis of Rule (P), we have

\[
\frac{D_s(\mu_A(x), p)}{D_s(\mu_A(x), m)} > \frac{\lambda_m}{\lambda_p} \iff \mu_A(x) < \alpha, \text{ where } 0 < \alpha < 1.
\]

(4.10)

The combination of two conditions leads to the following simplified rule:

\((\text{M})\) \hspace{1cm} \text{If } \beta < \mu_A(x) < \alpha, \text{ then take action } a_m, \text{ that is, } T_\tau(\mu_A)(x) = m, \)

where \( 0 < \beta < m \) is the solution to equation \( D_s(\mu_A(x), n) \lambda_n = D_s(\mu_A(x), m) \lambda_m \); and \( m < \alpha < 1 \) is the solution to equation \( D_s(\mu_A(x), p) \lambda_p = D_s(\mu_A(x), m) \lambda_m \).

\textbf{Rule (N):} If \( D_s(\mu_A(x), n) = 0 \), by monotonicity (MP\(_2\)), \( \mu_A(x) = 0 \). The two conditions of rule (N) must hold. That is, we take action \( a_n \) for \( \mu_A(x) = 0 \). If \( D_s(\mu_A(x), n) \neq 0 \), we can expressed the conditions in another way:

\[
\frac{D_s(\mu_A(x), m)}{D_s(\mu_A(x), n)} \geq \frac{\lambda_n}{\lambda_m} \land \frac{D_s(\mu_A(x), p)}{D_s(\mu_A(x), n)} > \frac{\lambda_n}{\lambda_p}.
\]

(4.11)

According to the analysis of rules (P) and (M), we know there exist two thresholds
$0 < \beta < m$ and $\gamma \in [0, 1]$ such that:

\[
\frac{D_s(\mu_A(x), m)}{D_s(\mu_A(x), n)} \geq \frac{\lambda_n}{\lambda_m} \iff \mu_A(x) \leq \beta, \quad \frac{D_s(\mu_A(x), p)}{D_s(\mu_A(x), n)} > \frac{\lambda_n}{\lambda_p} \iff \mu_A(x) < \gamma. \quad (4.12)
\]

Consider the relationship between $\gamma$ and $\beta$. If $\gamma > m$, by the fact that $\beta < m$, we can conclude that $\beta < \gamma$. The conditions of rule (N) becomes $\mu_A(x) \leq \beta$.

Suppose $\gamma \leq m$. Form the earlier discussion, we know that $\beta$ is the solution to equation $D_s(\mu_A(x), n)\lambda_n = D_s(\mu_A(x), m)\lambda_m$ and for $\mu_A(x) \leq \beta$, $D_s(\mu_A(x), n)\lambda_n < D_s(\mu_A(x), m)\lambda_m$ holds. According to monotonicity (MP$_5$) and assumption (II), for $\mu_A(x) \leq m$, $D_s(\mu_A(x), m)\lambda_m < D_s(\mu_A(x), p)\lambda_p$. We can conclude that the solution to equation $D_s(\mu_A(x), p)\lambda_p = D_s(\mu_A(x), n)\lambda_n$ cannot be less than $\beta$. That is $\beta \leq \gamma$. In summary, the condition $\mu_A(x) \leq \beta$ and $\mu_A(x) < \gamma$ can be simplified into $\mu_A(x) \leq \beta$. Finally, we have the following simplified rule:

\[
\text{(N) \quad If } \mu_A(x) \leq \beta, \text{ then take action } a_n, \text{ that is, } T_\tau(\mu_A)(x) = n,
\]

in which only one threshold is used.

By summarizing results of three rules, we can construct a transformation function $\tau$ by apply the following rules:

\[
\text{(P) \quad If } \mu_A(x) \geq \alpha, \text{ then take action } a_p, \text{ that is, } T_\tau(\mu_A)(x) = p,
\]

\[
\text{(M) \quad If } \beta < \mu_A(x) < \alpha, \text{ then take action } a_m, \text{ that is, } T_\tau(\mu_A)(x) = m,
\]

\[
\text{(N) \quad If } \mu_A(x) \leq \beta, \text{ then take action } a_n, \text{ that is, } T_\tau(\mu_A)(x) = n.
\]

Thus, we derive the general form of three-way approximation of a fuzzy set $\mu_A(x)$ defined by Equation (2.16).
Chapter 5

CONCLUSIONS AND FUTURE WORK

This thesis discusses three-way approximations of fuzzy sets. We examine formulations based on three principles and propose a more general formulation based on a semantic distance function. This chapter summarizes the main results of the thesis and points out potential future research.

5.1 Summary

We present a general framework for studying three-way, three-valued approximations of fuzzy sets. A three-way approximation is constructed by transforming a membership grade of an object to one of three grades \{n, m, p\}, where \(n\) and \(p\) denote two extreme membership grades and \(m\) denotes the middle membership grade. The problem of finding a best approximation is formulated as an optimization problem that minimizing the total degree of discrepancy between a three-way approximation and the original fuzzy set. We introduce a family of additive discrepancy measures and discuss a solution to obtain a pair of thresholds that leads to a minimum degree of discrepancy of each object. The problem of finding a pair of thresholds to get the
minimum total discrepancy is transformed to a problem of finding a pair of thresholds
to minimum individual object discrepancy.

We examine a class of transformation functions defined by a pair of thresholds
$(\alpha, \beta)$ with $0 < \beta < \alpha < 1$ on membership grades of a fuzzy set. More specifically,
if $\mu_A(x) \geq \alpha$, $\mu_A(x)$ is mapped to $p$; if $\mu_A(x) \leq \beta$, $\mu_A(x)$ is mapped to $n$; if $\beta < \\
\mu_A(x) < \alpha$, $\mu_A(x)$ is mapped to $m$. Within the general framework, we determine
an optimal pair of thresholds according to three principles namely, a principle of
uncertainty invariance, a principle of minimum distance, and a principal of least cost.
We examine the principle of uncertainty invariance under the formulation of shadowed
sets proposed by Pedrycz [29] and the formulation further developed by Tahayori,
Sadeghian and Pedrycz [36]. With repeat to the principle of minimum distance, we
introduce a distance-based formulation and obtain a pair of thresholds that leads
to a minimum distance. With respect to the principle of least cost, we look at the
decision-theoretic three-way approximation introduced by Deng and Yao [14]. Finally,
we provide a more general formulation based on the notion of a semantic distance
function. This thesis not only unifies existing studies on three-way approximation of
fuzzy sets, but also provides new results and additional insights.

Three-way approximation of fuzzy sets are a special model of three-way decision-
s [41]. The results reported in this thesis, in particularly the optimization based
framework, may be applied to studies on three-way decisions.

5.2 Future Work

The studies in the thesis can be extended in two directions. We can explore the use
of three principles introduced in Chapter 2 for three-way decision in general. We can
also study take the set-theoretic operators in the proposed framework.

An Exploration of Three Principles. In Chapter 3, we apply the three prin-
ciples to construct three-way approximations of fuzzy sets by optimization. Further studies on other models of three-way decisions, such as interval sets and quantitative rough sets may be made. We may explore more principles that can be used to construct three-way approximations of decision-theoretic rough set models and other three-way decision models.

A Study on Set-Theoretic Operators. In this thesis, we mainly talk about three-way approximations of fuzzy sets and the discrepancy measures between the original fuzzy set and the three-valued set. Further studies may be done to discuss the potential relation with the set-theoretic operators. For example, one may construct three-way approximations of the intersection of two fuzzy sets and examine its connection to the intersection of three-way approximations of two fuzzy sets.
REFERENCES


