Rough Set Models and Knowledge Acquisition for Imperfect Information Systems

Nguyen Do Van

A Dissertation submitted to
The Department of Information Science and Control Engineering
of Nagaoka University of Technology
in Partial Fulfilment of the Requirements for the Degree of
of Doctor of Engineering

Adviser: Professor Yamada Koichi

August 2014

Abstract

Rough set theory has been developed as a means to analyse vague description of objects. Objects characterized by attributes may be indiscernible based on the information available about them. Rough sets are approximation representations of a given set in the form of lower and upper approximations derived from crisp partitions. Rough set theory approach is important in the areas of machine learning, knowledge acquisition, decision analysis, and knowledge discovery from databases.

The original rough set approach is restricted to the case where objects attributes in information systems are described by precise values. Actual applications, however, often contain imperfect data including but not limited to missing, uncertain and imprecise values. Though numerous approach dealing with missing values have been published in the literature, lack of solutions to solve issues of uncertainty and imprecision still remains.

This research aims at proposing possible solutions for all of imperfect data mentioned above. The work contains introducing a representation of imperfect data, proposing two new rough set models and discussing methods for acquiring knowledge in imperfect information systems.

First of all, a representation of imperfect values is introduced. This representation must have ability to present any type of imperfect data. The solution chosen in this research is a combination of transforming missing, uncertain and imprecise values to probabilistic data.

Using the representation of imperfect data, a rough set model for imperfect data based on valued tolerance relations is then proposed. For this purpose, the research first suggests methods to obtain probabilities of matching - the probability that two objects are tolerant of each other on an attribute - for imperfect data. Combining

these probabilities with another index, we then propose a valued tolerance relation that can avoid problems stated in the literature of several rough set models.

The second rough set model proposed in this research is based on Dempster-Shafer theory. Several basic relations that are determined by comparing possible values sets of two objects on an attribute are first defined. Mass assignments for the occurrences of those basic relations are also calculated. Considering each attribute as a source of evidence and employing combination rules, mass assignments on a set of attributes are then determined. Last, based on belief and plausibility measures that are calculated from mass assignments for the occurrences of the basic relations, equivalence, tolerance and similarity relations among objects are defined.

Usually, a discernibility matrix is used to calculated reducts and core of an information system. To induce decision rules, an algorithm named LEM2 is a famous solution. However, it is evident that those approaches cannot be used in some cases. Therefore, finally, methods to obtain reducts and core and to induce decision rules in imperfect information systems are discussed.

Acknowledgements

First and foremost, I offer my sincerest gratitude to my supervisor, Professor YAMADA Koichi, who has supported me throughout my dissertation with his patience and knowledge whilst allowing me the room to work in my own way. His encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the research. I would also like to express my sincere thanks to the examiners of the dissertation, Prof.Fukumura, Prof.Yukawa, Prof.Marasinghe and Prof.Takahashi for their feedback to improve my dissertation

I am heartily thankful to Assistant Professor UNEHARA Muneyuki for his great advice in the research and comments throughout laboratory seminars, which assists me to find out the goal of the study.

I also would like to show my gratitude to The Ministry of Education, Culture, Sports, Science and Technology of Japan, who offers me the Monbukagakusho Scholarship. With the scholarship, I have an opportunity to study at Nagaoka University of Technology.

This dissertation would have remained a dream if I had not been provided useful lessons as well as permission to access the huge store of books and journals from the University. The staff of the University, especially Kokusai-ka, kindly helps me any time needed.

In my daily work I have been blessed with a friendly and cheerful group of the laboratory mates and friends in the University. They really provided me good arguments for my research.

Finally, I thank my parents for supporting me to come to Japan, my wife, my son and my little daughter for encouraging me to concentrate studying in Nagaoka.

DEDICATED

To my family

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Rough Sets in Knowledge Discovery

By the fast development of information technology, the volume of digital data is rapidly growing these days. It requires techniques that assist human to analyse and extract useful information from data. Using these techniques, previously unknown information, valid patterns and relationships in large data sets can be discovered. The information and knowledge extracted in this process can be used for applications ranging from market and investment analysis, fraud detection, and customer retention in production control to science exploration.

Knowledge Discovery in Database (KDD) is a process for extracting useful information (and it is also called Knowledge Acquisition). This process consists of some steps (Figure 1.1) including data preparation (selecting, cleaning, integration and transformation), data mining, and knowledge evaluation and representation. Basically, the process goes from data selection stage to knowledge representation stage. However, back track to previous stages may be sometimes required.

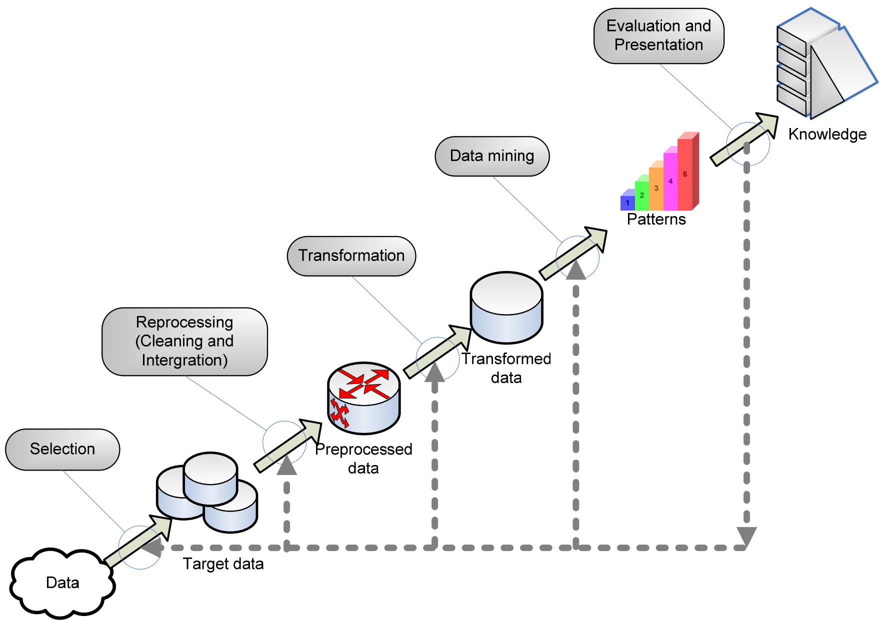

Figure 1.1: The Process of Knowledge Discovery in Database [16, 35]

Data mining is a particular step in KDD. Most data mining methods, as stated in [16, 17] are based on tried and tested techniques such as machine learning, pattern recognition, and statistics. The main goals of data mining [35] is to solve issues in classification, clustering, regression, feature extraction, association rules learning, summarization, etc. For dealing with problems in data mining various techniques are employed. Some examples of frequently used techniques are Decision tree [65], k-Nearest Neighbors algorithm [1], Naive Bayes [12], Support vector machine [9], Hierarchical, k-means [30], etc. The decision on what technique should be chosen depends on the requirement of solving issues.

Rough set approach is also a data mining solution for knowledge acquisition. Rough set theory was first introduced by Pawlak [60] in the early 1980s. It is a mathematical approach to present ambiguity, vagueness and uncertainty. Rough set theory is mainly used to analyse synthesize approximation of concepts from the acquired data. In knowledge discovery, rough set constitutes a sound basis by offering mathematical tools to discover patterns hidden in data. It can, according to [69], be used for feature selection, feature extraction, data reduction, decision rule generation, and pattern extraction.

The original model of rough set deals with correct and certain definition of objects in data sets based on equivalence relations. Two objects are considered as equivalent when their features are precisely equal to each other. Practical situations, however, are likely different. Bayes rough set [70, 71, 92] was introduced based on variable precision to allow some degree of uncertainty. A generalized definition of rough set [28, 88] was discussed for any relation rather than equivalence relation introduced in original rough set model. Fuzzy rough sets were also introduced [14, 15, 34, 40, 63, 84, 86] for approximating fuzzy sets which define members of a set in a range, rather than yes or no as it in original rough set. Moreover, instead of dealing with single table, there are also numerous research that are trying to expand the definition of rough set for relational databases that consist of multiple tables [32, 44, 75, 78].

Besides extending the definition of original rough set, scientists are working on solutions of defining relations between objects when data in information systems are not described by precise or discrete values. Finding solutions of dealing with missing value is most considerable concern of researchers in the field. Famous publications are found in [18, 21, 23, 26, 37, 38, 72, 73, 76, 83]. Information systems containing continuous values, apart from missing values, are dealt by approaches proposed in [5, 11, 56, 79]. Dai [10] and Guan [29] introduced fuzzy relation and maximal inclusion relation among objects in set-valued information system [58] in which an object attribute can be assigned with a set of values in attribute value domain.

1.2 Problem in Brief

Original rough set approach presupposes that all objects in an information system have precise and complete attribute values. Problems arise when information systems contain imperfect data including missing values, uncertainty and imprecision, which occasionally happens in the real world. It is thus necessary to develop a theory which may enable the classification of objects event if only partial information is available. Controversial rough set researches, however, mostly consider that imperfect data in information systems comes from missing values [18, 21, 23, 26, 37, 38, 72, 73, 76, 83]. Therefore, it may need a possible solution that could deal with multiple types of imperfect data.

1.3 Motivation of The Research

The goal of this research is studying rough set methods to archive knowledge in imperfect information systems. The research, therefore, targets to some objectives described as follows:

1. To introduce an representation of imperfect values.
2. To evaluate probability of matching between two object attribute values for an attribute. This probability can be used to define any valued tolerance/similarity relation.
3. To propose a new relation called extended tolerance relation in imperfect information systems.
4. To introduce a new relation based on Dempster-Shafter theory in imperfect information systems.

5. To obtain knowledge including reducts, cores and to induce decision rules from information systems which may have imperfect values.

1.4 Structure of The Dissertation

The rest of the dissertation is organized as follows:

Chapter 2 shows an overview of rough set theory. It starts with the definition and properties of the original model proposed by Pawlak. Two extended models (Bayes and Fuzzy) are then introduced. Generalized definitions of rough sets for an arbitrary binary relation among objects are also discussed.

Chapter 3 outlines some relations related to this study and points out what issues have not been solved. The main discussion is on incomplete information systems, which are mostly concerned in the field. The problem that object features are described partially is then discussed. That is object attributes not only being lost but also represented by uncertain or imprecision values. In order to solve the issue, this chapter also suggests a representation of imperfect data, which will be used for defining rough set models in the next two chapters.

Chapter 4 mainly discusses valued tolerance/similarity relation-based rough set approaches. A method for determining probability of matching between objects on an attribute is first proposed. This probability is then utilized for defining an extended tolerance relation in imperfect information systems.

Chapter 5 is for modelling rough set using Dempster-Shafer theory in imperfect information systems. The section starts with the notion of the theory of evidence. The tasks for defining the new rough set model are then step by step explained clearly.

Chapter 6 suggests methods to derive reducts and cores and to obtain decision rules in imperfect information systems using rough set models in Chapter 4 and 5.

The reason why methods discussed in the literature cannot be used is pointed out before introducing new techniques.

To end the dissertation, Chapter 7 makes a summary of this research with significant contributions and discusses unexpected limitations. Several open directions are also placed for the future work.

Some chapters of this dissertation have been published on international journal and presented in conference proceedings as written in the list of author's publications (Appendix A). A part of Chapter 3 is published in "*International Journal of Computer Applications, Vol. 89, No. 5, pp.1-8 (Mar. 2014)*". Chapter 4 is published in "*Proceeding of 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, The UK (2013)*", the probability of matching evaluation part of "*International Journal of Computer Applications, Vol. 89, No. 5, pp.1-8 (Mar. 2014)*", and the definition part of the extended rough set model in the journal of "*Advances in Fuzzy Systems, Volume 2013, Article ID 372091, (Oct 2013)*". Chapter 5 is published in "*Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 18, No. 3, pp.280-288, (May 2014)*". Chapter 6 is a combination of knowledge acquisition parts in "*International Journal of Computer Applications, Vol. 89, No. 5, pp.1-8 (Mar. 2014)*" and in the journal of "*Advances in Fuzzy Systems, Volume 2013, Article ID 372091, (Oct 2013)*".

Chapter 2

Overview of Rough Sets Theories

Originally, rough sets are defined in complete information systems in which object features are described by discrete and precise values. This chapter will first review rough set approach in Pawlak research [60]. We then survey some extensions of rough set definitions in probabilistic framework [92] and in the fuzzy set theory view point [13, 15]. There is also a discussion of original rough set for an arbitrary relation [28] instead of equivalence relation supposed in the first study of rough set theory.

2.1 Rough Set

2.1.1 Information Systems and Equivalence Relation

An information system is represented as a data table. Each row of this table represents an instance of an object such as people, things, etc. Information of every object is described by object attribute values.

An information system in the rough set study is formally defined as a pair $I = (U, A)$, where U is a non-empty finite set of objects called the universe and A is a non-empty finite set of attributes such that $f_a : U \rightarrow V_a$ for every $a \in A$ [60, 61]. The non-empty discrete value set V_a is called the domain of a. The original rough set theory deals with complete information systems in which $\forall x \in U$, $a \in A$, $f_a(x)$ is a precise value.

Any information system taking the form $I = (U, A \cup \{d\})$ is called a decision table where $d \notin A$ is called a decision and elements of A are called conditions. Let $V_d = \{d_1, ..., d_k\}$ denote the value set of the decision attribute, decision d then determines a set of partitions $\{C_1, C_2, ..., C_k\}$ of universe U, where $C_i = \{x \in U | f_d(x) = d_i\}$, $1 \leq i \leq k$. Set C_i is called the i-th decision class or concept on U. We assume that every object in U has a certain decision value in V_d.

An example of complete decision table is shown in Table 2.1. The universe is $U = \{x_1, x_2, \cdots, x_8\}$, the condition set is $A = \{Temperature, Headache, Nausea\}$ and decision d is Flu. In this table, x_1, x_4 and x_5 have exactly the same values on conditional attribute set $P = \{Temperature, Headache\}$. This case is (pair-wise) indiscernible using the available attributes. Based on equivalence/indiscernible relations among objects, the equivalence classes of all of the objects on P are $\{x_1, x_4, x_5\}$, $\{x_2\}$, $\{x_3\}$, $\{x_6, x_8\}$ and $\{x_7\}$.

Formally, in complete information systems, relation $EQU_P(x, y)$, $P \subseteq A$, denotes a binary relation between objects that are equivalent in terms of values of attributes in P [60]. The equivalence relation is reflexive, symmetric, and transitive. Let $E_P(x) = \{y \in U | EQU_P(y, x)\}$ be the set of all objects that are equivalent to x by P, which is then called an equivalence class. The family of all equivalence classes (or partitions) on U based on equivalence relation refers to as categories and is denoted by U/EQU_P.

[1] Source of table: [20].

Table 2.1: A complete decision table[1]

Cases	Condition			Decision
	Temperature	**Headache**	**Nausea**	**Flu**
x_1	high	yes	no	yes
x_2	very-high	yes	yes	yes
x_3	high	no	no	no
x_4	high	yes	yes	yes
x_5	high	yes	yes	no
x_6	normal	yes	no	no
x_7	normal	no	yes	no
x_8	normal	yes	no	yes

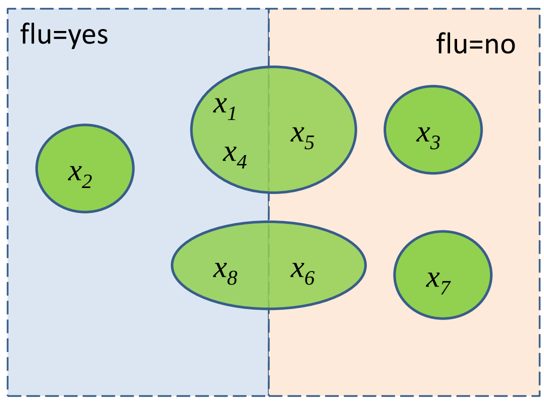

Figure 2.1: Approximating the set of patients using two conditional attributes *Temperature* and *Headache*

2.1.2 Approximation Space

Assume that we have to describe a group of patients $X \subseteq U$, who have flu in Table 2.1, by using conditional attribute subset P consists of $Temperature$ and $Headache$. From Figure 2.1, X cannot be exactly described in terms of P because the set may include or exclude objects that are indistinguishable on the basis of attributes P. For example, there is no way to represent set X by a set $\{x \in U | f_{Temperature}(x) = high \wedge f_{Headache}(x) = yes\}$. This is because x_1 and x_5 are equivalent to each other but x_5 in the concept of $Flu = no$. Target set X, however, can be approximated

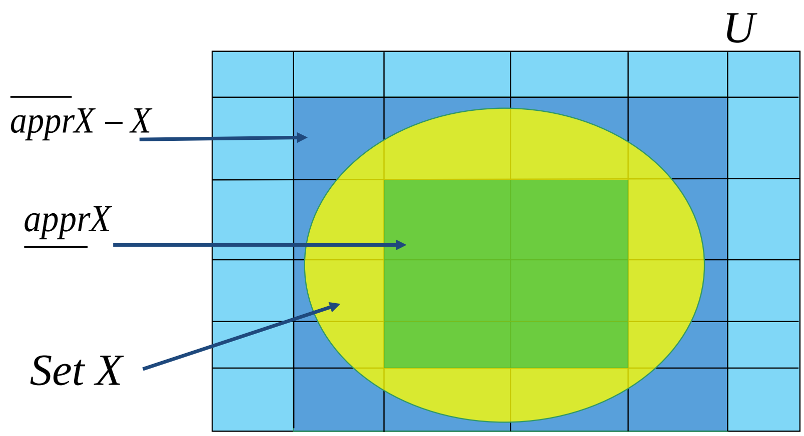

Figure 2.2: Illustrating approximations

using only the information contained within P by constructing the *lower* and *upper approximations* of X.

From equivalence classes, Pawlak [60, 61] defined an approximation space that contains lower and upper approximations denoted by $\underline{appr}X$ and $\overline{appr}X$, respectively, of set $X \subseteq U$ as follows:

$$\begin{aligned} \underline{appr}_P X &= \bigcup \{E_P(x) | x \in U, E_P(x) \subseteq X\} \\ &= \{x \in U | E_P(x) \subseteq X\} \qquad (2.1) \\ \overline{appr}_P X &= \bigcup \{E_P(x) | x \in U, E_P(x) \cap X \neq \emptyset\} \\ &= \{x \in U | E_P(x) \cap X \neq \emptyset\} \qquad (2.2) \end{aligned}$$

Set $bound_P X = \overline{appr}_P X - \underline{appr}_P X$ is named the *boundary* region of X. The set of $U - \overline{appr}_P X$ is called the outside region of X. Set X is said to be rough if the boundary region of X is none-empty. On the other hand, set X is crisp if the boundary of X is empty. Figure 2.2 illustrates universe U, object set X and approximations of X.

From the information system shown in Table 2.1, let $X = \{x \in U | f_{Flu}(x) = yes\}$ and $P = \{Temperature, Headach\}$. It can be inferred:

$$\underline{appr}_P X = \{x_2\},$$
$$\overline{appr}_P X = \{x_1, x_2, x_4, x_5, x_6, x_8\},$$
$$bound_P X = \{x_1, x_4, x_5, x_6, x_8\},$$
$$U - \overline{appr}_P X = \{x_3, x_7\}.$$

We get the following properties [61] of approximation space for any $X, Y \subseteq U$ directly from the definition of lower and upper approximations:

1. (a). $\underline{appr}(X) \subseteq X$,
   (b). $X \subseteq \overline{appr}(X)$,
2. (a). $\underline{appr}(\emptyset) = \emptyset$,
   (b). $\overline{appr}(\emptyset) = \emptyset$,
3. (a). $\underline{appr}(U) = U$,
   (b). $\overline{appr}(U) = U$,
4. (a). $X \subseteq Y \Rightarrow \underline{appr}(X) \subseteq \underline{appr}(Y)$,
   (b). $X \subseteq Y \Rightarrow \overline{appr}(X) \subseteq \overline{appr}(Y)$,
5. (a). $\underline{appr}(X \cup Y) \supseteq \underline{appr}(X) \cup \underline{appr}(Y)$,
   (b). $\overline{appr}(X \cup Y) = \overline{appr}(X) \cup \overline{appr}(Y)$,
6. (a). $\underline{appr}(X \cap Y) = \underline{appr}(X) \cap \underline{appr}(Y)$,
   (b). $\overline{appr}(X \cap Y) \subseteq \overline{appr}(X) \cap \overline{appr}(Y)$,
7. (a). $\underline{appr}(\underline{appr}(X)) = \underline{appr}(X) = \overline{appr}(\underline{appr}(X))$,
   (b). $\overline{appr}(\overline{appr}(X)) = \overline{appr}(X) = \underline{appr}(\overline{appr}(X))$,
8. (a). $\underline{appr}(X) = \sim \overline{appr}(\sim X)$,
   (b). $\overline{appr}(X) = \sim \underline{appr}(\sim X)$,

where $\sim X$ denotes a complementary set of X.

2.1.3 Rough Sets

The pair $\underline{appr}_P X$ and $\overline{appr}_P X$ composed of the lower and upper approximation is called a ***Rough Set***. The *accuracy* of the rough-set representation [60] of set X can be given as follows:

$$\alpha_P(X) = \frac{|\underline{appr}_P X|}{|\overline{appr}_P X|} \tag{2.3}$$

where $X \neq \emptyset$, $|Y|$ denotes the cardinality of a set Y.

The accuracy of the rough set representation of set X is the ratio of the number of object completely in X and the number of objects possibly belonging to X. Since $\emptyset \subseteq \underline{appr}_P X \subseteq \overline{appr}_P X$, we have $0 \leq \alpha_P(X) \leq 1$. The accuracy is a measurement of the how closely the rough set is approximating set X.

Set X is crisp with respect to P if the lower and upper approximations of X are equal. In this case, the boundary region is empty and $\alpha_P(X) = 1$. Set X, on the other hand, is *rough* with respect to P if the boundary region is not empty. In this case $\alpha_P(X) < 1$.

From lower and upper approximations of a set, rough set can be categorized as the following basic classes of rough sets, i.e., four categories of vagueness [60]:

- Set X is *roughly definable* if $\underline{appr}_P X \neq \emptyset$ and $\overline{appr}_P X \neq U$. This means that on attribute set P, there are objects which we can be certain belong to target set X, and there are also objects which we can definitively exclude from set X.

- Set X is *internally definable* if $\underline{appr}_P X \neq \emptyset$ and $\overline{appr}_P X = U$. This means that on attribute set P, there are objects which we can be certain belong to target set X, but there are no objects which we can definitively exclude from set X.

- Set X is *externally definable* if $\underline{appr}_P X = \emptyset$ and $\overline{appr}_P X \neq U$. This means that on attribute set P, there are no objects which we can be certain belong to target set X, but there are objects which we can definitively exclude from set X.

- Set X is *totally non-definable* if $\underline{appr}_P X = \emptyset$ and $\overline{appr}_P X = U$. This means that on attribute set P, there are no objects which we can be certain belong to target set X, and there are no objects which we can definitively exclude from set X. Thus, on attribute set P, we cannot decide whether any object is, or is not, a member of X.

In Table 2.1, the set $X = \{x \in U | f_{Flu}(x) = yes\}$ is roughly definable because of $\underline{appr}_P X \neq \emptyset$ and $\overline{appr}_P X \neq U$.

2.1.4 Reducts and Core

In information system $I = (U, A)$, a subset of conditional attributes set $P \subseteq A$ is called a reduct if the equivalence classes induced by P are the same as the equivalence classes induced by A and no attribute can be removed from set P without changing the family of equivalence classes U/EQU_A.

Set of attributes P formally is a reduct if and only if $U/EQU_P = U/EQU_A$ and $U/EQU_{P-\{a\}} \neq U/EQU_A$ for any attribute $a \in P$. Reducts of an information system are not unique. There can be several reducts if those attributes can preserve the family of equivalence classes of the information system.

A core is a set of attributes that consist of all necessary attributes for reducts. If a conditional attribute is removed from the core, this leads to a collapse in equivalence classes induced by any reduct. The core possibly is empty. In such case, there is no indispensable attribute.

2.1.5 Attribute Dependency

The next important aspect of rough set analysis is determining dependencies of attribute sets on other sets. A set of attributes P totally depends on a set of attributes Q, denoted as $Q \Rightarrow P$, if all values of attributes from P are uniquely determined by values of attributes from Q. In information system $I = (U, A)$, let

$U/EQU_P = \{E_P^i\}$, E_P^i is an equivalence class on U with respect to P, the dependency between $P, Q \subseteq A$ [61] can be stated as follows:

1. P depends on Q if and only if for any E_P^i exist E_Q^j such that $E_P^i \supseteq E_Q^j$.
2. P and Q are equivalent if and only if $P \Rightarrow Q$ and $Q \Rightarrow P$.
3. P and Q are independent if and only if neither $P \Rightarrow Q$ nor $Q \Rightarrow P$ hold.

To measure degrees of dependencies, a co-efficiency is introduced. The co-efficiency κ is defined as the degree that P depends on Q.

$$\kappa = \gamma(Q, P) = \sum_i \frac{|\underline{appr}_Q E_P^i|}{|U|}. \tag{2.4}$$

It can be said that P depends on Q in a degree κ, $(0 \leq \kappa \leq 1)$. If $\kappa = 1$, P depends totally on Q, and if $\kappa < 1$, P depends partially in a degree κ on Q. Obviously, P depends totally on Q when $\cup\{(x,y)|x,y \in U, EQU_Q(x,y)\} \subseteq \cup\{(x,y)|x,y \in U, EQU_P(x,y)\}$.

2.2 Bayes Rough Set

In the previous part, rough set theory [60, 61] is used as a tool to correctly and certainly derive classifications in information systems. However, most of practical data mining problems require identification of probabilistic pattern in data, typically in the form of probabilistic rules. In this part, an approach in which original rough set model is softened to allow some degree of uncertainty will be introduced. First, variable precision rough set model [92] is generated as an extension of original rough set model with parametric definitions of lower and upper approximations. Then it is modified to Bayesian rough set model [70, 71], which uses probability of observed set so-called prior probability as a parameter to estimate the chances of interesting events occur. The approaches are step by step introduced according to [70, 71, 92].

2.2.1 Variable Precision Model

Majority inclusion relation: Set $X \subseteq U$ is said to be included in set $Y \subseteq U$ if for all $x \in X$ implies $x \in Y$. Any object of X is, in other words, absolutely classified into Y if $X \subseteq Y$. There is thus no misclassification according to this definition. In real applications, however, it may be acceptable to allow some degree of misclassification. In order to do so, a measure $\zeta(X, Y)$ of the relative degree of misclassification of the set X with respect to set Y [92] is introduced as follows:

$$\zeta(X,Y) = \begin{cases} 1 - \dfrac{|X \cap Y|}{|X|} & \text{if } |X| > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

Clearly, X is included in Y if there is no misclassification. Formally,

$$Y \supseteq X \textit{ if only if } \zeta(X,Y) = 0.$$

Majority inclusion relation, however, allows some degree of misclassification. It is stated in [92] that the majority requirement implies that more than 50% of X elements should be in common with Y. The specified majority requirement imposes an additional restriction. The number of elements of X in common with Y should be not below a certain limit, for example 80%. These requirements may be added to the extended definition of inclusion relation by specifying an explicit limitation on a admissible level of classification error β must be within the range $0 \leq \beta < 0.5$. Formally, the β-majority inclusion relation is defined as follows:

$$Y \overset{\beta}{\supseteq} X \text{ if only if } \zeta(X,Y) \leq \beta. \tag{2.6}$$

Two useful properties of β-majority inclusion, according to the above definition, are listed as follows:

Proposition 2.2.1. *If $Y \cap Z = \emptyset$ and $Z \overset{\beta}{\supseteq} X$ then it is not true that $Y \overset{\beta}{\supseteq} X$.*

Proposition 2.2.2. *If $\beta_1 < \beta_2$ and $Y \overset{\beta_1}{\supseteq} X$ implies $Y \overset{\beta_2}{\supseteq} X$.*

See reference [92] for the proofs of the these propositions.

Set approximation in the VP-model: In information system $I = (U, A)$, based on equivalence EQU_P, $P \subseteq A$, universal U is partitioned into a collection of equivalence classes $U/EQU_P = \{E^1, E^2, \cdots, E^n\}$. In original rough set definition, lower approximation of set $X \subseteq U$ is a union of E^i such that E^i is included in set X. If β-majority inclusion relation is used instead of the tradition inclusion relation, the following generalized notion of β-lower approximation or β-positive region of the set $X \subseteq U$ can be obtained:

$$\underline{appr}_P^{\beta} X = \bigcup_{i=1..n} \{E^i \in U/EQU_P | X \overset{\beta}{\supseteq} E^i\} = \bigcup_{i=1..n} \{E^i \in U/EQU_P | \zeta(E^i, X) \leq \beta\}. \tag{2.7}$$

The β-upper approximation of set $X \subseteq U$ is defined as:

$$\overline{appr}_P^{\beta} X = \bigcup_{i=1..n} \{E^i \in U/EQU_P | \zeta(E^i, X) < 1 - \beta\}. \tag{2.8}$$

The β-boundary of set $X \subseteq U$ is defined as:

$$BN_P^{\beta} X = \bigcup_{i=1..n} \{E^i \in U/EQU_P | \beta < \zeta(E^i, X) \leq \beta\}. \tag{2.9}$$

The β-negative region of set $X \subseteq U$ is defined as a complement of the β-upper approximation

$$NEG_P^{\beta}X = \bigcup_{i=1..n} \{E^i \in U/EQU_P | \zeta(E^i, X) \geq 1 - \beta\}. \tag{2.10}$$

Directly from the above definitions, lower approximation of set X is the union of equivalence classes that can be included in X with misclassification degree less than β, while upper approximation is the collection of equivalence classes can be included in complement set $\sim X$ of X. If $\beta = 0$ then the original rough set model becomes a special case of VP-model because the majority inclusion relation becomes the original inclusion relation in this case.

Example 2.2.1. This example from [92] illustrates the variable precision rough set model. The universe $U = \{x_1, x_2, \cdots, x_{20}\}$ and the equivalence classes of the equivalence relation on attribute set A are shown as follows:

$$\begin{aligned}
E^1 &= \{x_1, x_2, x_3, x_4, x_5\}, \\
E^2 &= \{x_6, x_7, x_8\}, \\
E^3 &= \{x_9, x_{10}, x_{11}, x_{12}\}, \\
E^4 &= \{x_{13}, x_{14}\}, \\
E^5 &= \{x_{15}, x_{16}, x_{17}, x_{18}\}, \\
E^6 &= \{x_{19}, x_{20}\}.
\end{aligned}$$

Approximations of set $X = \{x_4, x_5, x_8, x_{14}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}\}$ for two accuracy levels $\beta_1 = 0$ and $\beta_2 = 0.25$ are derived as follows:

$$
\begin{aligned}
\underline{appr}_P^0 X &= E^6,\\
\overline{appr}_P^0 X &= E^1 \cup E^2 \cup E^4 \cup E^5 \cup E^6,\\
BN_P^0 X &= E^1 \cup E^2 \cup E^4 \cup E^5,\\
NEG_P^0 X &= E^3,\\
\underline{appr}_P^{0.25} X &= E^5 \cup E^6,\\
\overline{appr}_P^{0.25} X &= E^1 \cup E^2 \cup E^4 \cup E^5 \cup E^6,\\
BN_P^{0.25} X &= E^1 \cup E^2 \cup E^4,\\
NEG_P^{0.25} X &= E^3.
\end{aligned}
$$

2.2.2 Rough Set in Probabilistic Framework

Probabilistic frame work: Suppose that a prior probability function $\mathcal{P}(X) = |X|/|U|$ exists for every subset X of universal U, any subset of U that will be considered in this subsection then possibly occur with a uncertain degree $0 < \mathcal{P}(X) < 1$. It is assumed that an equivalence relation on U also exists with a finite set of equivalence classes $E^i \subseteq U/EQU_P$, such that $\mathcal{P}(E^i) > 0$. For each equivalence class E^i, we assign a conditional probability $\mathcal{P}(X|E^i) = |X \cap E^i|/|E^i|$. An extension of rough set model called Variable Precision Rough Set *(VPRS)* will be defined based on the notion of this frame work.

Rough set in probabilistic frame work: Directly from the conditional probability definition, E^i is a subset of set X if $\mathcal{P}(X|E^i) = 1$. The original rough set model is thus defined in the notion of probabilistic frame work [70] as follows:

$$POS(X) = \bigcup_i \{E^i | \mathcal{P}(X|E^i) = 1\}, \tag{2.11}$$

$$NEG(X) = \bigcup_i \{E^i | \mathcal{P}(X|E^i) = 0\}, \tag{2.12}$$

$$BN(X) = \bigcup_i \{E^i | \mathcal{P}(X|E^i) \in (0,1)\}. \tag{2.13}$$

Variable precision rough set (VPRS): It is stated in [71] that approximation space in the **VP**-model [92] is an extension of the rough set model aimed at increasing the discriminatory capabilities of the rough set approach by using parameter-controlled grades of conditional probabilities. This notion of VPRS [70, 71] is based on the lower and upper limit certainty thresholds l and u when defining approximation regions, satisfying $0 \leq l < \mathcal{P}(X) < u \leq 1$. Following is the definitions of approximation sets defined in [71, 93].

The u-positive region $POS^u(X)$ is controlled by the upper limit parameter u, which reflects the least acceptable degree of the conditional probability $\mathcal{P}(X|E^i)$ to include elementary set E^i in $POS^u(X)$:

$$POS^u(X) = \bigcup_i \{E^i|\mathcal{P}(X|E^i) \geq u\}. \tag{2.14}$$

The l-negative region $NEG^l(X)$ is controlled by the lower limit l, such that $0 \leq l < \mathcal{P}(X)$. $NEG^l(X)$ is an area where the occurrence of X is significantly - with respect to l - less likely than random guess $\mathcal{P}(X)$.

$$NEG^l(X) = \bigcup_i \{E^i|\mathcal{P}(X|E^i) \leq l\}. \tag{2.15}$$

The l-negative region $NEG^l(X)$ can be expressed as the $(1-l)$-positive region $POS^{(1-l)}(\sim X)$, where $\sim X$ is the complement of set X. The last considered region is (l, u)-boundary region, which is a *grey* area where there is no sufficient probabilistic bias towards neither X nor X.

$$BN^{l,u}(X) = \bigcup \{E|\mathcal{P}(X|E) \in (l, u)\}. \tag{2.16}$$

VPRS model, therefore, has ability to make approximation regions more flexible by using threshold l, u to control the acceptable degree of conditional probability of each equivalence class E^i on X. Original rough set model is clearly a special case of VPRS model, for $l = 0$ and $u = 1$.

Bayesian rough set model: In VPRS model, users have to determine what threshold is used to control degree of the acceptable degree of conditional probability. In [71, 93], the VPRS was modified to a so-called Bayesian rough set model (BRS). In this approach, prior probability is used as the parameter to control model derivation.

According to [93], the BRS positive region $POS^*(X)$ defines an area combined by equivalence classes where the conditional probability of each class is higher than the prior probability. The BRS negative region $NEG^*(X)$ defines an area of the universe formed by equivalence classes where the conditional probability of each classes is lower than the prior probability. The positive, negative and boundary regions are respectively defined as follows:

$$POS^*(X) = \bigcup_i \{E^i | \mathcal{P}(X|E^i) \geq \mathcal{P}(X)\}, \tag{2.17}$$

$$NEG^*(X) = \bigcup_i \{E^i | \mathcal{P}(X|E^i) \leq \mathcal{P}(X)\}, \tag{2.18}$$

$$BN^*(X) = \bigcup_i \{E^i | \mathcal{P}(X|E^i) = \mathcal{P}(X)\}. \tag{2.19}$$

Returning to Example 2.2.1, we have, the prior probability $\mathcal{P}(X) = 9/20 = 0.45$ and take it as the parameter of the set approximations. For each equivalence class $E^1, \cdots, E^6$, we have conditional probabilities respectively: 0.4, 0.33, 0.0, 0.5, 0.75, 1.0.

Set approximations are thus derived as follows:

$$\begin{aligned} POS^*(X) &= E^4 \cup E^5 \cup E^6, \\ NEG^*(X) &= E^1 \cup E^2 \cup E^3, \\ BN^*(X) &= \emptyset. \end{aligned}$$

By softening, Bayesian rough set can be applied effectively for data mining applications where acquisition of probabilistic rather than deterministic. A possible application was mentioned in [57]. In this research, authors applied the rough set method to Kansei engineering [46, 47] to develop customer oriented products. In this system, relational rules of embodying design attribute of products and human evaluation data such as sensory perception and feeling is used to extract human decision rules. As a solution, Bayesian rough set model was used because this model is much suitable for dealing with practical human evaluation data involving ambiguity of inconsistency.

2.3 Fuzzy Rough Set

Fuzzy rough sets were introduced by Dubois and Prade in 1990 [14, 15] as a fuzzy generalization of rough set. Recently, conventional issues of combining rough and fuzzy sets are discussed in many papers [34, 40, 63, 84, 86]. Fuzzy sets allow membership of elements in approximation sets in range rather than only yes or no in the original rough set model.

2.3.1 Fuzzy Set

A classical (crisp) set is normally defined as a collection of elements $x \in X$ that can be finite, countable or over countable. Each single element can be either belong to or not belong to a set $X' \subseteq X$. For a fuzzy set, a characteristic function allows various degrees of membership for elements of a given set. Then a fuzzy set $\mathcal{X}$ in X is a set of ordered pairs [94]:

$$\mathcal{X} = \{(x, \mu_{\mathcal{X}}(x)) | x \in X\}. \tag{2.20}$$

$\mu_{\mathcal{X}}(x)$ is called the membership function or grade of membership (also the degree of compatibility or degree of truth) of x in $\mathcal{X}$.

A fuzzy set [63] $\mathcal{X}$ on X is defined by a membership function $\mu_{\mathcal{X}} : X \rightarrow [0,1]$. A crisp set can be regarded as a special case of fuzzy sets in which the membership function is restricted to the extreme points $\{0,1\}$ of $[0,1]$. From the view point of fuzzy system, the next subsection will define fuzzy approximations of a fuzzy set based on a fuzzy relation between objects.

2.3.2 Rough Membership Functions

Let U denote a finite and non-empty set called the universe, and let R denote an equivalence relation EQU_P on U. R obviously is a reflexive, symmetric and transitive relation. If two objects x, y in U belong to the same equivalence class, i.e., we say that they are indistinguishable. The equivalence relation R partitions the set U into disjoint subsets. It defines the quotient set U/R consisting of equivalence classes of U on R.

Let $\underline{apprR}X$ and $\overline{apprR}X$ denote lower and upper approximation of $X \subseteq U$ on R. Those are called strong and weak membership functions of a rough set. Let μ_X and μ_R denote the membership functions of set X and of the set $\{(x,y) \in U \times U | R(x,y)\}$, respectively. The lower and upper approximations can be defined in the form of membership functions of object sets [84] as follows:

$$\mu_{\underline{apprR}X}(x) = \inf\{\mu_X(y) | y \in U, R(x,y)\}, \tag{2.21}$$
$$\mu_{\overline{apprR}X}(x) = \sup\{\mu_X(y) | y \in U, R(x,y)\}. \tag{2.22}$$

Approximations can also be defined in the form of membership functions of relations among objects.

$$\mu_{\underline{apprR}X}(x) = \inf\{1 - \mu_R(x,y) | y \notin X\}, \tag{2.23}$$
$$\mu_{\overline{apprR}X}(x) = \sup\{\mu_R(x,y) | y \in X\}. \tag{2.24}$$

For two special set $\emptyset$ and U the approximations are simply defined as $\mu_{\underline{R}U}(x) = 1$ and $\mu_{\overline{R}\emptyset}(x) = 0$. Based on the two equivalent definitions, lower and upper approximations may be interpreted as follows: An element x belongs to the lower approximation $\underline{apprR}X$ if all elements equivalent to x belong to X. In other words, x belongs to the lower approximation of X if any element not in X is not equivalent to x, namely, $\mu_R(x,y) = 0$. Likewise, x belongs to the upper approximation of X if $\mu_R(x,y) = 1$.

The weak and strong membership functions, thus, can be computed from the membership function of the reference set X and the set of pair (x,y) with relation R. For convenience, the strong and weak membership functions of a rough set can also be expressed [84] as follows:

$$\mu_{\underline{apprR}X}(x) = \inf\{\max(\mu_X(y), 1 - \mu_R(x,y)) | y \in U\}, \tag{2.25}$$
$$\mu_{\overline{apprR}X}(x) = \sup\{\min(\mu_X(y), \mu_R(x,y)) | y \in U\}. \tag{2.26}$$

2.3.3 Combination of Rough and Fuzzy Sets

The notion of rough fuzzy sets [14, 15] deals with the approximation of fuzzy sets in approximation space. A fuzzy relation $\mathcal{R}$ is a fuzzy subset on $U \times U$, let $\mu_{\mathcal{R}}$ denote the membership functions of the set of pair $(x,y) \in U \times U$ such that x in relation to y with respect to relation $\mathcal{R}$. A fuzzy relation may have three properties:

reflexivity: for all $x \in U$, $\mu_{\mathcal{R}}(x,x) = 1$,

symmetry: for all $x, y \in U$, $\mu_{\mathcal{R}}(x,y) = \mu_{\mathcal{R}}(y,x)$, transitivity:for all $x, y, z \in U$, $\mu_{\mathcal{R}}(x, z) \geq \min[\mu_{\mathcal{R}}(x, y), \mu_{\mathcal{R}}(y, z)]$.

The relation R will define the fuzzy equivalence class $[x]_{\mathcal{R}}$ of elements closing to x and the fuzzy equivalence class can be defined as follows:

$$\mu_{[x]_{\mathcal{R}}}(y) = \mu_{\mathcal{R}}(x, y). \tag{2.27}$$

For a fuzzy set $\mathcal{X}$, its approximations are called fuzzy rough set [14] and can be defined as follows:

$$\mu_{\underline{appr}\mathcal{R}\mathcal{X}}([x]_{\mathcal{R}}) = \inf\{\max[\mu_{\mathcal{X}}(y), 1 - \mu_{[x]_{\mathcal{R}}}(y)] | y \in U\}, \tag{2.28}$$
$$\mu_{\overline{appr}\mathcal{R}\mathcal{X}}([x]_{\mathcal{R}}) = \sup\{\min[\mu_{\mathcal{X}}(y), \mu_{[x]_{\mathcal{R}}}(y)] | y \in U\}. \tag{2.29}$$

They can be extended to a pair of fuzzy sets on the universe:

$$\mu_{\underline{appr}\mathcal{R}\mathcal{X}}(x) = \inf\{\max[\mu_{\mathcal{X}}(y), 1 - \mu_{\mathcal{R}}(x, y)] | y \in U\}, \tag{2.30}$$
$$\mu_{\overline{appr}\mathcal{R}\mathcal{X}}(x) = \sup\{\min[\mu_{\mathcal{X}}(y), \mu_{\mathcal{R}}(x, y)] | y \in U\}. \tag{2.31}$$

2.4 Rough Set on An Arbitrary Relation

In the previous two sections, we have discussed some methods of defining approximations of a set $X \subseteq U$ based on the majority inclusion notion of partitions to the set X and based on a model called Fuzzy Rough sets. Fuzzy sets allow the membership of elements in approximation sets in range rather than only yes or no in the original rough set model.

Those approximation methods are defined based on equivalence relation which is reflexive, symmetric and transitive. For the original rough set model, lower and upper approximations based on equivalence relation should be the same among the different three definitions: singleton, subset and concept definitions [60, 61]. In the case of non-equivalence relations, which may not be reflexive, symmetric nor transitive, approximation spaces defined by these methods may lead to variant results [28]. In this section, we will discuss approximations based on singleton, subset and concept approaches [28] with an arbitrary binary relation.

In information system $I = (U, A)$, let R be an arbitrary binary relation, the relation of object x with y, $x, y \in U$, with respect to attribute set $P \subseteq A$ is denoted by $R_P(x, y)$. For each object x, we define a neighbourhood that consists of successor and predecessor sets [28, 85]. Interpreting relation R as similarity, the successor set of x is the set of objects to which x is similar:

$$sucR_P(x) = \{y \in U | R_P(x, y)\} \tag{2.32}$$

The predecessor set of x is the set of objects which is similar to x:

$$preR_P(x) = \{y \in U | R_P(y, x)\} \tag{2.33}$$

Now, approximations based on singleton, subset and concept approaches with an arbitrary relation can be defined as follows:

Singleton lower approximation:

$$\underline{SingleAppr}_P(X) = \{x \in U | setR_P(x) \subseteq X\} \tag{2.34}$$

Singleton upper approximation:

$$\overline{SingleAppr}_P(X) = \{x \in U | setR_P(x) \cap X \neq \emptyset\} \tag{2.35}$$

Subset lower approximation:

$$\underline{SubsetAppr}_P(X) = \bigcup\{setR_P(x)|x \in U \wedge setR_P(x) \subseteq X\} \tag{2.36}$$

Subset upper approximation:

$$\overline{SubsetAppr}_P(X) = \bigcup\{setR_P(x)|x \in U \wedge setR_P(x) \cap X \neq \emptyset\} \tag{2.37}$$

Concept lower approximation:

$$\underline{ConceptAppr}_P(X) = \bigcup\{setR_P(x)|x \in X \wedge setR_P(x) \subseteq X\} \tag{2.38}$$

Concept predecessor upper approximation:

$$\overline{ConceptAppr}_P(X) = \bigcup\{setR_P(x)|x \in X \wedge setR_P(x) \cap X \neq \emptyset\} \tag{2.39}$$

where $setR_P(x)$ denotes either successor and predecessor neighbourhood sets of x.

The difference between subset and concept definitions may be missed easily. In subset definition, extended tolerance classes of all elements in the universal set are examined, while only elements in X are examined in the case of concept definition.

Obviously, singleton lower and upper approximations of X are subsets of the subset lower and upper approximations of X, respectively. The subset lower approximation is the same set as the concept lower approximation. The concept upper approximation, however, is a subset of the subset upper approximation.

Rough set could also be generalized with some other approaches [85, 88, 89]. Actually, the above three definitions are classified as constructive rough set formulations by Yao [85], where rough set formulations are divided into two different groups: constructive and algebraic methods. The notion of singleton definition is indeed the same as the element based definition suggested by Yao. While, subset definition is an expansion of concept definition and also undoubtedly is the same as the granule based definition in the Yao study. These definitions are special cases of the subsystem based definition by Yao when the covering is the set of neighbourhoods.

2.5 Summary

Based on available of information, two objects may be indiscernible/equivalent to each other. In some cases, it may not be able to describe an object set X precisely using attribute values because some members of X may be equivalent to objects that do not belong to X. Set X, in this case, is represented by approximations. Lower approximation is the set of objects that absolutely exists in X. Upper approximation, on the other hand, is the set of objects that possibly belongs to X.

The related concepts such as reduct, core, and attribute dependency are also introduced in this chapter. Reducts are the subset of attribute set such that the cardinality of attributes is minimal and on reducts the equivalence classes do not collapse. Core is the set of attributes that exist in all reducts. On the other words, core is the intersection of all reducts of an information system.

To be able to apply in actual systems, original rough set model is also extended to Bayes rough set and Fuzzy rough set models, which allows some uncertainty of the definition. In addition, a generalized rough set deal with an arbitrary binary relation is also discussed.

Chapter 3

Imperfect Information Systems

In the previous chapter, an overview of rough set theory has been introduced. The original rough set is defined based on equivalence relations. Some extensions of original rough set model as well as rough set definitions for other types of relation were also discussed. Those approaches were studied in complete information systems, in which objects attribute values are precise and discrete values. However, data in real applications sometimes are not described precisely. For example, in a case, data may be missing. In another case, values of objects are presented with partial confidence. Such kind of information could be interpreted as imperfect information system.

This chapter will discuss the problems, show the motivation and suggest a representation of imperfectness in information system. First, some related approaches in incomplete information systems, which is most concerned among imperfect data, are highlighted. We then discuss other possible types of imperfect data as well as motivations to deal with the problems. To prepare for two approaches that will be introduced in Chapter 4 and 5, we propose a representation of imperfect data in such systems.

Table 3.1: An example of a dataset with missing values[1]

Cases	Temperature	Headache	Nausea	Flu
x_1	high	*	no	yes
x_2	very-high	yes	yes	yes
x_3	*	no	no	no
x_4	high	yes	yes	yes
x_5	high	*	yes	no
x_6	normal	yes	no	no
x_7	normal	no	yes	no
x_8	*	yes	*	yes

3.1 Related Works

This section introduces some possible important rough set models for incomplete information systems.

Controversial rough set research mostly considers that imperfect data in information systems comes from missing values [18, 21, 23, 26, 37, 38, 72, 73, 76, 83]. There are many reasons why data is missing. Some attribute values are not recoded because they might not be necessary. For example the field "income" may be ignored in a questionnaire if the answer of "occupation" field is "undergraduate student". Data, on the other hands, may not exist even if it is important. For example data may not have been collected or it may have been deleted accidentally.

An information system with missing values is called incomplete information system [37, 38]. In incomplete information systems, Table 3.1 for example, objects may contain several unknown attribute values. Unknown values are denoted by an asterisk (*).

The notions with which incomplete data is dealt are stated in [26, 27]. In this report, the authors studied methods to handle missing attribute values in data mining

[1]Source of table:[26].

and conducted a classification of the investigated approaches as the several following description:

1. Most common attribute value: It is one of the simplest methods to deal with missing values. The value of attribute that occurs most often is selected as the value for all unknown values of the attribute [8].

2. Concept Most Common Attribute Value: The most common attribute value method does not pay any attention to the relationship between attributes and a decision. In this approach, the value of the attribute, which occur the most common within the concept, is selected to be the value for all the unknown values of the attribute. This method is also called maximum relative frequency method, or maximum conditional probability method (given concept).

3. Method of Assigning All Possible Values of the Attribute: In this method, an example with a missing attribute value is replaced by a set of new examples, in which the missing attribute value is replaced by all possible values of the attribute [48, 49]. If we have some examples with more than one unknown attribute value, we will do our substitution for one attribute first, and then do the substitution for the next attribute, etc., until all unknown attribute values are replaced by new known attribute values.

4. Method of Assigning All Possible Values of the Attribute Restricted to the Given Concept: The method of assigning all possible values of the attribute is not related with a concept. This method is a restriction of the method of assigning all possible values of the attribute to the concept, indicated by an example with a missing attribute value.

5. Method of Ignoring Examples with Unknown Attribute Values: This method is the simplest: just ignore the examples which have at least one unknown

attribute value, and then use the rest of the table as input to the successive learning process.

6. Method of Treating Missing Attribute Values as Special Values: In this method, we deal with the unknown attribute values using a totally different approach: rather than trying to find some known attribute value as its value, we treat "unknown" itself as a new value for the attributes that contain missing values and treat it in the same way as other values.

In fact, those classifications do not cover all methods of treating missing values. Depending on each system, a handling missing value method can be chosen based on characteristics and requirements of the system. If missing values do not cause any problem in relationship between objects, they can be ignored. On the other hand, database miners have to find which algorithm should be used.

In general, approaches deal with unavailable values based on one of the following two interpretations [19]. The first is "lost value" in which unknown values of attributes are already lost. Similarity relation [73] is one example of this semantics. The second is "do not care", which may be potentially replaced by any value in the domain. Such incomplete decision tables were broadly studied in numerous researches [37, 38]. Grzymala-Busse [21, 22, 23, 25] built a characteristic relation based on both "lost value" case and "do not care" case. The rest of the section will review some approaches which deal with incomplete information systems.

Tolerance relation

Given incomplete information system $I = (U, A)$, let relation $TOR_P(x, y)$, $P \subseteq A$ denote a binary relation between objects that are possibly equivalent in terms of values of attributes. The *tolerance relation* [37, 38] is defined by

$$TOR_P(x, y) \Leftrightarrow \forall a \in P, (f_a(x) = f_a(y)) \vee (f_a(x) = *) \vee (f_a(y) = *), \tag{3.1}$$

where $\vee$ denotes disjunction.

The relation is reflexive and symmetric, but does not need to be transitive. Let $T_P(x) = \{y \in U | TOR_P(y, x)\}$ be the set of objects that are in a relation with x in terms of P in the sense of the above tolerance relation. Due to the symmetric property, x is also tolerant toward elements in $T_P(x)$.

Rough sets based on the *tolerance relation* in incomplete information systems are defined in a way similar to those in complete information systems [60, 61]. Let $X \subseteq U$, $P \subseteq A$, $\underline{apprT}_P X$ is then the lower approximation [37, 38] of X in terms of P if and only if

$$\underline{apprT}_P X = \{x \in U | T_P(x) \subseteq X\}. \tag{3.2}$$

$\overline{apprT}_P X$ is the upper approximation of X in terms of P if and only if

$$\overline{apprT}_P X = \{x \in U | T_P(x) \cap X \neq \emptyset\}. \tag{3.3}$$

Now, we illustrate the above concepts with an incomplete decision table shown in Table 3.1. From this table, we can induce approximation space for X - group of people such that the value of *Flu* is *no* based on all condition attributes:

$$\underline{apprT}_A X = \{x_7\},$$
$$\overline{apprT}_A X = \{x_1, x_3, x_4, x_5, x_6, x_7, x_8\}.$$

The approximations clearly are quite poor. There are some objects which intuitively could be classified in X, while they are not in the lower approximation. Take, for instance, object x_6, we have its complete description, and intuitively there is no other object perceived as very tolerant to it. However, it is not included into the lower approximation of X. This is due to missing attribute values of objects x_8, which is actually tolerant to x_6 according to Equation (3.1).

Similarity relation

Stefanowski and Tsoukias stated that approximations obtained by the *tolerance relation* are quite poor from the viewpoint of the meaning of "approximation" [72, 73]. They assume that object x can be considered similar to another object y only if all known attribute values of x are the same as those of y. Such a relation is not symmetric. If one of a pair of objects has a more complete description than the other, the inverse relation will not hold. Formally, given incomplete information system $I = (U, A)$ and attribute set $P \subseteq A$, the *similarity relation* is defined as follows:

$$SIM_P(x, y) \Leftrightarrow \forall a \in P, (f_a(x) = *) \vee (f_a(x) = f_a(y)). \tag{3.4}$$

It is easy to observe that this relation is reflexive and transitive, although not necessarily symmetric. Now for each object we can induce two similarity sets: $S_P(x) = \{y \in U | SIM_P(y, x)\}$ is the set of objects similar to x - note that the arguments of SIM_P is not (x, y) - and $S_P^{-1}(x) = \{y \in U | SIM_P(x, y)\}$ is the set of objects to which x is similar. $S_P(x)$ and $S_P^{-1}(x)$ are clearly two different sets. From similarity sets, the authors introduce definitions of an approximation space of a set $X \subseteq U$ as follows:

$$\underline{apprS}_P X = \{x \in U | S_P^{-1}(x) \subseteq X\}, \tag{3.5}$$

$$\overline{apprS}_P X = \bigcup \{S_P(x) | x \in X\}. \tag{3.6}$$

By the definition of the similarity relation and the tolerance relation introduced in this section, we can see that the conditions for which similarity relation holds are a subset of the conditions for which tolerance relation holds (we can see that if

$SIM_P(x, y)$ then $TOR_P(x, y)$). Hence, *tolerance classes* of elements in U shall be "wider" than the respective *similarity classes* [72, 73].

Limited tolerance relation

Wang [76] proved that the *similarity relation* may result in some lost information. Objects $x = (*, 1, 2, 3, 4, 5, 6, 7, 8, 9)$ and $y = (0, *, 2, 3, 4, 5, 6, 7, 8, 9)$, for example, where elements in parentheses represent values of attributes defined by their positions, are tolerant according to (3.1) and intuitively similar to each other. They do not, however, satisfy the non-symmetric *similarity relation.* To avoid this problem, Wang [76] developed a novel *limited tolerance (LT) relation.*

Given incomplete information system $I = (U, A)$, attribute set $P \subseteq A$, and $O_P(x) = \{a | a \in P, f_a(x) \neq *\}$, the *limited tolerance relation* is defined on U as follows:

$$\begin{aligned} LTOR_P(x, y) \Leftrightarrow & (\forall a \in P, f_a(x) = f_a(y) = *) \vee \\ & ((O_P(x) \cap O_P(y) \neq \emptyset) \\ & \wedge (\forall a \in P, f_a(x) \neq * \wedge f_a(y) \neq * \rightarrow f_a(x) = f_a(y))), \end{aligned} \tag{3.7}$$

where $\wedge$ denotes conjunction.

In the formula, the condition that $f_a(x) \neq * \wedge f_a(y) \neq * \rightarrow f_a(x) = f_a(y)$ is equivalent to $f_a(x) = * \vee f_a(y) = * \vee f_a(x) = f_a(y)$. Thus, the two objects that satisfy $TOR_P(x, y)$ but not $LTOR_P(x, y)$ are only those satisfying $O_P(x) \cap O_P(y) = \emptyset$.

Generally speaking, two objects are in a *limited tolerance relation* if they are in one of two cases: The first case is that all attribute values of the two objects are missing and the second is a case where there is at least one attribute having an ordinary value for both objects and the two objects have the same value for these attributes. The *limited tolerance relation* is clearly reflexive and symmetric but not necessarily transitive.

Thus, *limited tolerance class* is defined by

$$LT_P(x) = \{y \in U | LTOR_P(y, x)\}. \tag{3.8}$$

Based on that, the approximation space is defined as follows:

$$\underline{apprLT}_P X = \{x \in U | LT_P(x) \subseteq X\}, \tag{3.9}$$
$$\overline{apprLT}_P X = \{x \in U | LT_P(x) \cap X \neq \emptyset\}. \tag{3.10}$$

Wang [76] also proved that the *tolerance relation* and the *similarity relation* are the two extremities for extending indiscernibility relation, and limited tolerance relation happens to be between the *tolerance relation* and the *similar relation*,

$$\underline{apprT}_P X \subseteq \underline{apprLT}_P X \subseteq \underline{apprS}_P X,$$
$$\overline{apprS}_P X \subseteq \overline{apprLT}_P X \subseteq \overline{apprT}_P X.$$

3.2 Problems from Imperfect Information and Motivation

In the beginning of the chapter, the problem of the original rough set that presupposes that all objects in an information system have precise and complete attribute values is addressed. Several methods of handling missing values are also reviewed. However, besides the missing values, there are many reasons why imperfect data are produced in datasets [45].

To be the first, *imprecision* is another type of possible imperfect data. Stored information is imprecise when it denotes a set of possible values and the real value is one of the elements of this set. Specific kinds of imprecise information include

Table 3.2: An example of a dataset with uncertainty[2]

Employees	**Deterministic Department**	**Stochastic Quality Bonus**	**Stochastic Sales**
Jon Smith	Toy	0.4[Great Yes] 0.5[Good Yes] 0.1[Fair Yes]	0.3[$30-34K] 0.7[$35-39K]
Fre Jones	Housewares	1.0[Good Yes]	0.5[$20-24K] 0.5[$25-29K]

disjunctive information, e.g., John is is either 31 or 32 years old; negative information, e.g., John is not 30 years old; range information, e.g., John is between 30 and 35 years old or John is over 30 years old. Such information may also have error margins, e.g., John is 30 give or take 1 year.

Another possible type of imperfect data is *uncertainty*[45]. Whereas the statement "John is either 31 or 32 years old" takes the form of imprecision, the statement "John is probably 32" or "John is 32 years old with a confidence of 0.6" denotes uncertainty. Both imprecise and uncertain values can be represented by probabilistic data [6]. Toyota may, for example, have demographic information indicating that customers living in a certain region are likely to purchase a Corolla with a probability of 0.7 or a Celica with a probability of 0.3. Table 3.2 illustrates an information system with probabilistic data.

The last type of imperfect data listed in [45] is *error*. Stored information is erroneous when it is different from the true information. Errors in given information that are identified can be removed and the rest is treated as information with missing values. In cases where errors are unidentified, however, the reliability of all of the information is lost. Approximations in rough set theory are derived from the information available, so, we do not deal with errors in this study. The

[2]Source of table: [2].

term “imperfect data” will, hereafter, represent the case of missing, uncertain and imprecise information.

To information systems containing missing, imprecise and uncertain values, it is inappropriate to apply a method that can deal only with missing values. One possible solution would be to combine the transformation of any type of imperfect data with probabilistic values [53, 55] and then to apply a probabilistic method [50, 51, 53, 72, 73]. Our motivation is thus to propose rough set model not only deal with missing value but also can solve the problem of uncertainty and imprecision which is a gap of rough set study on actual information. A representation of imperfectness including missing values, uncertainty as well as imprecision is first introduced. This representation should be useful to a single approach dealing with multiple type of imperfect information.

3.3 Imperfect Information Representation

An information system in the original rough set study is defined as a pair $I = (U, A)$, where U is a non-empty finite set of objects called the universe and A is a non-empty finite set of attributes such that $f_a : U \to V_a$ for every $a \in A$ [60, 61]. The non-empty discrete value set V_a is called the domain of a. The original rough set theory deals with complete information systems in which $\forall x \in U$, $a \in A$, $f_a(x)$ is a precise value.

Now, for an information system in which some attribute values of objects are missing and/or associated with probabilistic data, the attribute values of an object may be represented as follows:

Definition 3.3.1. In an imperfect information system $I = (U, A)$, let $t^x_{a,i} \subseteq V_a$ be the i-th set of overall “s” possible value sets of “x” on “a” and $p^x_{a,i} > 0$ be its probability.

Table 3.3: Dice game scores

Players	**Score**
Terry	3
David	6
Tom	*
Anna	8

Then the pair (T_a^x, P_a^x), where $T_a^x = \{t_{a,i}^x | 1 \leq i \leq s\}$, $P_a^x = \{p_{a,i}^x | \sum_i p_{a,i}^x = 1\}$, represents the imperfect value of object x on attribute a.

In the above, $t_{a,i}^x$ are not necessarily be mutually disjoint. Obviously, for this representation of imperfect values, it is able to present any type of imperfectness discussed in Section 3.2. A value is uncertain when any set of possible of values is singleton. In this case $|t_{a,i}^x| = 1$. Some types of missing values may have a pre-defined probability distribution and the imperfectness could be regarded as uncertainty. One example is a game of four people playing with dice. Their scores can be calculated based on the sum of two dice thrown for each of them. Table 3.3 shows their scores. In this table, as we can see that the score of Tom is unknown due to some reasons. The probability of each value that may be Tom's score, however, can be identified by a probability distribution for the sum of two dice. The probability that Tom's score is 7, for example, is 1/6. On the other hand, the probability that his score equals to 11 is 1/18. In this case $\forall v_i \in V_a, t_{a,i}^x = \{v_i\}, p_{a,i}^x = \lambda_a(v_i)$ when $f_a(x) = *$, where $\lambda_a(v)$ is the probability mass function on a.

A value is imprecise when there is only a set of multiple possible values and the probability of this set is also 1, formally $|T_a^x| = 1$, $p_{a,1}^x = 1.0$. A *precise* value and a *missing* value with no pre-defined probability distribution can be considered as two extreme kinds of imprecision. A value is precise if the set of possible values is singleton. In this case $|t_{a,1}^x| = 1$. Missing values without pre-defined probability distribution could be regarded as imprecise information where the set of possible values encompasses the entire attribute domain, such that $t^x_{a,1} = V_a$.

More importantly, if an object attribute value contains both uncertainty and imprecision, we can also use the imperfect representation in Definition 3.3.1 to show this type of value. To illusstrate, we represent the "*Stochastic sales*" value of Jon Smith in Table 3.2 as follows: $t^{Smith}_{Sales,1} = \{30, 31, 32, 33, 34\}$, $p^{Smith}_{Sales,1} = 0.3$ and $t^{Smith}_{Sales,2} = \{35, 36, 37, 38, 39\}$, $p^{Smith}_{Sales,2} = 0.7$.

3.4 Summary

In this chapter, related rough sets studies to deal with missing value are reviewed. There is also a discussion on a question "What is an imperfect information system?". Based on that, we introduce a representation of imperfect values. Consequently, the next three chapters will introduce two type of relations from imperfect description of objects using this representation as well as methods of acquire knowledge in such information systems.

Chapter 4

Valued Tolerance and Similarity Relations Based Rough Set

In studies of rough sets in incomplete information systems, probabilistic solutions have been introduced based on the possibility of "missing value" [18, 50, 51, 52, 72, 73]. Among them, some approaches [18, 72] suppose a priori assumption that there exists a uniform probability distribution on every attribute domain and compute valued tolerance (or similarity) classes based on joint probability distribution. This chapter firstly define a general method of determining a probability (probability of matching) that two objects may be tolerant of (similar to) each other on an attribute. The probability of matching will be defined based on the probability that two objects may take the same values on an attribute in the dataset. Based on probabilities on some attributes, a relation called extended tolerance relation will then be introduced.

4.1 Valued Tolerance/Similarity Relations Based Rough Set Models

A solution of defining valued tolerance/similarity relations can be stated as follows: first, the pair (T_a^x, P_a^x) that represents an imperfect value of object x is defined for each attribute a, and the probability that two objects are tolerant of (similar to) each other on the attribute is determined. The degree that two objects are tolerant of (similar to) each other on a set of attributes is then calculated, for example, using the joint probabilities assuming that all the attributes are independent of one another. This section will summarize concepts in valued tolerance/similarity relation definitions as well as a rough set approach based on this kind of relations. Problems of valued tolerance/similarity relation based rough set models in the related work are also addressed in this section.

For an information system, in which some attribute values of objects are missing and/or associated with uncertainty or imprecision, we define probabilities of attribute values. For a discrete attribute, ***probability of object attribute value*** denoted by $Pr_a(f_a(x) = v)$ represents the probability that object $x \in U$ takes value $v \in V_a$ on attribute $a \in A$. Two methods to estimate the probabilities of object attribute values will be discussed in the next section.

Based on the probabilities estimated, ***probability of matching*** between two objects $x, y \in U$ on attribute $a \in A$ denoted by $\theta_a(x, y)$ defines the probability that object x takes the same value as object y on attribute a. In [18, 72, 73], it is supposed that there is an uniform probability distribution on an attribute, and the probability of matching is defined as $\theta_a(x, y) = Pr_a(f_a(x) = v)) \bullet Pr_a(f_a(y) = v) = 1/|V_a|^2$ where v is a value in the domain of attribute a. The definition is clearly inadequate when we suppose the attribute values of both "x" and "y" are missing on "a". The definition of probability of matching is discussed and calculated in general in the next section.

From the probability of matching between two objects, we can induce the degree that $x, y \in U$ are tolerant of (similar to) each other on a set of attributes $P \subseteq A$. This degree is denoted by $\phi_P(x, y)$. The degree of tolerance/similarity can be defined as the probability that two objects have the same values on all attributes in set P and is calculated by joint probability $\phi_P(x, y) = \prod_{a \in P} \theta_a(x, y)$ assuming independence among attributes. Other methods of tolerance (similarity) degree definitions can be found in [52], which is also discussed in the section of extended tolerance relations.

Now, it is able to define a relation $R_P(x, y)$ between objects x and y by controlling the degree of tolerance(similarity) using threshold α, such that $R_P(x, y) \Leftrightarrow \phi_P(x, y) \geq \alpha$. Based on that, a neighbourhood, which consists of successor and predecessor sets, of an object [28, 85] is determined. Then rough sets can be defined as discussed in Chapter 2.

4.2 Probability of Matching

This section shows how to define the probability of matching between two objects for an attribute in imperfect information systems discussed in Chapter 3. This probability is used in order to define "Valued tolerance/similarity Relation based Rough Set" (VRRS) methods to this kind of information system. According to Section 4.1, probabilities of object attribute values need to be defined before calculating probability of matching.

4.2.1 Probability of Object Attributes Values

In general, if there is no information about probability distribution of attribute values, it is possible to make the hypothesis that the probability is determined by a uniform distribution.

Definition 4.2.1. In imperfect information system $I = (U, A)$, let the pair (T_a^x, P_a^x) present an imperfect attribute value of object x on a. Then the probability that object x takes value $v \in V_a$ on a can be calculated as follows:

$$Pr_a(f_a(x) = v) = \sum_i p_{a,i}^x \bullet \frac{|\{v\} \cap t_{a,i}^x|}{|t_{a,i}^x|}, \tag{4.1}$$

where $t_{a,i}^x$ is the i-th set possible value sets in T_a^x and $p_{a,i}^x$ is the probability of $t_{a,i}^x$.

In this equation, the probability mass distributed equally in possible value set $t_{a,i}^x$ will be added to the probability of attribute value if $v \in t_{a,i}^x$, such that $|\{v\} \cap t_{a,i}^x| = 1$. Obviously, in case of uncertainty where a probability distribution is given, it is not necessary to calculate the probability of object attribute values. In this case $Pr_a(f_a(x) = v) = p_{a,i}^x$ because $t_{a,i}^x = \{v\}$. In case of the missing value without any pre-defined probability distribution, $Pr_a(f_a(x) = v) = 1/|V_a|$ for any $v \in V_a$.

However, even no pre-defined probability distribution exist, we still can estimate the probability of attribute values in some cases. The next step will summarize two possible solutions discussed firstly in [53] with the adaptation of the representation of imperfect values in Definition 3.3.1.

4.2.2 Method of The Frequency of Attribute Value

The approach is based on the notion of "The most common method" - a method of handling missing values summarized by Grzymala-Busse [27, 28] - in which, missing values are replaced by the most common value of the attribute. This method of handling missing attribute values is implemented, e.g., in well-known machine learning algorithm CN2 [8].

Suppose the value domains are known. First, we define the probability that each value of an attribute appears based on frequencies of the available values for this attribute in dataset. The probability that a value $v \in V_a$ appears as a value of a certain object is define by

$$\rho_a(v) = \begin{cases} \dfrac{|V_a(v)|}{|U - V_a(?)|} & \text{if } V_a(?) \subset U, \\ \dfrac{1}{|V_a|} & \text{otherwise,} \end{cases} \tag{4.2}$$

where $V_a(v)$ and $V_a(?)$ are the sets of objects whose attribute value is "v" and the set of objects whose value on "a" is imperfect, respectively. The symbol "$\subset$" denotes a proper subset. As seen in the equation, the probability $\rho_a(v)$, $v \in V_a$ is defined by the ratio of the value v among objects whose values are not imperfect. If $V_a(?) = U$, that is, values of attribute a are imperfect in all objects, the equal probability distribution is given. The value of $\rho_a(v)$ is greater than zero if there is at least an object such that $f_a(x) = v$. Since it could be zero for many values if the size of U is small, the size of U should be large enough when using the approach.

Returning to the example in Table 3.1, the probabilities of attribute values are illustrated in Table 4.1. From this table, we can see that the value "high" of "Temperature" occurs more frequently than the other values. The most frequent values of "Headache" and "Nausea" happen to be "yes".

Now, we define the probability of object attribute values by the frequency of values in a dataset.

Table 4.1: Probability of attribute values

Attributes	**Values**	**Probability**
Temperature	very-hig	0.17
Temperature	high	0.50
Temperature	normal	0.33
Headache	yes	0.67
Headache	no	0.33
Nausea	yes	0.57
Nausea	no	0.43

Definition 4.2.2. In imperfect information system $I = (U, A)$, an attribute $a \in A$ and its domain V_a, $\rho_a(v)$ denotes the frequency of each value $v \in V_a$ in the dataset. Given object $x \in U$ with an imperfect value on a, the probability of object attribute value $Pr_a(f_a(x) = v)$ can be calculated as follows:

$$Pr_a(f_a(x) = v) = \sum_i p^x_{a,i} \bullet |\{v\} \cap t^x_{a,i}| \bullet \Psi(t^x_{a,i}),$$

where

$$\Psi(t^x_{a,i}) = \begin{cases} \dfrac{\rho_a(v)}{\sum_{v' \in t^x_{a,i}} \rho_a(v')} & \text{if } \sum_{v' \in t^x_{a,i}} \rho_a(v') \neq 0, \\ \dfrac{1}{|t^x_{a,i}|} & \text{otherwhise.} \end{cases} \tag{4.3}$$

On each possible range, if value $v \in t^x_{a,i}$, then the probability based on $t^x_{a,i}$ is determined by the proportion of probability on v to the whole range $t^x_{a,i}$. This proportion should take the equal probability when the probability on the whole range $t^x_{a,i}$ equals to zero.

The idea could be applied to the missing values with no predefined probability distribution. However, it should not be applied to attributes where uncertainty or a

Table 4.2: Probability of attribute values given a concepts

Attributes	**Values**	**Probability in concepts**	
		Flu=Yes	*Flu=No*
Temperature	very-high	0.33	0.00
Temperature	high	0.67	0.33
Temperature	normal	0.00	0.67
Headache	yes	1.00	0.33
Headache	no	0.00	0.67
Nausea	yes	0.67	0.50
Nausea	no	0.33	0.50

probability distribution is derived from a theoretical point of view, e.g. in the case of dice game mentioned before.

4.2.3 Method of The Frequency of Attribute Value Related To Concepts

This is an extension of the method in the previous subsection. Observing some systems, we sometimes recognize that attribute values might depend on some concepts. Supposed the value domains are known, the probability that a value $v \in V_a$ appears as a value of objects contained in a concept $X \subseteq U$ is defined as follows:

$$\rho_a(v)_X = \begin{cases} \dfrac{|V_a(v)_X|}{|X - V_a(?)_X|} & \text{if } V_a(?)_X \subset X, \\ \dfrac{1}{|V_a|} & \text{otherwise.} \end{cases} \tag{4.4}$$

where $V_a(v)_X$ and $V_a(?)_X$ are the set of objects in concept X whose attribute value is "v" and the set of objects whose value on "a" is imperfect, respectively.

Table 4.2 shows that flu relates to high and very-high temperature, headache and nausea. On the other hand non-flu supports the cases of low temperature and no headache.

In the same way as the previous method, it is possible to define probabilities of object attributes values by frequencies of values in the dataset. The part $\Psi(t^x_{a,i})$ in Equation 4.3 is replaced by

$$\Psi(t^x_{a,i}) = \begin{cases} \dfrac{\rho_a(v)_X}{\sum_{v' \in t^x_{a,i}} \rho_a(v')_X} & \text{if } \sum_{v' \in t^x_{a,i}} \rho_a(v')_X \neq 0, \\ \dfrac{1}{|t^x_{a,i}|} & \text{otherwhise.} \end{cases}$$

4.2.4 Obtaining Probability of Matching

This section will redefine the degree that two objects have the same value on an attribute if at least one of the two objects has missing, imprecise or uncertain values on their attributes. In general the probability of matching can be defined as the following definition:

Definition 4.2.3. Given information system $I = (U, A)$, on attribute $a \in A$ with its domain V_a, the probability that the value of object x is the same as the value of object y on a is given by:

$$\theta_a(x, y) = \sum_{v \in V_a} Pr_a(f_a(x) = v | f_a(y) = v) Pr_a(f_a(y) = v), \tag{4.5}$$

when $x \neq y$. Otherwise $\theta_a(x, y) = \theta_a(x, x) = 1$. Note that $\theta_a(x, y) = 1$, if the two objects x and y have the same precise value on a, while it is zero if they have different precise values. $Pr_a(f_a(x) = v | f_a(y) = v)$ denotes the conditional probability of $f_a(x) = v$ given $f_a(y) = v$. Hereafter, we assume that two events $f_a(x) = v$ and $f_a(y) = u$, $x, y \in U$, $a \in A$ are independent of each other for any $u, v \in V_a$.

The probability of matching for each type of missing, uncertain and imprecise values has been discussed in [53]. However, the concept of imperfectness introduced in this dissertation generalizes the way to calculate the probability of matching as follows:

$$\theta_a(x,y) = \begin{cases} \displaystyle\sum_{v\in\Lambda(x,y)} Pr_a(f_a(x)=v)Pr_a(f_a(y)=v) & \\ & \text{if } \Lambda(x,y)\neq\emptyset, \\ 0 & \text{if } \Lambda(x,y)=\emptyset, \end{cases} \tag{4.6}$$

when $x \neq y$, $\Lambda(x,y) = [\cup_i t_{a,i}(x)] \bigcap [\cup_i t_{a,i}(y)]$. Otherwise $\theta_a(x,y) = \theta_a(x,x) = 1$.

Obviously, the probability of matching between two objects equals to zero unless there are common possible values for both two objects. Otherwise, the sum of product of probabilities should be taken that the two objects coincide to have a common value.

In short, we have shown the method of calculating probability that two objects have the same attribute values in case of imperfect information. This probability can be used to define a valued tolerance/similarity relation, and then to obtain approximation space for any published VRRS approach.

4.2.5 Some Discussion for Continuous Values

In the previous subsection, we have discussed methods to obtain probabilities of matching in cases of discrete values. These methods will be used to define a tolerance relation in the next section. However, in some cases, continuous values also can be missing or described with uncertainty. Thus, we spend a small part to make an examination on continuous values. Hope that it could be useful to some extent.

In information system coming with continuous value, keeping the consistency of information systems, continuous attributes must be transformed into discrete ones. Some approaches discretize these attributes domains into ranges where each interval is mapped to a discrete value [5, 11, 56, 79]. In general, the targets of such studies are to find the minimum interval without weakening the discernibility in the dataset.

On continuous attributes containing imperfect data, indiscernibility relations are not available at all. There exists a way to deal with them using rough set technique. First discretizing the continuous data to discrete data [7], and then finding the attribute reduction using methods proposed in [20, 21, 23, 37, 38, 72, 73, 76, 83].

Let $I = (U, A)$ be a complete information system containing continuous values. Any pair (a, c), where continuous attribute $a \in A$, $c \in \mathbb{R}$, where $\mathbb{R}$ represents the set of all real numbers, will be called a cut on V_a. For $a \in A$, any set of cuts $\{(a, c_1^a), (a, c_2^a), ..., (a, c_k^a)\}$ on $V_a = [v_{min}^a, v_{max}^a) \subset \mathbb{R}$ defines a partition $V_a' = \{[c_0^a, c_1^a), [c_1^a, c_2^a), ...[c_k^a, c_{k+1}^a)\}$ where $v_{min}^a = c_0^a < c_1^a < ... < c_k^a < c_{k+1}^a = v_{max}^a$), and $V_a = [c_0^a, c_1^a) \cup [c_1^a, c_2^a) \cup ... \cup [c_k^a, c_{k+1}^a)$. Therefore, any set of cuts defines a new attribute domain V_a' on a and the equivalence between two objects on a [56] is defined as follows:

$$EQU_{\{a\}}(x, y) \Leftrightarrow \left(\text{iff } f_a(x), f_a(y) \in [c_i^a, c_{i+1}^a)\right). \tag{4.7}$$

On attributes associated with continuous values, two objects are equivalent if their attribute values fall in the same interval. If there is a missing, uncertain or imprecise value, the equivalence relation cannot be determined. We have to define the degree of tolerance (similar) instead. For continuous values, the probability that an attribute value of $x \in U$ on attribute $a \in A$ falls into an interval, say $[c_1, c_2) \subseteq V_a$, is given by $Pr_a(c_1 \leq f_a(x) < c_2)$. From probabilities of object attribute values, we are able to define a valued tolerance/similarity relation.

For continuous imperfect value, $t_{a,i}^x$ in Definition 3.3.1 should present for a interval into which the object attribute value falls. Thus the representation of imperfection in case of continuous value is defined as follows:

Definition 4.2.4. In imperfect information system $I = (U, A)$, on continuous attribute $a \in A$, let $t_{a,i}^x = [v_{i,1}, v_{i,2}]$, where $v_{i,1}, v_{i,2} \in \mathbb{R}$, $v_{i,2} > v_{i,1}$, be the i-th range of overall "s" possible range of object x on a and $p_{a,i}^x > 0$ be its probability. Then the pair (T_a^x, P_a^x), where $T_a^x = \{t_{a,i}^x | 1 \leq i \leq s\}$, $P_a^x = \{p_{a,i}^x | \sum_i p_{a,i}^x = 1\}$, represents imperfect values of object x on continuous attribute a.

If there is no information about probability distribution of attribute values, it is possible to make a hypothesis that the probability is determined by a uniform distribution. The probability that continuous values of a range falls in an interval, hence, could be defined as how large the interval cover the range. Now, the probability that an attribute value of $x \in U$ on attribute $a \in A$ falls into an interval can be defined as the following definition.

Definition 4.2.5. In imperfect information system $I = (U, A)$, let the pair (T_a^x, P_a^x) present an imperfect attribute value of object x on continuous attribute $a \in A$. Then the probability that an attribute value of x falls in the interval $[c_1, c_2) \subseteq V_a$ can be calculated as follows:

$$\begin{aligned} Pr_a(c_1 \leq f_a(x) < c_2) &= \lim_{\varepsilon \to 0} \sum_{[c_1,c_2) \cap [v_{i,1}, v_{i,2}] \neq \emptyset} p_{a,i}^x \bullet \frac{\min(c_2 - \varepsilon, v_{i,2}) - \max(c_1, v_{i,1})}{v_{i,2} - v_{i,1}} \\ &= \sum_{[c_1,c_2) \cap [v_{i,1}, v_{i,2}] \neq \emptyset} p_{a,i}^x \bullet \frac{\min(c_2, v_{i,2}) - \max(c_1, v_{i,1})}{v_{i,2} - v_{i,1}}. \end{aligned} \tag{4.8}$$

When a random variable takes values from a continuous range, in some cases, we have to do experiments to estimate probability distribution of the data. In other cases, the data is already described by a known probability distribution such as Gaussian, Laplace, Gamma distribution [43, 59]. In information system $I = (U, A)$, suppose a probability function $\lambda_a(v)$, $v \in V_a$ (that is called probability density function for continuous values), the probability that a continuous value falls in interval $[c_1, c_2]$ is determined by an integral:

$$\int_{c_1}^{c_2} \lambda_a(v) d(v).$$

For interval $[c_1, c_2) \subseteq V_a$, the probability is

$$\lim_{\varepsilon \to 0} \int\limits_{c_1}^{c_2-\varepsilon} \lambda_a(v)d(v).$$

The probability that an attribute value of $x \in U$ on attribute $a \in A$ falls into an interval is thus defined as the following definition.

Definition 4.2.6. In imperfect information system $I = (U, A)$, let the pair (T_a^x, P_a^x) present an imperfect attribute value of object x on continuous attribute $a \in A$ and $\lambda_a(v)$, $v \in V_a$ denotes the probability density function on a, the probability that an object attribute value of x falls in the interval $[c_1, c_2) \subseteq V_a$ is then calculated as follows:

$$Pr_a(c_1 \leq f_a(x) < c_2) = \sum_{[c_1,c_2) \cap [v_{i,1},v_{i,2}] \neq \emptyset} p_{a,i}^x \bullet \Psi(t_{a,i}^x),$$

where

$$\Psi(t_{a,i}^x) = \begin{cases} \dfrac{\lim\limits_{\varepsilon \to 0} \int\limits_{\max(c_1,v_{i,1})}^{\min(c_2-\varepsilon,v_{i,2})} \lambda_a(v)d(v)}{\int\limits_{v_{i,1}}^{v_{i,2}} \lambda_a(v)d(v)} & \text{if } \int\limits_{v_{i,1}}^{v_{i,2}} \lambda_a(v)d(v) \neq 0, \\ \dfrac{\min(c_2, v_{i,2}) - \max(c_1, v_{i,1})}{v_{i,2} - v_{i,1}} & \text{otherwhise.} \end{cases} \tag{4.9}$$

On each possible range, if $[c_1, c_2) \cap t_{a,i}^x \neq \emptyset$, the probability based on the interval $[c_1, c_2)$ is then determined by the proportion of the integral on the intersection of $[c_1, c_2)$ and $t_{a,i}^x$ to the integral on the whole range $t_{a,i}^x$. This proportion should take the equal probability when the integral on the whole range $t_{a,i}^x$ equals to zero.

Directly from the sum of probability on all possible range, it is able to obtain the probability of matching as the following equation:

$$\theta_a(x,y) = \begin{cases} \displaystyle\sum_{[c_j,c_{j+1})\cap\Lambda(x,y)\neq\emptyset} Pr_a(c_j \leq f_a(x) < c_{j+1})Pr_a(c_j \leq f_a(y) < c_{j+1}) & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \Lambda(x,y) \neq \emptyset, & \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \Lambda(x,y) = \emptyset, & \end{cases} \tag{4.10}$$

when $x \neq y$, $\Lambda(x,y) = [\cup_i t_{a,i}(x)] \bigcap [\cup_i t_{a,i}(y)]$. Otherwise $\theta_a(x,y) = \theta_a(x,x) = 1$.

Obviously, a probability of matching between two objects equals to zero unless there are common possible ranges for both two objects. Otherwise, a sum of product of probabilities should be taken that the two objects coincide to have a common range.

Equation (4.10) shows that in information systems containing continuous attributes with missing, uncertain or imprecise values, it is possible to use probability of matching for defining probabilistic based tolerance/similarity relations. On the other hand, in such kind of information systems, we may be able to define a distance function such as $distance_a(f_a(x), f_a(y)) = 1 - \theta_a(x,y)$ for defining similarity relation based on distance [74].

4.3 Extended Rough Set Model

Based on probabilities of matching between two objects on every attribute, it is possible to define a valued tolerance/similarity relation as discussed in the first section of this chapter. The simplest method is taking the product of probabilities on all attribute. In this section, a new valued tolerance relation will be introduced based on not only the probability of matching but also based on the existence of equivalence on some attributes.

4.3.1 Extended Tolerance Relation

To define whether two objects x and y are tolerant to each other or not, we introduce the concept tolerance degree between two objects by combining two relation indexes. One takes a binary value representing a binary equivalence relation defined by attributes with a known value in both the objects. The other is an index defined by attributes with the missing value in either of the objects. It is obtained from probability of matching assuming that $\theta_a(x, y)$ is independent of each other among attributes.

As discussed in Chapter 3, the *limited tolerance relation* was defined basically using attributes whose values are available in both x and y. We define a binary function that represents that LT relation can hold between the objects in the case of two objects have the same precise values on some attributes and utilize it.

Definition 4.3.1. In imperfection information system $I = (U, A)$, let $O_P(x) = \{a | a \in P, f_a(x) \neq ?\}$, $P \subseteq A$ and $x, y \in U$, where "?" denotes an imperfect value, the equivalence existence is defined by the following function:

$$\Theta_P(x,y) = \begin{cases} 1, & \text{if } (|O_P(x) \cap O_P(y)| > 0) \\ & \wedge(\forall a \in O_P(x) \cap O_P(y), f_a(x) = f_a(y)), \\ 0, & \text{otherwise,} \end{cases} \tag{4.11}$$

$\Theta_P(x, x) = 1$ is assumed in any case.

In incomplete information systems, it is clear that objects x, y have LT relation if $\Theta_P(x, y) = 1$, but $\Theta_P(x, y) = 1$ does not hold necessarily even if x, y have LT relation; for example, in case where for all $a \in P$, $f_a(x) = *, f_a(y) = *$ for different x and y.

Now, we define a tolerance degree between x and y by combining the equivalence existence with probability of matching defined in the previous section.

Definition 4.3.2. Let $I = (U, A)$ be an imperfect information system and attribute set $P \subseteq A$ and $x, y \in U$, a parameterized tolerance degree of x and y in terms of P is defined as follows:

$$\phi_P^\eta(x,y) = \begin{cases} 0, & \textit{if } \exists a \in P - (O_P(x) \cap O_P(y)), \theta_a(x,y) = 0, \\ \eta \prod\limits_{a \in P - (O_P(x) \cap O_P(y))} \theta_a(x,y) + (1-\eta)\Theta_P(x,y), & \text{otherwise}, \end{cases} \tag{4.12}$$

where η is a parameter taking a value in $(0, 0.5]$. If $O_P(x) \cap O_P(y) = P$, $\prod_{a \in P - (O_P(x) \cap O_P(y))} \theta_a(x,y) = 1$ is assumed. Thus, $\phi_P^\eta(x,y) = 1$ in this case. Obviously, $\phi_P^\eta(x,x) = 1$ and $\phi_P^\eta(x,y) = \phi_P^\eta(y,x)$. Then the reason why $\eta \in (0, 0.5]$ shall be explained below.

In (4.12), when $\Theta_P(x,y) = 1$, that is, when $f_a(x) = f_a(y)$ for all $a \in O_P(x) \cap O_P(y)$, it is satisfied that $1 - \eta < \phi_P^\eta(x,y) \leq 1.0$ if $\forall a \in P - (O_P(x) \cap O_P(y))$, $\theta_a(x,y) > 0$, otherwise $\phi_P^\eta(x,y) = 0$. Therefore, $\phi_P^\eta(x,y) = 1$ holds only in two cases; one is the case where $x = y$, and the other is the case where $O_P(x) \cap O_P(y) = P$ and $\forall a \in O_P(x) \cap O_P(y)$, $f_a(x) = f_a(y)$.

When $\Theta_P(x,y) = 0$, there are two cases; one is a case where there is $a \in O_P(x) \cap O_P(y) \neq \emptyset$ such that $f_a(x) \neq f_a(y)$. In this case, $\phi_P^\eta(x,y) = 0$. The other is the case where $O_P(x) \cap O_P(y) = \emptyset$. In this case, $0 \leq \phi_P^\eta(x,y) < \eta$, considering that $0 \leq \theta_a(x,y) < 1$ for $x \neq y$. Note that $\theta_a(x,y) = 1$ for only $a \in O_P(x) \cap O_P(y)$. Therefore, η could be understood as a value that separates the following cases:

(a) $\phi_P^\eta(x,y) > 1 - \eta$: $O_P(x) \cap O_P(y) \neq \emptyset$ and for all $a \in P$, $\theta_a(x,y) > 0$

(b) $\eta > \phi_P^\eta(x,y) > 0$: $O_P(x) \cap O_P(y) = \emptyset$, for all $a \in P$, $\theta_a(x,y) > 0$

(c) $\phi_P^\eta(x,y) = 0$: $\exists a \in P$, $\theta_a(x,y) = 0$.

In order to separate the cases between (a) and (b), η should satisfy $1 - \eta \geq \eta$. From those above, we have the constraint of $\eta \in (0, 0.5]$. If $\eta < 0.5$, $\phi_P^\eta(x,y)$ never

Table 4.3: Tolerance degree among objects[1]

	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0	0	0	0	0	0	0.221
x_2		0	0	0	0	0	0.721
x_3			0	0	0	0	0
x_4				0.533	0	0	0.949
x_5					0	0	0.148
x_6						0	0
x_7							0

takes a value between η and $1-\eta$. Hence, we define tolerance degrees by fixing $\eta = 0.5$, though $\phi^{\eta}_{P}(x,y)$ never takes the value of 0.5 as known from the conditions of (a) and (b):

$$\phi_P(x,y) = \begin{cases} 0, & \textit{if } \exists a \in O_P(x) \cap O_P(y) \neq \emptyset, f_a(x) \neq f_a(y), \\ \frac{1}{2} \prod_{\forall a \in P-(O_P(x)\cap O_P(y))} \theta_a(x,y) + \frac{1}{2}\Theta_P(x,y), & \text{otherwise}, \end{cases} \tag{4.13}$$

Tolerance degrees with $\eta = 0.5$ let us differentiate the three cases discussed before by seeing whether the degree is greater/less than 0.5 or whether it is greater than/equal to zero. This feature might be useful, because users can control conditions of tolerance based on equivalence existence and probability of matching with just a threshold value. It thus can be used in some valued tolerance/similarity relation based rough set models. This process shall be discussed in the next step.

Back to the example in Table 3.1, using imperfect representation in Definition 3.3.1, for the *Headache* attribute value of x_1 we have: $t^{x_5}_{Headache,1} = \{yes, no\}$, $p^{x_5}_{Headache,1} = 1.0$, $|T^{x_5}_{Headache}| = 1$. Consequently, $Pr_{Headache}(f_{Headache}(x_5) = yes) = 0.33$ and $Pr_{Headache}(f_{Headache}(x_5) = no) = 0.67$ if "*Method of the frequency of attribute value related to concepts*" is employed. Using the same way, we can also calculate

[1]Note that tolerance relations are symmetric so we need only half of the elements in the table.

probabilities of attribute values for all objects. Based on that, the tolerance degrees among objects in terms of all attributes according to (4.13) can be shown in Table 4.3.

In fact, we can choose another probability of matching on an attribute for (4.12) and (4.13). For example, we can choose $\theta_a(x, y) = 1/|V_a|$ as shown in [72]. The choice might depend on probability distribution of attribute values in each system.

The probabilistic terms in our tolerance degree look similar to those used by Stefanowski [73]. However, our approach uses probabilistic terms as pieces of evidence to derive a tolerance relation. Furthermore, this term is combined with *equivalence existence* to define the relation. On the other hand, in probabilistic approach proposed in [73], the authors suppose a *priori* assumption that there exists a uniform probability distribution on every attribute domain and compute tolerance classes based on the joint probability distribution. Their aim seems to define approximation spaces applicable in many cases. Such tolerance classes could be used in some applications, but we believe not in most.

Now, we define extended tolerance relation by controlling tolerance degree with a threshold.

Definition 4.3.3. Given imperfect information system $I = (U, A)$ and attribute set $P \subseteq A$ and given a threshold α, the extended tolerance relation is defined as follows:

$$ETR_P^{\alpha}(x, y) \Leftrightarrow \phi_P(x, y) \geq \alpha. \tag{4.14}$$

It is easy to observe that this relation is reflexive and symmetric but not necessarily transitive. From Table 4.3, if threshold $\alpha = 0.5$ is given, x_4 is tolerant of x_5 based on this relation.

By changing the threshold, for incomplete data, if $Pr_a(f_a(x) = v) > 0$ for all $v \in V_a, a \in P$ when $f_a(x) = *$, we are able to get the same results as those by the relations discussed in Section 3.1. These connections can be formalized by the following propositions:

Proposition 4.3.1. *Let $I = (U, A)$ be an incomplete information system. Given attribute set $P \subseteq A$ and $x \in U$, $Pr_a(f_a(x) = v) > 0$ for all $v \in V_a, a \in P$ when $f_a(x) = *$, if $\alpha \to 0$, then $ETR_P^{\alpha}(x, y) \Leftrightarrow TOR_P(x, y)$.*

Proof. When $\alpha \to 0$, $ETR_P^{\alpha}(x, y)$ is obtained as

$$\begin{aligned} ETR_P^{\alpha}(x, y) \Leftrightarrow & \phi_P(x, y) > 0 \\ \Leftrightarrow & \forall a \in P, \theta_a(x, y) > 0) \\ \Leftrightarrow & (\forall a \in P, (f_a(x) = *) \vee (f_a(y) = *)) \\ & \vee (\forall a \in O_P(x) \cap O_P(y) \neq \emptyset, f_a(x) = f_a(y)) \\ \Leftrightarrow & TOR_P(x, y). \end{aligned}$$

□

This proposition shows that with $\alpha \to 0$, the *extended tolerance relation* may get the same results as the *tolerance relation* in incomplete information systems if the probability of object attribute values are greater than zero for any values in the domain.

Proposition 4.3.2. *Let $I = (U, A)$ be an incomplete information system. Given attribute set $P \subseteq A$ and $x \in U$, $Pr_a(f_a(x) = v) > 0$ for all $v \in V_a, a \in P$ when $f_a(x) = *$, if $\alpha = 0.5$, then $ETR_P^{\alpha}(x, y) \Rightarrow LTOR_P(x, y)$ for any x, y and $ETR_P^{\alpha}(x, y) \Leftarrow LTOR_P(x, y)$ except the case such that $O_P(x) = O_P(y) = \emptyset$.*

Proof. When $\alpha = 0.5$, $ETR_P^{\alpha}(x, y)$ is obtained as

$$\begin{aligned} ETR_P^{\alpha}(x, y) \Leftrightarrow & \phi_P(x, y) \geq 0.5 \Leftrightarrow (\Theta_P(x, y) = 1) \wedge (\forall a \in P, \theta_a(x, y) > 0) \\ \Leftrightarrow & (O_P(x) \cap O_P(y) \neq \emptyset) \wedge (\forall a \in O_P(x) \cap O_P(y), f_a(x) = f_a(y)) \\ \Rightarrow & LTOR_P(x, y). \end{aligned}$$

Then it is evident that $LTOR_P(x, y) \Rightarrow ETR_P^{\alpha}(x, y)$ except the case where $O_P(x) = O_P(y) = \emptyset$. $\square$

This proposition notices that with $\alpha = 0.5$, the *extended tolerance relation* is an expansion of the *limited tolerance relation* in incomplete information systems if the probability of object attribute values are greater than zero for any values in the domain.

Proposition 4.3.3. *Let $I = (U, A)$ be an incomplete information system. Given attribute set $P \subseteq A$ and $x \in U$, if $\alpha = 1.0$, then $ETR_P^{\alpha}(x, y) \Leftrightarrow EQU_P(x, y)$ if $x \neq y$.*

Proof. Consider that $ETR_P^{\alpha}(x, y) \Leftrightarrow \phi_P(x, y) = 1$. As discussed before, $\phi_P(x, y) = 1$ holds only in the two cases where $x = y$ and where $O_P(x) \cap O_P(y) = P$ and for all $a \in P$,$f_a(x) = f_a(y)$, which is equivalent to $EQU_P(x, y)$. $\square$

This proposition shows that with $\alpha = 1.0$, the *extended tolerance relation* is able to get the same results as *equivalence relation* in incomplete information systems.

It should be noted that similarity/tolerance relations discussed in this dissertation are introduced to cope with imperfect information. However, we could also define those relations even in complete information tables. For example, the relation "*subclass-of*" is a similarity relation. It is clearly transitive, but not necessarily symmetric. We can also take the relation "*friend-of*" as an example of tolerance relation and examine its properties in the same way.

Definition 4.3.4. Let $I = (U, A)$ be an imperfect information system. Given attribute set $P \subseteq A$ and $x, y, z \in U$, if $\phi_P(y, x) > \phi_P(z, x)$, then y is more tolerant of x than z based on *extended tolerance relation*.

Proposition 4.3.4. *Let $I = (U, A)$ be an imperfect information system. Given attribute set $P \subseteq A$ and $x, y, z \in U$, if $O_P(x) \cap O_P(y) \neq \emptyset$, for all $a \in O_P(x) \cap O_P(y)$,*

$f_a(x) = f_a(y)$ and $O_P(x) \cap O_P(z) = \emptyset$ then y is more tolerant of x than z based on the extended tolerance relation.

Proof. Since $O_P(x) \cap O_P(y) \neq \emptyset$ and for all $a \in O_P(x) \cap O_P(y)$, $f_a(x) = f_a(y)$, from (4.13) we have $\phi_P(y,x) > 0.5$. We also have $\phi_P(z,x) < 0.5$ since $O_P(x) \cap O_P(z) = \emptyset$. Therefore, $\phi_P(y,x) > \phi_P(z,x)$. This is defined as y is more tolerant of x than z based on the *extended tolerance relation.* □

4.3.2 Neighbourhood and Approximations

Now, with a relation, we can derive a neighbourhood, which consists of *successor* and *predecessor sets*, of an object as discussed in Chapter 2. Due to symmetric property of *extended tolerance relation*, successor is the same set as predecessor. For this relation, therefore, we can introduce for any object $x \in U$ a tolerant set:

$$ET_P^\alpha(x) = \{y \in U | ETR_P^\alpha(y,x)\}. \tag{4.15}$$

From Table 4.3, given the threshold $\alpha = 0.1$, $P = A$, we have $ET_A^{0.1}(x_8) = \{x_2, x_4, x_5, x_8\}$.

Proposition 4.3.5. *Let $I = (U, A)$ be an imperfect information system, $P \subseteq A$, for all $x \in U$, if $\alpha \leq \beta$, then $ET_P^\alpha(x) \supseteq ET_P^\beta(x)$.*

Proof. Consider the following:

$$\begin{aligned}
ET_P^\alpha(x) &= \{y \in U | ETR_P^\alpha(y,x)\} \\
&= \{y \in U | \phi_P(y,x) \geq \alpha\} \\
&= \{y \in U | \phi_P(y,x) \geq \beta\} \cup \{y \in U | \alpha \leq \phi_P(y,x) < \beta\} \\
&= ET_P^\beta(x) \cup \{y \in U | \alpha \leq \phi_P(y,x) < \beta\} \supseteq ET_P^\beta(x)
\end{aligned}$$

□

This proposition shows that the cardinality of the tolerance class of x shall decrease if we increase the threshold to control the tolerance degree.

From tolerance classes, we can define approximations using singleton, subset and concept methods discussed in Chapter 2. Singleton lower approximation an upper approximation of an object set $X \subseteq U$ are defined as follows:

$$\underline{SingleAppr}_P^\alpha(X) = \{x \in U | ET_P^\alpha(x) \subseteq X\}, \tag{4.16}$$
$$\overline{SingleAppr}_P^\alpha(X) = \{x \in U | ET_P^\alpha(x) \cap X \neq \emptyset\}. \tag{4.17}$$

From Table 4.3, given threshold $\alpha = 0.6$ and $P = A$, we have approximation space for the concept $X = \{x \in U | f_{Flu}(x) = yes\}$:

$$\underline{SingleAppr}_P^{0.6}(X) = \{x_1, x_2, x_8\}, \tag{4.18}$$
$$\overline{SingleAppr}_P^{0.6}(X) = \{x_1, x_2, x_4, x_5, x_8\}. \tag{4.19}$$

The Subset and Concept approximation spaces are also defined as follows:

$$\underline{SubsetAppr}_P^\alpha(X) = \cup\{ET_P^\alpha(x) | x \in U \wedge ET_P^\alpha(x) \subseteq X\}, \tag{4.20}$$
$$\overline{SubsetAppr}_P^\alpha(X) = \cup\{ET_P^\alpha(x) | x \in U \wedge ET_P^\alpha(x) \cap X \neq \emptyset\}, \tag{4.21}$$
$$\underline{ConceptAppr}_P^\alpha(X) = \cup\{ET_P^\alpha(x) | x \in X \wedge ET_P^\alpha(x) \subseteq X\}, \tag{4.22}$$
$$\overline{ConceptAppr}_P^\alpha(X) = \cup\{ET_P^\alpha(x) | x \in X \wedge ET_P^\alpha(x) \cap X \neq \emptyset\} \tag{4.23}$$

Approximation spaces defined based on extended tolerance relation have some properties suggested by Pawlak [60, 61] as well as other properties. Table 4.4 shows which properties of the original rough set model are satisfied with singleton, subset, and concept definitions.

These properties within our approach can be proved as the same as those in the Grzymala-Busse and Wojciech Rzasa study [28] and the Pawlak research [61].

Table 4.4: Properties of approximations for the three definitions

	Properties	Singleton	Subset	Concept
1a	$\underline{appr}_P^\alpha(X) \subseteq X,$	$\surd$	$\surd$	$\surd$
1b	$X \subseteq \overline{appr}_P^\alpha(X),$	$\surd$	$\surd$	$\surd$
2a	$\underline{appr}_P^\alpha(\emptyset) = \emptyset,$	$\surd$	$\surd$	$\surd$
2b	$\overline{appr}_P^\alpha(\emptyset) = \emptyset,$	$\surd$	$\surd$	$\surd$
3a	$\underline{appr}_P^\alpha(U) = U,$	$\surd$	$\surd$	$\surd$
3b	$\overline{appr}_P^\alpha(U) = U,$	$\surd$	$\surd$	$\surd$
4a	$X \subseteq Y \Rightarrow \underline{appr}_P^\alpha(X) \subseteq \underline{appr}_P^\alpha(Y),$	$\surd$	$\surd$	$\surd$
4b	$X \subseteq Y \Rightarrow \overline{appr}_P^\alpha(X) \subseteq \overline{appr}_P^\alpha(Y),$	$\surd$	$\surd$	$\surd$
5a	$\underline{appr}_P^\alpha(X \cup Y) \supseteq \underline{appr}_P^\alpha(X) \cup \underline{appr}_P^\alpha(Y),$	$\surd$	$\surd$	$\surd$
5b	$\overline{appr}_P^\alpha(X \cup Y) = \overline{appr}_P^\alpha(X) \cup \overline{appr}_P^\alpha(Y),$	$\surd$	$\surd$	
6a	$\underline{appr}_P^\alpha(X \cap Y) = \underline{appr}_P^\alpha(X) \cap \underline{appr}_P^\alpha(Y),$	$\surd$		
6b	$\overline{appr}_P^\alpha(X \cap Y) \subseteq \overline{appr}_P^\alpha(X) \cap \overline{appr}_P^\alpha(Y),$	$\surd$	$\surd$	$\surd$
7a	$\underline{appr}_P^\alpha(X) = \underline{appr}_P^\alpha(\underline{appr}_P^\alpha(X)),$		$\surd$	$\surd$
7b	$\underline{appr}_P^\alpha(X) = \overline{appr}_P^\alpha(\underline{appr}_P^\alpha(X)),$			
7c	$\overline{appr}_P^\alpha(X) = \overline{appr}_P^\alpha(\overline{appr}_P^\alpha(X)),$			
7d	$\overline{appr}_P^\alpha(X) = \underline{appr}_P^\alpha(\overline{appr}_P^\alpha(X)),$			
8a	$\underline{appr}_P^\alpha(X) = \sim \overline{appr}_P^\alpha(\sim X),$	$\surd$	$\surd$	$\surd$
8b	$\overline{appr}_P^\alpha(X) = \sim \underline{appr}_P^\alpha(\sim X),$	$\surd$	$\surd$	$\surd$

$\surd$ indicates that the property is satisfied. $\underline{appr}_P^\alpha(X)$ and $\overline{appr}_P^\alpha(X)$ are lower and upper approximations that can be defined by singleton, subset, and concept methods.

Approximation spaces of those definition methods, in general, do not have properties 7a-7d. However, they are likely to satisfy the weaker versions of 7a-7d, which are defined by Yao [85].

Besides, our tolerance relation is controlled by thresholds of tolerance degrees. Therefore, new properties for thresholds can be introduced as shown below:

Proposition 4.3.6. *Let $I = (U, A)$ be an imperfect information system, $X \subseteq U$, and $P \subseteq A$. The following properties shall hold for arbitrary lower approximation $\underline{appr}_P^\alpha(X)$ and upper approximation $\overline{appr}_P^\alpha(X)$ defined by singleton, subset, and concept methods:*

$$(9a).\ \text{if } \alpha \le \beta \Rightarrow \underline{appr}_P^{\alpha}(X) \subseteq \underline{appr}_P^{\beta}(X),$$
$$(9b).\ \text{if } \alpha \le \beta \Rightarrow \overline{appr}_P^{\alpha}(X) \supseteq \overline{appr}_P^{\beta}(X).$$

Proof of (9a). Take an element $x \in \underline{appr}_P^{\alpha}(X)$. From any of the lower approximation definitions, $ET_P^{\alpha}(x) \subseteq X$ is derived. Since $ET_P^{\alpha}(x) \supseteq ET_P^{\beta}(x)$ according to Proposition 4.3.5, we get that $ET_P^{\beta}(x) \subseteq X$, and then $x \in \underline{appr}_P^{\beta}(X)$. Thus, if $x \in \underline{appr}_P^{\alpha}(X)$, then $x \in \underline{appr}_P^{\beta}(X)$ (note that $x \in ET_P^{\alpha}(x)$). □

Proof of (9b). Take an element $x \in \overline{appr}_P^{\beta}(X)$. From any of the upper approximation definitions, $ET_P^{\beta}(x) \cap X \neq \emptyset$ is derived. Since $ET_P^{\alpha}(x) \supseteq ET_P^{\beta}(x)$ according to Proposition 4.3.5, we get that $ET_P^{\alpha}(x) \cap X \neq \emptyset$, and then $x \in \overline{appr}_P^{\alpha}(X)$. Thus, if $x \in \overline{appr}_P^{\beta}(X)$, then $x \in \overline{appr}_P^{\alpha}(X)$ (note that $x \in ET_P^{\alpha}(x)$). □

Obviously, the greater threshold is the smaller upper approximation is obtained. While the increasing of the cardinality of lower approximation follows the rise of the threshold. Therefore, it is possible to widen or thin the boundary between lower and upper approximations of an objects set by changing the threshold.

4.4 Summary

This chapter studies a rough set theory for imperfect information systems based on valued tolerance/similarity relations and establishes a new model called extended tolerance relation based rough set model. Frequency of attribute values appearing in the decision table could be utilized to estimate the probability of matching among data items on an attribute. Tolerance degrees are then calculated based on a combination of existence of equivalence and probabilities of matching on some attributes. Given a threshold to control tolerance degrees, a tolerance relation is defined.

The approach is an extension of some rough set models and could solve the problem existing in tolerance relation of Kryszkiewicz in incomplete information systems. By adjusting the threshold, we are able to get the same results as tolerance, limited tolerance, and equivalence relations. The variable threshold also gives us a means to widen or thin the boundary region between lower and upper approximations. Actually, various lower and upper approximations are obtained using the approach, and users can choose a threshold that suits their requirements.

The approach is discussed on discrete values. However, it is possible to apply in imperfect information containing continuous values by altering the definition of equivalence existence. For this purpose, the equivalence relation shown in Equation (4.7) should be used.

Chapter 5

Rough Set Model Based on Dempster-Shafer Theory

As discussed, it is inappropriate to apply a method that deals only with missing values to information systems containing missing, imprecise and uncertain values. In the previous chapter, one possible solution would be to combine the transformation of any type of imperfect data with probabilistic values and then to apply a probabilistic method. Such a case, we must estimate probability of attribute values for each object. However, it is not always possible even to assume subjective probabilities when we know little knowledge about the domain. This chapter will propose a new method of rough set definition based on Dempster-Shafer theory that can deal with any type of imperfect data discussed in Chapter 3 without evaluating probability of attribute values.

The chapter firstly gives the basic notions of Dempster-Shafer Theory. It then introduces a set of basic relations among objects and mass belief functions of hypotheses of basic relations. Finally, a new rough set model based on Dempster-Shafer theory will be proposed.

5.1 Evidence Theory and Combination Rules

Dempster-Shafer theory

Dempster-Shafer theory (DST) is a mathematical theory of evidence [66]. It allows us to represent all kinds of imperfect data we discussed in Chapter 3, i.e., missing, imprecise, and uncertain data. In addition, it provides us with a tool for combining multiple evidence of relations obtained from independent sources of information.

Let Θ be a discrete and finite universal set. Function m, which is called a basic belief assignment (*bba*) or mass function, is defined as $m : 2^\Theta \to [0,1]$, where $m(\emptyset) = 0$ and $\sum_{\forall \mathbb{A} \subseteq \Theta} m(\mathbb{A}) = 1$ must be satisfied. 2^Θ is the power set of Θ [82].

The *bba* $m(\mathbb{A})$ represents the degree of belief that supports hypothesis $\mathbb{A}$. The mass of belief does not include a mass attributed to any subsets of $\mathbb{A}$ but, instead, measures the amount of belief of $\mathbb{A}$ itself. Subset $\mathbb{A}$ satisfying $m(\mathbb{A}) > 0$ is a focal element. A pair (m, F), where F is the set of all focal elements, is called the body of evidence.

In contrast to *bba*, belief function denoted by *Bel* of set $\mathbb{A}$ does include the mass of all subsets of $\mathbb{A}$. Another measure, plausibility function named *Pl*, is the sum of all of the masses of sets that intersect set of interest $\mathbb{A}$. Belief and plausibility measures are defined as follows:

$$Bel(\mathbb{A}) = \sum_{\mathbb{X} \subseteq \mathbb{A}} m(\mathbb{X}), \tag{5.1}$$

$$Pl(\mathbb{A}) = \sum_{\mathbb{X} | \mathbb{X} \cap \mathbb{A} \neq \emptyset} m(\mathbb{X}). \tag{5.2}$$

Belief and plausibility functions are connected by a dual property such that $Bel(\mathbb{A}) = 1 - Pl(\sim \mathbb{A})$. Furthermore, $Bel(\mathbb{A}) \leq Pl(\mathbb{A})$ for all $\mathbb{A} \in \Theta$. It is also possible to obtain basic belief assignments from belief measures with the following inverse function

$$m(\mathbb{A}) = \sum_{\mathbb{X} \subseteq \mathbb{A}} (-1)^{|\mathbb{A}|-|\mathbb{X}|} Bel(\mathbb{X}). \tag{5.3}$$

where $|\mathbb{A}|$ denotes the cardinality of $\mathbb{A}$.

Rule of combination

Suppose that there are two independent sources of information and that we must combine these pieces of evidence from two sets of mass assignment along with the sources. Dempster's rule derives common shared belief among multiple sources and ignores all conflicting (non-shared) belief through a normalization factor. The combination is calculated by the following equation

$$m(\mathbb{A}) = \frac{\sum_{\mathbb{X},\mathbb{Y}|\mathbb{X}\cap\mathbb{Y}=\mathbb{A}} m_1(\mathbb{X}) \bullet m_2(\mathbb{Y})}{1-K}, \tag{5.4}$$

where $\mathbb{A} \neq \emptyset$ and $K = \sum_{\mathbb{X},\mathbb{Y}|\mathbb{X}\cap\mathbb{Y}=\emptyset} m_1(\mathbb{X}) \bullet m_2(\mathbb{Y})$.

When three or more bodies of evidence are combined, $m_1(\mathbb{X}) \bullet m_2(\mathbb{Y})$ and $\mathbb{X} \cap \mathbb{Y}$ are replaced by $m_1(\mathbb{X}_1) \bullet m_2(\mathbb{X}_2) \bullet \cdots \bullet m_n(\mathbb{X}_n)$ and $\mathbb{X}_1 \cap \mathbb{X}_2 \cap \cdots \cap \mathbb{X}_n$, respectively [82].

There are also numerous alternative rules [13, 31, 80, 81, 82, 91] to be applied depending on different requirements.

5.2 Modelling Relations Based on Demspter-Shafer Theory

As shown in Chapter 3, imprecise data was represented by a set of possible values, and uncertain data was defined by a probability distribution on the attribute domain in the related work. The representation for multiple type of imperfectness was also introduced. In this section, the representation is also used as the body of evidence in the framework of Dempster-Shafer theory to develop a new model of relations.

An attribute value of this model is represented by body of evidence (m_a, F_a) that is a pair consisting a mass function and a set of focal elements on attribute $a \in A$. It may be represented later by (m_a^x, F_a^x), when an object x must be identified. To illustrate, when the value set of attribute a is $V_a = \{1, 2, 3, 4\}$, the missing value is represented as total ignorance, that is, $m_a(\{1, 2, 3, 4\}) = 1.0$. Examples of imprecise data are $m_a(\{1, 2, 3\}) = 1.0$, $m_a(\{1, 3\}) = 1.0$, $m_a(\{3, 4\}) = 1.0$, etc. Uncertain data are represented, for example, as $m_a(\{1\}) = 0.1$, $m_a(\{2\}) = 0.2$, $m_a(\{3\}) = 0.3$, and $m_a(\{4\}) = 0.4$. As shown by the explanation in the previous section, Dempster-Shafer theory could represent more complex imperfect data that mixes imprecise and uncertain data such as $m_a(\{1\}) = 0.5$, $m_a(\{2, 3\}) = 0.3$, $m_a(\{1, 2, 3, 4\}) = 0.2$.

What should be noted here is that Dempster-Shafer theory is an extension of possibility theory and probability theory [36]. This means that a fuzzy set, which is usually regarded as a possibility distribution, could be represented in the form of a body of evidence and that the imperfect data used in this model represents both imprecision and uncertainty and fuzziness. Fuzzy set $0.3/1 + 1.0/2 + 0.8/3$, for example, is given by mass functions $m(\{2\}) = 0.2$, $m(\{2, 3\}) = 0.5$ and $m(\{1, 2, 3\}) = 0.3$, where $0.3/1$ represents the membership value of "1" is 0.3, and "+" does the union operation.

Let $t_{a,i}^x \subseteq V_a$ be the i-th set of overall s possible value sets of x on a and $m_{a,i}^x = p_{a,i}^x > 0$ be its basic belief assignment. Pair (m_a^x, F_a^x), where $F_a^x = \{t_{a,i}^x | 1 \leq i \leq s\}$, represents the body of evidence on a. F_a^x is a *complete set* when $\sum_{1 \leq i \leq s} m_{a,i}^x = 1$.

As discussed in Chapter 3, a *precise* value and a *missing* value are also considered as two extreme kinds of imprecision. A value is precise when the set of possible values is a singleton and the *bba* is 1. In this case, $|F_a^x| = 1$, $|t_{a,1}^x| = 1$ and $m_{a,1}^x = 1$. Missing values without pre-defined probability distribution could be regarded as imprecise information where the set of possible values encompasses the entire attribute domain such that $|F_a^x| = 1$, $t_{a,1}^x = V_a$ and $m_{a,1}^x = 1$. In Chapter 4, a valued tolerance definition was also proposed. This relation is based on the probability of matching using probability of object attribute evaluation on $t_{a,i}^x$.

However, in this chapter, we will construct a set of basic relations on which we can define various equivalence, tolerance and similarity relations by combining some of basic relations.

Consider a case in which two objects x and y have an imperfect value of attribute a, that is, the values are represented by sets t_a^x and t_a^y of possible values, respectively. We then define several basic relations based on some situations listed as follows:

(1) If $t_a^x = t_a^y$ and are singletons, x and y are equivalent with respect to a and denoted by relation $\mathcal{E}_a(x, y)$.

(2) If $t_a^x = t_a^y$ and are not singletons, the true values of x, y are contained in both t_a^x and t_a^y. We thus say values of x and y are mutually inclusive for a, and denote the relation by $\mathcal{M}_a(x, y)$, where $\mathcal{M}_a(x, y)$ means that the true values of x, y are contained in t_a^x, t_a^y, respectively, except that t_a^x and t_a^y are singletons simultaneously.

(3) If t_a^y is a proper subset of t_a^x, the true value of y is contained in both t_a^x and t_a^y, but the true value of x is possibly, but not necessarily, contained in t_a^y. Therefore, we denote the relation between x and y with $\mathcal{I}_a(x, y)$ or $\mathcal{M}_a(x, y)$,

where $\mathcal{I}_a(x, y)$ means that the true value of y is contained in t_a^x but that the true value of x is not contained in t_a^y.

(4) If t_a^x is a proper subset of t_a^y, the true value of x is contained in t_a^y and the true value of y is possibly, but not necessarily, contained in t_a^x. We therefore denote the relation between x and y with $\mathcal{R}_a(x, y)$ or $\mathcal{M}_a(x, y)$. $\mathcal{R}_a(x, y)$ means that the true value of x is contained in t_a^y but that the true value of y is not contained in t_a^x.

(5) If t_a^x and t_a^y have a non-empty intersection but neither includes the other, then the true value of x may or may not be contained in t_a^y and vice versa. We therefore denote the relation between x and y with $\mathcal{I}_a(x, y)$ or $\mathcal{R}_a(x, y)$ or $\mathcal{M}_a(x, y)$ or $\mathcal{D}_a(x, y)$, where $\mathcal{D}_a(x, y)$ means that the true value of x is not contained in t_a^y nor is that of y contained in t_a^x.

(6) If t_a^x and t_a^y have an empty intersection, then the true value of x is not contained in t_a^y nor is that of y contained in t_a^x. The relation between x and y is therefore denoted by $\mathcal{D}_a(x, y)$.

Relations $\mathcal{E}_a(x, y)$, $\mathcal{M}_a(x, y)$, $\mathcal{I}_a(x, y)$, $\mathcal{R}_a(x, y)$ and $\mathcal{D}_a(x, y)$ are defined formally as follows:

$$\begin{aligned}
\mathcal{E}_a(x, y) \Leftrightarrow & (f_a(x) \in t_a^y) \wedge (f_a(y) \in t_a^x) \\
& \wedge (|t_a^x| = |t_a^y| = 1) && (5.5) \\
\mathcal{M}_a(x, y) \Leftrightarrow & (f_a(x) \in t_a^y) \wedge (f_a(y) \in t_a^x) \\
& \wedge \neg(|t_a^x| = |t_a^y| = 1) && (5.6) \\
\mathcal{I}_a(x, y) \Leftrightarrow & (f_a(x) \notin t_a^y) \wedge (f_a(y) \in t_a^x) && (5.7) \\
\mathcal{R}_a(x, y) \Leftrightarrow & (f_a(x) \in t_a^y) \wedge (f_a(y) \notin t_a^x) && (5.8) \\
\mathcal{D}_a(x, y) \Leftrightarrow & (f_a(x) \notin t_a^y) \wedge (f_a(y) \notin t_a^x) && (5.9)
\end{aligned}$$

Let $\Theta = \{\mathcal{E}, \mathcal{I}, \mathcal{R}, \mathcal{M}, \mathcal{D}\}$ be the set of all of the relation types where $\mathcal{E}$, $\mathcal{I}$, $\mathcal{R}$, $\mathcal{M}$ and $\mathcal{D}$ are used instead of $\mathcal{E}_a(x, y)$, $\mathcal{I}_a(x, y)$, $\mathcal{R}_a(x, y)$, $\mathcal{M}_a(x, y)$ and $\mathcal{D}_a(x, y)$, respectively, just for simplicity. Obviously, Θ is "exhaustive" and "exclusive," which means only that an element must always be true. In this case, any pair (x, y) satisfies one of these relations and only one holds for pair (x, y). It is therefore possible to define Θ as the frame of discernment of relations. Now, from the body of evidence, we calculate bba for possible hypotheses in 2^{Θ} as follows:

$$m_a(\{\mathcal{E}\}) = \sum_{\substack{t^x_{a,i}=t^y_{a,j} \\ |t^x_{a,i}|=|t^y_{a,j}|=1}} m^x_{a,i} \bullet m^y_{a,j} \tag{5.10}$$

$$m_a(\{\mathcal{M}\}) = \sum_{\substack{t^x_{a,i}=t^y_{a,j} \\ |t^x_{a,i}|\neq 1}} m^x_{a,i} \bullet m^y_{a,j} \tag{5.11}$$

$$m_a(\{\mathcal{M}, \mathcal{I}\}) = \sum_{t^y_{a,j}\subset t^x_{a,i}} m^x_{a,i} \bullet m^y_{a,j} \tag{5.12}$$

$$m_a(\{\mathcal{M}, \mathcal{R}\}) = \sum_{t^x_{a,i}\subset t^y_{a,j}} m^x_{a,i} \bullet m^y_{a,j} \tag{5.13}$$

$$m_a(\{\mathcal{M}, \mathcal{I}, \mathcal{R}, \mathcal{D}\}) = \sum_{\substack{t^x_{a,i}\cap t^y_{a,j}\neq\emptyset \\ t^x_{a,i}\not\subseteq t^y_{a,j};t^y_{a,j}\not\subseteq t^x_{a,i}}} m^x_{a,i} \bullet m^y_{a,j} \tag{5.14}$$

$$m_a(\{\mathcal{D}\}) = \sum_{t^x_{a,i}\cap t^y_{a,j}=\emptyset} m^x_{a,i} \bullet m^y_{a,j} \tag{5.15}$$

Masses will be assigned to 0 for the rest of the hypotheses in the power set. Formally, $m_a(\mathbb{A}) = 0$, $\forall \mathbb{A} \in 2^{\Theta} - \{\{\mathcal{E}\}, \{\mathcal{M}\}, \{\mathcal{M}, \mathcal{I}\}, \{\mathcal{M}, \mathcal{R}\}, \{\mathcal{M}, \mathcal{I}, \mathcal{R}, \mathcal{D}\}, \{\mathcal{D}\}\}$. In the relation of x with x itself, the mass function of the equivalence relation is assigned to 1 so that $m_a(\{\mathcal{E}_a(x, x)\}) = 1$.

Table 5.1 shows an example of imperfect decision table $I = (U, A \cup \{d\})$, where $U = \{x_1, \cdots, x_7\}$ and $A = \{a_1, a_2, a_3\}$. We assume $V_{a_1} = \{1, 2, 3\}$, $V_{a_2} = \{1, 2, 3, 4\}$,

Table 5.1: An information system with uncertainty and imprecision

U	a_1	a_2	a_3	d
x_1	1.0 {1}	1.0 {1}	1.0 {1}	d_1
x_2	1.0 {1,2}	1.0 {2}	1.0 {1}	d_1
x_3	1.0 {1}	1.0 {2}	1.0 {2}	d_1
x_4	0.6 {1} 0.4 {2,3}	0.7 {1,2} 0.3 {2,3}	0.3 {1} 0.7 {2,3}	d_1
x_5	0.9 {2,3} 0.1 {3}	0.5 {1,2} 0.5 {3}	0.2 {1,2} 0.8 {2,3}	d_2
x_6	1.0 {3}	1.0 {4}	1.0 {3}	d_2
x_7	1.0 {2}	1.0 {3,4}	1.0 {3}	d_2

{•} shows a set of possible values and the figure before the set represents the *bba* of the set.

$V_{a_3} = \{1, 2, 3\}$ and $V_d = \{d_1, d_2\}$. In this table, let us calculate the basic belief assignment for each hypothesis on the relation of x_4 with x_5 on attribute a_1. We first have $t^{x_4}_{a_1,1} = \{1\}$, $m^{x_4}_{a_1,1} = 0.6$, $t^{x_4}_{a_1,2} = \{2, 3\}$, $m^{x_4}_{a_1,2} = 0.4$, and $t^{x_5}_{a_1,1} = \{2, 3\}$, $m^{x_5}_{a_1,1} = 0.9$, $t^{x_5}_{a_1,2} = \{3\}$, $m^{x_5}_{a_1,2} = 0.1$. From equations (5.10) to (5.15) we then have $m_{a_1}(\{\mathcal{E}\}) = 0.0$, $m_{a_1}(\{\mathcal{M}\}) = 0.36$, $m_{a_1}(\{\mathcal{M}, \mathcal{I}\}) = 0.04$, $m_{a_1}(\{\mathcal{M}, \mathcal{R}\}) = 0.0$, $m_{a_1}(\{\mathcal{M}, \mathcal{I}, \mathcal{R}, \mathcal{D}\}) = 0.0$ and $m_{a_1}(\{\mathcal{D}\}) = 0.6$.

Proposition 5.2.1. *Let $I = (U, A)$ be an imperfect information system, and $\Theta = \{\mathcal{E}, \mathcal{I}, \mathcal{R}, \mathcal{M}, \mathcal{D}\}$ be the set of the basic relations defined by equations from* (5.5) *to* (5.9). *The sum of the masses of all hypotheses from Θ obtained by equations from* (5.10) *to* (5.15) *is then one.*

Proof. $\sum_{\forall \mathbb{A} \subseteq \Theta} m_a(\mathbb{A}) = m_a(\{\mathcal{E}\}) + m_a(\{\mathcal{M}\}) + m_a(\{\mathcal{M}, \mathcal{I}\}) + m_a(\{\mathcal{M}, \mathcal{R}\}) + m_a(\{\mathcal{M}, \mathcal{I}, \mathcal{R}, \mathcal{D}\}) + m_a(\{\mathcal{D}\}) = \sum_{ij} m^x_{a,i} \bullet m^y_{a,j} = \sum_i m^x_{a,i} \bullet \sum_j m^y_{a,j} = 1.$ □

We can see that the example shown above with x_4 and x_5 satisfies the proposition. Proposition 5.2.1 notices that the set of relations Θ and its basic belief assignments satisfy two conditions of Dempster-Shafer theory. We are therefore able to derive the belief and plausibility functions of hypotheses from Θ.

Proposition 5.2.2. *Given incomplete information system $I = (U, A)$ and set of attribute $P \subseteq A$, we have*

$$
\begin{aligned}
EQU_P(x, y) \Leftrightarrow & \forall a \in P, Bel_a(\{\mathcal{E}\}) = 1 && (5.16)\\
TOR_P(x, y) \Leftrightarrow & \forall a \in P, Bel_a(\{\mathcal{E}, \mathcal{M}, \mathcal{I}, \mathcal{R}\}) = 1 && (5.17)\\
SIM_P(x, y) \Leftrightarrow & \forall a \in P, Bel_a(\{\mathcal{E}, \mathcal{M}, \mathcal{I}\})) = 1 && (5.18)\\
LTOR_P(x, y) \Leftrightarrow & \forall a \in P, Bel_a(\{\mathcal{M}\})) = 1 \\
& \vee ((\forall a \in P, Bel_a(\{\mathcal{D}\}) = 0) \\
& \wedge (\exists a \in P, Bel_a(\{\mathcal{E}\}) = 1)) && (5.19)
\end{aligned}
$$

Proof. As discussed, in incomplete information systems, if the value of x on a is missing, then $|F_a^x| = 1$, $t_{a,1}^x = V_a$ and $m_{a,1}^x = 1$. The hypothesis $\{\mathcal{E}\}$ therefore represents the fact that $f_a(x) = f_a(y) \neq *$; $\{\mathcal{M}, \mathcal{I}\}$ notices that $f_a(x) = *, f_a(y) \neq *$; $\{\mathcal{M}, \mathcal{R}\}$ shows $f_a(x) \neq *, f_a(y) = *$ while $\{\mathcal{M}\}$ is $f_a(x) = *, f_a(y) = *$ and $\{\mathcal{D}\}$ is $f_a(x) \neq f_a(y) \neq *$. From definitions of these relations defined in Chapter 3, Section 3.1, the proposition is then easily proved. □

Proposition 5.2.2 indicates that the models of relations proposed here are replacements of those for incomplete information systems discussed in Chapter 3. In the next section, we define a new rough set approach in information associated with uncertainty and imprecise values. Note that missing value can be interpreted by imprecision.

5.3 Rough Set Approach Based on Dempster-Shafer Theory

Assuming that any attribute of an information system is independent of all of the other attributes, we consider information on each attribute as a source of evidence.

Table 5.2: Mass of hypotheses

Hypotheses	Sources from attributes		
	a_1	a_2	a_3
$\{\mathcal{E}\}$	0.00	0.00	0.00
$\{\mathcal{M}\}$	0.36	0.35	0.56
$\{\mathcal{M}, \mathcal{I}\}$	0.04	0.15	0.00
$\{\mathcal{M}, \mathcal{R}\}$	0.00	0.00	0.06
$\{\mathcal{M}, \mathcal{I}, \mathcal{R}, \mathcal{D}\}$	0.00	0.15	0.14
$\{\mathcal{D}\}$	0.60	0.35	0.24

Using a rule of combination in evidence theory, it is thus possible to calculate a mass function on Θ taking all attributes into account. We are then able to define a new rough set approach based on Dempster-Shafer theory.

Masses of all hypotheses for the relation between x_4 and x_5 are first calculated from sources of evidence on each attribute and shown in **Table 5.2**. It may now be possible to calculate a mass function on a given set of attributes using a combination rule [66]. However, we must notice that two objects are certainly distinguishable from each other on attribute set $P \subseteq A$ if they have a different value on an attribute in P. Two objects x, y are, for example, distinguishable if $m_a(\{\mathcal{D}\}) = 1$, even if $m_b(\{\mathcal{E}\}) = 1$ for any $b \in P - \{a\}$. We therefore use a combination rule incorporating the above unique property to calculate *bba* with a given set of attributes.

Definition 5.3.1. Given imperfect information system $I = (U, A)$ and attribute set $P \subseteq A$, the mass function of each hypothesis $\mathbb{A} \subseteq \Theta$ is defined as follows:

$$m_P(\mathbb{A}) = \begin{cases} 0, & \text{if } \mathbb{A} \neq \{\mathcal{D}\}, \exists a \in P | m_a(\{\mathcal{D}\}) = 1, \\ 1, & \text{if } \mathbb{A} = \{\mathcal{D}\}, \exists a \in P | m_a(\{\mathcal{D}\}) = 1, \\ m_P^{\mathfrak{R}}(\mathbb{A}), & \textit{otherwise.} \end{cases} \tag{5.20}$$

where $m_P^{\mathfrak{R}}(\mathbb{A})$ represents the *bba* of $\mathbb{A}$ on P by combination rule $\mathfrak{R}$. If Dempster's rule is used, for example, to combine evidence on $P = \{a, b\}$, $m_P^{DempsterRule}(\mathbb{A}) = \frac{\sum_{\mathbb{X},\mathbb{Y}|\mathbb{X}\cap\mathbb{Y}=\mathbb{A}} m_a(\mathbb{X}) \bullet m_b(\mathbb{Y})}{\sum_{\mathbb{X},\mathbb{Y}|\mathbb{X}\cap\mathbb{Y}\neq\emptyset} m_a(\mathbb{X}) \bullet m_b(\mathbb{Y})}$. Problems of Dempster's rule such as *loss of majority opinion* and *total mass to minority* are discussed in the literature [13, 31, 80, 81, 82, 91]. Depending on characteristics of each system, combination rule $\mathfrak{R}$ should be chosen from among the combination rules cited above.

Continuing with the example in Table 5.1, we choose Dempster's rule to induce the relation of x_4 with x_5 on attribute set $P = \{a_1, a_2, a_3\}$. The combination result obtained is $m_P(\{\mathcal{M}\}) = 0.629$, $m_P(\{\mathcal{M}, \mathcal{I}\}) = 0.005$ and $m_P(\{\mathcal{D}\}) = 0.366$. Masses of the rest of the hypotheses in Θ are equal to 0.

With mass assignment of any hypothesis $\mathbb{A} \subseteq \Theta$, we archive *Belief* and *Plausibility* measures of $\mathbb{A}$. Based on this, it is possible to decide which relation should be used between two objects. Undoubtedly, *Belief* measure is the amount of mass that directly supports a given hypothesis. If $Bel(\mathbb{A}) \geq Bel(\mathbb{B})$, then $\mathbb{A}$ is more certain than $\mathbb{B}$ as a hypothesis. If $Pl(\mathbb{A}) \geq Pl(\mathbb{B})$, however, then $\mathbb{A}$ has more potential than $\mathbb{B}$ [41]. In the next part, we therefore introduce *equivalence*, *tolerance* and *similarity* relations based on the scale of belief and plausibility for imperfect information systems.

Definition 5.3.2. Given imperfect information system $I = (U, A)$ and attribute set $P \subseteq A$, *equivalence*, *tolerance* and *similarity* relations between objects x and y based on Dempster-Shafer theory are defined as follows:

DS-Equivalence relation

$$DSE_P(x, y) \Leftrightarrow (Bel_P(\{\mathcal{E}\}) = 1) \tag{5.21}$$

Believable DS-Tolerance relation

$$BelDST_P(x, y) \Leftrightarrow (Bel_P(\{\mathcal{E}, \mathcal{M}, \mathcal{I}, \mathcal{R}\}) \geq Bel_P(\{\mathcal{D}\})) \tag{5.22}$$

Plausible DS-Tolerance relation

$$PlDST_P(x, y) \Leftrightarrow (Pl_P(\{\mathcal{E}, \mathcal{M}, \mathcal{I}, \mathcal{R}\}) \geq Pl_P(\{\mathcal{D}\})) \tag{5.23}$$

Believable DS-Similarity relation

$$\begin{aligned} BelDSS_P(x,y) \Leftrightarrow &(Bel_P(\{\mathcal{E},\mathcal{M},\mathcal{I}\}) \geq Bel_P(\{\mathcal{D}\})) \\ &\wedge (Bel_P(\{\mathcal{M},\mathcal{I}\}) \geq Bel_P(\{\mathcal{M},\mathcal{R}\})) \end{aligned} \tag{5.24}$$

Plausible DS-Similarity relation

$$\begin{aligned} PlDSS_P(x,y) \Leftrightarrow &(Pl_P(\{\mathcal{E},\mathcal{M},\mathcal{I}\}) \geq Pl_P(\{\mathcal{D}\})) \\ &\wedge (Pl_P(\{\mathcal{M},\mathcal{I}\}) \geq Pl_P(\{\mathcal{M},\mathcal{R}\})) \end{aligned} \tag{5.25}$$

In the definition, x is tolerant of y on attribute set P if the relation between x and y on P supports $\{\mathcal{E},\mathcal{M},\mathcal{I},\mathcal{R}\}$ rather than $\{\mathcal{D}\}$. Similarly, x is similar to y if the relation between them supports $\{\mathcal{I}\}$ rather than $\{\mathcal{R}\}$ and support for $\{\mathcal{D}\}$ is less than $\{\mathcal{E},\mathcal{M},\mathcal{I}\}$. There is no plausibility based definition for an *equivalence* relation because in this type of relation, attribute values of the two objects must be equal to each other precisely and completely on the whole set of attributes. The *equivalence* relation is clearly reflexive, symmetric and transitive. The *tolerance* relations are reflexive, symmetric but not need to be transitive. The *similarity* relations are not necessarily symmetric but they are reflexive and transitive.

Calculating a believe function from masses obtained in the last step for objects x_4 and x_5, we have $Bel_P(\{\mathcal{E},\mathcal{M},\mathcal{I},\mathcal{R}\}) = 0.634$, $Bel_P(\{\mathcal{E},\mathcal{M},\mathcal{I}\}) = 0.6341$, $Bel_P(\{\mathcal{E}\mathcal{M},\mathcal{R}\}) = 0.629$ and $Bel_P(\{\mathcal{D}\}) = 0.366$. Applying definition 5.3.2, x_4 and x_5 are in a *Believable DS-Tolerance* relation. In the case of a *Believable DS-Similarity* relation, x_4 is similar to x_5 but the reverse does not hold.

Proposition 5.3.1. *Given incomplete information system $I = (U, A)$, two objects are in a DS-Equivalence relation if and only if they are equivalent to each other. Two objects are in Believable DS-Tolerance or Believable DS-Similarity relations if they are in the tolerance or the similarity relations, respectively.*

Proof. Using the model in proposition 5.2.2, we have

$EQ_P(x,y) \Leftrightarrow \forall a \in P, Bel_a(\{\mathcal{E}\}) = 1$. This is equivalent to $Bel_P(\{\mathcal{E}\}) = 1$, hence, $EQ_P(x,y) \Leftrightarrow DSE_P(x,y)$.

$TOR_P(x,y) \Leftrightarrow \forall a \in P, Bel_a(\{\mathcal{E},\mathcal{M},\mathcal{I},\mathcal{R}\}) = 1$, hence, $Bel_P(\{\mathcal{E},\mathcal{M},\mathcal{I},\mathcal{R}\}) = 1$ and $Bel_P(\{\mathcal{D}\}) = 0$. Consequently, $Bel_P(\{\mathcal{E},\mathcal{M},\mathcal{I},\mathcal{R}\}) \geq Bel_P(\{\mathcal{D}\})$, therefore, $TOR_P(x,y) \Rightarrow BelDST_P(x,y)$.

$SIM_P(x,y) \Leftrightarrow \forall a \in P, Bel_a(\{\mathcal{E},\mathcal{M},\mathcal{I}\}) = 1$, hence, $Bel_P(\{\mathcal{E},\mathcal{M},\mathcal{I}\}) = 1$, $Bel_P(\{\mathcal{R}\}) = 0$ and $Bel_P(\{\mathcal{D}\}) = 0$. Thus, $Bel_P(\{\mathcal{E},\mathcal{M},\mathcal{I}\}) \geq Bel_P(\{\mathcal{D}\})$ and $Bel_P(\{\mathcal{E},\mathcal{M},\mathcal{I}\}) \geq Bel_P(\{\mathcal{E},\mathcal{M},\mathcal{R}\})$. Consequently, $Bel_P(\{\mathcal{E},\mathcal{M},\mathcal{I}\}) \geq Bel_P(\{\mathcal{D}\})$ and $Bel_P(\{\mathcal{M},\mathcal{I}\}) \geq Bel_P(\{\mathcal{M},\mathcal{R}\})$, therefore, $SIM_P(x,y) \Rightarrow BelDSS_P(x,y)$. □

With a relation, we can derive a neighbourhood that consists of successor and predecessor sets of an object [28, 85]. For a Demster-Shafer-based relation, we introduce two sets for any object $x \in U$ as follows:

The successor set of x is the set of objects to which x is similar

$$sucR_P^{DSR}(x) = \{y \in U | DSR_P(x,y)\}. \tag{5.26}$$

The predecessor set of x is the set of objects which is similar to x

$$preR_P^{DSR}(x) = \{y \in U | DSR_P(y,x)\}. \tag{5.27}$$

where DSR is either of the Dempster-Shafer-based relations introduced in definition 5.3.2. Obviously, $sucR_P^{DSE}(x,y) = preR_P^{DSE}(y,x)$, $sucR_P^{BelDST}(x,y) = preR_P^{BelDST}(y,x)$ and $sucR_P^{PlDST}(x,y) = preR_P^{PlDST}(y,x)$ due to the symmetric property of equivalence and tolerance relations.

From the neighbourhoods, it is possible to define approximations using singleton, subset and concept methods discussed in Chapter 2. Singleton lower approximation an upper approximation of a object set $X \subseteq U$ are defined as follows:

$$\underline{SingleAppr}_P^{DSR}(X) = \{x \in U | R_P^{DSR}(x) \subseteq X\}, \tag{5.28}$$
$$\overline{SingleAppr}_P^{DSR}(X) = \{x \in U | R_P^{DSR}(x) \cap X \neq \emptyset\}, \tag{5.29}$$
$$\underline{SubsetAppr}_P^{DSR}(X) = \cup \{R_P^{DSR}(x) | x \in U \wedge R_P^{DSR}(x) \subseteq X\}, \tag{5.30}$$
$$\overline{SubsetAppr}_P^{DSR}(X) = \cup \{R_P^{DSR}(x) | x \in U \wedge R_P^{DSR}(x) \cap X \neq \emptyset\}, \tag{5.31}$$
$$\underline{ConceptAppr}_P^{DSR}(X) = \cup \{R_P^{DSR}(x) | x \in X \wedge R_P^{DSR}(x) \subseteq X\}, \tag{5.32}$$
$$\overline{ConceptAppr}_P^{DSR}(X) = \cup \{R_P^{DSR}(x) | x \in X \wedge R_P^{DSR}(x) \cap X \neq \emptyset\}, \tag{5.33}$$

where $R_P^{DSR}(x)$ denotes either successor sets $sucR_P^{DSR}(x)$ and predecessor sets $preR_P^{DSR}(x)$.

Returning to the example in Table 5.1, we induce approximations for set of objects $X = \{x | x \in U, f_d(x) = d_1\} = \{x_1, x_2, x_3, x_4\}$ using successor sets defined in the *Believable DS-Similarity* relation. Successor sets of the objects in X are listed as follows:

$$
\begin{aligned}
sucR_P^{BlDSS}(x_1) &= \{x_1\},\\
sucR_P^{BlDSS}(x_2) &= \{x_2, x_4\},\\
sucR_P^{BlDSS}(x_3) &= \{x_3, x_4\},\\
sucR_P^{BlDSS}(x_4) &= \{x_2, x_3, x_4, x_5\}.
\end{aligned}
$$

Approximations of an object set, consequently, are obtained by either singleton, subset or concept definitions. Singleton approximations of set X are

$$\underline{SingleAppr}_P^{BlDSS}(X) = \{x_1, x_2, x_3\},$$
$$\overline{SingleAppr}_P^{BlDSS}(X) = \{x_1, x_2, x_3, x_4, x_5\}.$$

Rough set approximations defined by equations from (5.28) to (5.33) have properties from 1(a) to 5(a) in the list of properties for the original rough set. This properties within our approach are proved as the same as those in the Grzymala-Busse and Wojciech Rzasa study [28] and the Pawlak research [61]. For the rest of the properties, in general, these definitions do not hold. To satisfy such properties, approximations should have some modification as discussed in [28, 88].

5.4 Summary

In this chapter, we have proposed a new rough set model based on Dempster-Shafer theory for information systems that may have uncertain, imprecise and/or fuzzy values.

By introducing several basic relations between objects and assigning masses for any possible hypothesis of these relations, we may be come able to model relations studied in incomplete information systems in terms of Dempster-Shafer theory. Considering each attribute as a source of evidence, we then calculate mass assignments on an attribute set and introduce new relations including *equivalence*, *tolerance* and *similarity* relations based on Dempster-Shafer theory. These relations are used to determine the approximations of an object set in various ways including singleton, subset and concept methods.

One potential drawback of the proposed method is the computational complexity inherent in the Dempster-Shafer approach. The relation defined based on Dempster's rule of combination increases the computational cost exponentially with the number of evidence sources, which is the number of attributes in this approach. Studies such as [3, 67] or a new approximate combination rule could, however, make it easier to solve the problem in the near future.

Further work thus will target to discovering the best combination rule of evidence both for a sound decision on an appropriate choice of relations and for practical implementation.

Chapter 6

Knowledge Acquisition

In the two previous chapters, rough set models characterized by a valued tolerance relation and some Dempster-Shafer theory-based relations have been introduced to deal with imperfect information systems. This chapter will discuss how to apply such kind of rough set models in machine learning and data mining. Possible concerned applications are feature selection (reducts and core) and decision rule induction. In the discussion of each task, before proposing new approach, we review methods published in the literature and point out reasons why they may not be applied in the case.

In this chapter, for convenience, the relation between $x, y \in U$ is $R_P(x, y)$ with respect to attribute set $P \subseteq A$ in imperfect information system $I = (U, A)$ such that R_P is either the *extended tolerance relation* or Dempster-Shafer based relations. Methods to obtain reduct and core and to derive decision rules for imperfect information systems then will be introduced.

6.1 Reducts and core

6.1.1 Discernibility Matrices

The concept of reducts and cores was introduced by Pawlak [60] for complete information system. A useful method of deriving reducts and cores was introduced by Skowron in the form of discernibility matrices [68]. This representation has been applied in many rough set approaches. In the case of incomplete information systems [37], we consider a matrix of $|U| \times |U|$ having the next set as its elements at row x and column y:

$$\sigma_A(x, y) = \{a \in A | TOR_{\{a\}}(x, y) = false\}. \tag{6.1}$$

The entry $\sigma_A(x, y)$ is clearly the set of all of attribute attributes that discern objects x and y. The entry could be empty if x is tolerant of y with respect to the whole attribute set A.

Let $\vee\sigma_A(x, y)$ be a logical function representing $\vee_{a\in\sigma_A(x,y)}\sigma_A(x, y)$, where a is interpreted as a proposition that "attribute a can discern objects x and y". The logical function $\vee\sigma_A(x, y)$ gives knowledge about which attribute are necessary to discern x and y. For example, $\vee\sigma_A(x, y) = a_1 \vee a_2$, where $a_1, a_2 \in A$, means we need a_1 or a_2 to discern x and y. If $\sigma_A(x, y) = \emptyset$, then $\vee\sigma_A(x, y) = false$ and x and y are thus tolerant of each other. The next function is called discernibility function for object x.

$$\Delta(x) = \bigwedge_{y\in U} \bigvee \sigma_A(x, y). \tag{6.2}$$

For instance, if $\Delta(x) = (a_1 \vee a_2) \wedge a_3$, we need a_1 and a_3, or a_2 and a_3 to discern x from other objects.

Table 6.1: An example of incomplete decision table

Cases	a_1	a_2	a_3	a_4	d
x_1	2	1	*	*	d_2
x_2	*	1	1	1	d_1
x_3	1	1	1	*	d_2
x_4	1	1	2	1	d_2
x_5	2	1	2	*	d_2
x_6	*	2	1	1	d_3

Table 6.2: The discernibility matrix[1]

Cases	x_2	x_3	x_4	x_5	x_6
x_1		a_1	a_1		a_2
x_2			a_3	a_3	a_2
x_3			a_3	a_1, a_3	a_2
x_4				a_1	a_1, a_3
x_5				a_1, a_3	

The discernibility function to discern all objects from each other is consequently defined as:

$$\Delta = \bigwedge_{\substack{x,y \in U \times U \\ x \neq y}} \bigvee \sigma_A(x, y). \tag{6.3}$$

Table 6.1 is used to illustrate the method of discernibility matrices. From this table, the matrix shown in Table 6.2 is derived. Based on this, discernibility functions can be induced as follows:

$$\Delta(x_1) = a_1 \wedge a_1 \wedge a_2 = a_1 \wedge a_2,$$

[1]Note that discernibility matrix are symmetric so we need only half of the elements in the table.

$$
\begin{aligned}
\Delta(x_2) &= a_3 \wedge a_3 \wedge a_2 = a_2 \wedge a_3,\\
\Delta(x_3) &= a_1 \wedge a_3 \wedge (a_1 \vee a_3) \wedge a_2 = a_1 \wedge a_2 \wedge a_3,\\
\Delta(x_4) &= a_1 \wedge a_3 \wedge a_3 \wedge a_1 \wedge (a_2 \vee a_3) = a_1 \wedge a_3,\\
\Delta(x_5) &= a_3 \wedge (a_2 \vee a_3) \wedge a_1 \wedge (a_2 \vee a_3) = a_1 \wedge a_3,\\
\Delta(x_6) &= a_2 \wedge a_2 \wedge a_2 \wedge (a_2 \vee a_3) \wedge (a_2 \vee a_3) = a_2,\\
\Delta &= a_1 \wedge a_2 \wedge a_3 \wedge (a_1 \vee a_3) \wedge (a_2 \vee a_3) = a_1 \wedge a_2 \wedge a_3.
\end{aligned}
$$

Because $\Delta = a_1 \wedge a_2 \wedge a_3$, $P = \{a_1, a_2, a_3\}$ is thus only one reduct of this information system.

According to [61, Chapter 5], the method of obtaining reduct and core has many advantages, in particular it enables simple computation. This approach, however, cannot be applied to some rough set models, for example, with relations defined in [51, 52, 54, 76, 83]. In Table 6.1, using the limited tolerance relation [76], x_1 and x_2 are discerned by individual attribute a_1, a_3 or a_4. However, x_1 is tolerant of x_2 based on $A = \{a_1, a_2, a_3, a_4\}$, such that $LTOR_A(x_1, x_2) = true$. In the other words, we could not distinguish x_1, x_2 even a_1, a_3 or a_4 do exist.

There is another approach to obtain reducts and cores that could deal with missing values in incomplete information systems [39]. This method computes significances of every attributes based on the notion of rough entropy [4, 29]. This approach, however, faces the same problem regarding to the above explanation.

6.1.2 Reducts and Cores in Imperfect Information Systems

In this section, we shall propose a method to derive reducts and cores for imperfect information systems based on extended tolerance relation as well as Dempster-Shafer based relation. In Chapter 2, the definition of reducts and cores are discussed for complete information system. Reducts and cores in imperfect information system is also defined as follows:

Definition 6.1.1. A set of conditional attributes $P \subseteq A$ is a reduct of an imperfect information system, if the neighbourhoods induced by P are the same as the neighbourhoods induced by all of the attributes in set A and no attribute can be removed from P without changing the neighbourhoods.

To obtain reducts and cores, we introduce a comparison function of two attribute sets.

Definition 6.1.2. The comparison, a Boolean function between two relations in terms of two attribute sets $P, Q \subseteq A$ in an imperfect information system is defined as follows:

$$\omega(P, Q) = (\forall (x, y) \in U \times U, R_P(x, y) \Leftrightarrow R_Q(x, y)). \tag{6.4}$$

If $\omega(P, Q) = 1$, two relations developed from two different attribute sets P, Q make the same neighbourhoods. Now, for imperfect decision tables, relative reducts and cores are obtained based on a function called generalized decision function. A generalized decision function of object x is the set of decision of objects belonging to the successor or predecessor of x.

Definition 6.1.3. In imperfect decision table $I = (U, A \cup \{d\})$, $P \subseteq A$ is a set of conditional attributes and $F(V_d)$ is the power set of V_d, the function $\delta_P : U \to F(V_d)$ is defined as:

$$\delta_P(x) = \{d_i \in V_d | d_i = f_d(y), y \in setR_P(x)\}, \tag{6.5}$$

where $setR_P(x)$ denotes either successor or predecessor of x.

Definition 6.1.4. The comparison Boolean function between two relations in terms of attribute sets $P, Q \subseteq A$ in an imperfect decision table is defined as follows:

$$\omega'(P, Q) = (\forall (x, y) \in U \times \Gamma^x, R_P(x, y) \Leftrightarrow R_Q(x, y)), \tag{6.6}$$

where $\Gamma^x = \{z \in U | f_d(z) \notin \delta_P(x)\}$.

Proposition 6.1.1. *Attribute $a \in A$ is indispensable in A if and only if $\omega(A - \{a\}, A) = 0$ for imperfect information systems and $\omega'(A - \{a\}, A)$ for imperfect decision tables.*

Proof. Attribute a is indispensable in A if and only if $(\exists x \in U, setR_{A-\{a\}}(x) \neq setR_A(x))$ if and only if $(\exists (x, y) \in U \times U)$, $R_{A-\{a\}}(y, x) \not\Leftrightarrow R_A(y, x)$. Thus, $\omega(A - \{a\}, A) = 0$, from Definition 6.1.2, or $\omega'(A - \{a\}, A) = 0$, from Definition 6.1.4. $\square$

This proposition is applied in both imperfect information systems and imperfect decision tables. Function $\omega(A - \{a\}, A) = 0$ or $\omega'(A - \{a\}, A) = 0$ means that if a is removed from A, the neighbourhoods based on relation R in terms of $A - \{a\}$ are different from the neighbourhoods based on A. Attribute a is thus indispensable in the conditional attribute set A.

Definition 6.1.5. The core of A is the set of all indispensable attributes and defined by $core(A) = \{a \in A | \omega(A - \{a\}, A) = 0\}$ for imperfect information system and $core(A) = \{a \in A | \omega'(A - \{a\}, A) = 0\}$ for imperfect decision table.

Proposition 6.1.2. *A subset $P \subseteq A$ is a reduct of the imperfect information system (or decision table) if and only if*

(i) $\omega(P, A) = 1$ for imperfect information system (or $\omega'(P, A) = 1$ for imperfect decision tables).

(ii) for all $a \in P$, $\omega(P - \{a\}, P) = 0$ for imperfect information system (or $\omega'(P - \{a\}, P) = 0$ for imperfect decision tables).

Table 6.3: An imperfect table for knowledge acquisition

Cases	**Temperature**	**Headache**	**Nausea**	**Flu**
x_1	high	yes:0.9; no:0.1	no	yes
x_2	very-high	yes	yes	yes
x_3	high:0.3; normal:0.7	no	no	no
x_4	high	yes	yes	yes
x_5	high	yes:0.1; no:0.9	yes	no
x_6	normal	yes	no	no
x_7	normal	no	yes	no
x_8	very-high:0.3; high:0.7	yes	yes:0.7; no:0.3	yes

Proof. Following the definition of reducts stated at the beginning of this section, $P \subseteq A$ is a reduct if and only if

(i) neighbourhoods of any objects induced by P are the same as the neighbourhoods of these objects induced by all attributes in set A.

(ii) no attribute can be removed from set P without changing neighbourhoods of objects.

Based on Definitions 6.1.2 and 6.1.4, we have (i) $\Leftrightarrow$ $\omega(P, A) = 1$ for imperfect information systems (or $\Leftrightarrow$ $\omega'(P, A) = 1$ for decision tables). For the second condition, (ii) means all attributes in P are indispensable. Consequently, (ii) $\Leftrightarrow$ for all $a \in P$, $\omega(P - \{a\}, P) = 0$ for imperfect information systems (or $\Leftrightarrow$ for all $a \in P$, $\omega(P - \{a\}, P) = 0$ for imperfect decision tables), according to Proposition 6.1.1. $\square$

The information presented in Table 6.3 is used to illustrate the method. In this example, suppose that the extended tolerance relation method introduced in Chapter 4 is employed. The chosen threshold α is 0.1. Tolerance degrees among objects in terms of all attributes are calculated based on extended tolerance relation and then shown in Table 6.4. Note that tolerance relations are symmetric so we need

Table 6.4: Tolerance degrees among objects on all attributes

	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0	*0.515*	0	0	0	0	0.095
x_2		0	0	0	0	0	*0.605*
x_3			0	0	0	0	0
x_4				*0.550*	0	0	*0.745*
x_5					0	0	0.024
x_6						0	0
x_7							0

Table 6.5: Tolerance degrees among objects on {Headache,Nausea}

	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0	*0.550*	0	0	*0.950*	0	*0.135*
x_2		0	*1.000*	*0.550*	0	0	*0.850*
x_3			0	*0.135*	0	*0.850*	0
x_4				*0.550*	0	0	*0.850*
x_5					0	*0.950*	0.035
x_6						0	*0.650*
x_7							0

only half of the elements in the table. The italic styled numbers highlight the degrees that are greater or equal the threshold.

The tolerance degrees among objects when each attribute in {*Temperature, Headache, Nause*} is removed are also displayed in the tables from 6.5 to 6.7.

From the tables of tolerance degrees, it is easily seen that $w(A, A-\{Temperature\} = 0$ because $ETR_A^{0.1}(x_1, x_8) = false$, while $ETR_{A-\{Temperature\}}^{0.1}(x_1, x_8) = true$. Thus, $Temperature$ is indispensable. In the same way, $Headache$ and $Nausea$ are also indispensable. Therefore $\{Temperature, Headache, Nause\}$ is the core and also the unique reduct of this information system.

Table 6.6: Tolerance degrees among objects on {Temperature,Nausea}

	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0	*0.650*	0	0	0	0	*0.105*
x_2		0	0	0	0	0	*0.105*
x_3			0	0	*0.850*	0	0.032
x_4				*1.000*	0	0	*0.245*
x_5					0	0	*0.245*
x_6						0	0
x_7							0

Table 6.7: Tolerance degrees among objects on {Temperature,Headache}

	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0	0.015	*0.950*	*0.590*	0	0	*0.315*
x_2		0	0	0	0	0	*0.650*
x_3			0	*0.135*	0	*0.850*	0
x_4				*0.550*	0	0	*0.850*
x_5					0	0	0.035
x_6						0	0
x_7							0

6.2 Decision Rules

Rule induction is one of the most important knowledge discovery techniques in machine learning. A decision rule can be presented in the following expression:

$$r : \wedge_i (a_i = v) \rightarrow (d = w)$$

where $a_i \in A$, $v \in V_{a_i}$, and d and w is the decision attribute and a decision value respectively. Set $\mathcal{A}_r = \cup\{a_i\}$ is called condition set and attribute d is call decision of rule r. Hereafter, $f_a(r)$ and $f_d(r)$ represent the value of attribute $a \in \mathcal{A}_r$ and decision d, respectively, in r. $R_P(x, r)$, the same symbol for a relation between objects, is used to represent the relation between object x and the conditional part of r with respect to attribute set $P \subseteq A$.

In supervised learning, rules are obtained from information which consists of conditional and decisional attributes. However, due to imperfect data and/or some other reasons, rules may conflict with each other. In Table 3.1 for example, the rule from case x_4: $(Temperature = high) \wedge (Headache = yes) \wedge (Nause = yes) \rightarrow (Flu = yes)$ conflicts with the rule of x_5: $(Temperature = high) \wedge (Nause = yes) \rightarrow (Flu = no)$ if it is assumed that the missing value of *Headache* is "*yes*".

Rough sets, which describe a set of objects in the approximation space, play a vital role in rules induction. Rules induced from the certain region (lower approximation) and possible region (upper approximation) of a concept are called certain and possible rules respectively [24, 87]. The following subsections discuss the limitation of a famous algorithm and introduce a method to deriving certain and possible rules in imperfect information systems.

6.2.1 LEM2 Algorithm

Among published rule induction algorithms, LEM2 (Learning from Examples Module, version 2) of LERS (Learning from Examples using Rough Sets) is used commonly since it gives better results [24]. The algorithm is based on the idea of blocks of attribute-value pairs. For an attribute-value pair (a, v), a block $[(a, v)]$ is a set of all cases from U such that for attribute a has value v. This algorithm can be also used for some rough set approaches in incomplete information systems [21, 73, 87] in which objects belong to the block $[(a, v)]$ if their values on a are tolerant of (similar to) v. Let B be a non-empty lower or upper approximation of a concept represented by a decision-value pair (d, w). Let us say that the set B depends on a set T of attribute-value pairs if and only if

$$\emptyset \neq [T] = \cap_{(a,v)\in T}[(a, v)] \subseteq B \tag{6.7}$$

Table 6.8: An example incomplete table

Cases	a	b	d
x_1	a_1	$*$	d_1
x_2	$*$	b_1	d_1

From the equation, it could be believed intuitively that the less cardinality of T is the more objects T covers. Thus, set T is called a minimal complex of B if and only if B depends on T and no proper subset T' exists such that B depends on T'. LEM2 algorithm defines each rule based on a minimal complex.

However, as the same problem with the method of discernibility matrices to induce reducts and core, the belief mentioned above also does not hold in some cases such as relations defined in [51, 52, 54, 76, 83]. For example, using the limited tolerance relation [76], the pair $\{(a, a_1)\}$ covers only x_1, $\{(b, b_1)\}$ covers only x_2, yet $\{(a, a_1), (b, b_1)\}$ covers both x_1 and x_2 in Table 6.8, which violates the intuitive belief. Thus, in the next subsection, we will introduce a method to derive decision rules for not only valued tolerance/similarity relation and Dempster-Shafer based relation but also for other approaches in imperfect decision tables.

6.2.2 Obtaining Decision Rules in Imperfect Information Systems

Taking imperfect data representation introduced in Chapter 3, Section 3.3 into account, a method to obtain decision rules in imperfect information systems will be step by step introduced.

Definition 6.2.1. In imperfect decision table $I = (U, A \cup \{d\})$, let $t^x_{a,i}$ be the i-th possible value set of object $x \in U$ on attribute $a \in A$, a candidate rule set suggested from object x is denoted by $\mathcal{S}(x)$ and defined by the following equation:

$$\mathcal{S}(x) = \{r|(a \in P \subseteq A, P \neq \emptyset, f_a(r) \in \cup_i t^x_{a,i}) \wedge R_P(x, r)\}. \tag{6.8}$$

In the definition, possible conditional values of the candidate rule r are limited using the possible values sets of object x and the conditional part of the rule should be tolerant of (similar to) the object x with respect to attributes in $\mathcal{A}_r$. From this definition, the suggested rule set $\mathbb{S}(X)$ of an object set $X \subseteq U$ is defined by

$$\mathbb{S}(X) = \cup_{x \in X} \mathcal{S}(x) \tag{6.9}$$

On the other hand, for rule r, a set of objects that rule r covers is defined by the following equation

$$\mathcal{G}(r) = \{x|x \in U, R_{\mathcal{A}_r}(x, r)\} \tag{6.10}$$

Rule r is *optimal* if and only if no rule r' exists such that $\mathcal{A}_{r'} \subset \mathcal{A}_r$, $\forall a \in \mathcal{A}'_r$, $f_a(r) = f_a(r')$, and $\mathcal{G}(r') = \mathcal{G}(r)$. Let $\mathbb{G}(\mathcal{R}) = \cup_{r \in \mathcal{R}} \mathcal{G}(r)$ denotes the set of objects that rule set $\mathcal{R}$ covers, two new region concepts are defined as follows:

Definition 6.2.2. A rule set $\mathcal{R}$ is called a *lower covering* of a set of objects B if only if the following conditions are satisfied:

1. each member of $\mathcal{R}$ is a optimal.
2. $\forall r \in \mathcal{R}$, $\mathbb{G}(\mathcal{R} - \{r\}) \subset \mathbb{G}(\mathcal{R})$.
3. $\mathbb{G}(\mathcal{R}) \subseteq B$
4. and there is no rule set $\mathcal{R}'$ such that $\mathbb{G}(\mathcal{R}) \subset \mathbb{G}(\mathcal{R}') \subseteq B$

Definition 6.2.3. A rule set $\mathcal{R}$ is called an *upper covering* of a set of objects B if only if the following conditions are satisfied:

1. each member of $\mathcal{R}$ is a optimal.

2. $\forall r \in \mathcal{R}$, $\mathbb{G}(\mathcal{R} - \{r\}) \subset \mathbb{G}(\mathcal{R})$.

3. $\mathbb{G}(\mathcal{R}) \supseteq B$

4. and there is no rule set $\mathcal{R}'$ such that $\mathbb{G}(\mathcal{R}) \supset \mathbb{G}(\mathcal{R}') \supseteq B$

Based on the two definitions above, we suggest an algorithm to induce certain and possible rules by finding lower and upper coverings of a set of objects as follows:

Input: A set of object $X \subseteq U$

Output: lower covering $\underline{\mathcal{R}}$ and upper covering $\overline{\mathcal{R}}$ of X

Step 1:

Make a candidate rule set $L = \mathbb{S}(X)$;
Remove any rule in L which is not optimal;
$L' := \{r | r \in L, \mathcal{G}(r) \subseteq X\}$;
$L := L - L'$;
$\underline{\mathcal{R}} := \emptyset$;
$B := X$;

Step 2:

If $B = \emptyset$ or $L' = \emptyset$ then go to step 3;
Select the rule $r \in L'$ such that $\mathcal{G}(r) \cap B$ is the maximum;
If a tie occurs, select the rule with the smallest $|\mathcal{A}_r|$;
If another tie occurs, select the first rule;
$\underline{\mathcal{R}} := \underline{\mathcal{R}} \cup \{r\}$;
$B := B - (B \cap \mathcal{G}(r))$;
Remove from L' all rules r such that $B \cap \mathcal{G}(r) = \emptyset$;
Repeat the step 2;

Step 3:

$\overline{\mathcal{R}} := \underline{\mathcal{R}}$

Table 6.9: Candidate rules[2]

Rules	**Temperature**	**Headache**	**Nausea**	**Covered objects**
r_1			no	x_1, x_3, x_6, x_7, x_8
r_2		yes		$x_1, x_2, x_4, x_5, x_6, x_8$
r_3		yes	no	x_1, x_6, x_8
r_4		no		x_1, x_3, x_5, x_7
r_5		no	no	x_1, x_3
r_6	high			x_1, x_3, x_4, x_5, x_8
r_7	high		no	x_1, x_3, x_8
r_8	high	yes		x_1, x_4, x_5, x_8
r_9	high	yes	no	x_1, x_8
r_{10}	high	no		x_1, x_3, x_5
r_{11}	high	no	no	x_1, x_3
r_{12}			yes	x_2, x_4, x_5, x_7, x_8
r_{13}		yes	yes	x_2, x_4, x_5, x_8
r_{14}	very-high			x_2, x_8
r_{15}	high		yes	x_4, x_5, x_8
r_{16}	very-high		no	x_8

Step 4:

Remove from L all rules r such that $B \cap \mathcal{G}(r) = \emptyset$;

If $B = \emptyset$ or $L = \emptyset$ then stop;

Select the rule $r \in L$ such that $\mathcal{G}(r) - B$ is the minimum;

If a tie occurs, select the rule such that $\mathcal{G}(r) \cap B$ is maximum;

If a tie occurs, select the rule with the smallest $|\mathcal{A}_r|$;

If another tie occurs, select the first rule;

$\overline{\mathcal{R}} := \overline{\mathcal{R}} \cup \{r\}$;

$B := B - (B \cap \mathcal{G}(r))$;

Remove any rule $r \in \overline{\mathcal{R}}$ such that $\mathcal{G}(r) \subset \mathbb{G}(\overline{\mathcal{R}} - \{r\})$;

Repeat the step 4;

Returning to the information in Table 6.3, using extended tolerance relation, the threshold α is supposed to be 0.1,. It is then possible to induce certain and possible rules for the concept $X = \{x | f_{Flu}(x) = yes\}$.

[2]The absence of attribute values in a rule mean those value are not exist in conditional part of that rule.

At first, candidate rule set $\mathbb{S}(X)$ in which any rules that can satisfy one or more cases in Table 6.3 is determined using equations (6.8) and (6.9). Then rules, that are not optimal, will be removed. Rule $r' : (Temperature = very - high) \wedge (Headadge = yes) \rightarrow (Flu = yes)$, for example, is not optimal because it covers the same set $\{x_2, x_8\}$ with rule $r_{14} : (Temperature = very - high) \rightarrow (Flu = yes)$ while $\mathcal{A}_{r_{14}} \subset \mathcal{A}_{r'}$. The rules that are optimal are shown in Table 6.9. Then, to find certain rules, rules in a candidate set which cover only objects belonging to X should be chosen. Thus, the candidate set for certain rules is $L' = \{r_9, r_{14}, r_{16}\}$.

Going to the step 2 that is the step to induce certain rules, the maximal cardinality of $\mathcal{G}(r) \cap B$ is two. So the rule r_{14} that has the smallest number in cardinality of its conditional attribute part is selected such that $|\mathcal{A}_{r_{14}}| = 1$. The first certain rule presented by the first element of the lower covering, therefore, is:

$$(Temperature = very - high) \rightarrow (Flu = yes).$$

The above rule is now added to the certain rule set. All covered objects - x_2, x_8 - of this rule are also removed from the uncovered object set B, such that $B = B - \{x_2, x_8\} = \{x_1, x_4\}$. Rule r_{16} is then deleted from L' because it does not cover any object in $\{x_1, x_4\}$. Next, rule r_9 is chosen and it covers x_1. The next certain rule, thus, is:

$$(Temperature = high) \wedge (Headache = yes) \wedge (Nausea = no) \rightarrow (Flu = yes).$$

The step is stopped because $L' = \emptyset$.

Going through the step 3 and 4, the possible rule set of the concept X is obtained as follows:

$$(Temperature = high) \wedge (Headache = yes) \rightarrow (Flu = yes),$$
$$(Temperature = very - high) \rightarrow (Flu = yes).$$

From the above, it is possible to see that x_4 does not support any certain rule. This is because any candidate rule in $\mathcal{S}(x_4)$ would cover some other objects with *Flu=no*. This type of rule may be present in possible rules instead.

6.3 Summary

In this chapter, the methods to obtain reducts and core and to derive decision rules for imperfect information systems are introduced. These processes could avoid the problems of algorithms published in the literature as discussed in each section.

At first, some steps to obtain reducts and cores are defined. A potential drawback of the proposed approach is the computational complexity. However, an idea that is a combination between the method of discernibility matrices and this method would be a better solution. We are studying this direction and will complete it in the near future.

Besides, a method of obtaining decision rules from an imperfect decision table was also discussed and proposed. At the same time, the algorithm also can produce both certain and possible rules without calculating approximation space for any set of objects.

Chapter 7

Conclusion

7.1 Summary of The Research

This research studies rough sets in imperfect information systems and establishes two new models based on valued tolerance relation and Dempster-Shafer theory based relation. Techniques to acquire knowledge in imperfect decision tables are also suggested.

To begin with, the original rough set theory proposed by Pawlak and some extended rough set models have been introduced. The basic concept of rough sets is defined based on indiscernibility relations among data items. As a generalization, Bayes rough set model based on variable precision is suitable for dealing with practical human evaluation data. Another extension of rough set - fuzzy rough set - that is a combination of fuzzy sets and rough sets, benefits the use of level in approximation space definition by using membership functions. To be applied in real database, rough set definitions for any relations are also reported. An arbitrary relation may have one or more of reflexive, symmetric and transitive properties rather than all of properties like equivalence relations in original rough set model.

At the second point, the original rough set is also expanded to adapt requirements of incomplete data in information systems, in which values may be lost or unknown. As the simplest method, missing values are considered as the same as any possible values. Tolerance/similarity relations among items are then defined based on comparison between object attribute values for each attribute.

In real applications, however, information might be described imperfectly due to not only missing values but also due to uncertainty and imprecision. A value is uncertain when it is associated with probabilities. While, an object attribute value is imprecise when a set of possible values is given. To deal multiple types of imperfection in a single solution, this research suggests a representation of imperfect values.

Having studied several conventional methods to deal with missing values and researched some special database, we next propose a new rough set model based on valued tolerance relations. Frequencies of values appearing on a data set are used for estimating probabilities of matching among data items on an attribute. They are then employed for measuring tolerance degrees. Given a threshold for controlling uncertainty level, a tolerance relation is defined.

Apart from valued tolerance relations, a relation based on Dempster-Shafer theory is also introduced. Taking advantages of Dempster-Shafer theory that allow us to assign a probability mass to a set of events into account, we first defined several basic relations among objects on an attribute and determine mass assignment for occurrence of these basic relations. Considering each attribute as a source of evidence, a combination rule is then employed for calculating belief and plausibility functions on the whole attribute set. Some equivalence, tolerance and similarity relations among objects are then defined by comparing belief or plausibility of several hypotheses.

Finally, based on two rough set models introduced above, methods to acquire knowledge in imperfect information system are also introduced. Due to the fact that algorithms published in the literature cannot be used in not only two new models but also in some other approaches, techniques to derive reducts and core and to obtain decision rules should be redefined. In this discussion, the algorithm of deriving rules is able to obtain certain and possible rules without calculating approximations.

7.2 Contributions

The overall contribution of this research is introduction of rough set models in imperfect information systems from their definitions to their applications of knowledge discovering in such systems. The significant contributions could be stated as follows:

1. The first is introduction of an imperfect data representation into rough set models including missing, uncertain and imprecise values. It can also represent fuzzy data. The representation could be utilized for any study in imperfect information systems.

2. The second is proposing two rough set models which can define approximations in information systems containing multiple types of imperfection. Those approaches could open a new direction of rough set studies in imperfect information systems.

3. The third is solving an issue existing in controversial models for incomplete information systems. Using threshold for widening or thinning boundaries in the *extended tolerance* relation, the problem of the *tolerance relation* is then solved efficiently.

4. The last is suggesting techniques in knowledge discovery. The techniques allow not only two rough set models proposed in the research but other models also to

be used for obtaining reducts, core and for deriving decision rules in imperfect information systems.

7.3 Limitations

Studying rough sets in imperfect information systems is likely to bring much advantage to knowledge discovery processes. Nevertheless, there are still a few limitations in each discussion that would be studied in the future. At first, the Dempster-Shafer theory based relations may require a complex computation due to the combination rule of evidence. However, this is not in the scope of rough set study. We thus only hope that the problem inherent in Dempster-Shafer approach can be solved in the near future. Second, the computation of reducts and cores also takes time. We also addressed an idea of discernibility matrices notion as a solution to solve the problem.

7.4 Future Works

For further studying, we target three directions: solving the limitations existing in the current work, applying the proposed rough set models to other machine learning techniques, and implementing these models in a real application.

First of all, limitations of the current approach would give us the further work of the research. The computation time to obtain reducts and core should be cut down by applying the notion of the discernibility matrices. Formally, instead of determining the matrix of $|U| \times |U|$ having the next set as its elements at row x and column y by $\sigma_A(x, y) = \{a \in A | R_{\{a\}}(x, y) = false\}$, we could define the elements by $\sigma_A(x, y) = \{P | R_P(x, y) = false\}$, where $P \subseteq A$ is optimal such that there is

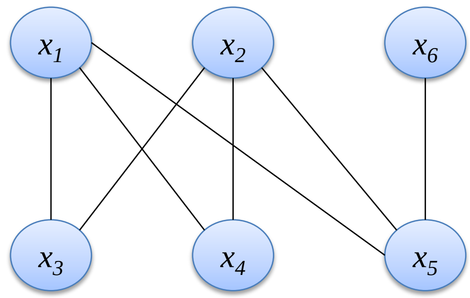

Figure 7.1: A social network

no $P' \subset P$, $R_{P'}(x,y) = false$. The new entry $\sigma_A(x,y)$ is the set of minimum attribute sets which discern objects x and y. It is also possible to reduce the time of the process to find out P for forming entry $\sigma_A(x,y)$ in the above equation. Let $B = \{a \in A | [\cup_i t_{a,i}(x)] \bigcap [\cup_i t_{a,i}(y)] = \emptyset\}$, objects x and y are clearly discerned by any $P \subseteq B$, $|P| = 1$. The remain task is thus examining subsets of $A - B$ for producing entry $\sigma_A(x,y)$.

The second target is using rough set models to deal with problems faced in social network study [33, 64, 77]. Let take an example of analysing roles and positions of people in a society. A role of each person is clarified by his connections to others. Intuitively, two people may have the same role or position in their society if they have relationships to the same group of people. A social network shown in Figure 7.1, for example, illustrates a society. This society is made up from a collection of six people from x_1 to x_6. Connections among individuals represent their relationship. In this society, x_1 and x_2 have the same role because they connect to x_3, x_4 and x_5. In social network technique, this phenomenon is named as *structural equivalence*.

However, human relations among individuals are often complicated. A person may have multiple connections to others. For example, Terry is a classmate of James at high school and they are colleagues because they are working for the same company.

Table 7.1: An information system of social connection

U	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0	0	1	1	1	0
x_2	0	0	1	1	1	0
x_3	1	1	0	0	0	0
x_4	1	1	0	0	0	0
x_5	1	1	0	0	0	1
x_6	0	0	0	0	1	0

In addition, human relationships are likely uncertain. For instance, we can assess a friendship based on how close the friendship between two people is or how frequently they get together.

In such kind of system, to analyse roles of people for multiple types of relations, it is possible to use rough set models discussed in this research. A group of people will present a universe of a system. Each people will also be a conditional attribute of the system. Table 7.1, for example, shows complete information system $I = (U, A)$ induced from the social network in Figure 7.1 where $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ and $A = \{x_1, x_2, x_3, x_4, x_5, x_6\}$. An object attribute value is 1 if there is a connection between two people, otherwise the value is 0. In this table, x_1 and x_2 are equivalent in terms of A. For complex human relations, it is possible to define an imperfect information system and introduce a so-called *structural tolerance relation* using this notion. We thus are able to classify groups of people playing the same role by inducing approximations.

The last vital target is finding out a possible application of the discussed models, hence these approaches become meaningful. In actual applications, there are many situations in which we have to describe examples with imprecise as well as uncertain representation rather than with singleton values. Affective images and impression that are used in Kansei engineering [46, 47] are good examples.

Table 7.2: Kansei information table for mobile phone design

U	a_1	a_2	a_3	a_4	D
x_1	2	1	0	0.8[metal],0.2[plastic]	deluxe
x_2	{1,2}	1	{0,1}	1.0[metal]	deluxe
x_3	2	{0,1}	1	0.7[metal],0.3[plastic]	deluxe
x_4	0	2	{1,2}	0.1[metal],0.9[plastic]	cute
x_5	1	0	2	0.2[metal],0.8[plastic]	cute
x_6	1	{0,1}	1	1.0[plastic]	sporty
x_7	1	0	{0,1}	0.1[metal],0.9[plastic]	sporty
x_8	2	0	0	1.0[plastic]	sporty

WEB-based form feature extraction system for mobile phone design [62, 90] is an example of applications. Some features of mobile phone design including body shape, partition, screen position, arrangement of number keys, and function keys are selected to evaluate. Numerous interviewees are then invited to give their felling about combination of features. The product knowledge representation system in [62] allows interviewees to chose just one value for each feature. However, human impression probably need more than that. They may chose something between round and sharp corners, for example, to evaluate each body shape from images provided.

To solve the problem above, it is possible to present impression data by uncertain or imprecision values. Let the product feature set be $A = \{a_1, a_2, a_3, a_4\}$, which denotes body shape, body ratio, bottom shape, material, respectively, and Kansei adjectives set be $D = \{deluxe, cute, sporty\}$. Each kind of body shape, body ratio, corner shape is shown by a picture and assigned with a number. Material types are plastic and metal. A possible example of evaluation of interviewees is shown in **Table 7.2**. Approximations are then derived based on rough set approach discussed in this research. Consequently, these approximations could be used in Kansei knowledge acquisition methods in [62, 90].

7.5 Last Summary

In this research, we tried to convince the readers why a discussion of rough set models is necessary for imperfect information systems. Two new approaches are also proposed along with methods for acquiring knowledge hidden in decision tables. Although there are still some problems encountered, we pointed out solutions and suggest applications of the research. We hope that the issues can be solved in the near future and the results become useful in actual applications.

Appendix A

Publications of the Research

International Journal

1. D.V. Nguyen, K. Yamada, M. Unehara : Rough Set Approach with Imperfect Data Based on Dempster-Shafer Theory, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 18, No. 3, pp.280-288, ISSN 1343-0130, Fuji Technology Press Ltd. (May 2014).

2. D.V. Nguyen, K. Yamada, M. Unehara : Rough Sets and Rule Induction in Imperfect Information Systems, International Journal of Computer Applications, Vol. 89, No. 5, pp.1-8, Published by Foundation of Computer Science, New York, USA (Mar. 2014).
http://dx.doi.org/10.5120/15495-4286

3. D.V. Nguyen, K. Yamada, M. Unehara : Extended Tolerance Relation to Define a New Rough Set Model in Incomplete Information Systems, Advances in Fuzzy Systems, Volume 2013, Article ID 372091, Hindawi Pub. (Oct. 2013).
http://dx.doi.org/10.1155/2013/372091

International Proceeding

1. D. V. Nguyen, K. Yamada, M. Unehara: On Probability of Matching in Probabilistiy Based Rough Set Definitions, IEEE-SMC2013, pp. 449-454, Manchester, UK (Oct. 14, 2013).
   http://dx.doi.org/10.1109/SMC.2013.82

2. D. V. Nguyen, K. Yamada, M. Unehara: Rough Set Model Based on Parameterized Probabilistic Similarity Relation in Incomplete Decision Tables, SCIS-ISIS 2012, pp. 577-582, Kobe, Japan (Nov. 21, 2012).

3. D. V. Nguyen, K. Yamada, M. Unehara: Knowledge reduction in incomplete decision tables using Probabilistic Similarity-Based Rough set Model, 12th International Symposium on Advanced Intelligent Systems (ISIS 2011), pp.147-150, Suwon, Korea (Sep. 29, 2011).

Local Proceeding

1. D. V. Nguyen, K. Yamada, M. Unehara: A rough set model based on probabilistic similarity measure for incomplete decision tables, Joint Workshop 2010 of Kanto and Hokushin'etsu Chapters of SOFT, pp.46-49 (2010/11/20).

Bibliography

[1] N. S. Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[2] D. Barbará, H. Garcia-Molina, and D. Porter. The management of probabilistic data. *IEEE Trans. on Knowl. and Data Eng.*, 4(5):487–502, October 1992.

[3] Jeffrey A. Barnett. Computational methods for a mathematical theory of evidence. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, pages 868–875, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc.

[4] Theresa Beaubouef, Frederick E. Petry, and G. Arora. Information-theoretic measures of uncertainty for rough sets and rough relational databases. *Inf. Sci.*, 109(1-4):185–195, 1998.

[5] Malcolm J. Beynon, Margaret Heffernan, and Aoife M. McDermott. Psychological contracts and job satisfaction: Clustering analysis using evidential c-means and comparison with other techniques. *Int. Syst. in Accounting, Finance and Management*, 19(4):247–273, 2012.

[6] Roger Cavallo and Michael Pittarelli. The theory of probabilistic databases. In *Proceedings of the 13th International Conference on Very Large Data Bases*, VLDB '87, pages 71–81, San Francisco, CA, USA, 1987. Morgan Kaufmann Publishers Inc.

[7] Michal R. Chmielewski and Jerzy W. Grzymala-busse. Global discretization of continuous attributes as preprocessing for machine learning. In *International Journal of Approximate Reasoning*, pages 294–301, 1996.

[8] Peter Clark and Tim Niblett. The cn2 induction algorithm. *Mach. Learn.*, 3(4):261–283, March 1989.

[9] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods.* Cambridge University Press, New York, NY, USA, 2000.

[10] Jianhua Dai and Haowei Tian. Fuzzy rough set model for set-valued data. *Fuzzy Sets and Systems*, 229(0):54 – 68, 2013. Theme: Computer Science.

[11] M.A. de Carvalho, C.H.V. de Moraes, G. Lambert-Torres, L.E. Borges da Silva, A.R. Aoki, and A Vivaldi. Transforming continuousattributes using ga for applications of rough set theory to control centers. *Intelligent System Application to Power Systems (ISAP)*, pages 1–8, 2011.

[12] Pedro Domingos and Michael Pazzani. On the optimality of the simple bayesian classifier under zero-one loss. *Mach. Learn.*, 29(2-3):103–130, November 1997.

[13] Didier Dubois and Henri Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4(3):244–264, 1988.

[14] Didier Dubois and Henri Prade. Rough fuzzy sets and fuzzy rough sets. In *International Journal of General Systems*, volume 17, pages 191–209. 1990.

[15] Didier Dubois and Henri Prade. Putting rough sets and fuzzy sets together. In Roman Slowinski, editor, *Intelligent Decision Support*, volume 11 of *Theory and Decision Library*, pages 203–232. Springer Netherlands, 1992.

[16] Usama M. Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17:37–54, 1996.

[17] Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurusamy. *Advances in Knowledge Discovery and Data Mining.* The MIT Press, February 1996.

[18] Yucai Feng, Wenhai Li, Zehua Lv, and Xiaoming Ma. Probabilistic approximation under incomplete information systems. In Zhongzhi Shi, K. Shimohara, and David Dagan Feng, editors, *Intelligent Information Processing*, volume 228 of *IFIP*, pages 73–80. Springer, 2006.

[19] Jerzy W. Grzymala-Busse. On the unknown attribute values in learning from examples. In Z.W. Ras and M. Zemankova, editors, *Methodologies for Intelligent Systems*, volume 542 of *Lecture Notes in Computer Science*, pages 368–377. Springer Berlin Heidelberg, 1991.

[20] Jerzy W. Grzymala-Busse. Data with missing attribute values: Generalization of indiscernibility relation and rule induction. In James F. Peters, Andrzej Skowron, Jerzy W. Grzymaa-Busse, Boena Kostek, RomanW. Swiniarski, and MarcinS. Szczuka, editors, *Transactions on Rough Sets I*, volume 3100 of *Lecture Notes in Computer Science*, pages 78–95. Springer Berlin Heidelberg, 2004.

[21] Jerzy W. Grzymala-Busse. Characteristic relations for incomplete data: A generalization of the indiscernibility relation. 3700:58–68, 2005.

[22] Jerzy W. Grzymala-Busse. Incomplete data and generalization of indiscernibility relation, definability, and approximations. In Dominik Slezak, Guoyin Wang, Marcin S. Szczuka, Ivo Dntsch, and Yiyu Yao, editors, *RSFDGrC (1)*, volume 3641 of *Lecture Notes in Computer Science*, pages 244–253. Springer, 2005.

[23] Jerzy W. Grzymala-Busse. A rough set approach to data with missing attribute values. In Guoyin Wang, James F. Peters, Andrzej Skowron, and Yiyu Yao, editors, *RSKT*, volume 4062 of *Lecture Notes in Computer Science*, pages 58–67. Springer, 2006.

[24] Jerzy W. Grzymala-Busse. Rule induction. In Maimon and Rokach [42], pages 249–265.

[25] Jerzy W. Grzymala-Busse, Patrick G. Clark, and Martin Kuehnhausen. Generalized probabilistic approximations of incomplete data. *International Journal of Approximate Reasoning*, 55(1, Part 2):180 – 196, 2014. Special issue on Decision-Theoretic Rough Sets.

[26] Jerzy W. Grzymala-Busse and Witold J. Grzymala-Busse. Handling missing attribute values. In Maimon and Rokach [42], pages 33–51.

[27] Jerzy W. Grzymala-Busse and Ming Hu. A comparison of several approaches to missing attribute values in data mining. In *Revised Papers from the Second International Conference on Rough Sets and Current Trends in Computing*, RSCTC '00, pages 378–385, London, UK, UK, 2001. Springer-Verlag.

[28] Jerzy W. Grzymala-Busse and Wojciech Rzasa. Definability and other properties of approximations for generalized indiscernibility relations. *T. Rough Sets*, 11:14–39, 2010.

[29] J.W. Guan and D.A. Bell. Rough computational methods for information systems. *Artificial Intelligence*, 105(12):77 – 103, 1998.

[30] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference and prediction.* Springer, 2 edition, 2008.

[31] T. Inagaki. Interdependence between safety-control policy and multiple-sensor schemes via dempster-shafer theory. *Reliability, IEEE Transactions on*, 40(2):182–188, Jun 1991.

[32] Masahiro Inuiguchi and Takuya Miyajima. Rough set based rule induction from two decision tables. *European Journal of Operational Research*, 181(3):1540–1553, September 2007.

[33] Matthew O. Jackson. Chapter 12 - an overview of social networks and economic applications. volume 1 of *Handbook of Social Economics*, pages 511 – 585. North-Holland, 2011.

[34] R. Jensen and Qiang Shen. Fuzzy-rough sets for descriptive dimensionality reduction. In *Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on*, volume 1, pages 29–34, 2002.

[35] Mehmed Kantardzic. *Data Mining: Concepts, Models, Methods and Algorithms.* John Wiley & Sons, Inc., New York, NY, USA, 2002.

[36] George J. Klir and Bo Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1995.

[37] Marzena Kryszkiewicz. Rough set approach to incomplete information systems. *Inf. Sci.*, 112(1-4):39–49, December 1998.

[38] Marzena Kryszkiewicz. Rules in incomplete information systems. *Inf. Sci.*, 113(3-4):271–292, 1999.

[39] Jiye Liang and Zongben Xu. The algorithm on knowledge reduction in incomplete information systems. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(1):95–103, February 2002.

[40] W.-N. Liu, JingTao Yao, and Yiyu Yao. Rough approximations under level fuzzy sets. In Shusaku Tsumoto, Roman Sowiski, Jan Komorowski, and JerzyW. Grzymaa-Busse, editors, *Rough Sets and Current Trends in Computing*, volume 3066 of *Lecture Notes in Computer Science*, pages 78–83. Springer Berlin Heidelberg, 2004.

[41] Y. Ma, J.S. Chandler, and D.C. Wilkins. On the decision making problem in dempster-shafer theory. In *International Symposium on Artificial Intelligence*, pages 293–302, Cancun, Mexico, 1992.

[42] Oded Maimon and Lior Rokach, editors. *Data Mining and Knowledge Discovery Handbook, 2nd ed.* Springer, 2010.

[43] S. L. Miller and D. Childers. *Probability and Random Process With Application to Signal Processing and Communications.* Elsevier Inc, 2004.

[44] R. S. Milton, V. Uma Maheswari, and Arul Siromoney. Studies on rough sets in multiple tables. In *Proceedings of the 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing - Volume Part I*, RSFDGrC'05, pages 265–274, Berlin, Heidelberg, 2005. Springer-Verlag.

[45] A. Motro. *Uncertainty management information systems: From needs to solutions.* Kluwer Acad., 1997.

[46] M. Nagamachi. Kansei engineering: a new ergonomics consumer-oriented technology for product development. *International Journal of Industrial Ergonomics*, 15:3–10, 1995.

[47] M. Nagamachi. Kansei engineering as a powerful consumer-oriented technology for product development. *Fundamental Reviews in Applied Ergonomics*, 33(3):289–294, 2002.

[48] Michinori Nakata and Hiroshi Sakai. Rough sets handling missing values probabilistically interpreted. In Dominik Slezak, Guoyin Wang, Marcin Szczuka, Ivo Dntsch, and Yiyu Yao, editors, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, volume 3641 of *Lecture Notes in Computer Science*, pages 325–334. Springer Berlin Heidelberg, 2005.

[49] Michinori Nakata and Hiroshi Sakai. Handling missing values in terms of rough sets. In *23rd Fuzzy System Symposium, Nagoya, Japan*, 2007.

[50] Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara. Knowledge reduction in incomplete decision tables using probabilistic similarity-based rough set model. In *12th International Symposium on Advanced Intelligent Systems (ISIS 2011)*, pages 147–150, 2011.

[51] Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara. Rough set model based on parameterized probabilistic similarity relation in incomplete decision tables. In *Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on*, pages 577–582, 2012.

[52] Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara. Extended tolerance relation to define a new rough set model in incomplete information systems. *Advances in Fuzzy Systems*, 2013. Article ID 372091.

[53] Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara. On probability of matching in probabilistiy based rough set definitions. In *IEEE-SMC2013*, pages 449–454, Manchester, The UK, 2013.

[54] Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara. Rough set approach with imperfect data based on dempster-shafer theory. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 18(3), 2014.

[55] Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara. Rough sets and rule induction in imperfect information systems. *International Journal of Computer Applications*, 89(5):1–8, 2014.

[56] Hung Son Nguyen. Discretization problem for rough sets methods. In *Proceedings of the First International Conference on Rough Sets and Current Trends in Computing*, RSCTC '98, pages 545–552, London, UK, UK, 1998. Springer-Verlag.

[57] Tatsuo Nishino, Mitsuo Nagamachi, and Hideo Tanaka. Variable precision bayesian rough set model and its application to kansei engineering. In JamesF. Peters and Andrzej Skowron, editors, *Transactions on Rough Sets V*, volume 4100 of *Lecture Notes in Computer Science*, pages 190–206. Springer Berlin Heidelberg, 2006.

[58] Ewa Orlowska and Zdzislaw Pawlak. Representation of nondeterministic information. *Theoretical Computer Science*, 29(12):27 – 39, 1984.

[59] A. Papoulis and S. U. Pillai. *Probability, random variables, and stochastic process*. McGraw-Hill, 2002.

[60] Zdzislaw Pawlak. Rough sets. *International Journal of Computer and Information Sciences*, 11:341–356, 1982.

[61] Zdzislaw Pawlak. *Rough Sets. Theoretical Aspects of Reasoning about Data.* Kluwer Acad., 1991.

[62] Fu qian Shi, Shou qian Sun, and Jiang Xu. Employing rough sets and association rule mining in kansei knowledge extraction. *Information Sciences*, 196:118–128, 2012.

[63] Anna Maria Radzikowska and Etienne E. Kerre. A comparative study of fuzzy rough sets. *Fuzzy Sets Syst.*, 126(2):137–155, March 2002.

[64] Garry Robins. A tutorial on methods for the modeling and analysis of social network data. *Journal of Mathematical Psychology*, 57(6):261 – 274, 2013. Social Networks.

[65] Lior Rokach and Oded Maimon. *Data Mining with Decision Trees: Theroy and Applications*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2008.

[66] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.

[67] Glenn Shafer and Roger Logan. Implementing dempster's rule for hierarchical evidence. *Artificial Intelligence*, 33(3):271 – 298, 1987.

[68] Andrzej Skowron and Cecylia Rauszer. The discernibility matrices and functions in information systems. In Roman Slowinski, editor, *Intelligent Decision Support*, volume 11 of *Theory and Decision Library*, pages 331–362. Springer Netherlands, 1992.

[69] Andrzej Skowron and Ning Zhong. Rough sets in kdd - tutorial notes. Technical report.

[70] Dominik Slezak and Wojciech Ziarko. Variable precision bayesian rough set model. In Guoyin Wang, Qing Liu, Yiyu Yao, and Andrzej Skowron, editors, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, volume 2639 of *Lecture Notes in Computer Science*, pages 312–315. Springer Berlin Heidelberg, 2003.

[71] Dominik Slezak and Wojciech Ziarko. The investigation of the bayesian rough set model. *International Journal of Approximate Reasoning*, 40(12):81 – 91, 2005. Data Mining and Granular Computing.

[72] Jerzy Stefanowski and Alexis Tsoukis. On the extension of rough sets under incomplete information. In Ning Zhong, Andrzej Skowron, and Setsuo Ohsuga, editors, *RSFDGrC*, volume 1711 of *Lecture Notes in Computer Science*, pages 73–81. Springer, 1999.

[73] Jerzy Stefanowski and Alexis Tsoukis. Incomplete information tables and rough classification. *Computational Intelligence*, 17(3):545–566, 2001.

[74] Jaroslaw Stepaniuk. Approximation spaces, reducts and representatives. In Lech Polkowski and Andrzej Skowron, editors, *Rough Sets in Knowledge Discovery 2*, volume 19 of *Studies in Fuzziness and Soft Computing*, pages 109–126. Physica-Verlag HD, 1998.

[75] Jaroslaw Stepaniuk. Rough sets in multi relational environment. In *Proceeding of. CS&P '06*, pages 376–387, 2006.

[76] Guoyin Wang. Extension of rough set under incomplete information systems. In *Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on*, volume 2, pages 1098–1103, 2002.

[77] S. Wasserman and G. Robins. Social network research: The foundation of network science. In Harris Cooper, editor, *APA Handbook of Research Methods in Psychology*. 2012.

[78] Jakub Wroblewski. Analyzing relational databases using rough set based methods. In *In Proceedings of IPMU 2000*, volume 1, pages 256–262, 2000.

[79] Guan Xin, Yi Xiao, and He You. Discretization of continuous interval-valued attributes in rough set theory and its application. In *Machine Learning and Cybernetics, 2007 International Conference on*, volume 7, pages 3682–3686, 2007.

[80] Ronald R. Yager. On the relationships of methods of aggregation of evidence in expert systems. *Cybernetics and Systems*, 16(1):1–21, 1985.

[81] Ronald R. Yager. On the dempster-shafer framework and new combination rules. *Information Sciences*, 41(2):93 – 137, 1987.

[82] Koichi Yamada. A new combination of evidence based on compromise. *Fuzzy Sets and Systems*, 159(13):1689 – 1708, 2008. Theme: Aggregation Operations and Methods.

[83] Xibei Yang, Xiaoning Song, and Xiaohua Hu. Generalisation of rough set for rule induction in incomplete system. *IJGCRSIS*, 2(1):37–50, 2011.

[84] Y. Y. Yao. Combination of rough and fuzzy sets based on -level sets. In *Rough sets and data mining: Analysis for imprecise data*, pages 301–321. Kluwer Academic, 1997.

[85] Y. Y. Yao. On generalizing rough set theory. In Guoyin Wang, Qing Liu, Yiyu Yao, and Andrzej Skowron, editors, *RSFDGrC*, volume 2639 of *Lecture Notes in Computer Science*, pages 44–51. Springer, 2003.

[86] Y. Y. Yao. Probabilistic rough set approximations. *Int. J. Approx. Reasoning*, 49(2):255–271, October 2008.

[87] Y. Y. Yao. Three-way decisions with probabilistic rough sets. *Information Sciences*, 2010(3):341–353, 2010.

[88] Y.Y. Yao. Relational interpretations of neighborhood operators and rough set approximation operators. *Information Sciences*, 111(14):239 – 259, 1998.

[89] Y.Y. Yao. Rough sets, neighborhood systems and granular computing. In *Electrical and Computer Engineering, 1999 IEEE Canadian Conference on*, volume 3, pages 1553–1558, May 1999.

[90] Lian-Yin Zhai, Li-Pheng Khoo, and Zhao-Wei Zhong. A rough set based decision support approach to improving consumer affective satisfaction in product design. *International Journal of Industrial Ergonomics*, 39(2):295 – 302, 2009.

[91] Lianwen Zhang. Advances in the dempster-shafer theory of evidence. chapter Representation, Independence, and Combination of Evidence in the Dempster-Shafer Theory, pages 51–69. John Wiley & Sons, Inc., New York, NY, USA, 1994.

[92] Wojciech Ziarko. Variable precision rough set model. *J. Comput. Syst. Sci.*, 46(1):39–59, February 1993.

[93] Wojciech Ziarko. Probabilistic approach to rough sets. *International Journal of Approximate Reasoning*, 49(2):272 – 284, 2008. Special Section on Probabilistic Rough Sets and Special Section on PGM06.

[94] Hans-Jürgen Zimmermann. *Fuzzy Set Theory and its Applications*. Springer, 2001.