2009

Knowledge Engineering from Data Perspective: Granular Computing Approach

Rushin Barot
San Jose State University

KNOWLEDGE ENGINEERING FROM DATA PERSPECTIVE :GRANULAR COMPUTING APPROACH

A Thesis

Presented to

The Faculty of the Department of Computer Science

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Rushin B. Barot

May 2009

APPROVED FOR THE DEPARTMENT OF COMPUTER SCIENCE

Dr. T.Y.Lin

Dr. Robert Chun

Dr. Soon Tee Teoh

APPROVED FOR THE UNIVERSITY

ABSTRACT

KNOWLEDGE ENGINEERING FROM DATA PERSPECTIVE :GRANULAR COMPUTING APPROACH

by Rushin B. Barot

The concept of rough set theory is a mathematical approach to uncertainly and vagueness in data analysis, introduced by Zdzislaw Pawlak in 1980s. Rough set theory assumes the underlying structure of knowledge is a partition. We have extended Pawlak's concept of knowledge to coverings. We have taken a soft approach regarding any generalized subset as a basic knowledge. We regard a covering as basic knowledge from which the theory of knowledge approximations and learning, knowledge dependency and reduct are developed.

DEDICATION

This research is dedicated to all the researchers who have contributed in the field Machine Learning. I got the opportunity to work on this research because of their contribution.

In addition, I would also like to dedicate this work to all the readers.

ACKNOWLEDGEMENTS

I would like to thank Dr. T. Y. Lin for his continuous support and guidance towards this project. Dr. T. Y. Lin inspired me to pursue rough sets, which eventually culminated in this project. I would also like to extend my thanks to Dr. Robert Chun and Dr. Soon Tee Tech for their expertise and advice. I would also like to thank my uncle Nitin Barot and aunty Nila Barot for their kind support, and my parents and sister for their motivation.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

KNOWLEDGE

Theory of knowledge has a long and rich history. Artificial Intelligence and logicians have been involved in various aspects of the knowledge. As per Pawlak, knowledge is deep-seated in the classificatory abilities of human beings and other species. Pawlak has introduced a rough set concept as a theoretical framework for discussions about knowledge, particularly when imprecise knowledge is of primary concern.

Pawlak assumes that knowledge is the ability to classify objects; so a knowledge is a mathematical partition. Thus knowledge is connected to varieties of classifications related to specific parts of the real or abstract world, called *universe of discourse* [Paw82].

Pawlak's method is limited to partition. But in real life, classifications have never been perfect. We consider the covering, in which the conditions of mutual disjointness are not required.

1.1 Knowledge based on covering

A collection of subset is called a partial covering. If the union of a partial covering is the whole space, then the partial covering is called covering. Suppose we have given finite set $U \neq \emptyset$ objects. Any subset $X \subseteq U$ of the universe will be called a concept in U. The given covering on U will be referred to as given Basic Knowledge about U.

Suppose we have some families of covering over U. i.e the collection $\mathcal{C} = \{C_1, C_2, ..., C_m\}$. Here $\mathcal{C}$ is called collection of families of subsets, where $C_1 = \{C_{11}, C_{12}, C_{13}, ..., C_{1n_1}\}$, $C_2 = \{C_{21}, C_{22}, C_{23}, ..., C_{2n_2}\}$ up-to
$C_m = \{C_{m1}, C_{m2}, C_{m3}, ..., C_{mn_m}\}$. Here C_j is a family of subsets, such that $C_{ji} \subseteq U$, $C_{ji} \neq \emptyset$ and $\bigcup C_{ji} = U$, where $j = 1, 2, ..., m$ and $i = 1, 2, ..., n$ with respect to j.From this covering $\mathcal{C}$ we can derive a new covering $\mathcal{G}$ by AND operation. Here $\mathcal{G} = \{G_1, G_2, ..., G_m\}$, where $G_j = \cap_{(finitely\ many\ i)} C_{ji} \neq \emptyset$.

So $(U, \mathcal{G})$ and $(U, \mathcal{C})$ will be referred to knowledge base. We are going to consider three knowledge base a) Pawlak Partition Knowledge Base (PPKBS), b) Pawlak Covering Knowledge Base (PCKBS) and c) General Covering Knowledge Base (GCKBS).

PCKBS only use all ($\cup$) to approximate an arbitrary concept while GCKBS use all ($\cup$) and ($\cap$).

CHAPTER 2

MACHINE LEARNING

2.1 Introduction

Classification and categories has been crucial concept in the knowledge theory. Categories are defined in available knowledge in the given knowledge base. In a knowledge base some categories can be defined however some categories cannot. In other words we will discuss about vagueness of the categories.

2.2 Rough Sets

Let $X \subseteq U$, and R be an equivalence relation. We can say that X is R-definable, iff X is the union of some R-basic categories;otherwise X is R-undefinable.[Paw82] The sets which can be defined in R by the given knowledge base are also called as R-exact set. The sets which cannot be defined in R by the given knowledge base are also called as R-inexact or R-rought sets. The figure 2.1 defines lower approximation and upper approximation.

2.3 Approximation of Set

This section describes theory of single covering or partition. Let knowledge base (U, P), (U, C) and (U, G) be approximation spaces for PPKBS, PCKBS and GCKBS respectively.

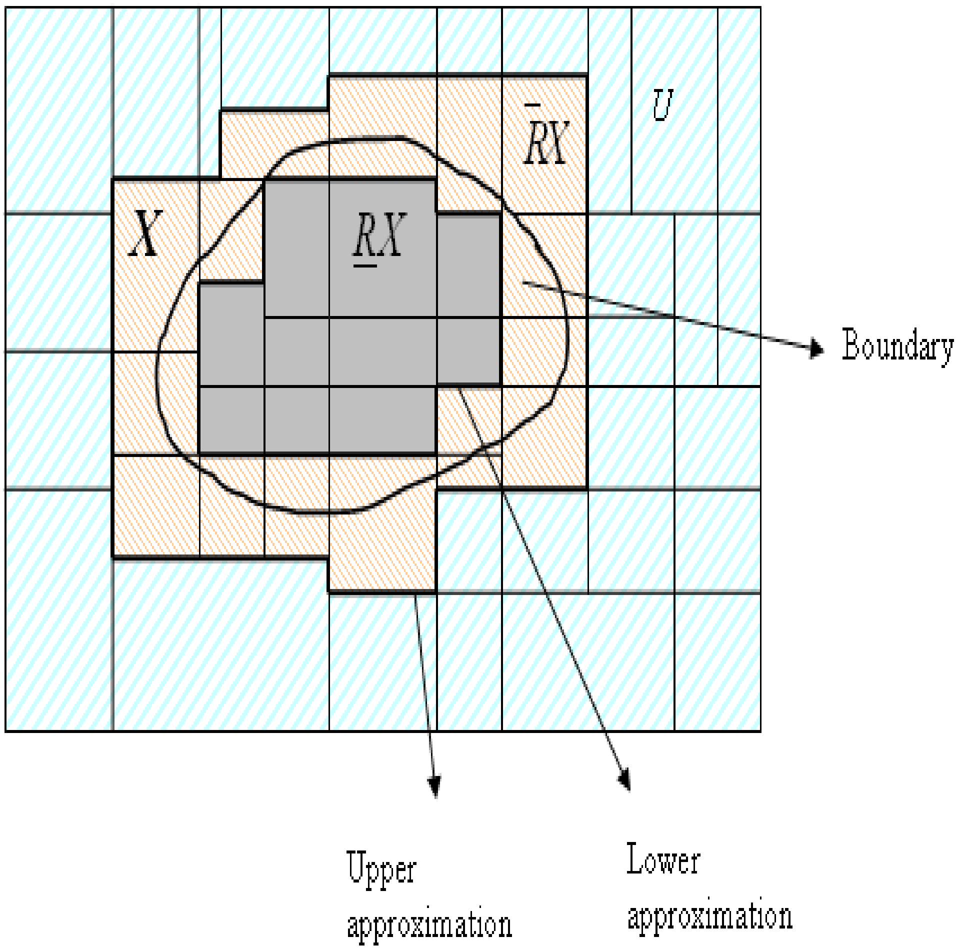

Figure 2.1: Rough Sets with lower bound and upper bound.

2.3.1 Approximation theory based on PPKBS

Pawlak uses approximation to express unknown concepts by available knowledge. Pawlak reveals approximation as learning. Its goal is to learn every concept in the universe. If a concept can be learnt precisely then it is called an exact concept. If we cannot learn a concept precisely then we have approximation.

As per Pawlak's theory, based on partition of certain universe U, which has a family of subset $P = \{P_1, P_2, ..., P_n\}$ be the partition. A partition is collection of mutually disjoint subsets, so for given knowledge base $K = (U, P)$. Let X be the subset of U.

$$\underline{P}X = \{z \mid \exists P_i \in P,\ s.t.\ z \in P_i \ and \ P_i \subseteq X\}$$
$$\overline{P}X = \{z \mid \forall P_i \in P,\ s.t.\ z \in P_i \ and \ P_i \cap X \neq \emptyset\}$$

called the P - lower and P-upper approximation of X respectively[Paw82].

2.3.2 Approximation theory based on PCKBS

In PCKBS, we have given knowledge base $K = (U, \mathcal{C})$ with each subset $X \subseteq U$ and a family of subset $C \in \mathcal{C}$. Here C_i is any member of covering C.

$$\underline{C}X = \{z \mid \exists C_i \in C,\ s.t.\ z \in C_i \ and \ C_i \subseteq X\}$$
$$\overline{C}X = \{z \mid \forall C_i \in C,\ s.t.\ z \in C_i \ and \ C_i \cap X \neq \emptyset\}$$

called the C - lower and C-upper approximation of X respectively.

2.3.3 Approximation theory based on GCKBS

In GCKBS, we have used AND and OR operation for basic knowledge. Our approximation is different from Pawlak. Suppose we are given a knowledge base $K = (U, \mathcal{G})$ with each subset $X \subseteq U$ and a family of covering $G \in \mathcal{G}$, where G is any family of subsets from $\mathcal{G}$ and G_i is a member of family of family of subsets.

$$\underline{G}X = \{z \mid \exists G_i \in G,\ s.t.\ z \in G_i \ and \ G_i \subseteq X\}$$
$$\overline{G}X = \{z \mid \forall G_i \in G,\ s.t.\ z \in G_i \ and \ G_i \cap X \neq \emptyset\}$$

called the G - lower and G-upper approximation of X respectively.

2.3.4 *

Theorem :

(1) Approximation space based on GCKBS is topological space.

(2) Approximation space based on PCKBS may not be topological space.

(3) Approximation based on PPKB is clopen topological space.

.

Case 1) Approximation spaces is a topological space generated by sub-base C or equivalently generated by base G.

Case 3) Equivalence class is based on topological spaces.

Pawlak's rough set theory is a special type of clopen(close and open) space. This is generalization of Pawlak's approximation theorem. We have demonstrated the concept using an example below.

2.3.5 Proposition

(1) X is R-definable if and only if $\underline{R}X = \overline{R}X$.

(2) X is rough with respect to R if and only if $\underline{R}X \neq \overline{R}X$.

Example

Suppose we have given the following set, $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$.
In our example we have used $\mathcal{C} = \{C_1\}$ and $C_1 = \{C_{11}, C_{12}, C_{13}, C_{14}, C_{15}\}$. Suppose we have given knowledge base (U, C_1) and we have family of subset of subset

$$c_{11} = \{x_1, x_2, x_3, x_4\}$$
$$c_{12} = \{x_3, x_4, x_5\}$$
$$c_{13} = \{x_5, x_6, x_2\}$$
$$c_{14} = \{x_6, x_7, x_8\}$$
$$c_{15} = \{x_7, x_8, x_1\}.$$

Here family of subset of subset is basic knowledge or covering base space. Now we have G_1 which is finite intersections of covering C_1 set. So

$$
\begin{aligned}
G_{1_1} &= c_{11} = \{x_1, x_2, x_3, x_4\} \\
G_{1_2} &= c_{12} = \{x_3, x_4, x_5\} \\
G_{1_3} &= c_{13} = \{x_5, x_6, x_2\} \\
G_{1_4} &= c_{14} = \{x_6, x_7, x_8\} \\
G_{1_5} &= c_{15} = \{x_7, x_8, x_1\}. \\
G_{1_6} &= c_{11} \cap c_{12} = \{x_3, x_4\} \\
G_{1_7} &= c_{11} \cap c_{13} = \{x_2\} \\
G_{1_8} &= c_{11} \cap c_{14} = \emptyset \\
G_{1_9} &= c_{11} \cap c_{15} = \{x_1\} \\
G_{1_{10}} &= c_{12} \cap c_{13} = \{x_5\} \\
G_{1_{11}} &= c_{12} \cap c_{14} = \emptyset \\
G_{1_{12}} &= c_{12} \cap c_{15} = \emptyset \\
G_{1_{13}} &= c_{13} \cap c_{14} = \{x_6\} \\
G_{1_{14}} &= c_{13} \cap c_{15} = \emptyset \\
G_{1_{15}} &= c_{14} \cap c_{15} = \{x_7, x_8\} \\
G_{1_{16}} &= C_{11} \cap C_{12} \cap C_{13} = \emptyset \\
& \textit{and so on.}
\end{aligned}
$$

which is also called categories of covering. Therefore, distinct elements of set

$G_1 = \{G_{1_1}, G_{1_2}, G_{1_3}, G_{1_4}, G_{1_5}, G_{1_6}, G_{1_7}, G_{1_8}, G_{1_9}, G_{1_{10}}, G_{1_{13}}, G_{1_{15}}\}$.

We can see that our basic knowledge is much bigger than Pawlak's basic knowledge. Pawlak's knowledge base is special case in covering base.

Let $K = (U, C_1)$ be the knowledge base over U and G_1 can be generated from C_1. G_1 is finite intersections from covering C_1. Let $X_1 = \{x_1, x_2, x_3, x_4\}$ and $X_2 = \{x_1, x_6, x_7\}$.

$$\underline{G}X_1 = G_{1_1} \cup G_{1_6} \cup G_{1_7} \cup G_{1_9} = \{x_1, x_2, x_3, x_4\}$$
$$\overline{G}X_1 = G_{1_1} \cup G_{1_6} \cup G_{1_7} \cup G_{1_9} = \{x_1, x_2, x_3, x_4\}$$

here X_1 is called an exact set.

$$\underline{G}X_2 = G_{1_9} \cup G_{1_{13}} = \{x_1, x_6\}$$
$$\overline{G}X_2 = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$$

CHAPTER 3

REDUCTION OF KNOWLEDGE

3.1 Introduction

A fundamental problem is whether the whole knowledge is always necessary to define some categories available in the knowledge being considered. This problem arises in many practical applications and will be referred to as knowledge reduction. Knowledge dependency plays key role in reduction of knowledge. Here we have considered several covering or partition as knowledge space.

3.2 Reduct and Core of Knowledge

There are two fundamental concepts, a reduct and the core. A reduct of knowledge essentially reduces the knowledge without compromising the knowledge base. From Reduct we can also generate original knowledge base. In short, reduct is essential part of knowledge, if we loose data from reduct we cannot reproduce the original knowledge, whereas the core is the most important part of the knowledge.

Let $\mathbf{G} = (R, S)$ be the family of covers. If $\bigcap(\mathbf{G}) \neq \bigcap(\mathbf{G} - \{R\})$ than R is indispensable in $\mathbf{G}$ and R is dispensable if $\bigcap(\mathbf{G}) = \bigcap(\mathbf{G} - \{R\})$.

3.2.1 Proposition

$$CORE(\mathcal{G}) = \bigcap RED(\mathcal{G})$$

Where $RED(\mathcal{G})$ is the family of the all reducts of $\mathcal{G}$.

Example

Let $\mathcal{G} = \{G_1, G_2, G_3\}$ be the family of covers. where

$$U(G_1) = \{\{x_1, x_2, x_7\}, \{x_3, x_4, x_6\}, \{x_5, x_6\}, \{x_1\}\}$$
$$U(G_2) = \{\{x_1, x_3, x_7\}, \{x_1, x_2, x_5\}, \{x_4, x_6\}, \{x_4\}\}$$
$$U(G_3) = \{\{x_1, x_4, x_7\}, \{x_1, x_6\}, \{x_1, x_2\}, \{x_3, x_5\}, \{x_4, x_6\}\}$$

So that

$U(\mathcal{G}) = \bigcap \mathcal{G} = \bigcap\{G_1, G_2, G_3\} =$

$\{\{x_1, x_7\}, \{x_1, x_2\}, \{x_4, x_6\}, \{x_1\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}\}$

Here the relation G_1 is indispensable in $\mathcal{G}$ since,

$$U(\mathcal{G} - \{G_1\}) = \{\{x_1, x_2\}, \{x_1, x_7\}, \{x_4, x_6\}, \{x_1\}, \{x_2, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}\} \neq U(\mathcal{G})$$

Similarly for relation G_2 ,

$$U(\mathcal{G} - \{G_2\}) = \{\{x_1, x_2\}, \{x_1, x_7\}, \{x_4, x_6\}, \{x_1\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}\} = U(\mathcal{G})$$

so cover G_2 is dispensable in $\mathcal{G}$and for cover G_3 we have

$$U(\mathcal{G} - \{G_3\}) = \{\{x_1, x_7\}, \{x_1, x_2\}, \{x_4, x_6\}, \{x_1\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}\} = U(\mathcal{G})$$

hence the relation G_3 is dispensable in $\mathcal{G}$.

Here classification deduced by the covers G_1, G_2, G_3 is same as the classification deduced by the cover G_1 and G_2 or G_1 and G_3. So the reduct of the knowledge base $U(G_1, G_2) \neq U(G_1)$ and $U(G_1, G_2) \neq U(G_2)$ concludes that G_1 and G_2 are not dependent. We have two reducts of $\mathcal{G}$ which are $\{G_1, G_2\}$ and $\{G_1, G_3\}$.

As per proposition, $\{G_1, G_2\} \cap \{G_1, G_3\} = \{G_1\}$ is a core of the $\mathcal{G}$.

3.3 Reduction of Categories

A reduct of knowledge is its essential part, which suffices to define all basic concepts occurring in the considered knowledge, whereas the core is in a certain sense its most important part.

Reduct of family of sets : Let $F = \{C_1, C_2, ..., C_n\}$, be a family of sets such that $C_i \subseteq U$.

We say that C_i is dispensable in F, if $\bigcap(F - \{C_i\}) = \bigcap F$; otherwise the C_i is indispensable in F.

The family F is independent if all of its components are indispensable in F; otherwise F is dependent.

The family $H \subseteq F$ is a reduct of F, if H is independent and $\cap H = \cap F$.

The family of all indispensable sets in F will be called the core of F, denoted CORE(F).

3.3.1 Proposition

$$CORE(F) = \bigcap RED(F)$$

where RED(F) is the family of all reducts of F[**?**].

Example

Suppose we have family of sets $F = \{C_1, C_2, C_3\}$, where

$$C_1 = \{C_{11}, C_{12}\}$$

$$C_2 = \{C_{21}, C_{22}\}$$

$$C_3 = \{C_{31}, C_{32}\}$$

where

$$C_{11} = \{x_1, x_2, x_3, x_4\}$$

$$C_{12} = \{x_4, x_5\}$$

$$C_{21} = \{x_1, x_2, x_5\}$$

$$C_{22} = \{x_3, x_4\}$$

$$C_{31} = \{x_1, x_2, x_4\}$$

$$C_{32} = \{x_3, x_4, x_5\}$$

Hence $\cap F = C_1 \cap C_2 \cap C_3 = \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_4\}, \{x_5\}\}$ because the relations

$$\cap(F - \{C_1\}) = C_2 \cap C_3 = \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_4\}, \{x_5\}\}$$

$$\cap(F - \{C_2\} = C_1 \cap C_3 = \{\{x_1, x_2, x_4\}, \{x_3, x_4\}, \{x_4, x_5\}, \{x_4\}\}$$

$$\cap(F - \{C_3\} = C_1 \cap C_2 = \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_3\}, \{x_4\}\}.$$

In order to find reducts of the family of $F = \{C_1, C_2, C_3\}$. From the covering C_1, C_2, C_3 we can say that C_1 is dispensable in family of F, hence the reduct of F are $\{C_1, C_2, C3\}$ and $\{C_2, C_3\}$. Thus two reducts of family F, namely $\{C_1, C_2, C3\}$ and $\{C_2, C_3\}$ and $\{C_1, C_2, C3\} \cap \{C_2, C_3\} = \{C_2, C_3\}$ is the core of family F.

CHAPTER 4

DEPENDENCY OF KNOWLEDGE

4.1 Introduction

In knowledge dependency we have shown the semantic aspects of dependency. In the following, we need the notion of knowledge dependency; it has two notions, weak dependency(WD) and strong dependency(SD). Without specification, we define this knowledge dependency as follows ;

(1) Knowledge Q_X depends on knowledge P_X iff $P_X \subseteq Q_X$ is called weak dependency(WD).

(2) Knowledge Q_X depends on knowledge P_X iff $\bigcup P_X = Q_X$ is called strong dependency(SD).

where the inclusion($\cap$) and union($\cup$) will be clearly defined in each section. Here X can be Partition(P), Pawlak Covering(C) or General covering(G).

The following are three cases of knowledge dependency.

4.1.1 Knowledge Dependency based on PPKBS

As per Pawlak's theory, dependency for partition has the following property. Let $K = (U, \mathcal{P})$ be knowledge base and let $P, Q \subseteq \mathcal{P}$. In this case, X is partition in given definition of knowledge dependency.

We have observed following properties of PPKBS

A) Knowledge P and Q are PPKBS. WD = SD.

B) $P_1 \cap P_2 \Rightarrow P_1$ and $P_1 \cap P_2 \Rightarrow P_2$.

Suppose equivalence classes from a partition $P_1 = \{P_{11}, P_{12}, ..., P_{1n}\}$. Let P_{1i} be any equivalence class from a family of set P_1, where index $i = 1, 2, ..., n$. $P_2 = \{P_{21}, P_{22}, ..., P_{2m}\}$, where P_{2j} be the any equivalence class from a family of set P_2, where index $j = 1, 2, ..., m$. Partitions P_1 and P_2 are over U.

Let say we choose a equivalence class P_{1i} from P_1. It is a partition, so there is no other same elements in any other equivalence class in P_1. Now if we check all the equivalence classes of P_2. There has to be some equivalence classes where $P_{2j} \cap P_{1i} \neq \emptyset$. We have k number of equivalence classes from P_2, so $k \leq j$, which is $\bigcup_{k=1}^{n} P_{2k}$.

$P_{1i} \subset (\bigcup_{k=1}^{n} P_{2k})$, $P_{1i} \cap (\bigcup_{k=1}^{n} P_{2k}) = P_{1i}$, so that we can say that $P_1 \cap P_2 \Rightarrow P_1$. We can say same as for $P_1 \cap P_2 \Rightarrow P_2$.

C) $P_1 \Rightarrow P_1 \cup P_2$ and $P_2 \Rightarrow P_1 \cup P_2$.

$P_1 \cup P_2$ is a smallest partition that contains P_1 and P_2. Let X be the equivalence class of new partition $P_1 \cup P_2$ such that P_{1j} contains in X, so $P_{1j} \subset X$. We can say that $P_1 \Rightarrow P_1 \cup P_2$, same way we can show for $P_2 \Rightarrow P_1 \cup P_2$.

The following example will demonstrate the above properties.

Example

Suppose we have a family for subsets P_1, P_2, where $P_1 = \{P_{11}, P_{12}\}$ and $P_2 =$

$\{P_{21}, P_{22}\}$.

$$P_{11} = \{x_1, x_2, x_3\}$$
$$P_{12} = \{x_4, x_5, x_6\}$$
$$P_{21} = \{x_1, x_2\}$$
$$P_{22} = \{x_3, x_4, x_5, x_6\}$$

so followings are intersection properties

$$P_{11} \cap P_{21} = \{x_1, x_2\}$$
$$P_{11} \cap P_{22} = \{x_3\}$$
$$P_{12} \cap P_{21} = \emptyset$$
$$P_{12} \cap P_{22} = \{x_4, x_5, x_6\}$$

$$P_{11} \cap P_{21} \subset P_{11}$$
$$P_{11} \cap P_{22} \subset P_{11}$$
$$P_{12} \cap P_{21} \subset P_{12}$$

same as

$$P_{11} \cap P_{21} \subset P_{21}$$
$$P_{11} \cap P_{22} \subset P_{22}$$
$$P_{12} \cap P_{22} \subset P_{22}$$

so that we can say that $\forall P_i, \exists P_i \cap P_j \subset P_i$.

$$(P_{11} \cap P_{21}) \cup (P_{11} \cap P_{22}) = P_{11}$$

$$(P_{22} \cap P_{21}) \cup (P_{21} \cap P_{22}) = P_{12}$$

$$(P_{11} \cap P_{21}) = P_{21}$$

$$(P_{11} \cap P_{22}) \cup (P_{12} \cap P_{22}) = P_{22}$$

We can say that both weak dependency and strong dependency holds true.

$$P_{11} \cup P_{21} = \{x_1, x_2, x_3\}$$

$$P_{11} \cup P_{22} = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$

$$P_{12} \cup P_{21} = \{x_1, x_2, x_4, x_5, x_6\}$$

$$P_{12} \cup P_{22} = \{x_3, x_5, x_6\}$$

In the above example we got only partition $(P_1 \cup P_2) = \{\{x_1, x_2, x_3, x_4, x_5, x_6\}\}$, which satisfy that for all equivalence classes of partitions P_1 and P_2 is subset of equivalence class of partition $(P_1 \cup P_2)$.

4.1.2 Knowledge Dependency based on PCKBS

Let $K = (U, \mathcal{C})$ be knowledge base and let $C, D \subseteq \mathcal{C}$. In PCKBS, X is Pawlak's theory covering in given knowledge dependency definition.

We observe following properties for PCKBS :

D) $C_1 \cap C_2 \Rightarrow C_1$ and $C_1 \cap C_2 \Rightarrow C_2$

Suppose all the member's from $C_1 = \{C_{11}, C_{12}, ..., C_{1n}\}$. Let C_{1i} be the any member from a family of set C_1, where index $i = 1, 2, ..., n$. As covering C_1 and C_2 are over U. $C_2 = \{C_{21}, C_{22}, ..., C_{2m}\}$, where C_{2j} be the any member a family of set C_2, where index $j = 1, 2, ..., m$.

Let say we choose one member C_{1i} from C_1. It is covering so there may or may not be other member in C_1. If we check all the members of C_2. There has to be some members, where $C_{2j} \cap C_{1i} \neq \emptyset$. Consider k number of members from C_2, where $k \leq j$, which is $\bigcup_{k=1}^{n} C_{2k}$. $C_{1i} \subset (\bigcup_{k=1}^{n} C_{2k})$, $C_{1i} \cap (\bigcup_{k=1}^{n} C_{2k}) = C_{1i}$. We can say that $C_1 \cap C_2 \Rightarrow C_1$. We can show same as $C_1 \cap C_2 \Rightarrow C_2$.

E) $C_1 \Rightarrow C_1 \cup C_2$ and $C_2 \Rightarrow C_1 \cup C_2$

$C_1 \cup C_2$ is a smallest covering that contains C_1 and C_2. Let X be the member of new covering $C_1 \cup C_2$ such that C_{1j} contains in X. So $C_{1j} \subset X$. So we can say that $C_1 \Rightarrow C_1 \cup C_2$. We can prove for $C_2 \Rightarrow C_1 \cup C_2$.

These property can be easily understand from the below example.

Example

Suppose we have a family of covering C_1 and C_2, where $C_1 = \{C_{11}, C_{12}\}$ and $C_2 = \{C_{21}, C_{22}\}$.

The elements of subsets C_{11}, C_{12}, C_{21} and C_{22} are below:

$$C_{11} = \{x_1, x_2, x_3\}$$

$$C_{12} = \{x_2, x_4, x_5, x_6\}$$

$$C_{21} = \{x_1, x_2, x_6\}$$

$$C_{22} = \{x_3, x_4, x_5, x_6\}$$

So followings are intersection of subsets of covering

$$C_{11} \cap C_{21} = \{x_1, x_2\}$$
$$C_{11} \cap C_{22} = \{x_3\}$$
$$C_{12} \cap C_{21} = \{x_2, x_6\}$$
$$C_{12} \cap C_{22} = \{x_4, x_5, x_6\}$$

$$(C_{11} \cap C_{21}) \cup (C_{11} \cap C_{22}) = C_{11}$$
$$(C_{12} \cap C_{21}) \cup (C_{12} \cap C_{22}) = C_{12}$$
$$(C_{11} \cap C_{21}) \cup (C_{12} \cap C_{21}) = C_{21}$$
$$(C_{11} \cap C_{22}) \cup (C_{12} \cap C_{22}) = C_{22}$$

Therefore, strong dependency holds true for the Pawlak's covering.

$$C_{11} \cup C_{21} = \{x_1, x_2, x_3, x_6\}$$
$$C_{11} \cup C_{22} = \{x_1, x_2, x_3, x_4, x_5, x_6\}$$
$$C_{12} \cup C_{21} = \{x_1, x_2, x_4, x_5, x_6\}$$
$$C_{12} \cup C_{22} = \{x_2, x_3, x_4, x_5, x_6\}$$

Lets say $C_{31} = \{x_1, x_2, x_3\}$. We check for some category $C_2 \subseteq C_{31}$. In our case we don't find any categories in C_2. We include new element in C_{31}, which can satisfy for any categories in $C_1 \subseteq C_3$ and $C_2 \subseteq C_3$. Continuing this process, we get the new covering C_3 which satisfies for both covering C_1 and C_2. So in above example we got new cover $C_3 = \{\{x_1, x_2, x_3, x_4, x_5, x_6\}\}$.

4.1.3 Knowledge Dependency based on GCKBS

Let $K = (U, \mathcal{G})$ be knowledge base and let $G_1, G_2 \subseteq \mathcal{G}$. Here G_1 and G_2 are finite intersections of family of covering C and D respectively. X is General Covering in the definition of knowledge dependency.

More Precisely, knowledge G_2 is derivable from knowledge G_1, if all member of G_2 can be defined in terms of some member of knowledge G_1. If G_2 is derivable from G_2, we will also say that G_2 depends on G_1 and can be written as $G_1 \Rightarrow G_2$.

We observe following properties :

F) $G_1 \cap G_2 \Rightarrow G_1$ and $G_1 \cap G_2 \Rightarrow G_2$

Suppose all the member from $C_1 = \{C_{11}, C_{12}, ..., C_{1n}\}$. Let C_{1i} be the any member from a family of set C_1, where index $i = 1, 2, ..., n$. Covering C_1 and C_2 are over U. $C_2 = \{C_{21}, C_{22}, ..., C_{2m}\}$, where C_{2j} be the any categories a family of set C_2, where index $j = 1, 2, ..., m$. G_1 is finite intersections of C_1 and G_2 is finite intersections of C_2.

Let say we choose one member G_{1i} from G_1. It is covering so there may or may not be other member in G_1, which satisfy $G_{1j} \subset G_{1i}$. However we know that there is at least one member who satisfy $G_{1i} \subset G_1$. If we check all the members of G_2. There have to be some members, where $G_{2j} \cap G_{1i} \neq \emptyset$. Conder k number of members from G_2, where $k \leq j$, which is $\bigcup_{k=1}^{n} G_{2k}$.

$G_{1i} \subset (\bigcup_{k=1}^{n} G_{2k})$. Consequently $G_{1i} \cap (\bigcup_{k=1}^{n} G_{2k}) = G_{1i}$. We can say that $G_1 \cap G_2 \Rightarrow G_1$. Therefore, we can prove same as $G_1 \cap G_2 \Rightarrow G_2$.

G) $G_1 \Rightarrow G_1 \cup G_2$ and $G_2 \Rightarrow G_1 \cup G_2$.

$G_1 \cup G_2$ is a smallest covering that contains G_1 and G_2. Let X be the member of new covering $G_1 \cup G_2$ such that G_{1j} contains in X. Consequently $G_{1j} \subset X$. Therefore, we can say that $G_1 \Rightarrow G_1 \cup G_2$. Similarly, we can say for $G_2 \Rightarrow G_1 \cup G_2$.

CHAPTER 5

KNOWLEDGE REPRESENTATION

5.1 Introduction

A n-dimensional Euclidean space is the Cartesian product of n copies of real numbers. A unit point is a point whose co-ordinates are all 1 but one single $0, (1, ..., 0, 1, 1, ..., 1)$. These unit pointes can be viewed as vertices. Here we will explain in depth about n-simplex.These unit points will be regarded as vertices. We will use them to illustrate the notion of n-simplex.

Let us examine the $n-simplexes$, when $n = 0, 1, 2, 3$. 0-simplex $\triangle$ $(\mho_0)$ consist of a vertex $\mho_0$, which is a point in Euclidean space. 1-simplex $\triangle(V_0, V_1)$ consists of two points V_0, V_1. These two unit points can be explained as (V_0, V_1) an open segment in Euclidean space; However we did not include its end points. 2-Simplex $\triangle$ (V_0, V_1, V_2) consists of three points V_0, V_1, V_2. These three point can be explained with vertices V_0, V_1, and V_2 as an open triangle, which also does not include its vertices and edges. 3-Simplex $\triangle$ (V_0, V_1, V_2, V_3) has a four points V_0, V_1, V_2, V_3 that considered as an open tetrahedron. Again, its boundaries has not been included.

Formally, we can say that the points in Euclidean space can be any kind of objects but could not be vertices.

Definition 1: A n-simplex, denoted by $\triangle$ ($V_0, ..., V_n$) is a set of independent abstract vertices $V_0, ..., \{V_n\}$. A $\mathcal{G}-subset$ of a n-simplex is a G-simplex $\triangle$ $(V_{j_0}, ..., V_{j_a})$

whose vertices are a subset of $V_0, ..., V_n$ with cardinality $\mathcal{G} + 1$.

Definition 2: Simplex can be defined as single vertex has any set. Closed condition can be defined as simplex without empty set of subset. Here we conceives simplical complex with vertices and subsets of finite numbers of set.

$\mathcal{G} - simplex$ is simplex S who has exact one more vertices. There are also the dimension of s so that we could write $dims = G$. and it is also interpreted as non-closed simplical complex.

Let us calculate any set of elements k+1 can be explained as set of abstract vertices. We can define simplex as maximal only iff the simplex can not be a face of other simplexes. The k-complex can be defined as the maximal dimension of the constituting simplexes is k.

5.2 Granular Structure

Definition 3: Global granular space can be defined as a granular model set (U, B) where B is family of subsets $X_j, j \in J$. and U is knowledge is connected to varieties of classifications related to specific parts of the real or abstract world. Here J can be regard as an index set.

Examples on Intervals Granular Model of a Covering

$U_1 = (-3, 3)$, the open interval of real line. B_1 is the covering ,$(-3, 1), (-1, 3), (-2, 2)$

We name these intervals by x,y,z in the order listed. This covering is not a partition, we have the following intersections 1. $x \cap y = (-1, 1)$

2. $x \cap z = (-2, 1)$

3. $y \cap z = (-1, 2)$

4. $x \cap y \cap z = (-1, 1)$

So B_1 generated a family of non-empty granules, $B = a, b, c, ab, ac, bc, abc$, where $\cap$ has been compressed.

Now we will consider the concept of the closed triangle, where it has two-dimensional simplical complex. Now by interpretation of points $x = (1, 0, 0), y = (0, 1, 0) and z = (0, 0, 1)$, where each intersection can be viewed as open segment of the two points. We can consider open segment xy, xz and yz and all three points can be viewed as open triangle.

Example Granular model of a partition.

Let us have another example of B U_2 = (-3,3) ,

B_2 = (-3,-2),[-2,1), [-1,1),[1,2),[2,3)

We have x,y,z,a,b as clo-open intervals respectively and this partitions has intersection which are no-empty.

We can define quotient structure as a set of granules in discrete set.

5.3 The Granular Structure of a Covering

Let $B = B_j | j = 0, 1, ..., n$ is the finite set of cover U. In order to understand this we will consider set B, which is combinatorial topology:

1. $V_{granule} = B_j | j = 0, 1, 2, ...n$ is the granular vertices which is interpreted as abstract vertices.

2. A subfamily $\{B_{j_k}, k \in J\}$ can be called as a granular-simplex iff it has no-empty intersection or granules . In notations, $B_{j_k}, kinJ = B_{j_k} | \cap_k \in_J B_{j_k} \neq \emptyset$

$S_{granule}$ can be interpreted as set of abstract simplex.

Theorem The pair $(V_{granule}, S_{granule})$ is an Abstract Simplical Complex.

Granular Simplical Complex(GSC) can be easily understand by the family of non- empty granular points who has no-empty intersections to meet closed object. It is also considered as important part of $V_{granule}$.

Theory GSC can be considered as quotient structure iff B is a partition and if B is a partition, $V_{granule} = S_{granule}$ and GSC has only 0-simplexes.

This can be proof by B is a partition, then the family $B_{j_k}, k \in J = B_{j_k} | \cap_k \in J_{j_k} \neq \emptyset = \emptyset$, if the cardinality of $J \geq 2$. GSC is a set who has 0-simplexes.

5.4 Knowledge Representation Theory

The knowledge representation of a covering is a mapping. $K : U \to \beta; \rho \to \cap_i B_j$ where $\rho \in B_j$ for maximum number of j. The idea has be represented in the table. Let U can be a set with a covering β which has 12 basic granules, that is $\beta = \{a, b, c, d, e, f, g, h, w, x, y, z\}$.

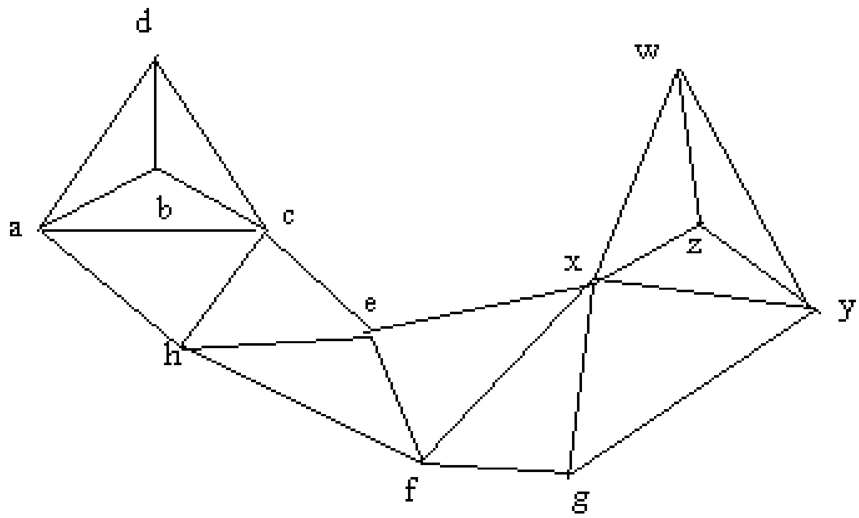

Figure 5.1: A complex with twelve vertexes.

As we can see from the figure that each vertex can be interpreted as granules. if granules has edge connection iff two vertex has not null intersection. Similarly if three vertex has non-empty intersection and three points are connected then it is an open triangles.

From the figure we can derive below intersections

$$a \cap b \cap c \cap d \neq \emptyset; w \cap x \cap y \cap z \neq \emptyset; \tag{5.1}$$

$$a \cap c \cap h \neq \emptyset; c \cap e \cap h \neq \emptyset; \tag{5.2}$$

$$e \cap f \cap h \neq \emptyset; f \cap e \cap x \neq \emptyset; \tag{5.3}$$

$$f \cap g \cap x \neq \emptyset; x \cap y \cap g \neq \emptyset; \tag{5.4}$$

The result can be shown in the table. For example, you can verify all the intersections given above from the table.

Table 5.1: Table :1 The structure of the covering β

IDG Connect

Universe	Granules	Contents of granules
1	a	1,13,14
2	b	2,13
3	c	3,13,14,15
4	d	4,13
5	e	5,14,15,16
6	f	7,16,18
7	g	9,18,19
8	h	5,14,15,16
9	w	11,20
10	x	8,17,18,19,20
11	y	10,19,20
12	z	12,20
13	$a \cap b \cap c \cap d$	1
14	$a \cap c \cap h$	2
15	$c \cap e \cap h$	3
16	$e \cap f \cap h$	4
17	$f \cap e \cap x$	5
18	$f \cap g \cap x$	6
19	$x \cap y \cap g$	7

CHAPTER 6

CONCLUSION

Rough Set Theory is a successful discipline in knowledge engineering. We extended Partition to Covering and Granulation.

However extension is a quantum jump:

- Approximation is viewed as learning theory.
- Classical Knowledge Representation is extended from set theory to algebraic theory.

BIBLIOGRAPHY

[AHU74] A.V. Aho, J.E. Hopcroft, and J.D. Ullman, *The design and analysis of computer algorithms*, Addison-Wesley, 1974.

[BF81] A. Barr and E.A. Feigenbaum, *The handbook of artificial intelligence*, Addison-Wesley, 1981.

[BM77] G. Birkhoff and S. MacLane, *A survey of modern algebra*, Macmillan, 1977.

[Hob85] J.R. Hobbs, *Granularity*, Proceedings of the Ninth Internation Joint Conference on Artificial Intelligence, 1985, pp. 432–435.

[Lee83] T.T. Lee, *Algebraic theory of relational databases*, The Bell System Technical Journal **62** (1983), no. 10, 3159–3204.

[LH96] T.Y. Lin and M. Hadjimichael, *Non-classificatory generalization in data mining*, Proceedings of the 4th Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery, 1996, pp. 404–412.

[Lin88] T.Y. Lin, *Neighborhood systems and relational database*, Proceedings of CSC'88, 1988, p. 725.

[lin89] *Neighborhood systems and approximation in database and knowledge base systems*, 1989.

[Lin96] T.Y. Lin (ed.), *A set theory for soft computing*, 1996.

[LZ04] T.Y. Lin and L. Zadeh, *Special issue on granular computing and data mining*, International Journal of Intelligent systems **19** (2004), no. 7, 565–566.

[LZDO98] T.Y. Lin, N. Zhong, J. Duong, and S. Ohsuga, *Frameworks for mining binary relations in data*, Rough sets and Current Trends in Computing (A. Skoworn and L. Polkowski, eds.), LNCS 1424, Springer-Verlag, 1998, pp. 387–393.

[Paw82] Z. Pawlak, *Rough sets*, International Journal of Information and Computer Science **11** (1982), no. 15, 341–356.

[Paw91] ———, *Rough sets–theoretical aspects of reasoning about data*, Kluwer Academic Publishers, 1991.

[SK56] W. Sierpenski and C. Krieger, *General topology*, University of Toronto Press, 1956.

[Szy02] C. Szyperski, *Component software: Beyond object-oriented programming*, Addison-Wesley, 2002.

[Zad79] L.A. Zadeh, *Fuzzy sets and information granularity*, Advances in Fuzzy Set Theory and Applications (N. Gupta, R. Ragade, and R. Yager, eds.), North-Holland, 1979, pp. 3–18.

[Zim91] H. Zimmerman, *Fuzzy set theory –and its applications*, Kluwer Acdamic Publisher, 1991.

[ZZ92] B. Zhang and L. Zhang, *Theory and applications of problem solving*, North-Holland, 1992.