

July 2011

ACQUISITION AND EXTRACTION OF DYNAMIC KNOWLEDGE USING D-ROUGH SET

Brojo Kishore Mishra

Dept. of Computer Science MITS, Rayagada, mishra.bkm@gmail.com

Susanta Kumar Das

Dept. of Computer Science Berhampur University, dr.dassusanta@yahoo.co.in

Follow this and additional works at: <https://www.interscience.in/ijcsi>



Part of the [Computer Engineering Commons](#), [Information Security Commons](#), and the [Systems and Communications Commons](#)

Recommended Citation

Kishore Mishra, Brojo and Kumar Das, Susanta (2011) "ACQUISITION AND EXTRACTION OF DYNAMIC KNOWLEDGE USING D-ROUGH SET," *International Journal of Computer Science and Informatics*: Vol. 1 : Iss. 1 , Article 4.

DOI: 10.47893/IJCSI.2011.1003

Available at: <https://www.interscience.in/ijcsi/vol1/iss1/4>

This Article is brought to you for free and open access by the Interscience Journals at Interscience Research Network. It has been accepted for inclusion in International Journal of Computer Science and Informatics by an authorized editor of Interscience Research Network. For more information, please contact sritampatnaik@gmail.com.

ACQUISITION AND EXTRACTION OF DYNAMIC KNOWLEDGE USING D-ROUGH SET

Brojo Kishore Mishra
Dept. of Computer Science
MITS, Rayagada.
mishra.bkm@gmail.com

Swarupananda Bissoyi
Development Lead
Samsung India S/w operation
Bangalore.

Susanta Kumar Das
Dept. of Computer Science
Berhampur University,
dr.dassusanta@yahoo.co.in

Abstract

Theory of knowledge has a long and rich history. Various aspects of knowledge are widely discussed issues at present, mainly by logicians and artificial intelligence (AI) researchers. It is one of the concepts, used to build intelligence system. Many soft computing tools are available for extraction, acquisition and validation of knowledge. Rough set is one such tool, mainly used for classification and extraction of knowledge. Rough Set Theory was proposed by Pawlak in 1982 as a tool for knowledge Extraction. However, when knowledge extraction is studied, we observed that most of the knowledge is static in nature. For analyzing Knowledge having dynamic in nature, Pawlak's Rough Set Theory must be reconsidered. Dong Ya Li (et. al.) has already proposed the concept of dynamic Rough Set in 2007. We here, further analyze this concept and try to find out some more properties of it. Dynamic Rough Set (D-rough set) is a common form of Pawlak's Rough Set as Pawlak's rough set can be considered as a special case of D-rough set. D-rough set is based on concepts, such as elementary transfer coefficient. D-rough set and D-Approximate set can be used for studying and analyzing dynamic knowledge. Further, we study and analyze the properties mentioned by Bussee. Grzymala-Busse has established some properties of approximation of classifications. These results are irreversible by nature. Pawlak has noted that these results of Busse establish that the two concepts, approximation of sets and approximation of families of sets (or classifications) are two different issues and that the equivalence classes of approximate classifications cannot be arbitrary sets. He has further stated that if we have positive example of each category in the approximate classification then we must have also negative examples of each category. In this paper, we have mentioned these aspects of the theorems of Busse and tried to study their properties, when D-rough and D-Approximate set has been incorporated. Lastly, we had provided the physical interpretation of each one of them.

Keywords: Knowledge, Rough Set, D-rough set, D-approximate set, Approximation of families of set.

I Introduction

Rough Set Theory is a mathematical formalism for representing uncertainty that can be considered as an extension of the classical set theory. It has been used in many different research areas, including those related to inductive machine learning and reduction of knowledge in knowledge-based systems. We can observe the following about the rough set approach:

- Introduction of efficient algorithms for finding hidden patterns in data,
- Determination of optimal sets of data (data reduction),
- Evaluation of the significance of data,
- Generation of sets of decision rules from data,
- Easy-to-understand formulation,
- Straightforward interpretation of obtained results,
- Suitability of many of its algorithms for parallel processing.

Rough set theory, proposed by Pawlak in 1982, can be seen as a new mathematical approach to vagueness. The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). Objects characterized by the same information are indiscernible (similar) in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis of rough set theory. In rough set approach indiscernibility is defined relative to a given set of functional (attributes).

The basic assumption of rough set theory as put forth by Pawlak is that human knowledge about a universe depends upon their capability to classify its objects. Classifications of a universe and equivalence relations defined on it are known to be interchangeable notions. So, for mathematical reasons equivalence relations were considered by Pawlak to define rough sets. A pair of crisp sets, called the lower and upper approximations of the set, represents a rough set. The lower approximation of a rough set comprises of those elements of the universe, which can be said to belong to it definitely with the available knowledge. The upper approximation on the other hand comprises of those elements, which are possibly in the set with respect to the available information. The concept of rough sets was primarily concerned with the study of intelligent systems characterized by insufficient and incomplete information.

Any set of all indiscernible (similar) objects is called an elementary set and forms a basic granule of knowledge about the universe. Any union of some elementary sets is referred to as crisp set- otherwise the set is rough. As per the study of Busse's

Concepts [2, 3, 4],

For algorithmic approach we divide the attributes into two types:

1. Conditions
2. Decisions (or Actions)

Objects are described by values of conditions, while classifications of experts are represented by values of decisions. For a set of conditions of the information system and a given action d of an expert, lower and upper approximations of a classification, generated by d , may be computed in a straight forward way, using their simple definition. Such approximations are the basis of rough set theory. From these approximations, certain and possible rules may be determined for action d , again in a straightforward way. Induced rules are categorized into certain and possible.

II Definitions

Let U be an universe of discourse and R be an equivalence relation over U . By U/R we denote the family of all equivalence classes of R , referred to as categories or concepts of R and the equivalence class of an element $x \in U$ is denoted by $[x]_R$.

Definition 1 By a knowledge base, we understand a relational system $K = (U, \mathfrak{R})$, where U is as above and \mathfrak{R} is a family of equivalence relations over U .

Definition 2 For any subset $P(\neq \phi) \subset \mathfrak{R}$, the intersection of all equivalence relations in P is denoted by $IND(P)$ and is called the *indiscernibility relation over P* . We define $IND(K) = \{ IND(P): P(\neq \phi) \subset \mathfrak{R} \}$.

Definition 3 Let $X \subseteq U$ and R be in $IND(K)$. The sets $\underline{R}X$ and $\overline{R}X$ are called the *R-lower* and *R-upper approximations* of X respectively and are defined as follows:

$$\underline{R}X = \{x \in U: [x]_R \subseteq X\}, \quad \overline{R}X = \{x \in U: [x]_R \cap X \neq \phi\}.$$

Definition 4 For $X \subseteq U$, the *R-boundary* of X is denoted by $BN_R(X)$ and is defined as

$$BN_R(X) = \overline{R}X - \underline{R}X.$$

Definition 5 A set $X \subseteq U$ is said to be *rough with respect to R* if and only if $\underline{R}X \neq \overline{R}X$; that is,

$BN_R(X) \neq \phi$. X is said to be *R-definable* if and only if $\underline{R}X = \overline{R}X$ or $BN_R(X) = \phi$.

It may be noted that R-definable sets are crisp sets with respect to R . Many properties of the lower and upper approximations of rough sets, union of rough sets and intersection of rough sets have been obtained.

Definition 6 Let $x \in U$ and $X \subseteq U$. We say x is *certainly in X* with respect to R if and only if $x \in \underline{R}X$ and x is *possibly in X* if and only if $x \in \overline{R}X$.

III Approximation of Classifications

Classifications of universes play central roles in basic rough set theory. We define below a classification formally.

Definition 1 Let $F = \{X_1, X_2, \dots, X_n\}$ be a

family of non empty sets defined over U . We say that F is a *classification of U* if and only if

$$X_i \cap X_j = \phi \text{ for } i \neq j \text{ and } \bigcup_{i=1}^n X_i = U.$$

Definition 2 Let F be as above and R be an equivalence relation over U . Then $\underline{R}F$ and $\overline{R}F$ denote respectively the *R-lower* and *R-upper approximations* of the family F and are defined as

$$\underline{R}F = \{ \underline{R}X_1, \underline{R}X_2, \dots, \underline{R}X_n \},$$

$$\overline{R}F = \{ \overline{R}X_1, \overline{R}X_2, \dots, \overline{R}X_n \}.$$

We assume that F is a classification of U and R is an equivalence relation over U .

Grzymala-Busse has established some properties of approximation of classifications. These results are irreversible by nature. Pawlak has noted that these results of Busse establish that the two concepts, approximation of sets and approximation of families of sets (or classifications) are two different issues and that the equivalence classes of approximate classifications cannot be arbitrary sets. He has further stated that if we have positive example of each category in the approximate classification then we must have also negative examples of each category. In this article, we further analyze these aspects of theorems of Busse and provide physical interpretation of each one of them by taking a standard example.

One primary objective is to extend the results of Busse by obtaining necessary and sufficient type theorems and show how the results of Busse can be derived from them. The results of Busse we discuss

here are in their slightly modified form as presented by Pawlak.

Theorems on approximation on classifications

In this section, we shall establish two theorems which have many corollaries generalizing the four theorems established by Busse and presented in slightly modified forms by Pawlak. We shall also provide interpretations for most of these results including those of Busse and illustrate them through a simple example of toys.

We shall use the following notations for representational convenience:

$$N_n = \{1, 2, \dots, n\}.$$

For any $I \subset N_n$, I^c is the complement of I in N_n .

Theorem 1 For any $I \subset N_n$, $\bar{R}(\bigcup_{i \in I} X_i) = U$ if

$$\text{and only if } \underline{R}(\bigcup_{j \in I^c} X_j) = \phi.$$

Corollary 1 Let $F = \{X_1, X_2, \dots, X_n\}$ be a classification of U and let R be an equivalence relation on U and $I \subset N_n$. If $\bar{R}(\bigcup_{i \in I} X_i) = U$ then

$$\underline{R}X_j = \phi \text{ for each } j \in I^c.$$

Corollary 2 For each $i \in N_n$, $\bar{R}X_i = U$ if and only if $\underline{R}(\bigcup_{j \neq i} X_j) = \phi$.

Taking $I = \{i\}^c$, in Theorem 3.1 we get

Corollary 3 For each $i \in N_n$, $\underline{R}X_i = \phi$ if and only if $\bar{R}(\bigcup_{j \neq i} X_j) = U$.

Corollary 4 If there exists $i \in N_n$ such that $\bar{R}X_i = U$ then for each $j (\neq i) \in N_n$, $\underline{R}X_j = \phi$.

Corollary 5 If $\bar{R}X_i = U$ for all $i \in N_n$ then $\underline{R}X_i = \phi$ for all $i \in N_n$.

Theorem 2 For any $I \subset N_n$, $\underline{R}(\bigcup_{i \in I} X_i) \neq \phi$ if and only if $\bigcup_{j \in I^c} \bar{R}X_j \neq U$.

Corollary 6 For $I \subset N_n$, if $\underline{R}(\bigcup_{i \in I} X_i) \neq \phi$ then $\underline{R}X_j \neq U$ for each $j \in I^c$.

Corollary 7 For each $i \in N_n$, $\underline{R}X_i \neq \phi$ if and only if $\bigcup_{j \neq i} \bar{R}X_j \neq U$.

Taking $I = \{i\}^c$ in Theorem 3.2 we get

Corollary 8 For all $i, 1 \leq i \leq n$, $\bar{R}X_i \neq U$ if and only if $\underline{R}(\bigcup_{j \neq i} X_j) \neq \phi$.

By Corollary 3.7, $\underline{R}X_i \neq \phi \Rightarrow \bigcup_{j \neq i} \bar{R}X_j \neq U$ for each $j \neq i, 1 \leq j \leq n$.

Corollary 9 If there exist $i \in N_n$ such that $\underline{R}X_i \neq \phi$ then for each $j (\neq i) \in N_n$, $\bar{R}X_j \neq U$.

Corollary 10 If for all $i \in N_n$, $\underline{R}X_i \neq \phi$ holds then $\bar{R}X_i \neq U$ for all $i \in N_n$.

IV D – Rough set and D – Approximate set

Rough Set theory is generally used for static knowledge, without considering the elementary transfer outward or inward X. Where X is categories as any static set under the universe U. As in real situations, information system have dynamic characteristic. Thus, we can define this dynamic situation as either addition or disappearance of information at a certain rate. Thus, Li. Et. al. proposed the concept of dynamic rough set theory which is a generalization of Pawlak's rough set theory. We can define that element is either moving inward or outward of category X.

Definition 1 Suppose (U, \mathfrak{R}) is a non empty knowledge base, and $X \subseteq U$ and $T \subseteq \mathfrak{R}$. For $x \in U$, we define

$$\rho_{(x,T)}^-(x) = \frac{\text{card}([x]_T \setminus X)}{\text{card}([x]_T)}, \text{ as } x \in X.$$

$$\rho_{(x,T)}^+(x) = 1 - \frac{\text{card}([x]_T \setminus X)}{\text{card}([x]_T)}, \text{ as } x \in \square X.$$

$\rho_{(x,T)}^-(x)$ and $\rho_{(x,T)}^+(x)$ are known as outward and inward transfer coefficient of element x about T respectively.

Definition 2. Suppose (U, \mathfrak{R}) is a non empty knowledge base, and $X \subseteq U$ and $T \subseteq \mathfrak{R}$. $M_T^+(X)$ is called inflated dynamic main set of X about T, which stands for inflated D-main set defined by

$$M_T^+(X) = \{x \mid x \in (\sim X), d_T^+(X) \leq \rho_{(x,T)}^+(x) < 1\}$$

$A_T^+(X)$ is said to be inflated dynamic assistant set of X about T, which stands for inflated D-assistant set, defined by

$$A_T^+(X) = \{x \mid x \in (\sim X), 0 \leq \rho_{(x,T)}^+(x) < d_T^+(X)\}$$

X_T^+ is called inflated dynamic set of X about T, which stands for inflated D-set defined by

$$X_T^+ = X \cup M_T^+(X)$$

$\rho_{(x,T)}^+(x)$ is said inflation degree of X about T defined by

$$\rho_T^+(X) = \frac{\text{card}(M_T^+(X))}{\text{card}(X)}$$

Definition 3. Suppose (U, \mathfrak{R}) is a non empty knowledge base, and $\rho \subseteq \mathfrak{R}$ and $T \subseteq \mathfrak{R}$. X_T^+ is an inflated D – Set of $X \subseteq U$. Then Lower approximation set and upper approximation set of X_T^+ on ρ , are respectively called inflated D – Lower approximated set inflated and D – Upper approximated set of X on ρ , which is respectively denoted as $\underline{\rho}_T^+(X) = \{x \mid x \in U, [x]_\rho \subseteq X_T^+\}$, and $\overline{\rho}_T^+(X) = \{x \mid x \in U, [x]_\rho \cap X_T^+ \neq \emptyset\}$.

When $\underline{\rho}_T^+(X) = \overline{\rho}_T^+(X)$ holds, X_T^+ is ρ -inflated dynamic definable set of X, which stands for ρ -inflated D – definable set.

When $\underline{\rho}_T^+(X) \neq \overline{\rho}_T^+(X)$ holds, X_T^+ is ρ -inflated dynamic rough set of X, which stands for ρ -inflated D – rough set.

Positive region $POS_\rho(X_T^+)$, negative region $NEG_\rho(X_T^+)$ and boundary region $BN_\rho(X_T^+)$ of X_T^+ about ρ , are respectively called inflated D – positive region, inflated D – negative region and inflated D – boundary region of X on ρ , defined respectively by the following three equations.

$$POS_{(\rho,T)}^+(X) = \underline{\rho}_T^+(X),$$

$$NEG_{(\rho,T)}^+(X) = U \setminus \overline{\rho}_T^+(X),$$

$$BN_{(\rho,T)}^+(X) = \overline{\rho}_T^+(X) \setminus \underline{\rho}_T^+(X).$$

V Conclusions

Most of the knowledge in real life is uncertain or imprecise in character. If one tries to do away with these then much of the desired knowledge is lost. Busse, for the first time developed theories to deal

with such knowledge base through rough set approach. Also, he has taken inconsistency into account. Automation of his approaches has been carried out by us successfully in this project. It is worth noting that we have developed some of the algorithms ourselves and made enhancement of the theory whenever required.

A new approach to knowledge acquisition under uncertainty can be build based on rough set theory. The real world phenomena are consisting of information system, where inconsistencies are included can be tested through this model. Such inconsistencies are present because of different actions of the same expert for different objects described by the same values of conditions. Different actions of different experts for the same object are another source of inconsistency. For a set of conditions of the information system, and a given action of an expert, lower and upper approximations of a classification, generated, may be computed in a straightforward way, using their simple definitions. Such approximations are the basis of rough set theory. From these approximations, certain and possible rules may be determined for action.

The above concept of D- rough set theory was proposed by Dong Ya li et. al (IEEE FSKD 2007), can be used and implmented with Busse's Properties. Busse's properties are based on Pawlak's rough set theory which deals with static information. As the real world demand - knowledge to be dynamic, Implementing D – rough concept with Busse's property will lead to generation of rule and inductive condition to solve problems related to uncertainty.

The rule generation using rough sets has been studied by many authors and different types of efficient algorithms have been developed by them. These new methods can be implemented. However, it is worth noting that each algorithm has its specific area of application. So, one can develop a package of such rule generation algorithms in future. Also, reduction in the number of attributes in a knowledge base makes the storage and analysis more efficient. This can be added in future.

Reference

1. Dubois, D. and Prade, H., Twofold fuzzy sets and rough sets, fuzzy sets and systems 17(1985).
2. J.W.Grzymala-Busse., On the reduction of knowledge representation systems, Proc. 6th international workshop on expert systems and their applications, Aviguan, France, April 28-30 (1986), pp. 463-478 .
3. J.W.Grzymala-Busse., Knowledge acquisition under uncertainty – a rough set

- approach, journal of intelligent and robotic systems, (1988), pp. 3-16.
4. J.W.Grzymala-Busse., Rough set strategies to Data with missing attribute values, found and novel approaches in data mining, T.Y.L in, S.Ohsuga, C.J.Lian and X.hu (Eds). Studies in computational Intelligence, vol.9 .special edition, (2006), pp. 197-212.
 5. Mishra, B.K., Knowledge Acquisition under Uncertainty – A Rough Set Approach, M.Tech – project Dissertation, Berhampur University, 2008.
 6. Pawlak, Z., Rough sets, International journal of computer and information sciences ,II (1982),pp.341-356.
 7. Pawlak, Z., Rough classification, International journal Man-Machine studies, 24 (1983), pp. 469-483.
 8. Pawlak, Z., Rough sets. - Basic notions, Institute Comp.Sci. Acad. Sci. Rep. No. 431, Warsaw(1981).
 9. Pawlak, Z., Classification of objects by means of attributes, Institute Comp. Sci. Polish Acad. Sci. Rep. No.429, Warsaw,(1981).
 10. Pawlak, Z., Rough Sets, Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, 1991.
 11. Pawlak,Z. and Skowron,A., Rudiments of rough sets, Information Sciences-An International Journal.Elsevier,177(1)(2007),3-27.
 12. Pawlak,Z. and Skowron,A., Rough sets: Some extensions, Information Sciences-An International Journal.Elsevier,177(1)(2007),28-40.
 13. Pawlak,Z. and Skowron,A., Rough sets and Boolean reasoning, Information Sciences- An International Journal.Elsevier,177(1)(2007),41-73.
 14. Tripathy, B.K, Mohanty, D. and Ojha, J., On Approximation of Classification and Rough Comparison of Rough Sets, communicated to Foundations of Computing and Decision Sciences.
 15. Tripathy, B.K and Mitra. A, “On Some Basic properties of Rough sets and Applications”, Accepted with modification for publication in ‘Transactions on Rough Sets’ Springer Publications.
 16. Dong Ya Li and BAo Qing Hu, “A Kind of Dynamic Rough Sets”, FSKD 2007, IEEE.