Scientific
Research

Realization of Rough Set Approximation Toplogical Operations Based on Formal Concept Analysis

Huilai Zhi

School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
Email: zhihuilai@126.com

Abstract

There is an intimate correlation between rough set theory and formal concept analysis theory, so rough set approximations can be realized by means of formal concept analysis. For any given multiple valued information system, the realization of rough set approximation operation has two major steps, firstly convert the information system from multiple valued one to single valued formal context, secondly realize rough set approximation operations aided by concept lattice, which is equivalent to a query operation under some necessary conditions.

Keywords

Rough Set, Approximation Operation, Formal Context, Concept Lattice

1. Introduction

Rough set theory (RS), first described by a Polish computer scientist Zdzisław I. Pawlak [1], is a formal approximation of a crisp set (*i.e.*, conventional set) in terms of a pair of sets which give the lower and the upper approximation of the original set. In the standard version of rough set theory, the lower- and upper-approximation sets are crisp sets, but in other variations, the approximating sets may be fuzzy sets. Formal concept analysis (FCA) is a principled way of automatically deriving ontology from a collection of objects and their properties [2]. The term was introduced by Rudolf Wille in 1984, and built on applied lattice and order theory that was developed by Birkhoff and others in the 1930s.

Research about the relationship between formal concept analysis and rough set theory has gained some development. Kent [3] and Yao [4] bring upper approximation and lower approximation in rough set theory into formal concept analysis, and discuss many approximation operators based on formal concept analysis. Qu Kaishe [5] focuses on the theory research about the correlation between formal concept analysis and rough set. He

first reveals limitations of data analysis and processing in the rough set theory, and then provides a way for the synthesis of formal concept analysis and rough set theory by using nominal scale.

In these studies, the link between rough set theory and concept lattice is pointed out, but in practice how to realize the operation of rough set by using concept lattice is still left as a problem. Therefore, in this article, we focus on the way to realize rough set approximation topological operations based concept lattice, which includes upper approximation and lower approximation.

2. About Rough Set Theory and FCA

2.1. Basic Concepts in Rough Set Theory

Indiscernibility Relation is a central concept in rough set theory, and is considered as a relation between two objects or more, where all the values are identical in relation to a subset of considered attributes. Indiscernibility relation is an equivalence relation, where all identical objects of set are considered as elementary [6].

Approximations is also other an important concept in Rough Sets Theory, being associated with the meaning of the approximation topological operations [7]. The lower and the upper approximations of a set are interior and closure operations in a topology generated by the indiscernibility relation. Below is resented and described the types of approximations that are used in Rough Sets Theory.

a) Lower Approximation (B')

Lower Approximation is a description of the domain objects that are known with certainty to belong to the subset of interest. The Lower Approximation Set of a set X, with regard to R is the set of all of objects, which certainly can be classified with X regarding R, that is, set B'.

Formally, $B'(X) = \{u \in U | [u]_R \subseteq X\}$

b) Upper Approximation (B^*)

Upper Approximation is a description of the objects that possibly belong to the subset of interest. The Upper Approximation Set of a set X regarding R is the set of all of objects which can be possibly classified with X regarding R, that is, set B^*.

Formally, $B^*(X) = \{u \in U | [u]_R \cap X \neq \varnothing\}$, $\varnothing$ denote an empty set.

c) Boundary Region (BR)

Boundary Region is description of the objects that of a set X regarding R is the set of all the objects, which cannot be classified neither as X nor $-X$ regarding R. If the boundary region is a set $X = \varnothing$, then the set is considered "Crisp", that is, exact in relation to R; otherwise, if the boundary region is a set $X \neq \varnothing$ the set X "Rough" is considered. In that the boundary region is BR.

Formally, $BR(X) = B^*(X) - B'(X)$.

2.2. Basic Concepts in Formal Concept Analysis

Formal concept analysis, which is proposed by R. Wille, is founded on a basis of order theory and lattice theory, and is a mathematical structure which depicts relationship between objects and attributes according to basic information provided by data base. Formal concept analysis has been successfully used in many fields, and to some extend it has been treated as a means of external cognition [8].

Definition 1: Given a formal context $K = (U, A, I)$, $X \subseteq U$, $B \subseteq A$, if X and B satisfy $f(X) = B$, $g(B) = X$, and $f(X) = \{m \in A | \forall g \in X, gIm\}$, $g(B) = \{g \in U | \forall m \in B, gIm\}$, we call the pair (X, B) is a concept of formal context K. X is called extent of concept (X, B), B is called intent of concept (X, B). $L(U, A, I)$ or $L(K)$ is the set which contains of the concepts of K [2].

Definition 2: Given a formal context $K = (U, A, I)$, $(X_1, B_1), (X_2, B_2) \in L(K)$, if $X_1 \subseteq X_2$ or $B_2 \subseteq B_1$, then call (X_1, B_1) is a son concept of (X_2, B_2), and denote as $(X_1, B_1) \leq (X_2, B_2)$. Apparently, $L(K)$ forms a lattice, which is called concept lattice [2].

Given two concepts (X_1, B_1), (X_2, B_2) of a formal context $K = (U, A, I)$, we have: If $X_1 \subseteq X_2$, then $f(X_2) \subseteq f(X_1)$, for $X_1, X_2 \subseteq U$; If $B_1 \subseteq B_2$, then $g(X_2) \subseteq g(X_1)$, for $B_1, B_2 \subseteq A$.

3. Realization of Approximation Operation Based on FCA

Realization of rough set approximation operation can be divided into two steps: first, convert multiple valued

information system to multiple valued formal context, and then covert multiple value formal context to single value formal context; second, realize rough set approximation operation by using the technique of formal concept analysis based on concept lattice.

3.1. Convert Multiple Valued Is from to Single Valued Formal Context

In an information system $IS = (U, A)$, A is the set of attributes, for every attribute a_i, if the value of a_i belongs to $\{0,1\}$ in which 1 represents an object has this attribute and 0 represents an object doesn't has this attribute, then this information system corresponds to an single value formal context. Otherwise, if the value of a_i belongs to a set V_i, and the size of V_i is greater than 2, then we have to convert this attribute to several attributes according to the size of V_i. For example, if there is a attribute "shape", and the value of "shape" belongs to {triangle, round, rectangle}, then attribute "shape" can convert to three attributes, namely "triangle", "round" and "rectangle", and the value of every attributes belongs to $\{0,1\}$.

Definition 3: Let $IS = (U, A)$ be an information system, if the value of an attribute a_i belongs to a set V_i, and the size of V_i is $n\,(n \geq 2)$, then convert a_i to a series of attributes $\{b_1, b_2, \cdots, b_n\}$, and call $\{b_1, b_2, \cdots, b_n\}$ are derived attributes of a_i. After carrying out this converting process, attributes set A convert to a new attributes set M, then M is called the derived attributes of A.

Definition 4: Let $IS = (U, A)$ be an information system, assume $G = U$, M is the derived attributes of A, I is the relation between U and M, the we call $K = (G, M, I)$ is the single value formal context derived from information system IS.

3.2. Realization of Approximation Based on Concept Lattice

For the sake of narrative convenience, we agree several symbols in this paper. $L(K)$ denotes Concept lattice of formal context K and $B(K)$ denotes all concepts of $L(K)$. Lower case letters represents concepts, for example, we use letter c denote a concept, and let $Ext(c)$ and $Int(c)$ represent extension and intension of concept c respectively.

Definition 5: Let $IS = (U, A)$ be an information system, $K = (G, M, I)$ is the single value formal context derived from information system IS, assume $N \subseteq M$, $J = (G \times N) \cap I$, we call formal context $H = (G, N, J)$ is sub-context of K regarding N.

Theorem 1: Let $IS = (U, A)$ be an information system, $K = (G, M, I)$ is the single value formal context derived from information system IS, $X \subseteq U$, $B \subseteq A$, $N \subseteq M$, attribute B is derived from attribute A, $H = (G, N, J)$ is sub-context of K regarding N, $L(H)$ is the concept lattice of H, Upper Approximation:

$$B^*(X) = \left\{ x \in Ext(c) \middle| Ext(c) \cap X \neq \varnothing, c \in B(G, N, J), |Int(c)| = |B| \right\},$$

Lower Approximation:

$$B'(X) = \left\{ x \in Ext(c) \middle| Ext(c) \subseteq X, c \in B(G, N, J), |Int(c)| = |B| \right\}.$$

Proof: According to the definition of concept in FCA, extent set $Ext(c)$ has attributes $Int(c)$, and size of $Int(c)$ equals to the size of B, then $Ext(c)\big|_{c \in B(G,N,J)}$ forms the equivalence classes of the B-indiscernibility relation, which means $Ext(c)\big|_{c \in B(G,N,J)}$ equals to $[x]_B$. According to definition of upper and lower approximation, this theorem obviously holds.

Theorem 2: Let $IS = (U, A)$ be an information system, $K = (G, M, I)$ is the single value formal context derived from information system IS, assume $X \subseteq U$, $B \subseteq A$, $B(K)$ contains all the concepts in concept lattice $L(K)$, Upper Approximation is:

$$B^*(X) = \left\{ x \in Ext(c) \middle| Ext(c) \cap X \neq \varnothing, c \in B(G, M, I), \forall a_l \in B, m_{li} \in Int(c) \right\},$$

Lower Approximation is:

$$B'(X) = \left\{ x \in Ext(c) \middle| Ext(c) \subseteq X, c \in B(G, M, I), \forall a_l \in B, m_{li} \in Int(c) \right\}.$$

Proof: According to the definition of concept in FCA, extent set $Ext(c)$ has attributes $Int(c)$, and $\forall a_l \in B, m_{li} \in Int(c)$, then $Ext(c)\big|_{c \in B(G,M,I)}$ forms the equivalence classes of the B-indiscernibility relation, which means $Ext(c)\big|_{c \in B(G,M,I)}$ equals to $[x]_R$. According to definition upper and lower approximation, this theorem obviously holds.

3.3. An Example

In an information system $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$, three are equivalent classes, namely

$$U/R_1 = \{\{1, 3, 7\}, \{2, 4\}, \{5, 6, 8\}\},$$

$$U/R_2 = \{\{1, 5\}, \{2, 6\}, \{3, 4, 7, 8\}\},$$

and

$$U/R_3 = \{\{2, 7, 8\}, \{1, 3, 4, 5, 6\}\}.$$

Let $X = \{4, 6, 7, 8\}$, $B = \{R_2, R_3\}$. Now let's compute $B^*(X)$ and $B'(X)$ by using Theorem 1 and Theorem 2.

In the above information system, R_2 and R_3 divide U into three equivalent classes, thus the size of the value range of R_2 and R_3 is 3, and then we must convert these two attributes form multiple value to single value. The derived single value context is shown in **Table 1**. By using Godin algorithm [9], we can build the concept lattice of a given single valued formal context.

1) Compute $B^*(X)$ and $B'(X)$ by using Theorem 1

Under the attributes constraint $B = \{R_2, R_3\}$, we get sub-context K_1 of formal context K, which includes the columns labeled by R_2 and R_3. Concept lattice $L(K_1)$ is shown in **Figure 1**.
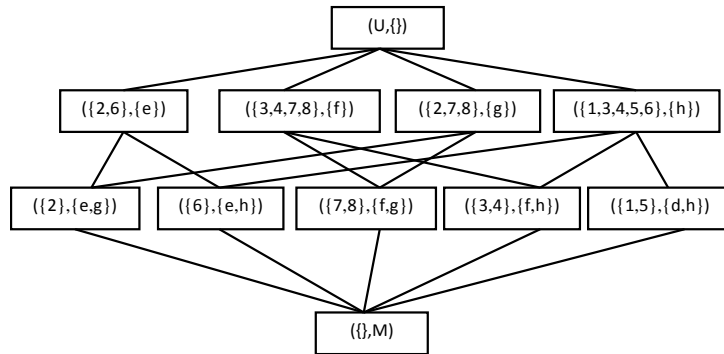

Figure 1. Concept lattice $L(K_1)$.

Table 1. Single valued formal context K.

	R_1			R_2			R_3	
	a	b	c	d	e	f	g	h
1	*			*				*
2		*			*		*	
3	*					*		*
4		*				*		*
5			*	*				*
6			*		*			*
7	*					*	*	
8			*			*	*	

In concept lattice $L(K_1)$, the concepts whose extent satisfy the constraint $\left|Ext(c)\right| = \left|B\right| = \left|\{R_2, R_3\}\right| = 2$ are $(\{2\}, \{e, g\}), (\{6\}, \{e, h\}), (\{7, 8\}, \{f, g\}), (\{3, 4\}, \{f, h\})$, and $(\{1, 5\}, \{d, h\})$. According Theorem 1, we have $B^*(X) = \{6, 7, 8, 3, 4\}$, $B'(X) = \{6, 7, 8\}$, and $BR(X) = \{3, 4\}$.

2) Compute $B^*(X)$ and $B'(X)$ by using Theorem 2

First R_2 convert to single vale set $\{d, e, f\}$ and R_3 convert to single vale set $\{g, h\}$. Besides $X \subseteq B^*(X)$, for the father concept $(\{3, 4\}, \{f, h\})$ of concept $(\{4\}, \{b, f, h\})$, the intent satisfy $f \in \{d, e, f\}$ and $h \in \{g, h\}$, according to Theorem 2, objects 3 and 4 belongs to $B^*(X)$.

Finally, we get $B^*(X) = \{6, 7, 8, 3, 4\}$, $B'(X) = \{6, 7, 8\}$ and $BR(X) = \{3, 4\}$.

By using different theory, we get the same result, and to some extent this proves the correctness of our method.

4. Conclusions

There is an intimate link between rough set theory and concept lattice, so rough set approximation operation can be realized being aided by FCA. From this perspective, realizing rough set operation aided by FCA can be seen as a kind of generalization of the concept lattice queries that meet certain conditions.

Realization of rough set approximation operation on concept lattice can be divided into two steps: first, convert multiple value information system to multiple value formal context, and then covert multiple value formal context to single value formal context; second, realize rough set approximation operation aided by formal concept analysis on concept lattice.

The method is expected to be of further use for calculating approximation of rough set that with real values, interval numbers and other self-defined types.

Acknowledgements

References

[1] Pawlak, Z. (1991) Rough Sets—Theoretical Aspects of Reasoning about Data. Kluwer Academic, Dordrecht.

[2] Ganter, B. and Wille, R. (l999) Formal Concept Analysis: Mathematical Foundation. Springer-Verlag, New York.

[3] Kent, R.E. (1994) Rough Concept Analysis. In: Ziarko, W.P., Ed., *Rough Sets and Fuzzy Sets Knowledge Discovery*, Springer-Verlag, London, 248-255. http://dx.doi.org/10.1007/978-1-4471-3238-7_30

[4] Yao, Y.Y. (2004) Concept Lattices in Rough Set Theory. In: Dick, S., Kurgan, L., Pedrycz, W. and Reformat, M., Eds., *Proceedings of the* 2004 *Annual Meeting of the Noah American Fuzzy Information Processing Society*, Banff, 27-30 June 2004, 796-801.

[5] Qu, K.S., Zhai, Y.H., Liang, J.Y., *et al*. (2007) Representation and Extension of Rough Set Theory Based on Formal Concept Analysis. *Journal of Software*, **18**, 2174-2182. http://dx.doi.org/10.1360/jos182174

[6] Pawlak, Z. (1998) Granularity of Knowledge, Indiscernibility and Rough Sets. *The* 1998 *IEEE International Conference on Fuzzy Systems Proceedings—IEEE World Congress on Computational Intelligence*, Anchorage, 4-9 May 1998, 106-110.

[7] Wu, C., Yue, Y., Li, M., *et al*. (2004) The Rough Set Theory and Applications. *Engineering Computations*, **21**, 488-511. http://dx.doi.org/10.1108/02644400410545092

[8] Scaife, M. and Rogers, Y. (1996) External Cognition: How Do Graphical Representations Work. *International Journal of Human Computer Studies*, **45**, 185-213.

[9] Godin, R., Missaoui, R. and Alaoui, H. (1995) Incremental Concept Formation Algorithms Based on Galois (coNcept) Lattices. *Computational Intelligence*, **11**, 246-267. http://dx.doi.org/10.1111/j.1467-8640.1995.tb00031.x

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.