

Using Rough Set to Construct the Enterprise Information Management System

Xing Wei^{1, a}

¹ Nanjing Normal University Taizhou College, Taizhou, China

^a wx43061068@163.com

Keywords: enterprise management; rough set; information management system

Abstract. Rough set theory as a new hotspot in the field of artificial intelligence, it can effectively deal with incomplete and uncertain knowledge representation and reasoning. Rough set theory is built on the basis of the classification mechanism, it will be classified understand in a particular space on the equivalence relation, equivalence relations constitute the division of space. The paper puts forward using rough set to construct the enterprise information management system. The experiment shows the CPU Time in the attribute numbers, indicating that Jelonek is superior to rough set in building enterprise information management system.

Introduction

As society enters the era of network information, economic integration, the development of any enterprise or department are closely related with the complex knowledge and information. While traditional database management technology to provide access and query capabilities, the face of a wealth of data and poor knowledge "phenomenon, the enterprises have been impossible to stick to the existing knowledge and management rules, and the urgent need to effectively respond to large-scale data. In order to effectively solve the knowledge acquisition this bottleneck problem, researchers proposed the concept of knowledge discovery (KDD). As an important step in the process of knowledge discovery, data mining (DM) to explore the potential pattern in the data to identify business operators may have neglected nugget of knowledge to make forward-looking, knowledge-based decision-making suggestions [1]. Collected data from a real system with noise is not precise enough or incomplete. purely mathematical assumptions to eliminate or avoid this uncertainty, the effect is often not ideal, the other hand, if you look at it, the appropriate handling of such information, often contribute to the solution of the actual system.

A core element in the rough set theory is to study the attribute reduction; attribute reduction can simplify the classification standards, so that people gain a better understanding of the substance of classified. Typically, all the properties in the information system S is not as important as the attribute reduction can be found in a smaller set of attributes BA , makes it possible to use the A Description of a collection of objects can be used B describes the inevitable, thus removing the redundant attributes. Rough set theory, it is a carved incompleteness and uncertainty of the mathematical tools to effectively analyze inaccurate, inconsistent, and incomplete, such as a variety of incomplete information, but also data analysis and reasoning to discover hidden knowledge, to reveal potential law. Rough set theory is built on the basis of the classification mechanism, it understand the classification of the equivalence relation on the specific space, while the equivalent relations constitute the division of space.

Rough set theory knowledge and understanding for the division of the data, each divided into a collection known as the concept of rough set theory main idea is to use known knowledge base, knowledge base known to be imprecise or uncertain knowledge of the (approximate) depicts the theoretical and the other to deal with uncertain and imprecise theory on the most significant difference is that it does not require any a priori information required to handle data collection. The paper puts forward using rough set to construct the enterprise information management system.

The research of rough set in information management

On Rough Set theory, rough sets Pawlak knowledge representation theory, it is a deal with ambiguity and uncertainty of new mathematical tools. In this theory, his first upper approximation, lower approximation, the concept of reduction, nuclear boundary region, the nature of these concepts is to emphasize the reduction of information systems. At the same time, these theories makes the rough set (RS) as the rise in recent years database knowledge discovery (of KDD, known as the theoretical basis of the data mining). This approach for artificial intelligence (AI) and cognitive science are very important, especially in machine learning, knowledge acquisition, decision analysis, to discover knowledge from the database, expert systems, decision support systems, inductive inference, model the field of recognition is more important.

Rough set theory extension of the classical set theory for classification of knowledge embedded in the collection, whether as part of the collection consisting of an object of a set X need to determine the basis of existing knowledge can be divided into three cases: object (a) a certainly belongs to the set X ; (2) object a sure does not belong to set X ; (3) object a may belong or may not belong to the set X the collection division of the close dependence of the domain knowledge at our disposals relative rather than absolute [2]. to a limited set of non-empty set U called the universe, and I in the U family of equivalence relations, namely, knowledge about U , the binary pair $K = (U, I)$ called an approximation space (approximation space).

The rough set theory and the other to deal with uncertain and imprecise theory on the most significant difference is that any a priori information, so the uncertainty description of the problem or deal with can be said that the need to provide the required processing data collection is more objective, this theory can not contain a mechanism to deal with imprecise or uncertain original data, so this theory and probability theory, fuzzy mathematics and theory of evidence to deal with uncertain or imprecise theory are highly complementary. Therefore, the relationship between rough set theory and other theories is also one of the priorities of the Rough Set Theory. Based on rough set theory applied research is mainly concentrated in the attribute reduction, rule acquisition based on rough sets, computational intelligence algorithms based on rough sets, as is shown by equation 1.

$$SNR = 10 \times \log_{10} \frac{\sum_{i=1}^M \sum_{j=1}^N F(x_i, y_j)^2}{\sum_{i=1}^M \sum_{j=1}^N [R(x_i, y_j) - F(x_i, y_j)]^2} \quad (1)$$

Data mining method, rough set theory to support data mining system, a number of steps, because of its easy to understand people's experience of forms of knowledge, but is more suitable for application in the enterprise management decisions, so 20 years in dealing with complex systems has been widely used and developed rapidly. Rough set theory is a new mathematical tools of dealing with vagueness and uncertainty, compared with the past to resolve the ambiguity of knowledge theory and methods, it does not require any outside required to handle data collection before experience, but only based on the classification ability of the observational data to address the inaccuracy of the data analysis and processing, as is shown by equation2.

$$x = \sum_{i=1}^n x_i + \prod_{j=1}^m x_j' \quad (2)$$

Let x be an object in U , X a subset of U , $I(x)$, said the collection of all objects can not be resolved with x , in other words, the equivalence class determined by x , ie I (each object in the x) and x has the same characteristics as attributes (attribute), a collection of approximation of X I (Lower approximation) is defined as: $I^*(X) = \{x \in U: I(x) \subseteq X\}$ $I^*(X)$ is actually the largest collection of objects based on existing knowledge to judge certainly belong to X .

From the above definition, can be seen in rough set theory "vague" (uncertainty) (on Vague) and "uncertain" relationship between these two concepts: "ambiguous" to describe the collection refers to collection of boundary; and "uncertainty" to describe the elements in the collection, refers to whether an element belongs to a collection is uncertain is presented by one figure as equation 3.

$$\hat{\gamma}^{(\beta)}(m|m-1) = \Phi_w(m-1)\hat{X}^{(\beta)}(m-1, M) \quad (3)$$

Upper approximation $\text{apr}(X)$ that contains X , the smallest definable set in the subsystem $\sigma(U/R)$. With the definition of lower and upper approximations, it is easy to get the other definition of the concept in rough set theory, such as the boundary region, the positive domain and negative domain. Three is defined as the combination of the other theories extended rough set model to create a link. 3.2 Let $R \subseteq U \times U$ is an arbitrary binary relation on the domain of element-based extension of model, defines an extension of the approximation space $\text{apr} = (U, R)$. From the collection point of view, the use of non-equivalence relation obviously can be extended rough set-based elements defined in 3.1. For example, where the equivalence class of elements $x [x]_R$ as a neighborhood of x is obtained based on neighborhood rough set model. Neighborhood relations $R_s(x)$ is only required to satisfy the reflexive, not required to meet the symmetry or transitivity. Then the element-based definition of the equivalence class $[x]_R$ $R_s(x)$ non-equivalence relations instead of on the rough set model based on the non-equivalence relation $R_s(x)$.

For attribute reduction of information system, you can use the basic algorithm or approximation algorithm. The basic algorithm to construct the discernibility matrix, and then to distinguish between matrix derived based on the discernibility function, after applying the absorption law can be simplified, making it the disjunctive normal form, then each master contains the style of the original system a reduction. Example: with an enterprise information system $S = (U, A)$, of which $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ represents the set of objects, $A = \{a, b, c, d\}$ represents the property set database are using the basic algorithm for reduction: first construct discernibility matrix of information systems, and then gives the discernibility function, as is shown by equation 4.

$$b = \ln(B_y / B_x); b \in (-\infty, +\infty) \quad (4)$$

Rough set theory is a very practical subject, from birth to now, although only 10 years, but in many fields has achieved fruitful results, such as approximate reasoning, digital logic analysis and simplification, the establishment of predictable models, decision support, control algorithm acquisition, machine learning algorithms and pattern recognition, and so on [3].

Then go to the binary relation relations generated by the binary approximation space by the constructive method to export a collection of approximation operators exactly is defined by the axiomatic method given operator. Special axiom of approximation operators guarantee the existence of some special type of binary relation, so that these relations by the constructor to produce a given operator; the contrary, the binary relation derived constructor approximation operators must meet certain axioms, these axioms by the algebraic method to produce a given binary relation. The rough set axiomatic method the first Lin.

Rough set theory, mathematical tools of probability theory, fuzzy sets, evidence theory, after a processing of uncertainty. Are to varying degrees in many practical systems, there are uncertainties; the collected data often contain noise, inaccurate or incomplete, as is shown by figure 1.

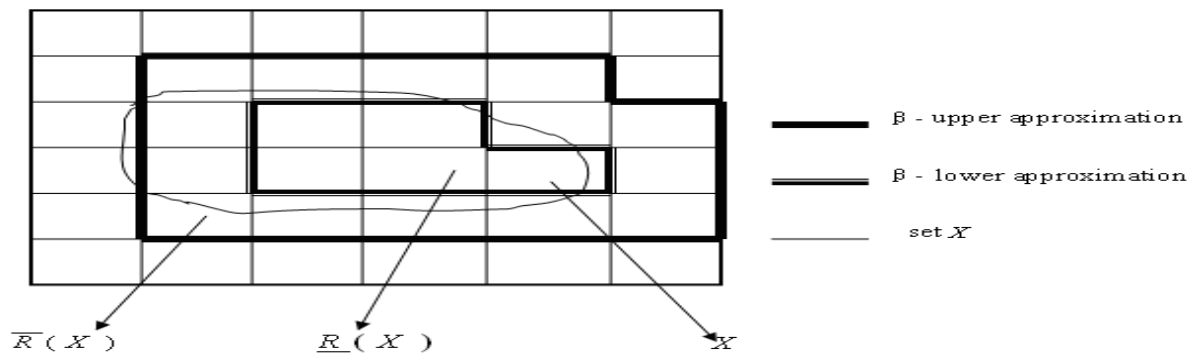


Figure. 1 Sketch map of variable precision rough set

Rough sets and fuzzy sets can handle incomplete (imperfect) data, but in different ways, focus on describing the information ambiguity (vagueness) the degree of fuzzy sets, rough sets is emphasized that the data can not discern, imprecise (imprecision) and ambiguous. language as metaphor in the image processing, when discourse image clarity, rough Sets emphasis on composition of the image pixel size, while the fuzzy sets are stressed pixels of different gray rough set research the relationship between a collection of different classes of objects, focusing on classification; fuzzy set is the study of the relationship of the different objects belonging to the same class of membership, focuses on the degree of membership. Rough sets and fuzzy sets are two different theories, but not against each other, they can complement each other in dealing with imperfect data, as is shown by equation5.

$$\hat{\gamma}^{(n)}(m,s) = \hat{\gamma}^{(n)}(m,s-1) + \overline{K}^{(n)}(m,s)[z(m,s) - \Psi_w(m,s)\hat{\gamma}^{(n)}(m,s-1)] \quad (5)$$

Review the basic concepts of rough sets based on rough set approach to description of two studies: the constructor method and the axiomatic method. And from the collection point of view and the operator point of view to explain the rough set. Concept, define the set of concepts, in order to describe the knowledge; you first need to know what is the concept. From the classic point of view, each concept contains its connotation and denotation. In order to give a specific description of the concept of connotation and denotation, we consider a simple knowledge representation system, that the information table. The information table is a group of objects, a collection of objects through a set of attributes to describe.

Text to start from the basis of rough set theory, the author of two parts: (1) starting from the identification matrix Skowron can, through discussion of the dependence between attribute importance and attribute, the use of conditional entropy to construct a consistent data the heuristic algorithm of attribute reduction; instance and UCI database proved the effectiveness of the algorithm; and error correction.(2) starting from the variable domain of the point of view, the improved discernibility matrix can not differentiate between class definition, combined with the discussion of attribute importance and attribute inter-dependence, we construct a variable domain of the heuristic algorithm and instance and UCI database proved the effectiveness of the algorithm.

The development of enterprise information management system based on rough set

Effective reduction method can simplify the enterprise information system to ease the pressure on the enterprise information management, decision analysis. In fact, in the face of massive high-dimensional data size, more effective and efficient attribute reduction algorithm for the extraction of knowledge. The idea of an intelligent information systems to business users, has a certain practical significance [4]. Rough set is a more promising method of dealing with uncertainty, I believe that will be applied in more fields, rough set theory is still in the continued development of, and as the founder of the Rough Set Theory ZPaw and lak pointed out, there are some theoretical problems to be solved, such as for inexact reasoning rough logic (rough logic) method, rough set theory and nonstandard analysis (Non standard analysis) and non-parametric statistics between relations.

Similar description to help up and down can also be given and PRS1 equivalent on the Rough Sets. Another definition, known as the P2-rough set collection. That is, $\text{Def}(U, L(A)) = \sigma(U / E(A))$ ordered pair $\text{apr} = (U, E(A))$ is called a Pawlak approximation space, referred to as the approximation space. Therefore, the language $L(A)$ of all define the denoted, $\text{Def}(U, L(A)) = \text{Def}(\text{apr}).U / E(A)$, can construct a σ algebra, that is, σ (the $U / E(A)$), which contains the empty set \varnothing and the equivalence relation $E(A)$ consisting of equivalence classes, and pay and complement operators is closed, as is shown by equation6.

$$\hat{\gamma}^{(\eta)}(m,s) = \hat{\gamma}^{(\eta)}(m,s-1) + \bar{K}^{(\eta)}(m,s)[z(m,s) - \Psi_w(m,s)\hat{\gamma}^{(\eta)}(m,s-1)] \quad (6)$$

Attribute reduction algorithm has been some success, however, not yet a recognized and efficient reduction algorithm. Through the study of some of the existing reduction algorithm, the paper presents a reduction algorithm based on the improvement of the attribute frequency. The algorithm is a heuristic algorithm, built in the discernibility matrix calculation on the basis of. Improved algorithm is based on Hu's algorithm Jelonek algorithm on the basis of calculation of the discernibility matrix, to ensure that the final be able to find the decision-making information systems a reduction, compared Jelonek algorithm, the computation time significantly reduced. Data mining based on rough set theory is in its infancy, data mining and rough set theory research, there are many problems worth exploring, the paper will be a combination of both study certainly there are many imperfections, related work remains to be further study. Our system runs on a Windows XP-based PC with an Intel Pentium 5 CPU at 3.8 GHz. The system is developed using Java on Java SDK 1.4.2 platform, as is shown by figure2.

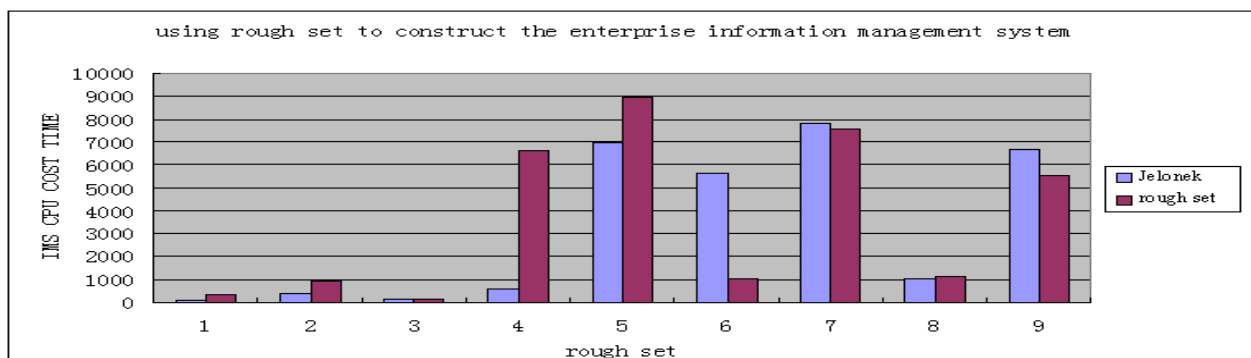


Figure. 2 Development of enterprise information management system compared result of Jelonek and rough set

In this model, we are given with a degree of membership and the weight of the information system, the processing of the noise, the division of the approximation space, the dependence of the decision attribute on condition attribute, attribute reduction, the association rule mining steps the establishment of research, and the numerical example of the model is feasible. The extended model of rough set to overcome the classical rough set classification is too strict, too sensitive to noise, some of the hidden rules in the border loss defects.

Summary

The data is the processing center of the core business of most enterprises, a variety of external sources of data within the system to a large increase. In reality, attribute reduction, information systems for the mitigation of large-scale data sets for enterprise management and decision-making pressure, with a certain degree of practical significance. The paper puts forward using rough set to construct the enterprise information management system.

References

- [1] Liu Jiming, "A New Plant Leaf Classification Method based on Neighborhood Rough Set", AISS, Vol. 4, No. 1, pp. 116 ~ 123, 2012.
- [2] Qiming Li, Yimin Chen, Yunhua Zhang, Yan Liu, Dongyang Zhao, "Grasp Gesture Recognition Based on Rough Sets Theory in AR System", JDCTA, Vol. 5, No. 1, pp. 226 ~ 234, 2011
- [3] Gao Ren, Tang Long, Wu Juebo, "A Novel Recommender System Based on Fuzzy Set and Rough Set Theory", AISS, Vol. 3, No. 4, pp. 100 ~ 109, 2011.
- [4] Xiaohua Liu, "Fault Diagnosis of Fan based on Improved SSVM and Rough Set", IJACT, Vol. 3, No. 6, pp. 329 ~ 336, 2011.