Engineering Application Way of Faults Knowledge Discovery Based on Rough Set Theory

Zhao Rongzhen¹, Li Chao², Deng Linfeng¹
¹ Key Laboratory of Digital Manufacturing Technology & Application of the Ministry of Education, School of Mechanical and Electronic Engineering, Lanzhou Univ.of Tech. Lanzhou, 730050 China
² College of Petrochemical Technology Lanzhou Univ.of Tech. Lanzhou, 730050 China

E-mail: zhaorongzhen@lut.cn

Abstract. For the knowledge acquisition puzzle of intelligence decision-making technology in mechanical industry, to use the Rough Set Theory (RST) as a kind of tool to solve the puzzle was researched. And the way to realize the knowledge discovery in engineering application is explored. A case extracting out the knowledge rules from a concise data table shows out some important information. It is that the knowledge discovery similar to the mechanical faults diagnosis is an item of complicated system engineering project. In where, first of all-important tasks is to preserve the faults knowledge into a table with data mode. And the data must be derived from the plant site and should also be as concise as possible. On the basis of the faults knowledge data obtained so, the methods and algorithms to process the data and extract the knowledge rules from them by means of RST can be processed only. The conclusion is that the faults knowledge discovery by the way is a process of rising upward. But to develop the advanced faults diagnosis technology by the way is a large-scale knowledge engineering project for long time. Every step in which should be designed seriously according to the tool’s demands firstly. This is the basic guarantees to make the knowledge rules obtained have the values of engineering application and the studies have scientific significance. So, a general framework is designed for engineering application to go along the route developing the faults knowledge discovery technology. Keywords: Rough Set Theory; Data classification; Knowledge discovery; Data Protection; Mechanical fault diagnosis.

1. Introduction
The information technology is playing the very important roles in machinery industry now [1-2]. One of typical examples is the device of condition monitoring and faults diagnosis in large-scale rotating machinery [3-4]. It is a set of information system to protect rotating machinery safety running. To improve the quality of its faults decision-making, the many methods of artificial intelligence have been set continuously into the device during last two decades [3,5-6]. And the various shortcomings in them were being overcome by try at the same time. As we know, the capacities of machine intelligence like experts system depend completely on the knowledge rules in their knowledge bases. Whatever kinds of information devices, the people especially wish that the rules in their knowledge base should be better and the base is better to have almost all fully knowledge to its universe. And the
device should be also able to make automatically its intelligence level go up a new advanced degree continuously by machine learning manner. Unfortunately, the knowledge acquisition in intelligence devices has been a nasty puzzle for so long [3]. The reasons resulted in the defects are that not only is expert’s resource very shortage, but also the representations of professional knowledge derived from domain’s experts carry also very fuzziness and uncertainty and so on [4]. Since to break down the defects has been difficult so far now, as a matter of fact they are severely blocking the development in information technologies of machinery industry. So, the new way to advance the machine intelligence level of machinery equipment by means of some new information theories should been explored further.

Up to now, for the goal to solve the knowledge acquisition puzzle in faults diagnosis of rotating machinery, the comprehensive utilizatons of machine’s running information, especially the some abnormal information in its running history, have started to attract the attentions of faults diagnosis domain. It was considered as a right and challenging direction on developing the faults diagnosis technologies [1,3-4]. One of the expected goals is that the device to implement the faults decision-making should be able to upgrade its performances into a new advanced level by automatic machine learning like human being. Obviously, the artificial intelligence will yet play the important role for the prospects. Two kinds of new information technologies originated in data mapping in artificial intelligence and developing to the layer of data driven technology currently, which are respectively the Data Mining (DM) and the Knowledge Discovery in Database (KDD) [2], are considered as to hold the potential power to solve the puzzle of knowledge acquisitions. The researches to them begin to take a mainstream tendency in artificial intelligence domain now [2,6-9].

Both DM and KDD are cognizant of the expansive prospect to develop the machine intelligence by data processing [2]. On the point of the tools used to implement two items of tasks, RST created by Z. Pawlak [10-11] provides an effective new tool by [12]. But to put the above things into to solve the knowledge acquisition puzzle like the faults diagnosis of rotating machinery, the some questions will be expected to study as fast as possible [2,6,8]. So, for the advanced degree to develop the faults diagnosis technology, the something carried to use RST as a kind of tools of faults knowledge discovery in engineering applications will be probed into in this paper.

The paper was organized as follows. The next section presents a few of basic notions in RST. The relation between the tool’s demands for data and the knowledge processing in faults knowledge discovery will be concerned about in third section. Subsequently, the way to implement the faults knowledge discovery will be explored. Then the next section, aimed to the generalization of the faults response characteristics of large-scale rotating machinery, the way of engineering applications with regard to the faults knowledge discovery will make out by. The conclusions will be given obtained at the last.

2. The some conceptions in RST

The RST is a kind of mathematics tool specialized to process the data set with symbols [10-11]. According to its concepts in [12], the tool faces with a knowledge representation system, i.e. an information system. It can be also described as a pair of $S= (U, A)$. In where, $U$ named the universe denotes a set of a non-empty and finite objects; $A$ is a set of non-empty attributes which it is use to describe the status of all objects in a universe. The tool’s functions on mathematics meaning can be summed up simply as follows. It is its main capabilities that to solve the classifications of a set of symbols data and to implement a reasoning decision-making to its data subset on the same as universe. To implement the classifications to a special universe with the tool can be described as follows. In a general way, due to have just the limited available information to the universe can be used, and some objects in the universe based on observation may be indiscernible, so the RST takes for granted that the classification is come true. Secondly, the concepts with regard to reasoning decision-making can be found in [11]. A couple of approximate concepts to a subset of all objects are used to deduce the decision-making as precisely as possible, and the boundary area between them shows out some uncertainty with regard to decision-making results.
The above concepts in RST reflect the follow messages. Namely in nature, to use RST as a kind of analysis tools of symbols data set to implement the knowledge discovery, the work must orderly include the two items of tasks [4].

- First, to implement a kind of classification to \( U \). If every class represents a knowledge granule in \( U \), then all granules in \( U \) build into one knowledge base of \( U \). Distinctly, the base consists of by a kind of granular structure. While every one of all concepts in the base denotes a knowledge granule essentially.
- According as the obtained knowledge granules, a reasoning decision-making with some indeterminate degree to a subset \( X \subseteq U \) can be deduced from orderly. Without doubt, the results obtained have often some biased relative to their primitive situations.

The above conclusions have been obtained by in our early research [4]. An important result from them had found out. It is that both the mode of data’s expressions with regard to \( U \) and the rules to separate \( U \) into knowledge granules are two key factors to solve the classifications. But the reasoning decision-making with indeterminate must relied on the classification’s results to deduce. Since both the above issues and the concepts of knowledge discovery in database have very consanguineous relations [6,8-9], the junction between the classifications and the knowledge processing in next sections will be analyzed via a case.

3. A case of knowledge discovery derived from the RST

3.1. A table with symbols data

Table 1 is a designed data set with discrete symbols. It will be used to explore the mode implementing classifications and knowledge discovery to \( U \). Where, \( U \) presents a set of all objects in a specific universe. It can be any objects of the studied problems, for example a group of objects or a set of concepts as and so. Attributes set \( A \) is used to describe the status of all objects in a universe \( U \). It is consisted of two parts. They are two subsets of \( C=\{a_1,a_2,a_3\} \) named condition attributes set and \( D=\{d\} \) of decision attributes set. Those symbols data in the table’s centric area forms an attributes value domain. In the value domain the data are denoting some distribution status of objects in \( U \) by the expression of attributes values.

| Universe \( U \) | Attributes set \( A \) |
|------------------|------------------------|
|                  | Condition Attributes set \( C \) | Decision Attributes set \( D \) |
|                  | \( a_1 \) | \( a_2 \) | \( a_3 \) | \( d \) |
| \( x_1 \)       | 2       | 1       | 3       | 1       |
| \( x_2 \)       | 3       | 2       | 1       | 2       |
| \( x_3 \)       | 2       | 1       | 3       | 1       |
| \( x_4 \)       | 2       | 2       | 3       | 2       |
| \( x_5 \)       | 1       | 1       | 4       | 3       |
| \( x_6 \)       | 1       | 1       | 2       | 3       |
| \( x_7 \)       | 3       | 2       | 1       | 2       |
| \( x_8 \)       | 1       | 1       | 4       | 3       |
| \( x_9 \)       | 2       | 1       | 3       | 1       |
| \( x_{10} \)    | 3       | 2       | 1       | 2       |
| \( x_{11} \)    | 2       | 2       | 1       | 1       |
| \( x_{12} \)    | 1       | 3       | 2       | 1       |

According to the equivalence relations, or the equivalence classes concepts created in RST [11], implement a kind of operation processing to Table 1. Both the classes and the knowledge granules obtained was show in Table 2.
3.2. A kind of classifications results

To implement the classifications to the symbols data in Table 1, Table 2 show one result. Distinctly, at the transform from Table 1 to Table 2, the objects with the same expression are gathered into one class.

**Table 2.** The result obtained by classifications

| Universe | Attributes set A | Condition Attributes Subset | Decision-making Subset |
|----------|-----------------|-----------------------------|------------------------|
| **C**    | **a₁** | **a₂** | **a₃** | **d** |
| **a₁**   | 2     | 1     | 3     | 1    |
| **a₂**   | 3     | 2     | 1     | 2    |
| **a₃**   | 1     | 1     | 4     | 3    |
| **a₄**   | 2     | 2     | 1     | 1    |
| **a₅**   | 1     | 3     | 2     | 1    |

To look into and compare of the two tables, then the other concept implied in Table 2 can be also described as follows:

- **Class 1:** \( X₁=\{x₁, x₃, x₄, x₉\} \);
- **Class 2:** \( X₂=\{x₂, x₇, x₁₀\} \);
- **Class 3:** \( X₃=\{x₅, x₆, x₈\} \);
- **Class 4:** \( X₄=\{x₁₁\} \);
- **Class 5:** \( X₅=\{x₁₂\} \).

If the above points are associated with KDD and DM in artificial intelligence technology sequentially [2,9], Table 2 reflected out a structure of knowledge granules or a kind of data patterns contained in the classification of Table 1 yet. In where, meaning of the symbols in Table 2 is the same as in Table 1. The table shows that the original data in Table 1 was partitioned into five classes, i.e. \( U=\{X₁, X₂, X₃, X₄, X₅\} \). But by the operation, the some objects with the same attributives values are gathered inside one class. In other word, the objects in every class have the same characterizations and every class has the same description on attributes values.

3.3. A group of knowledge rules obtained

In accordance with fundamental conceptions of knowledge engineering, a set of knowledge granules implied in Table 2 can be pick up automatically. According to the scheme of knowledge description in knowledge engineering, which is like “if Condition then Decision”, a group of knowledge rules extracted by the data analysis from Table 1 can also be described as:

- **Rule 1:** if \((a₁, 2)\land(a₂, 1)\land(a₃, 3)\) then \((d, 1)\).
- **Rule 2:** if \((a₁, 3)\land(a₂, 2)\land(a₃, 1)\) then \((d, 2)\).
- **Rule 3:** if \((a₁, 1)\land(a₂, 1)\land(a₃, 4)\) then \((d, 3)\).
- **Rule 4:** if \((a₁, 2)\land(a₂, 2)\land(a₃, 1)\) then \((d, 1)\).
- **Rule 5:** if \((a₁, 1)\land(a₂, 3)\land(a₃, 2)\) then \((d, 1)\).

Where, the operator “\(\land\)” refers to the “And” relation in set algebra. From the perspective of engineering application of KDD, obviously, Rule1~Rule5 builds up a base of knowledge rules. And they are denoting the five knowledge granules in Table 2. On the points of view from knowledge discovery, in fact the above process implements distinctly once knowledge discovery to Table 1.

4. Explore the way of engineering application

4.1. Characteristics in Table 1
Based on the classifications concepts of RST, a case to implement once knowledge discovery to a data set had been provided in sections 3. Some conclusions from them can be summarized out. The first is that both the difficult degree of knowledge discovery and the application values of the discovered rules depend entirely on the data structure and the symbols status as well as their distribution in Table 1. Without doubt, if from the perspective of easy to realize the knowledge discovery to observe Table 1, the attribute’s number in attributes set \( A \) should be fewer. And the symbols number in attributes value domain should also be the same as fewer properly. This can guarantee that to make the scale of the extracted knowledge rules be as far as possible a few less and the rules be ease of use as we expect it [4]. Otherwise, the discovered rules may is too small and no application value. So, like Table 1, the data to implement knowledge discovery must be a set of data as simple as possible.

Secondly, our goal is to realize the knowledge discovery by analyzing data so as to solve the puzzle of knowledge acquisition. And looking forward to the discovered rules can reach an expert’s high in professional fields as far as possible. But from the transformation relationships inside Table 1 and Table 2 as well as the already acquired knowledge rules, it can get that the so-called knowledge discovery is actually the results for some small knowledge points to be merged into the large knowledge particles by RST as a kind of tool to process a set of data. So, to develop the knowledge discovery technology of professional fields, a concise data table should not only be obtained in advance, and the data in table should also contain something to look forward to extract the knowledge rules. This gives rise to important new problems that is both how to describe the professional knowledge and how to preserve properly the knowledge into a table like Table 1 with data mode.

In addition to the above two issues besides, to the specific knowledge discovery of professional fields, the other concepts involved in Table 1 must also solved beforehand. For example the definitions of a universe \( U \) and the symbols in attributes values domain of Table 1 etc. Apparently, each knowledge discovery task on engineering application will produced a serious of new problems waiting to be solved by.

4.2. Discussion about engineering applications

The knowledge discovery based on RST needs a scientific datasheet to satisfy the demands of RST as the tool of knowledge discovery. In which, the expected knowledge described with data mode must contains among them. The information system in Table 1 shows out that both the attribute’s number in attributes set \( A \) and the symbols in attributes’ values domain should be all as few as possible. Otherwise, the knowledge discovery may will failure. So, the engineering application with regard to knowledge discovery based on RST put forward a new challenge to us.

Here, to solve the knowledge acquisition puzzle in faults diagnosis of rotating machinery by knowledge discovery will be used as an example, and the implementation framework to achieve the expected goals will be discussed. As we know, the rotating machinery is a class of very important equipment in chemical industry and electric power etc. Because they are all essentially nonlinear dynamic systems and the parameters in running environment are very dynamic and random changes, the responses of different faults at runtime presents the strong variability. More serious is that even the same kinds of fault, its responses in different time are yet polymorphism. So, to describe accurately the faults state with data mode is not ease tasks. That it is a very difficult problem about fault’s quantification description has been quite a lot of research confirmed.

4.3. Way of Engineering application with regard to faults knowledge discovery

To acquire the knowledge rules using data as a kind of knowledge resource has been proved to be feasible in concepts and methods. But for along the direction to realize a sure knowledge discovery task, the other preconditions also should be satisfied. Obviously, one of the most important things is to delimit a universe category specifically to the task. It should be a set of some concepts or a group of objects with same features and so on. Second, to have a proper attributes set as succinctly as possible is necessary in order to easy to implement the knowledge discovery. It should have the concise
characteristics as like attributes set $A$ of Table 1. This is the preconditions for security to make data classification easily realize. Third, the symbols number in data expressions of objects in universe $U$ is expected as far as possible a few less. This is the basic conditions to make the knowledge described with data can be gathered into classifications. In short, the above items aimed to a determined universe concept should be considered as a priority.

A few of opinions from the above discussion can be obtained. It is that the engineering application of knowledge discovery based on RST should be implemented on a concrete universe, for example the faults diagnosis of rotating machinery. The first work must be preserved the faults knowledge with data into an information system. On the basis of to get the original and massive faults data, both methods and algorithms study with regard to faults knowledge discovery can be carried out widely. So, to the faults knowledge discovery in faults diagnosis of rotating machinery, a uniform framework and its implementing steps can be planned as two parts from Figure 1 to Figure 2.

Figure 1. A framework to save the original faults diagnosis knowledge with scientific pattern

Figure 2. A scientific model to process the professional knowledge in mechanicals engineering domain based on RST

Figure 1 shows a process to implement the scientific protection for the faults knowledge expected. It includes both to extract the quantitative features of faults status according to the fault diagnosis research ideas and how to describe the faults decision-making with a group of knowledge vector by knowledge engineering concepts. The target of the process is to expect a kind of desire data resource with the faults knowledge obtained. Figure 2 provides a path to implement the faults knowledge discovery. The path started from an original information system with the expected knowledge. And the intermediate needs to experience the data classification such as. To obtain a knowledge rules base that can be used to instruct the engineering applications was considered as to work a phase of the project.
From engineering application point of view, both the data processing model and the framework in Figure 1 and Figure 2 indicate a new way to solve the faults knowledge acquisition puzzle by data analysis.

5. Discussions

Obviously, the way indicates a developing direction for intelligence decision-making technology in faults diagnosis of rotating machinery. But because the features of faults information carry generally a considerable degree of uncertainty, so in practice to save the faults knowledge acquisition puzzle with data is a hard problem like the bottleneck. This shows that the faults knowledge discovery on engineering application is not only a system engineering projects, but it is a very difficult task waiting to solve. So, the knowledge discovery based on RST put forward the new researches tasks to us. Both the scientific protections of faults knowledge and to make them form a kind of precious data resources have been given a boost to a new high degree. Since it is the basis guarantees to makes the obtained rules have the application value. The new demands put forth new tasks by.

6. Conclusions

RST provides an effective new tool for us to solve the puzzle of knowledge acquisition in engineering domain. It is proved that the tool can solve the puzzles of faults knowledge acquisition in faults diagnosis of rotating machinery. But it results in some new basic scientific problems waiting to be solved. They are that the faults knowledge must be protected with data mode properly and the methods and the algorithms with regard to faults knowledge discovery must be processed based on the effective faults data resource from plants etc. So, a framework implementing a faults knowledge discovery in engineering application is designed out by. Both the way and steps implemented in the faults knowledge discovery in faults diagnosis of rotating machinery are planned out. But if to go along the way, some requirements emerged put forward a new challenge to us.

Acknowledgments

This work was supported by a grant from the National Natural Science Foundation of China (No. 50875118) and the found of postgraduate supervisor of the Gansu Education Department (0903-11) as well as the project of basic scientific research and operating expenses for universities in Gansu Province ‘Testing dynamic information and intelligent fault decision technology of complex mechanical system’.

References

[1] Stephan Ebersbach, Zhongxiao Peng. Expert system development for vibration analysis in machine condition monitoring [J]. Expert Systems with Applications, 34 (2008) 291-299.
[2] E. Triantaphyllou, T. W. Liao, S.S. Iyengar. A focused issue on data mining and knowledge discovery in industrial engineering [J]. Computers & Industrial Engineering, 43 (2002) 657-659. Preface.
[3] Venkatasubramanian, Venkat; Rengaswamy, Raghunathan; Kavuri Surya N.; et al. A review of process fault detection and diagnosis: Part III: Process history based methods [J]. Computers and Chemical Engineering, 2003, 27(3): 327-346.
[4] Zhao Rongzhen. Investigation of elementary problems on diagnosis knowledge discovery of a rotor-bearings system with double spans based on theory of rough sets [D]. Xi’an: Xi’an Jiaotong University, 2006.1.
[5] Shusaku Tsumoto. Mining diagnostic rules from clinical databases using rough sets and medical diagnostic model [J]. Information Sciences, 162 (2004) 65-80.
[6] Chao-Ton Su, Long-Sheng Chen, Yuebwern Yih. Knowledge acquisition through information granulation for imbalanced data [J]. Expert Systems with Applications, 31 (2006) 531-541.
[7] Yuhua Qian, Jiye Liang, Deyu Li, et al. Measures for evaluating the decision performance of a decision table in rough set theory [J]. Information Sciences, 178 (2008) 181-202.
[8] Geng, Z., Zhu, Q. Rough set-based heuristic hybrid recognizer and its application in fault diagnosis [J]. Expert Systems with Applications (2008), doi:10.1016/j.eswa.2008.01.020.
[9] Yang Bingru. Five important issues in the present field of knowledge discovery [J]. Engineering Science (Chinese), 2008, 11(4):76-83.
[10] Z. Pawlak, A. Skowron. Rough sets: Some Extensions [J]. Information Sciences, 177(2007)28-40.
[11] Z. Pawlak, A. Skowron. Rudiments of rough sets [J]. Information Sciences, 177 (2007) 3-27.
[12] K. Thangavel, A. Pethalakshmi, Dimensionality reduction based on rough set theory: A review [J]. Appl. Soft Comput. J. (2008), doi:10.1016/j.asoc.2008.05.006.