An Integrated Methodology of Rough Set Theory and Grey System for Extracting Decision Rules

Hossam A. Nabwey[1,2] and Mahdy S. El-Paoumy[1,3]

[1]*Department of Mathematics, Faculty of Science, Salman Bin Abdulaziz University, Al-Kharj, Saudi Arabia*
[2]*Department of Engineering Basic science, Faculty of Engineering, Menofia University, Menofia, Egypt*
[3]*Department of Mathematics, Faculty of Commerce, Alazhar University, Egypt*

Abstract

Grey system theory and rough set theory are two different mathematical tools that are used to deal with uncertain or incomplete information, and yet they are relevant and complementary to a certain degree. The appropriate hybrid of the two theories can overcome the shortages of their definitions and applications and thus has more powerful functions. This paper proposes An Integrated Methodology that extracting decision rules based on combining grey system and rough set theory. The effectiveness of the proposed methodology was verified by application of this methodology to discover grade rules of electrical transformer evaluation.

Keywords: *Rule Induction, data mining, knowledge discovery in database, Knowledge Acquisition, rough set theory, Grey System, power transformer evaluation*

1. Introduction

The ability of acquiring decision rules from empirical data or the environment is an important requirement for both natural and artificial organisms. For example, in an intelligent system, decision rules can be extracted by performing inductive learning [1]. Many techniques such as decision tree learning [2], neural network learning [3], and genetic algorithm-based learning [4], have been developed to carry out such a task. With the rapid advent of IT technology, it has often been said that we live in the "information age". As a result, the ability to automatically and efficiently extract knowledge from the huge amount of raw data has become an important research area. In reality, the raw data gleaned from a specific environment may contain uncertainty, that is, the data may be imprecise or incomplete. Imprecise data refer to information that is fuzzy or even conflicting. For example, the opinions about the performance of a machine as assessed by two engineers can be different. This will introduce inconsistency in the knowledge concerning the performance of the machine. On the other hand, incomplete data refer to missing data in the data records and may be caused by the unavailability of equipment or oversight of operators. This imprecise and incomplete nature of raw data is obviously the greatest obstacle to the task of rule extraction. Over the past decades, many theories and techniques have been developed to deal with uncertainty in rule induction, for example, fuzzy set theory [5] and the Dempster-Shafer theory of belief functions.

Rough set theory, which was introduced by Pawlak (1982) in the early 80's, provides a novel and powerful way of dealing with vagueness and uncertainty. It focuses on the discovery of patterns in imprecise data and can be used as a basis to perform formal reasoning, machine learning and rule discovery [6]. In less than two decades, rough set theory has rapidly established itself in many real-life applications such as medical diagnosis, control algorithm acquisition and process control, and information retrieval [7].

Compared with other theories in dealing with uncertainty, RST has some irreplaceable advantages, but still has some one-sidedness and shortcomings. For example, it greatly reduces the new predictive capability due to the over-fitting for data and cannot deal with the classification of multi-attribute decision-making; the border region that the RS depicts is relatively simple, such as equivalence relations based on the classification of RS is certain, while there is not belonging or containing to a certain extent; it can identify the random rule only supported by a few examples; it has no appropriate approach to deal with the raw data which are ambiguity. According to probability statistics made by some scholars, each method has its own scope of application, and there is no method that can solve all the problems [8].

Grey system theory was founded by Professor Deng Julong, a well-known Chinese scholar, in 1982; the grey system theory introduced is described as a mathematical tool to handle small sample and poor information. The systems with partially known and partially unknown information are viewed as grey systems, and valuable information is extracted or produced by mining the partially known information.

RST and grey system theory have a certain degree of relevance and complementarily. Both of them make it possible to find a model from the data made fuzzy by too many details through reducing the precision of data expression. At the same time, neither of them requires any transcendental knowledge. RST deals with no overlap of the sort of roughness and the rough concept and focuses on the indiscernibility between objects, while grey system theory deals with the grey fuzzy sets that are “explicit extension but ambiguous intension” and emphasizes the uncertainty of poor information. The rough membership degree of RSs is directly acquired by calculating from the data analyzed, and it can depict objectively the uncertainty of the data.

In practical application, we often combine several techniques to construct a “hybrid” approach [9], so that it could complement each other to overcome the limitations of individual techniques, avoiding the existing method’s disadvantages or weaknesses when used separately; such a hybrid system is superior to the use of a single method, so in this paper a hybrid model of RS and grey system have been discussed and it was used to discover grade rules of electrical transformer evaluation.

2. Basic Concepts of Rough Sets Theory

The rough sets theory was proposed by Pawlak (1982) as a novel and powerful mathematical tool for reasoning about imprecision, vagueness and uncertainty [10]. The philosophy of rough sets theory is based on the idea of classification. The ability to classify is a fundamental feature of any living organism, a robot or an agent, which, in order to behave rationally in the external world, must constantly classify concrete or abstract objects such as entities, events, processes, and signals. In order to do so, one has to ignore minor differences between objects, thus forming classes of objects that are not noticeably different. These indiscernible classes can be viewed as elementary concepts used by an agent to build up its knowledge about reality. Elementary concepts can be combined into compound concepts that are uniquely defined in terms of elementary concepts. Any union of elementary sets is called a crisp {or precise) set. However, the granularity of knowledge results in situations in which some notions cannot be expressed precisely within the available knowledge and can be defined only approximately. Such sets are referred to as rough {vague, imprecise).

2.1. Information System and Indiscernibility Relation

For algorithmic reasons, the information regarding the objects is supplied in the form of a data table, whose separate rows refer to distinct objects (actions), and whose columns refer to different attributes considered. Each cell of this table indicates an evaluation (quantitative or qualitative) of the object placed in that row by means of the attribute in the corresponding column.

Formally, a data table is the 4-tuple $S = \langle U, Q, V, f \rangle$, where U a finite set of objects (universe), $Q = \{q_1, q_2, q_3 \ldots\ldots q_m\}$ is a finite set of attributes, Vq is the domain of the attribute q, $V = \bigcup_{q \in Q} V_q$ and $f : \mathrm{UxQ} \rightarrow \mathrm{V}$ is a total function such that $f(x,q) \in \mathrm{V_q}$ for each $q \in \mathrm{Q}, x \in U$, called information function.

Therefore, each object x of U is described by a vector (string) $Des_q(x) = [f(x,q_1), f(x,q_2), \ldots . f(x,q_m)]$ called description of x in terms of the evaluations of the attributes from Q; it represents the available information about x.

To every (non-empty) subset of attributes P is associated an indiscernibility relation on U, denoted by IP:

$$I_p = \{(\mathrm{x,y}) \in UxU : f(x,q) = f(y,q) \forall q \in P\} \quad (1)$$

If $(x,y) \in I_p$, it is said that the objects x and y are P-indiscernible. Clearly, the indiscernibility relation thus defined is an equivalence relation (reflexive, symmetric and transitive). The family of all the equivalence classes of the relation I_p is denoted by U / I_p and the equivalence class containing an element $x \in U$ by $I_p(x)$. The equivalence classes of the relation I_p are called P-elementary sets. If P = Q, the Q-elementary sets are called atoms.

2.2. Approximations

Let S be a data table, X a non-empty subset of U and $\phi \neq P \subseteq Q$. The P-lower approximation and the P-upper approximation of X in S are defined, respectively, by:

$$\underline{P}(X) = \{x \in U : I_P(x) \subseteq X\} \quad (2),$$

$$\overline{P}(X) = \bigcup_{x \in X} I_P(X) \quad (3)$$

The following ratio defines an accuracy of the approximation of X, $X \neq \phi$, by means of the attributes from P :

$$\alpha_P(X) = \frac{|\underline{P}(X)|}{|\overline{P}(X)|} \quad (4)$$

Where $|Y|$ indicates the cardinality of a (finite) set Y.

Obviously, $0 \le \alpha_P(X) \le 1$; if $\alpha_P(X) = 1$, X is an ordinary (exact) set with respect to P; if $\alpha_P(X) < 1$, X is a rough (vague) set with respect to P.

Another ratio defines a quality of the approximation of X by means of the attributes from P:

$$\gamma_P(X) = \frac{|\underline{P}(X)|}{|X|} \quad (5)$$

The quality $\gamma_P(X)$ represents the relative frequency of the objects correctly classified by means of the attributes from P. Moreover, $0 \le \alpha_P(X) \le \gamma_P(X) \le 1$, and $\gamma_P(X) = 0$ iff $\alpha_P(X) = 0$,while $\gamma_P(X) = 1$ iff $\alpha_P(X) = 1$.

2.3. Reduction of Attributes

Another issue of great practical importance is that of "superfluous" data in a data table. Superfluous data can be eliminated, in fact, without deteriorating the information contained in the original table [11].

Let $P \subseteq Q$ and $p \in P$. It is said that attribute p is *superfluous* in P if $I_P = I_{P-\{p\}}$; otherwise, p is indispensable in P.

The set P is *independent* (orthogonal) if all its attributes are indispensable. The subset $P^{\backslash}$ of P is a *reduct* of P (denotation Red (P)) if $P^{\backslash}$ is independent and $I_{P^{\backslash}} = I_P$.

More than one reduct of P may exist in a data table. The set containing all the indispensable attributes of P is known as the core. Formally

$$Core_{\gamma}(P) = \bigcap \mathrm{Red}_{\gamma}(P) \quad (6)$$

3. The Basic Concepts of the Grey System Theory

Grey system theory was pioneered by well-known Chinese scholar, Professor Julong Deng. It is used to deal with the problem of uncertainty in less data little sample, which is designed as greyness, thus the system of what having greyness is said to be grey system; accordingly, there are whitening systems: complete information, and black systems: devoid of information. One of the main tasks of the grey system theory is to find the mathematical relationship and the changing rules among the factors or within one factor based on the behavior feature data of social, economic, ecological system, and so on. Furthermore, grey system theory holds that any random process is a grey quantum that changes within certain ranges and time zones, so a random process is viewed as a grey process.

3.1. Grey Number

Grey number is an expression of the grey system behavior character, and it is the basic "unit" or "cell" of the grey system. A grey number is such a number whose exact value is unknown but the range within which the value lies is known. In applications, a grey number is an uncertain number of an interval or a set of numbers [12].

3.2. Whitening of Grey Number

A grey number can usually be whitened through a whitening weight function, which is usually designed based on the given information, and there is no fixed pattern. Typical weight functions of whitenization $f\,[a1,\, b1,\, b2,\, a2]$ are shown in Figure 1.

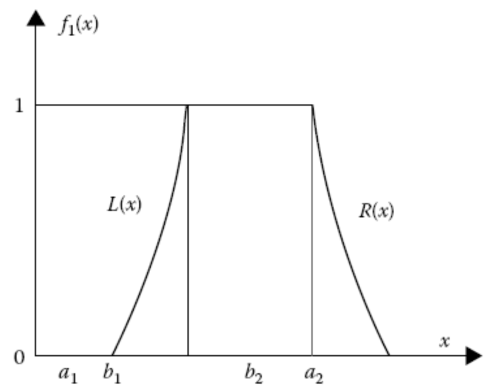

Figure 1

Assumed that

$$f_1(x)=\begin{cases} L(x) & \mathrm{x}\in[a_1,b_1) \\ 1 & \mathrm{x}\in[b_1,b_2] \\ R(x) & \mathrm{x}\in(b_2,a_2] \end{cases} \qquad (7)$$

Where

$L(x)$ is an increasing function of the left

$R(x)$ is a right down function

$[b_1,b_2]$ is the peak district

a_1, a_2 are the starting point and ending point respectively

b_1, b_2

are turning points, the grey number $\mathrm{x}\in[a_1,a_2]$

The continuous functions with fixed starting and ending points, and values increasing on the left and decreasing on the right are called typical weight functions of whitenization.

3.3. Grey Degree

According to the grey system theory,

$$g^{\circ} = \frac{2|b_1 - b_2|}{b_1 + b_2} + \max\left\{\frac{|a_1 - b_1|}{b_1}, \frac{|a_2 - b_2|}{b_2}\right\} \tag{8}$$

g° is called the grey degree of the grey number x for the typical weight function of whitenization, as shown in Figure 1. The expression (8) of the grey degree consists of two parts, in which the first one represents the impact of the size of the peak district on grey degree and the other represents the impact of the coverage of the $L(x)$ and $R(x)$ on grey degree.

According to this definition, for the grey degree of a grey number, the greater the peak area and the larger the coverage area of $L(x)$ and $R(x)$, the greater the grey degree.

4. The Algorithm Structure

The main structure of the algorithm is.

Step 1: Input the information system S of the discretization from raw data.

Step 2: construct the decision table from the information system by applying the following steps:

Step 2.1: Establish an information matrix with given attribute value.

Step 2.2: Calculate cluster weight η_j^k (or given by experts) with the given whitenization weight function f_j^k

Step 2.3: From formula $\sigma_i^k = \sum_{j=1}^{m} f_j^k(x_{ij})\eta_j^k$ we can figure out cluster coefficient, and further determine object belonging to grey class.

Step 2.4: Construct decision table with decision attribute value regarded as the object belonging to grey class.

Step 3: Determine the levels of grey degree of grey numbers are defined as follows:

- The grade of white numbers is defined as

$$g_c = 0 : \underline{\mu}_x(x) = 0 \Rightarrow \underline{g}_c = 0 \text{ and } \overline{\mu}_x(x) = 1 \Rightarrow \overline{g}_c = 0$$

- Grade 1 of grey numbers is defined as

$$g_c = 1 : \underline{\mu}_x(x) \in (0, 0.1] \Rightarrow \underline{g}_c = 1 \text{ and } \overline{\mu}_x(x) \in [0.9, 1) \Rightarrow \overline{g}_c = 1$$

- Grade 2 of grey numbers is defined as

$$g_c = 2 : \underline{\mu}_x(x) \in (0.1, 0.2] \Rightarrow \underline{g}_c = 2 \text{ and } \overline{\mu}_x(x) \in [0.8, 0.9) \Rightarrow \overline{g}_c = 2$$

- Grade 3 of grey numbers is defined as

$$g_c = 3: \underline{\mu}_x(x) \in (0.2, 0.3] \Rightarrow \underline{g}_c = 3 \text{ and } \overline{\mu}_x(x) \in [0.7, 0.8) \Rightarrow \overline{g}_c = 3$$

- Grade 4 of grey numbers is defined as

$$g_c = 4: \underline{\mu}_x(x) \in (0.3, 0.4] \Rightarrow \underline{g}_c = 4 \text{ and } \overline{\mu}_x(x) \in [0.6, 0.7) \Rightarrow \overline{g}_c = 4$$

- Grade 5 of grey numbers is defined as

$$g_c = 5: \underline{\mu}_x(x) \in (0.4, 0.5] \Rightarrow \underline{g}_c = 5 \text{ and } \overline{\mu}_x(x) \in [0.5, 0.6) \Rightarrow \overline{g}_c = 5$$

- The grade of black numbers is defined as

$$g_c > 5: \underline{\mu}_x(x) = \overline{\mu}_x(x) \quad \underline{g}_c = \overline{g}_c > 5$$

Step 4: for each g_c compute the g_c-lower approximation, and g_c- upper approximation of X with I_p in S by

$$\underline{apr}_P^{g_c}(X) = \cup \left\{ \frac{|I_P(x)| \cap X}{|I_P(x)|} \le \overline{g}_c \right\} \tag{9}$$

$$\overline{apr}_P^{g_c}(X) = \cup \left\{ \frac{|I_P(x)| \cap X}{|I_P(x)|} > \overline{g}_c \right\} \tag{10}$$

Step 5: compute the reduct at different levels of grey degree of grey numbers

Step 6: compute the measure of classification quality by

$$\gamma_P^{g_c}(P, D) = \frac{\left| \left\{ \frac{|X| \cap I_P(x)}{|I_P(x)|} \le \overline{g}_c \right\} \right|}{|U|} \tag{11}$$

Step 7: Finally extract the corresponding rule sets.

5. Case Study

In this section, an example is given to show how the proposed algorithm can be used to generate diagnostic rules for the power transformer. According to the historical fault data of the power transformer, the fault decision table is shown in Table1. Here, the condition attributes are grades of concentration of dissolved gases in the insulation oil, such as Hydrogen (H_2), Methane (CH_4), Carbon Dioxide (CO_2), Ethane (C_2H_6), Ethylene (C_2H_4), and Acetylene (C_2H_2). The decision attribute (D) is the fault class of the transformer, where “0” represents the fault of local discharge, “1” represents the fault of low-energy discharge and “2” represents the fault of low-temperature superheat.

According to the proposed method in the paper, the reduct at different levels of grey degree of grey numbers and the corresponding rule sets obtained are in Table 2.

Table 1. A Decision Table for the Transformer Fault

U	H_2	CH_4	CO_2	C_2H_6	C_2H_4	C_2H_2	D
X1	normal	normal	high	normal	normal	high	2
X2	high	Very High	high	normal	normal	high	2
X3	high	Very High	normal	high	high	normal	1
X4	high	Very High	normal	high	high	normal	1
X5	high	Very High	Very High	normal	normal	high	2
X6	normal	Very High	high	normal	high	high	0

Table 2. g_c-Reduct and the Corresponding Rule Sets Generated

Level	**g_c-Reduct**	**Rules**	**Support**	**Classification Quality**
$g_c = 1$	$\{H_2, CH_4, CO_2\}$	**If** H_2=normal ^ CH_4 = normal ^ CO_2=high **Then** D = 2	1	$\gamma_P^{g_c}(P,D)=1$
		If H_2=high ^ CH_4 = very high ^ CO_2=high **Then** D = 2	1	
		If H_2=high ^ CH_4 = very high ^ CO_2= normal **Then** D = 1	2	
		If H_2=high ^ CH_4 = very high ^ CO_2= very high **Then** D = 2	1	
		If H_2=normal ^ CH_4 = very high ^ CO_2=high **Then** D = 0	1	
$g_c = 3$	$\{C_2H_6, C_2H_4\}$	**If** C_2H_6 =normal ^ C_2H_4 = normal **Then** D = 2	4	$\gamma_P^{g_c}(P,D)=1$
		If C_2H_6 = high ^ C_2H_4 = high **Then** D = 1	2	
$g_c = 4$	$\{CH_4, C_2H_6\}$	**If** CH_4 = normal ^ C_2H_6 = normal **Then** D = 2	1	$\gamma_P^{g_c}(P,D)=1$
		If CH_4= very high ^ C_2H_6 = normal **Then** D = 2	3	
		If CH_4= very high ^ C_2H_6 = high **Then** D = 2	2	

6. Conclusion

The paper has demonstrated, using a simple example, that the appropriate hybrid of the Rough Set theory and grey system theories can overcome the shortages of their definitions and applications For example, a rough membership function can be used to complement the shortcomings of the definitions of the grey degree of grey numbers, while knowledge of different grey grades can be obtained from grey degree of grey numbers. Besides, we can compare the relative dominating degree of reduction attribute by applying the associated grey degree. Many features of that approach can be obtained such as:

- The hybrid algorithm is sure to get real reducts, it gives more different reducts.
- It can discover If-Then rules from very large, complex databases.
- It can flexibly select biases for search control.
- The classification Quality is rather high and satisfactory in practical applications.

Acknowledgements

We thank the President of Salman Bin Abdulaziz University, Deanship of Scientific Research at Salman Bin Abdulaziz University, Dean of College of Science and Humanitarians studies and all the Staff of Department of mathematics for their continuous support, encouraging and for their useful and fruitful discussion. This paper was supported by Salman Bin Abdulaziz University under the grant number 1432/ ه/ 13

References

[1] S. K. M. Wong, W. Ziarko and Y. R. Li, "Comparison of rough-set and statistical methods in inductive learning", International Journal of Man-Machine Studies, vol. 24, **(1986)**, pp. 53-72.
[2] J. R. Quinlan, "Induction of decision trees", Machine Learning, vol. 1, **(1986b)**, pp. 81-106.
[3] L. V. Fausett, “Fundamentals of Neural Networks: Architectures, Algorithms, and Applications”, Englewood Cliffs, NJ, U.S.A., Prentice-Hall, **(1994)**.
[4] D. E. Goldberg, “Genetic Algorithms in Search, Optimisation and Machine Learning, Reading, Mass”, U.S.A., Addison-Wesley, **(1989)**.
[5] L. A. Zadeh, "Is probability theory sufficient for dealing with uncertainty in AI: A negative view", in L. N. Kanal, J. F. Lemmer, eds. Uncertainty in Artificial Intelligence, **(1986)**, pp. 103-116. New York, U.S.A., North Holland Press.
[6] H. A. Nabwey, “A Probabilistic Rough Set Approach to Rule Discovery”, International Journal of Advanced Science and Technology, vol. 30, (2011) May, pp. 25-34.
[7] F. Min, Q. Liu and C. Fang, “Rough sets approach to symbolic value partition”, Internat. J. Approx. Reason, **(2008)**.
[8] Z. Pawlak and A. Skowron, “Rough sets and Boolean reasoning”, Information Sciences, vol. 177, **(2007a)**, pp. 41–73.
[9] S. M. Shaaban and H. A. Nabwey, in B. Murgante, *et al.*, (Eds.): ICCSA 2012, Part II, LNCS 7334, **(2012)**, pp. 316–330.
[10] M. Arabani, “Evaluation of rough set theory for decision”, Scientia Iranica, vol. 13, no. 2, **(2006)**, pp. 152–158.
[11] H. A. E. Mohamed, “An Algorithm for Mining Decision Rules Based on Decision Network and Rough Set Theory”, in T. -h. Kim, *et al.*, (Eds.): UCMA 2011, Part I, CCIS 150, **(2011)**, pp. 44–54.
[12] J. Deng, “Basic Method of Grey System”, 2nd edn. Wuhan, China: Huazhong University of Science and Technology Press, **(2005)**.