Differential Evolution for Rule Extraction and Its application in Recognizing Oil Reservoir

Jin-ling Li[1, 2], Hai-xiang Guo[1, 3, 4], Ke-jun Zhu[1], De-yun Wang[1], Jing-lu Hu[4]
[1]School of Economics and Management
China University of Geosciences
Wuhan, 430200, P. R.China
[2]Jiangcheng College
China University of Geosciences
Wuhan, 430074, P. R.China
[3]Key Laboratory of Tectonics and Petroleum Resources
China University of Geosciences
Ministry of Education
Wuhan Hubei 430074, P. R.China
[4]Graduate School of Information
Production and Systems, Waseda University
Hibikino 2-7, Wakamatsu-ku, Kitakyushu-shi
Fukuoka 808-0135, Japan
faterdumk0732@sina.com

JDIM
Journal of Digital Information Management

ABSTRACT: *The method of rule extraction based on differential evolution (DE-Rule) is proposed in this paper. The style of rule is IF-THEN. Connection word is AND in the antecedent of rule. The consequence of rule is class label. The above rule is encoded and expressed to individual in population of differential evolution. Then the population is iterated through differential mutation and binomial crossover. Optimal rule set is obtained after decoding the optimal individual. Finally, DE-Rule, RS-Rule and ANN-GA-Cascades-Rule are used to recognize oil reservoir (dry layer, water layer, inferiority layer and oil layer) in the Jianghan oil field. Data of well oilsk81 is training data and data of well oilsk83 is testing data. The results show that performance of accuracy and interpretability in DE-Rule is the best.*

Categories and Subject Descriptors:
I.2.10 [Artificial intelligence]; **I.4.10 [Data Mining]**

General Terms:
Evolution Algorithm, Fuzzy Rule

Keywords: Differential Evolution, Rule Extraction, Oil Reservoir

Received: 2 August 2013, Revised 8 September 2013, Accepted 13 September 2013

1. Introduction

Huge amounts of raw data generated in the procedure of oil exploration, which covers certain information that could become knowledge and even be formed to IF-THEN rules, thus is helpful for better decision-making. Artificial neural networks and genetic algorithms are both commonly used for extracting rules. When the input and output variables as well as the fuzzy partition of the variables become too many, the rules will be exponentially growing [1]. Using genetic algorithms to extract rules [2] could, on one hand achieve the global optimization search, but would also, on the other hand, be hard to get the expression of chromosome and conformation of fitness function. Some researchers proposed rule extraction methods combining artificial neural networks (ANN) and genetic algorithm (GA). These methods fall into two categories. One is optimizing parameters using GA in the process of extracting rules by ANN, the other is selecting rules by GA in the process of extracting rules by ANN. Wang Gang[3] used ANN to extract rules and GA to optimize parameters, e.g. connecting weights of ANN , membership functions. Li Yang extracted rules using ANN, and reduced rules using GA. However, there are some drawbacks which are listed as follows: (1) Antecedent of rule cannot be reduced. (2) Accuracy for recognizing is the single object. (3) Coding work for the whole process of extracting rules is complicated.

According to the above drawbacks , some improved works are proposed in this paper: (1) Antecedent of rule is reduced by GA-FCM [5] that find some key attributes or features for recognizing oil reservoir from data sets. (2) At

the same time, two main goals are considered for rule extraction: one is maximizing accuracy and the other one is minimizing complexity, which means that a good interpretability for rules. However, the two goals are often conflicted with each other. Based on the tradeoff of accuracy and interpretability, the main consideration of rule sets should be maximization of the accuracy and minimization of the complexity. Then in the process of rules selection, the following three main performances should be considered [3,4,5]: $f_1(S)$, $f_2(S)$ and $f_3(S)$ where S denotes the rule set, $f_1(S)$ denotes the number of correctly classified training samples by S, $f_2(S)$ denotes the number of fuzzy rules in S, and $f_3(S)$ denotes the total number of antecedent conditions of fuzzy rules in S. The larger value of $f_1(S)$ indicates the higher recognition accuracy of fuzzy rules set, and the smaller value of $f_2(S)$, $f_3(S)$ indicates the better interpretability of fuzzy rules set.
(3) To propose a new method for extracting rule based on differential evolution which can be easily coded.

2. Differential Evolution Algorithm

Differential Evolution (DE) was firstly proposed by Price and Storn [9,10,11] in 1995. We now point out some of the reasons why the researchers have been looking at DE as an attractive optimization tool. These reasons includes: (1) Compared to most other EAs, DE is much more simple and straightforward to implement. Main body of DE takes four to five lines to code in any programming language. Simplicity to code is important for practitioners from other fields, since they may not be experts in programming and are looking for an algorithm that can be simply implemented and tuned to solve their domain-specific problems. (2) The gross performance of DE in terms of accuracy, convergence speed and robustness still makes it attractive for applications to various real-world optimization problems. (3) The number of control parameters in DE is very few (Cr, F, and NP in classical DE). Thus DE has been successfully applied to various domains of science and engineering such as signal processing [12], chemical engineering [13], pattern recognition [14], microbiology [15], and image processing [16].

3. Rule Extraction Based on Differential Evolution Algorithm

3.1 Individual structure of DE-rule

(1) Individual structure of DE-rule is shown in Figure.1. $X_{i,G}$ is the i^{th} individual at the Gi^{th} generation. Each individual $X_{i,G}$ includes the NR rules. The value of NR is determined by the specific problem.

(2) Each rule in individual consists three parts: rule control, antecedence and class label. Rule control determines whether the rule should be taken into account for classification. Namely, if the real value contained in the field is higher than a real value Rule Threshold (RT), then

(3) Each antecedence of rule includes four parts: antecedence control, connection type, constant $C1$ and $C2$. The first field is called antecedence control which determines whether or not the antecedence is present in the rule. Also in this case, a real-valued parameter Ante Threshold (AT) is defined for DE-Rule, and the antecedence is active only if the value in antecedence control is higher than AT. The parameter AT plays an important role, since it has influence on the size of the achieved rules: the higher its value, the lower the number of antecedence that, on average, will be active in any rule, hence the more compact the rules. The second field which is called connection type represents the type of relation operator connecting the variable (feature) and the constant value(s). Seven different operators have been chosen: $<$, $\leq$, $=$, $>$, $\geq$, IN, OUT. The first five operators need only one constant value only, i.e.$C1$, while the latter two need two constant values $C1$and $C2$. IN means that the variable is within the range expressed by $C1$ and $C2$ in their order of appearance (a check is carried out throughout the algorithm so that $C1 \leq C2$). OUT, instead, means that the variable is outside the range $[C1, C2]$. The third and fourth fields contain the real values for the constants $C1$ and $C2$, respectively. The connection type is computed by starting from the connection type value and by rounding it to its ceiling value. For example, a real value of 4.12 in that field corresponds to ceiling (4.12) = 5, where ceiling () is the ceil function, so that value of 4.12 represents the fifth operator in their list. Since the order for the operators has been set as IN, $<$, $\leq$, $=$, $>$, $\geq$, OUT, that value actually codes for a $>$ operator. The class related to the rule is computed in a similar way, so that a value of 1.68 in that field means that the rule refers to class label ceiling (1.68) = 2.

Thus, the total length NG for each individual in the population can be given by $NG = NR \times (1 + (4 \times NV) + 1)$.

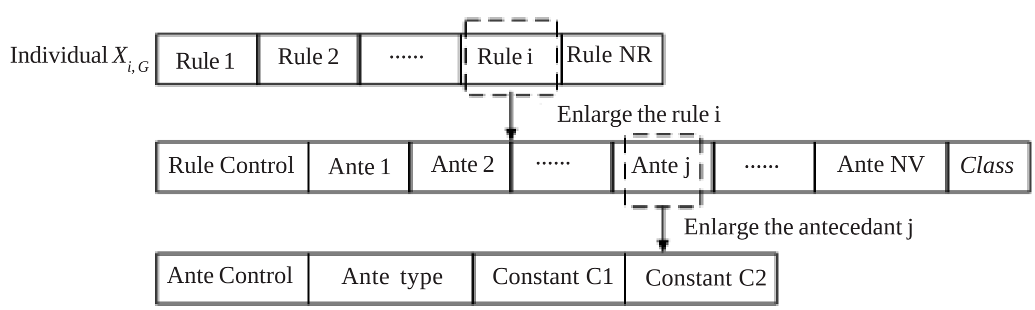

Figure 1. Individual Structure of DE-Rule (Antecedence for Short is Ante)

3.2 Example of individual structure in DE-rule

As an example, Figure.2 reports a DE individual for a problem with two variables and two classes. Thus, for each rule we need $1 + (2 * 4) + 1 = 10$ fields. In this example, each individual is supposed to contain a set of three rules. Therefore, each DE individual will contain $10 * 3 = 30$ real values. Let us suppose that $NR = 3$, $RT = 0.5$, $AT = 0.5$, $NV = 2$, the order of connection type is $IN, <, \leq, =, >, \geq, OUT$. The value of ante type is 1 and corresponding connection type is IN. The value of ante type is 2 and corresponding connection type is $<$, the rest followed by analogy. The value of rule control is $0.68 \geq RT = 0.5$, so rule 1 is active. The first antecedence control of rule 1 is $0.91 \geq AT = 0.5$, so the antecedence is active. The first connection type is 2.13 and ceiling (2.13) is 3, the corresponding connection type is $\leq$. The second antecedence control of rule1 is $0.63 \geq AT = 0.5$, the second antecedence control is active. The second connection type is 4.60. Ceiling (4.60) is 5 and its corresponding type is $>$. The class label is 1.14 and ceiling (1.14) is 2. The corresponding classification is 2.The second rule is not active , since its rule control field contains 0.24. For the third rule, rule control = 0.83, so this rule is active and refers to class1 since class field contains 0.62. In it the antecedence for the first variable is not active (Ante control = 0.34), whereas that for the second variable is active (Ante control = 0.76) and its relation operator is $\leq$, encoded by 2.29.

Therefore, the DE individual in the Figure 2 decodes the following set of rules:

IF $(x_1 \leq 6.14)$ AND $(x_2 > 8.59)$ THEN class label is 2 .

IF $(x_2 \leq 3.12)$ THEN class labe1 is 1.

3.3 How to Resolve the Conflicts Among Rules of DE

Rules are randomly generated based on DE. There must be some conflicts among these rules. We will give a specific example to illustrate the conflicts and corresponding method used for dealing with the conflicts.

Suppose that rule set consists two rules with two features (variables), and the range of feature is $\in [0, 1]$.

Rule 1: IF $(0.2 \leq x_1 \leq 0.7)$ AND $(x_2 \leq 0.4)$ THEN class label is 1.

Rule 2: IF $(0.4 \leq x_1 \leq 0.9)$ AND $(0.1 \leq x_2 \leq 0.7)$ THEN class label is 2.

The rule set can be expressed in 2D plane coordinate which is shown in Figure 3.

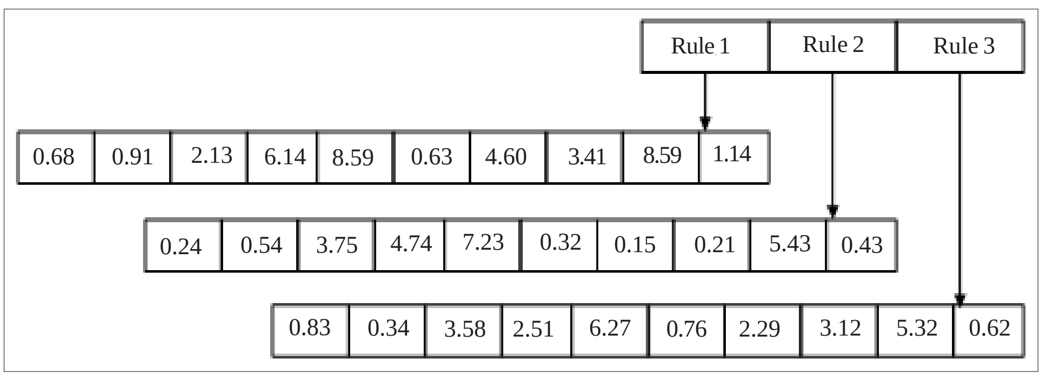

Figure 2. Example of Individual Structure in DE-Rule

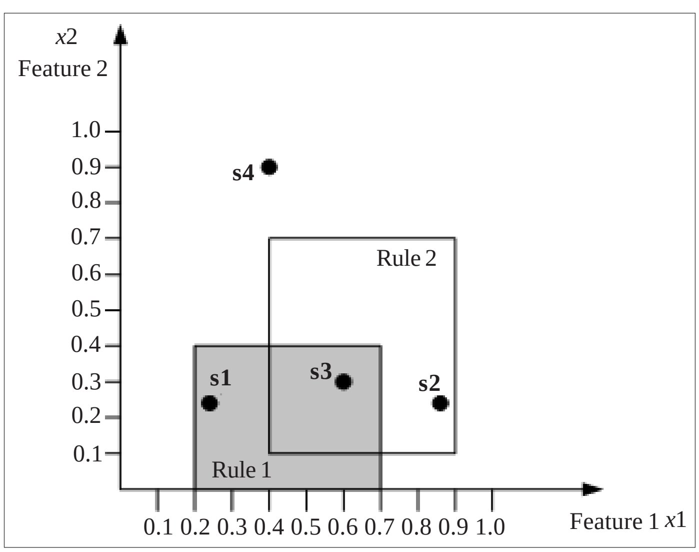

Figure 3. 2D plane coordinate of rule set

3.3.1 Y-Y conflict
From Figure.3, $s3$ (0.6, 0.3) belongs to Y-Y conflict. $s3$ (0.6, 0.3) can be represented by rule 1 and rule 2 respectively. Although $s3$ can be represented by rule 1 and rule 2 at the same time, membership degree of $s3$ belonging to rule 2 is larger than membership degree of $s3$ belonging to rule 1. According to the criterion of maximal membership degree, $s3$ should be represented by rule 2 and $s3$ belongs to class 2.

3.3.2 N-N conflict
From Figure.3, $s4$ (0.4, 0.9) belongs to N-N conflict. No rule can represent $s4$ (0.4, 0.9). Membership degree of $s4$ belonging rule 2 is larger than membership degree of $s4$ belonging to rule 1. $s4$ belongs to class 2.

3.3.3 Criterion for resolving conflicts
Membership degree is used for dealing with conflicts of rules. Distance between data and centre of rule denotes membership degree. From Figure.3, center of rule 1 is (0.45, 0.2), center of rule 2 is (0.65, 0.4); membership degree of s3 belonging to rule 1 is

$$\mu_1 = \frac{1}{\sqrt{(0.6-0.45)^2 + (0.3-0.2)^2 + 1}} = 0.847,$$

membership degree of $s3$ belonging to rule 2 is

$$\mu_2 = \frac{1}{\sqrt{(0.6-0.65)^2 + (0.3-0.4)^2 + 1}} = 0.863,$$

according to the criterion of maximal membership degree, $s2$ belongs to class 2.

3.4 Fitness Function
For the i^{th} individual in the population, its fitness $f(i)$ is computed as the percentage of correctly classified cases over the training set.

$$f(i) = \frac{N_c}{N_{tr}} \times 100\%$$

where N_c is the number of cases in the training set that are correctly classified by the set of rules represented by the individual i, and N_{tr} is the total number of cases in the training set.

4. Case Study

4.1 Data From Jianghan Oilfield
Data of oilsk81 and oilsk83 are logging data in Jianghan oilfield as shown in Table1 and Table 2. There is listed part of data and if you need all data, you can refer the reference [17]. Training data is oilsk81 and testing data is oilsk83. From Table 1 and 2, the correlative well log features with recognizing oil-bearing formation are acoustic travel time (AC), compensated neutron logging (CNL), resistivity (RTY), porosity (POR), oil saturation (SO), permeability (PERM). AC can be obtained by acoustic logging equipments, analyzing the property that the sonic propagation varies when it comes to different rocks and fluids. Generally, acoustic travel time would increase dramatically if there is oil vapor in the void. Various effects of interaction between CNL and other substances can be used to study rock formation properties of the cross section. RTY is a physical quantity that represents resistance properties of various substances, as well as a main parameter that judges fluids properties of reservoir. Commonly, resistance increases if there is oil vapor while decreases if there is water in reservoir. POR is defined as the ratio of the void space in a rock to the bulk volume of that rock multiplied by 100 to express in percent. SO is defined as ratio of void volume occupied by crude oil to total void volume of rock in oil reservoir. Allowable capability of fluid passing to the rock in some difference of pressure is called permeability (PERM), whose measurement is the rate of permeation. The recognizing oil-bearing formation means to recognize the characters of each layer in reservoir. These characters include dry layer, water layer, inferior layer and oil layer. So our goal is to obtain predictive accuracy in the experiment of recognizing oil-bearing formation and the task is classification. The key features are AC and SO obtained from the feature set through GA-FCM [5]. Features reduction can help improve the interpretation of rule set.

Layer	AC (μs/m)	CNL (%)	RTY (Ω.m)	POR (%)	SO (%)	PERM (m.μm2)	*Class*
1	195	7.5	13.0	6.0	0	0	Dry
2	225	10.0	7.3	11.0	0	0	Water
3	230	14.0	5.5	12.0	0	0	Water
4	220	9.0	25.0	9.0	56	1.3	Oil
5	225	8.0	30.0	9.0	58	2.3	Oil
6	210	7.0	26.0	6.0	0	0	Dry
7	195	4.0	36.0	5.5	0	0	Dry
8	210	6.0	130.0	7.0	48	0.7	Inferiority
9	202	6.0	55.0	7.0	52	0.8	Inferiority
10	195	4.5	50.0	6.0	0	0	Dry

Table 1. The Data of Oilsk81(Part of Data is Listed)

Layer	AC (μs/m)	CNL (%)	RTY (Ω.m)	POR (%)	SO (%)	PERM (m.μm2)	*Class*
1	225	10.0	4.0	10.0	0	0	Water
2	226	10.0	5.0	10.5	0	0	Water
3	220	8.5	6.6	9.5	0	0	Water
4	205	7.5	50.0	7.0	36	0.7	Inferiority
5	208	5.0	18.0	7.0	35	0.8	Inferiority
6	225	7.0	15.0	9.0	50	1.2	Oil
7	190	2.0	53.0	3.0	0	0	Dry
8	188	4.0	70.0	2.0	0	0	Dry
9	203	8.0	27.0	7.0	40	0.8	Inferiority
10	211	8.5	9.0	7.5	50	5.0	Oil

Table 2. The Data of Oilsk83 (Part Data is Listed)

The algorithm of extracting rules	$f_1(S)$ Accuracy	$f_2(S)$ Rule number	$f_3(S)$ Antecedence number
RS-Rule[18]	96%	10	20
ANN-GA-Cascades-Rule [19]	96%	7	13
DE-Rule	96%	4	8

Table 4. Comparing with Accuracy and Interpretability (Testing data is oilsk83)

4.2 Results Comparison and Analysis

Rule set for reservoir identification is obtained by decoding the optimal individual in DE-Rule.

IF $177.3263\ 2 \le AC \le 216.4935$ and $So < 40.2605$ then the layer is dry.

IF $AC > 221.35$ and $So < 17.27$ then the layer is water.

IF $AC \le 225.5848$ and $29.7645 < So < 17.27$ then the layer is inferiority.

IF $AC \ge 205.4050$ and $So > 35.0229$ then the layer is oil.

According to Table 4, we list the multi-objective optimizing (MOPs) expressions which make tradeoff of accuracy and interpretability of rule set. *DE*-Rule is the best performance among RS-Rule, ANN-GA-Cascades-Rule and *DE*-Rule.

MOP-1: Max $f_1(S)$ and Min $f_2(S)$: *DE*-Rule is better than ANN-GA-Cascades-Rule and RS-Rule.

MOP-2: Max $f_1(S)$ and Min $f_2(S)$: *DE*-Rule is better than ANN-GA-Cascades-Rule and RS-Rule.

MOP-3: Max $f_1(S)$, Min $f_2(S)$ and Min $f_3(S)$: *DE*-Rule is better than ANN-GA-Cascades-Rule and RS-Rule.

5. Conclusion

The paper proposed an rule extraction algorithm based on differential evolution. DE-Rule actually has two parts. One part is rule which is used to partition the space and the other is differential evolution which is used to search the optimal rule. DE-Rule benefits to expand the theory frame of DE, and DE-Rule for reservoir identification is also enrich application category of DE. Future work will combine DE-Rule with other intelligent methods to optimize the parameters such as RT, AT, F and so on, which tries to improve performance like convergence speed of DE-Rule.

6. Acknowledgements

This work is supported by the National Natural Science Foundation of China No.71103163,71103164, 71301153 , by Program for New Century Excellent Talents in University, No. NCET-13-1012, by Research Foundation of Humanities and Social Sciences of Ministry of Education of China No.10YJC790071, by the Fundamental Research Founds for National University, China University of Geosciences(Wuhan) No. CUG120111, CUG110411, G2012002A, by China Postdoctoral Science Foundation Grant No. 20090461293, by special grade of the financial support from China Postdoctoral Science Foundation Grant No. 201003670, by the open foundation for the research center of resource environment economics in China University of Geosciences (Wuhan) and by the open foundation for Key Laboratory of Tectonics and Petroleum Resources (China University of Geosciences), Ministry of Education No. TPR-2011-11.

References

[1] Tsukimoto, H. (2000). Extracting rules from trained neural networks. *IEEE Trans. Neural Networks*,11 (2) 377-389.

[2] Lim, M. H., Rahardja, S., Gwee, B. H. (1996). A GA Paradigm for Learning fuzzy Rules. *Fuzzy Sets and Systems*, (82) 177-186.

[3] Wang Gang, Wang Bennian. (2008). Data Mining Study Based on Fuzzy Neural Networks and Genetic Algorithms. *Computer Technology and Development*,18 (2) 119-122.

[4] Li Yang, Gao Zhengguang, Li Qiyan. (2004). A Data Mining Architecture Based on ANN and Genetic Algorithm. *Computer Engineering*, 30 (6) 155-156.

[5] Zhu Kejun, Su Shunhua, Li Jinling. (2005). Optimal Number of Clusters and the Best Partition in Fuzzy C-mean. *Systems Engineering-Theory* & *Practice*, 3, p.52-61.

[6] Ishibuchi, H., Murata, T., Turksen, I. B. (1997). Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. *Fuzzy Sets Systems*, 89 (2) 135-150.

[7] Ishibuchi, H., Nakashima, T., Murata, T. (2001). Three-objective genetics-based machine learning for linguistic rule extraction. *Inform Sci.*, 136 (1–4) 109-133.

[8] Ishibuchi, H., Yamamoto, T. (2004). Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. *Fuzzy Sets Systems*, 141 (1) 59-88.

[9] Storn, R., Price, K. V. (1995). Differential evolution: A simple and efficient adaptive scheme for global optimization over continuous spaces. ICSI, USA, Tech. Rep. TR-95-012, [Online]. Available:http://icsi.berkeley.edu /~storn/litera.html.

[10] Storn, R., Price, K. V. (1996). Minimizing the real functions of the ICEC 1996 contest by differential evolution. *In*: Proc. IEEE Int. Conf. Evol. Comput., p. 842-844.

[11] Storn, R. (1996). On the usage of differential evolution for function optimization. *In*: Proc. North Am. Fuzzy Inform. *Process. Soc.*, p.519-523.

[12] Das, S., Konar, A. (2006). Design of two dimensional IIR filters with modern search heuristics: A comparative study. *Int. J. Comput. Intell. Applicat*, 6 (3) 329-355.

[13] Wang, F. -S., Jang, H. -J. (2000). Parameter estimation of a bio-reaction model by hybrid differential evolution. *In*: Proc. IEEE Congr. Evol. Comput. Piscataway, *NJ: IEEE Press*, 1, p. 410-417.

[14] Das, S., Abraham, A., Konar, A. (2008). Adaptive clustering using improved differential evolution algorithm. *IEEE Trans. Syst., Man, Cybern. A*, 38 (1) 218-237, Jan.

[15] Noman, N., Iba, H. (2007). Inferring gene regulatory networks using differential evolution with local search heuristics. *IEEE/ACM Trans. Comput. Biol. Bioinform*, 4 (4) 634-647, Oct.

[16] Omran, M., Engelbrecht, A. P., Salman, A. (2005). Differential evolution methods for unsupervised image classification. *In*: Proc. 7th Congr. Evol. Comput. (CEC-2005), Piscataway, *NJ: IEEE Press*, 2, p. 966-973.

[17] Guo Haixiang, Liao Xiuwu, Zhu Kejun. (2011). Optimizing Reservoir Features in Oil Exploration Management Based on Fusion of Soft Computing. *Applied Soft Computing.* 11 (1) 1144-1155, JAN.

[18] Guo Haixiang, Zhu Kejun, He Yong. (2005). Extract of Fuzzy Rules of Economic Increase Based on Fuzzy Clustering and Rough Set. *Chinese Journal of Management*, 4 (2) 437–440

[19] Hai-xiang Guo, Ke-jun Zhu, Si-wei Gao. (2009). Extracting Fuzzy Rules Based on Fusion of Soft Computing in Oil Exploration Management. *Expert Systems with Applications*, 36 (2) 2081-2087.