A HYBRID METHOD FOR EXTRACTING CLASSIFICATION RULES

Chuanli Zhuang [1,2], Zetian Fu [1,2], Daoliang Li [2*]

1. *College of Economics & Management, China Agricultural University, 100083, Beijing,China*
2. *Key Laboratory of Modern Precision Agriculture System Integration, College of Engineering, China Agricultural University 100083, Beijing,China*

Abstract: Neural networks is considered the most powerful classifier and rough set theory is thought of the best to reduce attributes and to generate rules. The combination of neural networks and rough set is very useful for knowledge acquires. Integrating of the advantages of two approaches and having solved the data continuous problem, this paper presents a hybrid method to extract classification rules. There are three models in our method, in first model, neural networks was employed to classify the data sets. In the second model, the continuous attributes are discretized and the self-organizing neural network was applied to ensure result consistent before and after the discretization. In the third model, rough sets theory was used to reduce attributes and generate the rules. The proposed approach was applied on abandoned mine wastes data and the extracted rules was testified based on the analysis of case studies, the result show that the method was reasonable.

Key words: extracting rules, rough sets, self-organizing artificial neural networks, discretization, classification

1. INTRODUCTION

One of the important issues in development expert system is knowledge acquisition from expert. Recently, in order to automate this problem,

* Corresponding author: tel: +86-10-62736717, Fax: +86-10-62737679, E-mail: dliangl@cau.edu.cn

machine-learning techniques have been developed to extract knowledge from database. Among proposed approaches deriving rules from training examples is most common. Given a set of examples a learning program tries to induce rules that describe each class.

Neural networks are considered the most powerful classifier for their low classification error rates and robustness to noise. There are some kinds of neural network. Self-organizing is a special kind of neural networks without teaching. It has the function of self-organizing. Through training by itself, it can classify the input samples automatically. But neural networks lack explanation facilities for their knowledge. The knowledge is buried in their structures and weights. It is often difficult to extract rules from a trained neural network.

Rough sets theory, a mathematical tool to deal with vagueness and uncertainty of information, was first introduced by Pawlak in the 1980's. It has been applied in a lot of applications such as machine learning, knowledge discovery, and expert system since then and has been proved to be very effective. It constitutes a framework for inducing minimal decision rules. These rules in turn can be used to perform a classification task. However, it only deals with the classificatory of data tables and it required the data continuous.

The combination of rough sets and neural networks is very nature for complementary features. In this paper, integrating of the advantages of two approaches, we present a hybrid method to extract classification rules. Different from those previous works, the neural network in this method was used twice; it is not only a classifier but also a checker. First time, neural networks is used as a classifier to classify the data set, second, the self-organizing neural networks is applied as a checker to testify the consistent of discrete result besides classification. After this the rough set can work well to reduce attribute and generate rules.

The remainder of this paper is organized as follows: in section 2, we make a brief description about rough set theory and the definition of classification accuracy and coverage based on this theory. Extracting classification method is described in Section 3. A sample was presented in section 4, and finally concluding remarks are given in section 5.

2. THE THEORY OF ROUGH SETS

2.1 Information system

An information system is a 4-tuple $S=\{U,A,V,f\}$,where U is a finite set of objects, called the universe; A is a finite set of attributes, $V=U_{a\in A}V_a$ is a domain of attribute a. and $f:U\times A\rightarrow V$, called an information function such that $f(x,a)\in V_a$, $\forall a\in A,\forall x\in U$.

In the classification problems an information system is also seen as a decision table assuming that $A=C\bigcup D$ and $C\bigcap D=\phi$,where C is a set of condition attribution and D is a set of decision attributes

2.2 indiscernibility

Let $S=\{U,A,V,f\}$ be an information system, every $P\subseteq A$, generates a indiscernibility relation

$ind(P)$ on U ,which is defined as follows:

$ind(P)=\{(x,y)\in U\times U:f(x,a)=f(y,a),\forall a\in P\}$, $U/ind(P)=\{C_1,C_2\cdots C_K\}$ is a partition of U by P ,every C_i is an equivalence class. For $\forall x\in U$,the equivalence class of x in relation $U/ind(P)$ is defined as follows:

$$[x]_{IND(P)}=\{y\in U:f(y,a)=f(x,a),\forall a\in P\}$$

2.3 P _ lower approximation & P _upper approximation

Let $P\subseteq A$, $X\subseteq U$, the P _lower approximation of X (denote by $P_*(X)$) and the P _upper approximation of X (denote by $P_*(X)$) are defined as follows:

$$P_*(X)=\{Y\in U:[y]_{ind(P)}\subseteq X\},$$

$P^*(X)=\{Y\in U:[y]_{ind(P)}\bigcap X\neq\phi\}$ $P_*(X)$ is the set of all objects from U which can be certainly classified as elements of X employing the set of attributes P. $P^*(X)$ is the set of objects of U which can be possibly classified as elements of X using the set of attributes P.

Let P,Q $\subseteq$ A , the positive region of classification $U/IND(Q)$ with respect to the set of attributes P ,or in short P -positive region of Q is defined as :

$$POS_P(Q)=\bigcup_{X\in U/IND(Q)}\underline{P}X$$

$POS_P(Q)$ contains all objects in U that can be classified to one class of the classification $U/IND(Q)$ by attributes P. The dependency of Q on P is defined as:

$$\gamma_P(Q) = \frac{card(POS_P(Q))}{card(U)}$$

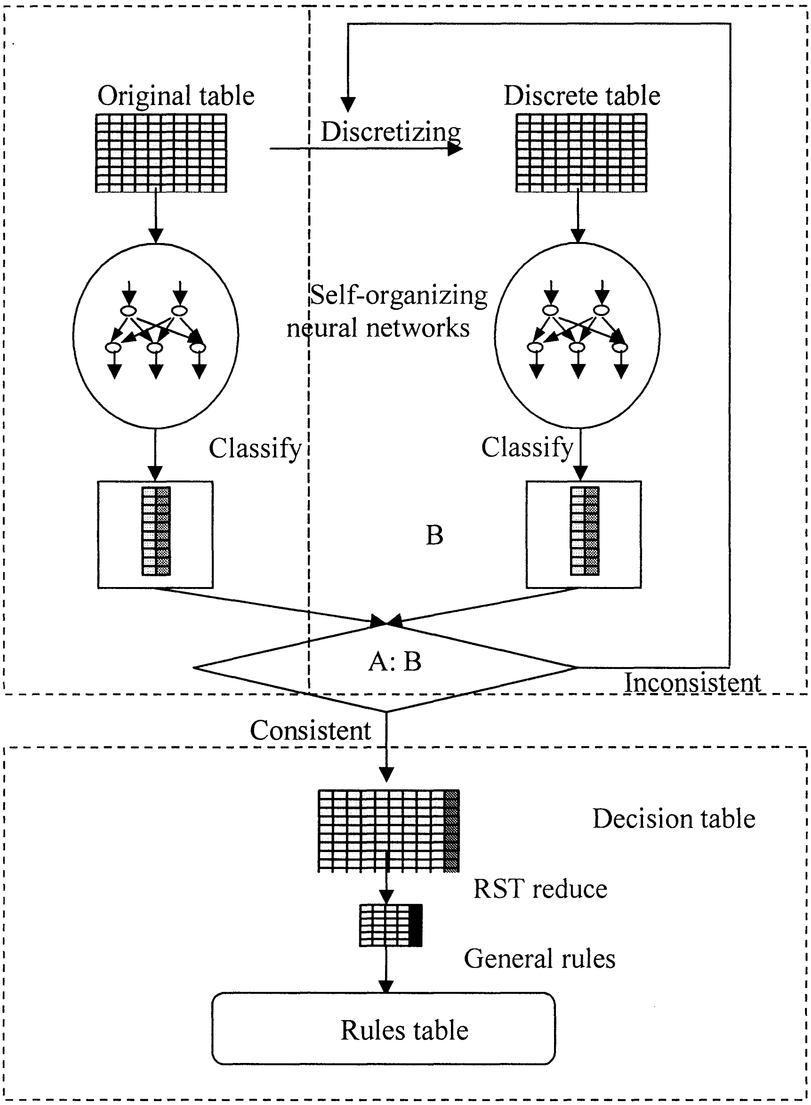

Fig. 1show the procedures of our approach

2.4 reduct

An attributes a is said to be dispensable in P with respect to Q ,if $\gamma_P(Q) = \gamma_{P\text{-}\{a\}}(Q)$; otherwise a is indispensable attribute in P with respect to Q.

Let $S = \{U, C \bigcup D, V, f\}$ be a decision table, the set of attributes P ($P \subseteq C$) is a reduct of attributes C , which statisfies the following conditions:

$\gamma_P(D) = \gamma_C(D)$ and $\gamma_P(D) \neq \gamma_{P'}(D)$, $\forall P' \subset P$. A reduct of condition attributes C is a subset that can discern decision classes with the same discriminating capability as C, and none of the attributes in the reduct can be eliminated without decreasing its discriminating capability.

2.5 Classification accuracy and coverage

If the equal-classes of $U / ind(C)$ and $U / ind(D)$ are presented by X_i , γ_j , defines decision rules as: r_{ij} : $X_i \rightarrow \gamma_j$, $\gamma_j \bigcap X_i \neq \phi$ .And Classification accuracy and coverage is defined as:

$$\mu(X_i, \gamma_j) = \frac{card(\gamma_j \bigcap X_i)}{card(X_i)}$$

$$\eta(X_i, \gamma_j) = \frac{card(\gamma_j \bigcap X_i)}{card(\gamma_j)}$$

Where $card(A)$ denotes the cardinality of a set A , $\mu(X_i, \gamma_j)$ denotes a classification accuracy of X_i as to classification of γ_j , $\eta(X_i, \gamma_j)$ denotes a classification coverage of X_i as to classification of γ_j , respectively.

3. METHODOLOGY

Our hybrid method of extracting classification Rules of non-decision attribute data sets consists of three major models:

Model 1: Classification

Using the clustering function of self-organizing neural networks, the category of each data is acquired. The goal of this step is to obtain the decision attribute of data sets. If the data sets have decision set, this model can be skipped. There are four kinds of self-organizing algorithms. They are self-organizing competitive learning (SCL) neural networks, self-organizing feature map (SFM) neural networks and the frequency-sensitive competitive learning (FSCL) neural network. There is some software to implement implement the algorithm of self-organizing neural networks, such as artificial neural networks (ANN) toolbox of Matlab, ANN toolbox of Microsoft Office programmer, etc.

Model 2: discreteness

To reduce and extract rules from decision table by rough set, the attribute of each data in decision table should be discrete. This step is to discretize the continuous attributes. Firstly, Using discretizing algorithm, the discrete table is obtained then using self-organizing neural networks to testify the result consistent before and after the data discretized attribute. If they are consistent, then the decision table was obtained by combining the discrete table with the classification and goes to mode 3, else rediscretize the condition attributes. The usually discrete algorithms are equal-distance, equal-frequency, naïve scalar and semi-naïve scalar, etc.

Model 3: reduce the decision table and general rules

The last model is to generate the rules. Applying rough set theory, a reduct of decision table is obtained and the final knowledge-a rule set is generated from the reduced decision table.

4. APPLICATION

In this section, an example is given to show how the proposed methodology can be used to generate the classification rules from data set without decision attributes. Table 1 shows the climate-geography features of 12 abandoned mine wastes. There is no appropriate classification for them yet. There are 8 attributes Annual average temperature (℃), Annual Sunshine hours(h/a), Annual rainfall (mm), Accumulated temperature (℃), Percent of sand (%), Altitude (m), Annual precipitation (mm), Annual average wind velocity, denoted by a, b, c, d, e, f, g and h respectively.

Table 1. Original table

Area	a	b	C	d	e	f	g	h
SZ	6.9	2779	440	2476	30	1400	1570	3.6
YQ	10.9	2767	576	3325	72	1500	1362.5	2.8
TF	6.8	2350	650	3400	18	250	1667	3.6
HQ	13.3	2307	631	4391	56	480	2072	3.5
TL	0.1	2850	383	2525	68	1200	1784	3.7
QA	10.1	2782	735	3200	44	400	1100	2.5
HN	3.1	2427	475	2750	39	260	900	3.6
FS	6.5	2386	800	2800	12	230	1600	3.5

YKS	7.3	3100	385	3000	35	1100	2000	3.7
JZ	14.9	2300	604	4700	76	500	800	2.9
CF	5.0	2950	368	2400	40	1050	2100	3.7
LC	7.3	2518	663	3600	62	1058	1614	4.0

The followings are the deatil

Step 1. As a matter of convenience, we have normalized the input data firstly.

$x_{ij} = \dfrac{x_{ij}^{'} - x_j^{\min}}{x_j^{\max} - x_j^{\min}}$ $i = 1 \cdots 8, j = 1 \cdots 12$. Where $x_{ij}^{'}$ is non-normalized elements of input table 1, $x_j^{\max}$, $x_j^{\min}$ are maximal and minimal non-normalized elements of $j-\text{column}$, respectively. Using self-organizing competitive learning function newc() to build a self-organizing neural networks. Set 8 input neurons and 3 output neurons, Kohonen learning rate and Conscience learning rate defined as default. The result was show in table 2

Table 2 the result of classification

Category	Colliery discard land
I	SZ、TL、YKS、CF、
II	YQ、HQ、JZ、LC
III	TF、QA、HN、FS

The classification is reasonable in view of the similar geographical feature of each category.

Step 2

Using the equidistance algorithm to discretize each attribute in table 1, the distance equal the one third of range. The discretization table (table 3) was obtained.

Table 3 the result of distcreted data

Area	a	b	c	d	e	f	g	h
SZ	2	2	1	1	1	3	2	3
YQ	3	2	2	2	3	3	2	1
TF	2	1	2	2	1	1	3	3
HQ	3	1	2	3	3	1	3	3
TL	1	3	1	1	3	3	3	3
QA	3	2	3	2	2	1	1	1
HN	1	1	1	1	2	1	1	3

FS	2	1	3	1	1	1	2	3
YKS	2	3	1	1	2	3	3	3
JZ	3	1	2	3	3	1	1	1
CF	1	3	1	1	2	3	3	3
LC	2	1	3	2	3	3	2	3

Using self-organizing to classify these 12 discreted data, the result is same as the before. So combing the condition attributes with the result of classification, the decision table obtained (table4).

Table 4. The decision table

Area	a	b	c	d	e	f	g	h	class
SZ	2	2	1	1	1	3	2	3	1
YQ	3	2	2	2	3	3	2	1	2
TF	2	1	2	2	1	1	3	3	3
HQ	3	1	2	3	3	1	3	3	2
TL	1	3	1	1	3	3	3	3	1
QA	3	2	3	2	2	1	1	1	3
HN	1	1	1	1	2	1	1	3	3
FS	2	1	3	1	1	1	2	3	3
YKS	2	3	1	1	2	3	3	3	1
JZ	3	1	2	3	3	1	1	1	2
CF	1	3	1	1	2	3	3	3	1
LC	2	1	3	2	3	3	2	3	2

Step 3 reduce decision table and generate rules

using ROSE software which developed by Professor S. Wilk and the staff of Institute of Computing Science of Poznan University of Technology , set the least RHS accuracy 50% and least RHS coverage 100%, the rules table was acquired(table 5).

Table 5. rules table

rules	accuracy， coverage
rule 1. (b = 3) => (class = 1)	[75%, 100%]
rule 2. (a = 2) & (c = 1) => (class = 1)	[50%, 100%]
rule 3. (a = 1) & (f = 3) => (class = 1)	[50%, 100%]
rule 4. (a = 1) & (g = 3) => (class = 1)	[50%, 100%]
rule 5. (c = 1) & (f = 3) => (class = 1)	[100%, 100%]
rule 6. (c = 1) & (g = 3) => (class = 1)	[75%, 100%]
rule 7. (d = 1) & (f = 3) => (class = 1)	[100%, 100%]
rule 8. (d = 1) & (g = 3) => (class = 1)	[75%, 100%]

rule 9. (e = 2) & (f = 3) => (class = 1)	[50%, 100%]
rule 10. (e = 2) & (g = 3) => (class = 1)	[50%, 100%]
rule 11. (f = 3) & (g = 3) => (class = 1)	[75%, 100%]
rule 12. (d = 3) => (class = 2)	[50%, 100%]
rule 13. (a = 3) & (b = 1) => (class = 2)	[50%, 100%]
rule 14. (a = 3) & (c = 2) => (class = 2)	[75%, 100%]
rule 15. (a = 3) & (e = 3) => (class = 2)	[75%, 100%]
rule 16. (b = 1) & (e = 3) => (class = 2)	[75%, 100%]
rule 17. (c = 2) & (e = 3) => (class = 2)	[75%, 100%]
rule 18. (c = 2) & (h = 1) => (class = 2)	[50%, 100%]
rule 19. (d = 2) & (e = 3) => (class = 2)	[50%, 100%]
rule 20. (d = 2) & (f = 3) => (class = 2)	[50%, 100%]
rule 21. (d = 2) & (g = 2) => (class = 2)	[50%, 100%]
rule 22. (e = 3) & (f = 1) => (class = 2)	[50%, 100%]
rule 23. (e = 3) & (g = 2) => (class = 2)	[50%, 100%]
rule 24. (e = 3) & (h = 1) => (class = 2)	[50%, 100%]
rule 25. (a = 2) & (f = 1) => (class = 3)	[50%, 100%]
rule 26. (b = 1) & (d = 1) => (class = 3)	[50%, 100%]
rule 27. (b = 1) & (e = 1) => (class = 3)	[50%, 100%]
rule 28. (c = 3) & (f = 1) => (class = 3)	[50%, 100%]
rule 29. (d = 2) & (f = 1) => (class = 3)	[50%, 100%]
rule 30. (d = 1) & (f = 1) => (class = 3)	[50%, 100%]
rule 31. (e = 1) & (f = 1) => (class = 3)	[50%, 100%]
rule 32. (e = 2) & (f = 1) => (class = 3)	[50%, 100%]
rule 33. (e = 2) & (g = 1) => (class = 3)	[50%, 100%]

Step 4 rules testify

Author collected the data of other three abandoned mine wastes from the web. This data is incomplete. However it can be classified by the rules. the result was show in table 6

Table 6. the results of test

Area	FX	HLH	JC
a	7.5(2)	0.2(1)	11.2(3)
b	2750(2)	3163(3)	2563(1)
c	540(2)	381.5(1)	650(2)
d	3400(2)	-	4000(2)
e	30(1)	-	-
f	500(1)		-
g	1746(3)	1543.8(2)	1558(3)

h	3.7(3)	-	-
Rules	25,29,31	1	13,14
Class	3	1	2

5. CONCLUSION

In this paper, we have presented a hybrid approach to extract classification rules through integrating the advances of self-organizing neural networks and rough sets. Using the clustering function of self-organizing neural networks, the method classifies the data all continuous attributes discretized by some discrete algorithms in succession, then using self-organizing neural testify the similarity of result of classification. At last the rules and the rudect was obtained using rough sets theory if the similarity is accepted.

The difficult problem of extracting classification rules of neural networks is solved through this hybrid method though the black box mechanism of neural network classification.

Discrete algorithms are used to make the self-organizing neural networks and rough set be applied effective. Using self-organizing networks to testify the result of discretization, this method can ensure the rules extracted from the discrete decision table consistent with original data.

The application of extracting classification rules on abandoned mine wastes case show that the method was reasonable and effective.

Although the proposed method works well in extracting classification rules, there are is still much work to be done in this field. Our method need to try some time on find the discretization algorithm to obtain the right one. In the future we will attempt to find the appropriate discrete algorithm quickly.

ACKNOWLEDGEMENTS

The authors would like to thank Guo Xiangyun, Wang Ying and Zhu Xiang for their helpful to collect data and constructive comments.

REFERENCE

[1]. Xu Dong, WuZheng. Analysis and design for ANNbased on MATLAB6.x. *Xidian University Press*.2002

[2]. Wen Xin, ZhouLu, etc. Design for ANN Based on MATLAB. Science press, 2001

[3]. Xie Hongqua, ChenQiuji. A Research on Mine Land Reclamation Conditions with the Help of SOM Internet Principle[J]. Journal of Beijing vocational technical institute of industry NO.1. Vol. 3 Jan 2004,31-33

[4]. Xiangdong Ma, Qiang Fu. Applying Self-Organizing Competition Artificial Neural Networks to Classify the Soil[J]. Nature and Science, 1(1), 2003,

[5]. V.V Gafiychuk etc. Analysis of Data Clusters Obtained by Self-organizing Methods[J]. Physica A 341(2004) 547-555

[6]. Renpu Li, Zheng-ou Wang, mining classification rules using rough set and neural networks [J],European journal of operation research 157 (2004) 439-448

[7]. Zhang wenxiu, etc. theory and method of rough sets [M] science publisher 2003.1

[8]. Tzung-Pei Hong , etc. Learning rules from incomplete training example by rough sets expert system with application 22(2002) 265-293

[9]. T.-P. Hong, etc. Learning a coverage set of maximally general fuzzy rules by rough sets. Expert system with application 19(2000) 97-103

[10]. Shusaku Tsumoto. Extract of experts' decision rules from clinical databases using rough set model.intelligent data analysis, 2(1998)215-227

[11]. Hilario lÒpez García,Iván Machón González. Self-organizing map and clustering for wastewater treatment monitoring. Artificial intelligence. 17(2004)215-225