Research on Rough Set Based Network Pattern Recognition Model

Yang JIANNING, Lin KUN
Yunnan Normal University Business School,
Kunming, 650106, China

Received: 25 September 2013   /Accepted: 22 November 2013   /Published: 30 December 2013

Abstract: According to high development of web technologies, the amount of data has been increased rapidly. Data mining and pattern recognition of web data has become more and more important for both individuals and companies. What’s more, an increasingly number of users trend to do more their daily work to internet, which makes it more difficult to analyze and mine these large amount of data. Conventional network pattern recognition model can’t handle these big data effectively. Therefore, this paper designed a new rough set based network pattern recognition model, presented the Bayesian network structure and designed the rough set classified model for classifier. Finally, this paper designed a set of experiment to demonstrate the performance and quality of rough set based network pattern recognition model, the result shows it can work effectively with a higher accuracy and performance. 

Keywords: Pattern Recognition, Rough Set, Bayesian Network.

1. Introduction

With a widely in-depth application of variety of intelligent expert, troubleshooting, decision support systems, currently pattern recognition is widely attention and research. Made a lot of results based on the traditional, maturity classification and recognition, intervention of fuzzy mathematical thinking, ANN wealth of statistical models, appearing Evolutionary algorithm and a number of excellent algorithms, support vector machine, proposed rough set theory and some new methods. The statistical pattern recognition research and application full of vitality [1]. However, because pattern recognition involves many complex issues, existing theories and methods still have inadequacies in solving these problems, especially in recent years, with the emergence of the mass of information and incomplete information: in one side, more description of data attributes, that is increasing the dimension of the data, increasing the amount of training data, more and more time in learning the traditional pattern recognition methods, on the other side, the collected information may be missing many key attribute values, if traditional pattern recognition use samples in learning and training, unable to capture accurately rules hidden in data, besides, restrictions on the level and learning conditions, learning outcomes often have a certain noise which affecting stability of the way and the pattern recognition model [2, 9-11]. Finding new theories and methods based on existing pattern recognition theory is becoming a research hotspot and difficult in currently pattern recognition.
2. Network Security and Pattern Recognition

2.1. Pattern Recognition Introduction

Generally the performance of a time and spatial distribution of information is called mode. The role and purpose of the pattern recognition is correctly classified specific message as a category, according to some concept design a classifier based on available data, and then, learning good classifier to predict the unknown sample. There are two basic pattern recognition methods, they are statistical pattern recognition method and structure pattern recognition method [3].

Bayesian network identification model is established on Bayesian network classifier, it’s a typical classification model based on statistical methods and an effective representation and probabilistic reasoning model for uncertain environment knowledge, theoretical basis is Bayesian theorem. Prior probability of the event linked with posterior probability, calculate the sample belongs to a class posterior probability through the existing evidence, with maximum posterior probability of the class is the class of the object [4]. Since it is mainly used to measure the weight of probability and describe the correlation between data, thereby solving the inconsistencies between data, even a separate issue.

The pattern recognition target architecture is showed as Fig. 1, it’s consisted of five components: windows, documents, links, images and forms.

![Fig. 1. The network architecture of pattern recognition target.](image)

Compared with other pattern recognition methods, Bayesian network pattern recognition method has the advantages of mildness and fault tolerance, it may do a good deal with incomplete information in identify issues and it has been one of the most effective models in uncertain knowledge representation and reasoning [5, 6]. But the application of current Bayesian network pattern recognition method still exists difficult in classifier structure and parameter learning. How to improve the efficiency of its structure learning quickly and effectively estimate the parameters and stability of the application have become the main direction of the majority of scholars recently. Paper research these issues paper about Bayesian network based on rough set model and pattern recognition methods.

2.2. Pattern Recognition Model

According the method learning classification to training samples, we can divide classification model into two categories: passive classification and active classification. Passive learning model chosen training samples at random, passive acceptance of these samples information. This passive study show significant deficiencies: 1) Processing the training sequence samples often make learning classifier having an order correlation, overly sensitive to data; 2) If encounter noise samples, this noise should been spread always, affect the progress in classification; 3) Lack of the ability of comprehensive no annotated sample information. Active classification model is active in training sample selection, concentrated on the test case in candidate sample, and these examples will be added to the training set in a certain way. Generally, active learning in the initial training set of active learning, the number of samples with marked are rarely [7, 8]. It uses these training samples with categories labeled to learning a classification model, then apply a certain selection strategy, select the best sample from candidate samples(without category labels) and add to the training set. Revised the classifier parameters, choosing the best next sample, until the candidate sample set is empty or it reaches a certain standard.

In active Bayesian classifier, select the instance of the class conditional posterior entropy classifier that instance and measure certainty degree of current classification, using minimum and maximum entropy active learning method, first, choose a maximum and minimum conditional entropy candidate sample from the candidate set (min example: max example), then add these two samples to the training set simultaneously. The largest sample of class conditional entropy of is very likely the strange unknown sample, add the largest sample of class conditional entropy, allows classifier emphasis a sample with special information early; the smallest class conditional entropy is the determine sample in classifier. It is also more accurate in classifier, reduce the error propagation caused by the joining of an uncertainty sample.

The pattern network of credit deceive problem diagram is showed in Fig. 2.

Literature specializing in establishing and learning of active Bayesian network classifier, literature practice active Bayesian network classifier to the detection of unknown malicious executable code recognition, literature use active Bayesian network classifier for speech recognition. However, this method also has the following disadvantages, the
feature extraction of active Bayesian classifier is limited to a fixed length (Window value), the detection rate and accuracy of model should to be improved, when modeling, it’s cost large computation and time [9].

Fig. 2. The pattern network of credit deceive problem.

Generally, Dynamic Bayesian network classifiers have two characteristics, 1) Network topology is the same within each time slice, between the film and the film is connected through a similar arc. 2) The T frame network is only linked with the T-1 and the T-i-1 network. Therefore, this method is more difficult to achieve algorithm, it needs long time to training Network structure and parameters.

2.3. Challenges of Current Model

Rough set theory is a mathematical tool to deal with imprecise information and uncertain information, it immunity to data in some extent. Compare with a hard partition of traditional recognition, classified by approximately reduction method of rough set, it doesn’t require any information except collection of data for processing which greatly simplifies the computational complexity; Bayesian network classifier is a reasoning method based on probabilistic uncertainty, with the characteristics of mildness and Fault Tolerance, both are have the ability to classified and deal with incomplete information. But if direct use of these two methods in the absence of key attributes it may not meet the actual needs of pattern recognition due to lower judgment.

Thesis attempts to combine Rough set method and Bayesian networks, using rough set streamline the data and Bayesian network training streamlined data, construct Rough set – Bayesian network pattern recognition model, this model not only improve the pattern classification accuracy and the ability to deal with incomplete data but enhance Bayesian network classifiers decision-making efficiency. This method can overcome the weakness of rigid reasoning in rough set. Avoid drawbacks of computational complexity brought by simple network classifier. It’s a mildness, fault tolerance, and practical method of pattern recognition.

3. Rough Set Based Network Pattern Recognition Model

3.1. Bayesian Network Structure

A complete statistical pattern recognition model is mainly composed of four parts: data gets, pretreatment, feature extraction, select and classification decision. Generally data acquisition and pretreatment are the research of digital signal processing and graphics on topics, and the Papers data rare from the standard database, regardless of feature extraction and selection, this chapter will focus on classification decision of Bayesian network pattern recognition model.

Bayesian theory groundbreaking work from paper of Reverend Thomas Bayes “About the chance of problem solving comments”, who is a mathematician and theologian in the eighteenth century, article was first published by his friend in 1763 after his death.

Bayesian network is a directed acyclic graph model, it encodes a relation about a set of variables conditional probability, is more popular in the field of artificial intelligence knowledge representation methods of uncertainty.

3.1.1. Definition of Bayesian Networks

A Bayesian network is a directed acyclic graph, composed by a representative of the variable nodes and a directed edge connecting these nodes, where each node is marked by the quantitative probability information, expressed as $B = \{G, P\}$ ($G$ is represented as a variable domain $GAG$, $P$ is a corresponding set of conditional probability collections).

If a collection of random variables $V$ (With limited variables $n$), $G$ is directed acyclic graph, $E$ is Collection of directed edges, $P$ is Collection of conditional probability distribution, here is a Bayesian network model represented by mathematical symbols:

$$B = \{G, P\} = \{V, E, P\}$$  \hspace{1cm} (1)

Note

$$G = \{V, E\}, V = \{V_1, V_2, \ldots, V_n\}$$

$$E = \{V_iV_j|V_i, V_j \in V\}$$

$$P = \{P(V_i|V, V_2, \ldots, V_{i-1})|V_i \in V\}$$

Bayesian network is a directed acyclic model based on probabilistic reasoning. It may represent complex relationship between the variables in specific issue as a network structure, reaction the variables dependency in the problem domain through the network model, for uncertain knowledge representation and reasoning.
A Bayesian network is mainly composed of two parts, corresponding to qualitative description and quantitative description in problem areas they are Bayesian network structure and network parameters.

### 3.1.2. Bayesian Network Structure

Bayesian network structure is a directed acyclic graph (DAG), consists of a set of nodes and a set of directed edges. Each node in centralized nodes represents a random variable and variables can be an abstract of any problem, used to represent the phenomenon of interest, component, status or property and so on, meaning a certain physical and practical significance. Directed edges represent dependencies or causality between variables, the arrow of directed edges represents the direction of causal influence (by the parent node to the child node), if there is no connection between the nodes edge, means that the corresponding variable is conditional independence between nodes, if all nodes pointing to $X$, then we named these nodes as parents node. Despite the node $X$ pointing to node $Y$ which has a directed edges frequent is used to denotes $X$ caused of $Y$, but in Bayesian network, this is not the only explanation for the directed edge. For example, $Y$ may be Associated with $X$, but not caused by $X$. Although Bayesian network can represent causal relationships, they are not limited to indicate causality.

### 3.1.3. Bayesian Network Parameters

Another part of the Bayesian network response the partial probability distribution of the variable correlation, that is network parameters (Probability parameters), usually called Conditional Probability Table (CPT), the table lists each node relative to its parent node all possible conditional probabilities. Bayesian networks have agreed node $X_i$ as the parent node, $X_i$ conditional independence with any child nodes. Probability values represent strength of the association or confidence between child nodes and its parent node, node probability for no parent node is a priori probability. Bayesian network structure is the result of data instance abstract and a macroscopic description of the problem areas. The probability parameter is exact expressing the association strength between variables, belonged to quantitative description.

Bayesian networks establish an arc between connection nodes, then joint probability distribution of Bayesian network with n nodes:

$$ P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Pa_i) \quad (2) $$

Node $Pa_i$ is the collection of parent node $X_i$.

#### 3.2. Variable Precision Model Design

A central problem of Rough set theory is classification analysis. The limitations of rough set is that the classification it handles must be entirely correct or affirmative, because it classification is strict accordance to the equivalence class, so it is exact classification, that is contains or does not contain Another limitation is that objects it processes are known and all these conclusions obtained from the model apply only to the set of objects. However, in practical applications, often needs to apply some conclusion from small object to the massive set of objects.

Variable precision Rough set model is an extension from Rough set model, it introduce $\beta (0 \leq \beta \leq 0.5)$ based on the basic rough set model, which allows a certain degree of misclassification rate exists. Improve the concept of approximation space, the other hand, it also conducive to use rough set theory found irrelevant data from the relevant data. Certainly, the main task of variable precision rough set model is resolving problems which are no function between property and uncertain data classification. If $\beta = 0$, Rough set model is a special case of Variable precision rough set model.

Definition, set $U$ as a finite universe, $X, Y \subseteq U$

$$ c(X, Y) = \begin{cases} 1 - |X \cap Y|/|X| & X \cap Y \neq \emptyset \\ 0 & |X| = 0 \end{cases} \quad (3) $$

$|X|$ represents the cardinality $X$, named $c(X, Y)$ is relative misclassification rate about $X$ set to $Y$ set. Obviously, $0 \leq c(X, Y) \leq 1$, $c(X, Y) = 0$. And only if $X \subseteq Y$.

Definition, set $0 \leq \beta < 0.5$, majority inclusion relation definition

$$ Y \supseteq X \iff c(X, Y) \leq \beta \quad (4) $$

“Majority” means the number of common elements contains to $X,Y$ is 50 % more than elements in $X$.

Definition 5, if $(U, R)$ is the approximation space, on the domain $U$ is a non-empty finite set, $R$ Equivalence to $U$, $U / R = \{E_1, E_2, \ldots, E_n\}$ is the equivalence or basic consist to $R$, if $X \subseteq U$,

$$ R_\beta X = U \{E \in U \mid R / X \supseteq E \} = U \{E \in U \mid c(E, X) \leq \beta \} \quad (5) $$

$$ \overline{R}_\beta = U \{E \in U \mid 1 / c(E, X) \leq 1 - \beta \} \quad (6) $$
If $R_{\beta}X = \overline{R_{\beta}X}$, then $X$ to $\beta$ is define, or $X$ to $\beta$ is rough variable precision rough set.

Definition 6 If $U$ is a finite non-doctinaire domain, $X \subseteq U$, $R \subseteq U \times U$ is arbitrary binary relations to $U$. $A = (U, R)$ is called Generalized Approximation Space. For any $X \in U$.

Definition upon approximation and lower approximation belong to $X$ about myopia space $A = (U, R)$ is:

$$R_{\beta}X = \bigcup \{x \in U | X \supseteq R_{\beta}(X)\} = \bigcup \{x \in U | c(R_{\beta}(X), X) \leq \beta\}$$

(7)

$$\overline{R_{\beta}X} = \bigcup \{x \in U | c(R_{\beta}(X), X)(1 - \beta)\}$$

(8)

In which $\beta$ indicates the approximate level of accuracy of the set. If $R_{\beta}X = \overline{R_{\beta}X}$, $X$ on the approximation space $A$ called $\beta$ -generalized definable, Or $X$ on the approximation space $A$ called $\beta$-generalized rough.

If $R$ is equivalence relation, $\beta$ – generalized variable precision rough set model reduces to the classical variable precision rough set model.

3.3. Rough Set Classified Model

Bayesian network learning is the basis to establish Bayesian network classifier, learning algorithm of Bayesian network directly determines the effect of Bayesian network classifiers feasibility, practicality and efficacy. They are the determinants of Bayesian network classifiers quality. Although Bayesian network learning algorithm has achieved many results, but there still are shortcomings in algorithms, such as: training volume, computational complexity and more stringent restrictions in using. Rough set theory is without loss of information under the premise of the original information system express the same knowledge with the minimum conditions attributes set. It is the simplest form to keep the same classification ability of information systems. Through Rough set reduction the knowledge, not only eliminate redundant information, reduce the amount of computation, improve recognition speed, but also prevent significant information loss by the reduction, affect the accuracy of identification.

The structure model of rough set classifier is showed as below Fig. 3.

3.3.1. Rough Set-Based Bayesian Network Structure Learning

Bayesian network structure learning goal is to get a network structure, the network structure can be described the probability distribution of the training data set. Search the optimal structure is actually a process of searching the best individual at all possible network structure space, but Bayesian network structure may with $n! \cdot 2^n$ species on the set of variables $X$, search space is huge, full search is the problem of $NP$, and local optimal network structure may also have multiple; meanwhile, if the property contains missing information or more information is missing, redundant attributes highlighted in the role of classification, and decline the rate of model correctly judge, the predicted results of $BAN$ classifier is very unstable. So Rough set dimensionality reduction helps to reduce the amount of computation during training structure and improve the reliability of the priori information.

3.3.2. Rough Set-Based Bayesian Network Parameter Learning

Got Rough set based on Bayesian network structure by algorithm structure learning, it relates to another important issue of classification, that is parameter estimation. Calculate the conditional probability of each node according to the training data, the accuracy of parameter estimation is directly affecting the classification performance. There are many parameter estimation ways, commonly used methods are maximum likelihood method, Bayesian parameter learning method, $EM$ (expectation-maximization) and so on. With the case of completeness and incompleteness training sample data, the parameter learning is not the same.

3.4. Design of Classifier

When learning to get through the rough set based Bayesian network structure and parameters, you can create a rough set – a Bayesian network classifier.

When the sample data is complete, by the rough set reduction algorithm based on rough set and Bayesian theory shows: the identification pattern recognition objects $x$:

$$\forall \Phi \rightarrow \Psi \in \text{Dec}(S) \text{ and } x \in \Phi \Rightarrow \Psi,$$

(1)
If $0(\pi s(\Psi / \Phi))(j, x)$ to be identified will correspond to different modes of class $c_i(1 \leq j \leq n)$ in a plurality of decision rules, mode the possibility of a variety of classification $x$, then $x$ cannot be directly on the mode classification.

Definition $Dec(S)$ is the decision-making algorithm of $S$, \( \{ \Phi \rightarrow \Psi_j \}^n_{j=1} \in Dec(S) \)

\[ \forall x \in \Phi, \text{ to be determined } i, i \in \{1, \ldots, n\}, \]

cause $x \in \Phi_i$. Consider $x$ pattern recognition, from the requirement to minimize misclassification departure, according to the classical criteria for minimum error rate Bayesian decision theory, can be obtained based on rough set minimum error rate Bayesian decision criteria are as follows:

If $\pi_s(\Psi_j / \Phi) = \max_{j=1, \ldots, n} \pi(\Psi_j / \Phi)$, then $x \in \Phi_j$.

Through (10) formula can be obtained:

If $\pi_s(\Psi_j / \Phi) \pi_s(\Psi_i) = \max_{j=1, \ldots, n} \pi(\Psi_j / \Phi) \pi(\Psi_i)$, then $x \in \Phi_j$.

Knowledge of probability theory will be used to describe (3.12) formula is:

If $P(c_i|x_1, x_2, \ldots, x_k) = \max_{j=1, \ldots, n} P(c_i|x_1, x_2, \ldots, x_k)$

then $x \in c_i$.

2) When the sample data is not complete when, after reduction of rough and after the identification mode $X$ can be both corresponding to the same class $c_i$ multiple rules may also correspond to different class $c_j (1 \leq j \leq n)$ multiple rules.

Support by the formula indicates:

\[ \sigma_s(\Phi_i, \Psi_j) = \frac{\sup ps(\Phi_i, \Psi_j)}{\text{Card}(U) \pi_s(\Phi_i / \Psi_j) \pi_s(\Psi_j)} \]

$\Phi_i \rightarrow \Psi_j$ measure of the decision-making rules in the importance of centralized decision-making, and make full use of a priori information of the sample data, it can be used to identify the mode of processing $X$ corresponding to the same class $c_i$ multiple rules, namely:

If $\sigma_s(\Phi_i, \Psi_j) = \max_{\pi_s(\Phi_i, \Psi_j)}$, then $x \in \Phi_j$.

Through (13) formula can be obtained:

If $\pi_s(\Phi_i, \Psi_j) = \max_{\pi_s(\Phi_i, \Psi_j)}$, then $x \in \Phi_j$.

In the same class multiple issues are making corresponding rules, the rest is between different types of identification problem, for which only need to use the sample data and complete the identification method can solve.

4. Experiment and Validation

4.1. Experiment Environment and Scope

To verify the feasibility and effectiveness of the method, under conditions of Windows XP and PHP/2.4 hardware environment, with Matlab7.0 programming, learning from multiple UCI data sets, achieved the three classifiers: NB, BAN, RBAN, and for different completeness of sample data, learning on multiple data sets, count the accuracy of each classification rate and horizontal comparison.

According to the four stages of statistical pattern recognition method, general steps of Bayesian network based on rough set pattern recognition are as follows:

1) Inputting training data;
2) Data preprocessing, includes data relative concentration and decentralized;
3) Simple data attributes of Rough set
4) The training of RBAN classifiers
5) Pretreatment and reduction of test samples
6) Test sample data entry, classification and statistical analysis of test results

4.2. Test Results

For the case of data complete and incomplete, the results of the experiment are as follows:

1) Data complete, this paper uses three data sets from UCI: Iris data, Waveform and Wine data, evidence for incomplete information system, their specific information as shown in Table 1.

First random fifty data from the above three data sets as a test set, then each remaining data were used as training set. Using the training data set training the three classifiers, then testing the test data, the final test results are shown in Table 2. Where accuracy (%) and time (s) is the average of twenty experiments, the parameter value ($\alpha$, $\beta$) of classification RBAN from Iris data, Waveform and Wine data are: (0.25, 0.1), (0.2, 0.1), (0.25, 0.1).
Table 1. The data of complete information.

<table>
<thead>
<tr>
<th>Date</th>
<th>Number of samples</th>
<th>Condition attributes</th>
<th>Number of categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris</td>
<td>180</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Waveform</td>
<td>320</td>
<td>23</td>
<td>2</td>
</tr>
<tr>
<td>Wine</td>
<td>180</td>
<td>13</td>
<td>4</td>
</tr>
</tbody>
</table>

From the experimental results: 1) The accuracy of classification RBAN is higher than the other two methods, the recognition speed is slower than classifier NB, effectively reducing the complexity of solving Bayesian classifier recognition method. 2) Recognition accuracy for Iris data and Wine data is better than Waveform data. 3) Recognition accuracy of classification RBAN for Waveform data higher than Iris data and Wine data.

2) Data incomplete, this paper uses three data sets *hepatitis data, processed Switzerland data* and *processed via data* from UCI, conduct empirical for incomplete information system, their specific information as shown in Table 3.

First random fifty data from the above three data sets as a test set, then each remaining data were used as training set. Using the training data set training the three classifiers, then testing the test data, the final test results are shown in Table 4. Where accuracy (%) and time (s) is the average of twenty experiments, the parameter value \(\alpha, \beta\) of classification RBAN from hepatitis data, processed Switzerland data and processed via data are: \((0.25, 0.15), (0.2, 0.1), (0.25, 0.1)\).

Table 2. The comparison of accuracy (%) and speed (s) of complete information.

<table>
<thead>
<tr>
<th>Method Data</th>
<th>The number of test samples</th>
<th>Classifier NB Accuracy</th>
<th>Classifier NB Time</th>
<th>Classifier BAN Accuracy</th>
<th>Classifier BAN Time</th>
<th>Classifier RBAN Accuracy</th>
<th>Classifier RBAN Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris</td>
<td>80</td>
<td>93.12</td>
<td>4.2</td>
<td>93.46</td>
<td>5.5</td>
<td>98.26</td>
<td>4.2</td>
</tr>
<tr>
<td>Waveform</td>
<td>80</td>
<td>94.33</td>
<td>7.6</td>
<td>95.04</td>
<td>14.2</td>
<td>92.43</td>
<td>8.7</td>
</tr>
<tr>
<td>Wine</td>
<td>80</td>
<td>92.81</td>
<td>6.2</td>
<td>92.35</td>
<td>12.4</td>
<td>95.17</td>
<td>7.6</td>
</tr>
</tbody>
</table>

Table 3. The data of incomplete information.

<table>
<thead>
<tr>
<th>Information</th>
<th>Number of samples</th>
<th>Condition attributes</th>
<th>Number of categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total number</td>
<td>Maximum missing number</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>153</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>Processed Switzerland</td>
<td>121</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>Processed va</td>
<td>205</td>
<td>23</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 4. The comparison of accuracy (%) and speed (s) of incomplete information.

<table>
<thead>
<tr>
<th>Method Data</th>
<th>The number of test samples</th>
<th>Classifier NB Accuracy</th>
<th>Classifier NB Time</th>
<th>Classifier BAN Accuracy</th>
<th>Classifier BAN Time</th>
<th>Classifier RBAN Accuracy</th>
<th>Classifier RBAN Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepatitis</td>
<td>50</td>
<td>82.37</td>
<td>6.3</td>
<td>78.13</td>
<td>9.6</td>
<td>98.26</td>
<td>7.4</td>
</tr>
<tr>
<td>Processed Switzerland</td>
<td>50</td>
<td>72.45</td>
<td>8.5</td>
<td>70.67</td>
<td>11.3</td>
<td>88.24</td>
<td>9.7</td>
</tr>
</tbody>
</table>

From the experimental results: 1) The less classification categories, the higher accuracy rate of decision, mainly due to the fewer categories, various types getting more fully training and the classification is more accuracy; 2) The more default attribute value, the more inaccurate classification; 3) The accuracy of classification RBAN is higher than the other two methods, the recognition speed is slower than classifier NB, mainly due to classifier dimensionality reduction effectively, removing the redundant information, improves the reliability of the information.

5. Conclusion

In based on the theory of Rough set Bayesian, using the simple algorithm definition of the relevant knowledge in thesis learning the structure and parameters from Bayesian network which is based on Rough set, got Rough set-Bayesian network classifier, and analysis sample data for completeness and incompleteness, through experiments illustrate the effectiveness and feasibility of Rough set-Bayesian network classifier. Compared with the traditional classification NB and BAN, whether complete or...
incomplete information system. **RBAN** Classifier has a higher recognition accuracy, recognition speed slightly slower than the **NB** classification, but significantly accelerated than **BAN**.

General relations of Rough set model and Variable precision rough set model are all expanded from classical rough set theory. General relations of rough set model overcomes the classical rough set model defects caused by the equivalence relation, can better deal with the incompleteness of information, variable precision rough set model by allowing a certain degree of presence of misclassification rate, it has better immunity to data noise. Paper combine general relations of rough set model and variable precision rough set model together, not only deal with incomplete information model information incompleteness but also overcome the data noise, meanwhile by variable precision Rough Sets reduction algorithm $\beta$ and $\alpha-\beta$ for attribute reduction heuristic, using Bayesian network classifiers for pattern recognition, avoiding the problem of finding the smallest reduction NP-hard, so paper proposed method achieved better recognition results, and better recognition efficiency.

**References**


