Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(7):2264-2270

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

System model of college students' network behavior research based on rough sets

Jiqin Jiang

Chongqing City Management College, No.151 Daxuecheng South Rd., Shapingba, Chongqing, China

---

ABSTRACT

Taking internal students of vocational college as research sample, this paper utilizes basic theory of rough sets to discuss problems of discretization and reduction process in the application of rough sets. On the basis of knowledge discovery, it proposes and designs a type of college students' network research system model based on rough sets exploring the basic thoughts of applying network management method to do analysis on network behavior. Experimental results indicate that the college students' network research system model which is based on rough sets keeps obvious effects on entropy discretization algorithm and genetic reduction algorithm.

Keywords: Rough sets; College student; Network behavior

INTRODUCTION

According to the display of "31st China Internet Development Report" released by China Internet Network Information Center (CNNIC), the number of Chinese netizen reaches 564 million and Internet penetration is 42.1% in which the rate of netizens from 20 to 29 years old reaches 30.4%, student group covers 25.1% and online seizure duration each week per capita is up to 20.5 hours[1]. The Internet has extended into each social field and presents influences on various aspects of human life. Fact indicates that in the face of virtual world constructed by Internet, contemporary college students show high degree of recognition and participation enthusiasm which becomes an important platform for them to acquire knowledge, exchange ideas and do entertainment. It is college student that is affected the most and accepts Internet the fastest. They enjoys the convenience from Internet. Meanwhile it brings some negative impacts. Many researchers also notice the serious influences and begin to do researches on relevant issues of network behavior, network morality, network psychology, etc. Through questionnaire survey and combining with statistic package SPSS analysis or by network technology, this research method solves the monitoring of emergency network behavior[2,3]**.** Network technology is less used to analyze network behavior of college students.

Rough Sets is a type of theoretical method proposed by Professor Z. Pawlak of Warsaw University of Technology in Poland in the early 1980s which researches incomplete and uncertain knowledge and studies the expression, learning and conclusion of data[4]. At present rough sets has become one of the most active research fields in computer science and kept developing in various application areas for example medical data analysis, geography, vibration analysis, language identification, approximate classification, fault diagnosis and cost forecast, etc.

Therefore this paper takes college students of vocational colleges as research sample and uses the basic theory and method of rough sets to propose the basic thoughts of utilizing network management method to do research on network behavior and design a type of college students' network research system model based on rough sets.

BASIC THEORY OF ROUGH SETS

According to the involved content of system model research, several basic concepts in rough sets are briefly introduced[5].

2.1 ***Information System***

Knowledge expression in rough sets generally applies Information Table or Information System which could be represented as quadruples *S*=（*U,A,V,f*）. Here *U* is called discourse domain, *A* is total attribute, *A*=*C*U*D, C* is finite set composed of condition attributes and *D* is finite set composed of decision attributes. *V* stands for the set made up of total $a\in A$ range. *f* is an information function which provides every attribute of each object with one information value.

2.2 ***Indiscernible Relation***

Indiscernible relation is the basis of rough sets. In terms of some elements in Discourse Domain *U*, there may exists connection among a number of same information. Therefore a type of indiscernible relation lies among these elements on the view of the known information. In other words, some objects in *U* keep several same or equal attribute value thus these objects could not be discerned just according to these attribute values. Therefore they cannot be discerned from this point. In terms of the given attribute set $R\in A$, the binary equivalent relation *R* is formed which is also called *R* Indiscernible Relation.

2.3 ***Attribute Reduction and Core***

For the unknown knowledge contained in a number of data which is found through utilizing rough sets, not all the existing condition attributes are essential. In fact, some of them are redundant. The original classification result may not be influenced after deleting these redundant attributes. On the other hand, redundant data has to be effectively deleted in order to find knowledge from a huge number of data, improve the discovery efficiency and reduce the noise disturbance. Therefore Data Reduction must be done on the data whose purpose is to delete the redundant data in system under the condition of keeping original classification capacity of information system and leading basic attribute of approximate space to be complete.

Attribute reduction is an important function in rough sets. In information table, one condition attribute corresponds with a piece of equivalence relation. All condition attributes would do a partition on the whole discourse domain *U*/*C.* Also all decision attributes do a partition on the whole discourse domain *U*/*D.* The target of attribute reduction is to delete unnecessary or unimportant attributes under the condition of remaining classification capacity unchanged so that part of essential condition attributes would keep the same classification capacity with that of decision attribute *D*. Intersection of reduction is called Core.

2.4 ***Significance of Attribute***

Significance of attribute is a key concept for problems of discretization and reduction. Generally significance of attribute could be defined like this. As *Y* is the classification derived by decision attribute set *D,* significance of attribute subset *X'* in the attribute set *X* may be defined as $R_X(Y)-\ R_{X\setminus X'}(Y)$ (Here $X'\supseteq X$. If attribute set *X* is the whole condition attribute set, this may be called "Significance of Attribute Subset *X'* " in short.). It stands for the influences on approximate classification *Y* when attribute subset *X'* is taken out of attribute set *X.*

NETWORK BEHAVIOR RESEARCH SYSTEM MODEL BASED ON ROUGH SET

Network behavior research based on rough sets is to do analysis on college students' network behavior through training sample data experiment so that it may provide guidance for the management of college students' network behavior and help realize the goal to specify their network behavior. For this purpose, the following basic work has to be done:

(1) Be sure of the requirements of network behavior research. Acquire the original data of college students' network behavior under the existing network environment.
(2) Design the general model of college students' network behavior research system based on rough sets.

According to the knowledge discovery process based on rough sets and combining the research target, we design a type of college students' network research system model based on rough sets (as shown in Figure 1.).

Firstly through requirements analyzing, network behavior research method based on rough sets clarifies the main problems of network behavior research to determine the original data of research sample. Secondly according to the research requirements of network behavior, it obtains sample original data through checking the monthly online history in the billing system database of campus network to design and establish decision tables including group everyday network behavior, group web site network behavior, special individual network behavior, etc. Thirdly in accordance with rough sets methods, it does a series of operation like preprocessing, discretization and attribute reduction on the decision tables to generate decision rule set. Fourthly it extracts the concentrated rules to get results or knowledge of network behavior research. Finally it applies them to the management tests and analyses of network behavior in order to provide decision-making guidance for network behavior management.

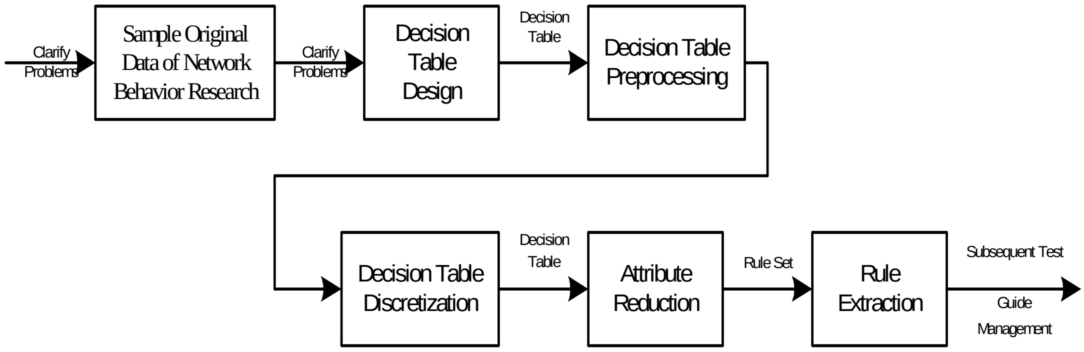


Figure 1. Network Behavior Research System Model Based on Rough Sets

The whole designed system keeps 6 functional modules including data acquisition function, decision table design function, decision table preprocessing function, decision table discretization function, attribute reduction function and rule extraction function. Actually network behavior research based on rough sets is a knowledge discovery and application process. Although each functional module is relatively Independent during the process of realizing network behavior research system, its operation runs with the process shown in Fig.1. Each functional module designs various network behavior research strategies and processing methods according to its requirements. Users choose different research contents and methods according to their specific requirements to realize the research targets. System development environment: Windows 7+vs2010+.net framework 4.0.

EXPERIMENT AND ANALYSIS OF COLLEGE STUDENTS' NETWORK BEHAVIOR RESEARCH

According to the basic condition of the billing system in the college, 4,825,729 pieces of Internet access records are acquired in which IP records of non-student network activity areas like office area and teachers' family area are filtered out.

4.1 *Design of Decision Table*

Research experiment adopts rough sets experiment tools RSES whose object processing needs realizing by the component design of information system. According to research requirements, the design of decision table mainly takes everyday network behavior of research group as accordance. Functional diagram of decision table design is Figure 2.

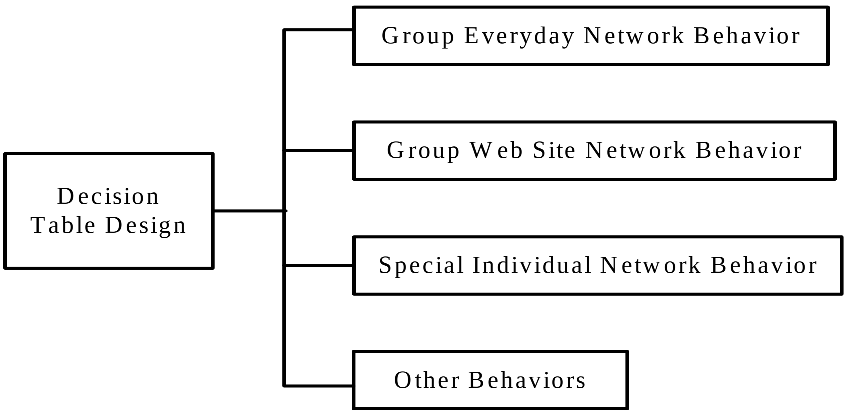


Figure 2. Functional Diagram of Decision Table Design

Object determination of decision table: This design takes students' everyday access time quantum as objects of information system (This model defines 1 hour as 1 time quantum. It may be reset according to specific situation.).

Attribute determination of decision table: This design takes access type namely network behavior as condition attribute of information system in which students' behavior at each time quantum would be set to be decision attribute.

Then the information system namely decision table is designed. It is for the research serving students' network behavior at each time quantum during the 24 hours per day. In this tetrad S=< U, A, V, f >, U stands for all time quantums per day which takes hour as one unit. V is the range of attribute a. Information function f describes students' network behavior during a certain time quantum.

4.2 *Decision Table Preprocessing*

Preprocessing function mainly designs problems of deleting repetition, completing data, encoding attribute, analyzing consistency, etc [6]. After preprocessing is finished, decision attribute value begins to do encoding description in which rest is 0, having classes is 1, others is 2 and self-study is 3. Then we acquire the decision table in which RSES is imported (seen in Figure 3.). In this figure Row"24/15"and Value $O:1$- $O:24$ describes conditions during 24 hours of a day and Row "15" describes the decision attribute value after encoded.

Table: Network Behavior_ Time Quantum

24 / 15	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
O:1	0.33	0.21	2.62	0.11	1.01	3.22	0.38	0.15	0.03	1.31	0.02	0.02	0.13	0.12	0
O:2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
O:3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
O:4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
O:5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
O:6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
O:7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
O:8	1.64	0.61	1.63	0.33	0.64	5.55	1.84	7.94	0.54	2.03	0.07	0.00	0.77	0.53	1
O:9	19.65	13.57	81.22	4.48	4.38	12.43	6.22	15.32	1.31	3.92	0.39	0.72	0.42	0.03	1
O:10	18.36	12.63	92.11	2.50	6.44	16.20	8.21	18.98	3.28	5.21	1.93	3.30	8.52	0.18	1
O:11	16.33	15.33	113.31	3.02	6.23	19.32	5.98	11.02	3.56	3.95	1.64	2.74	4.92	0.15	1
O:12	14.89	14.33	101.29	2.99	7.41	18.33	12.71	13.20	2.97	4.79	1.43	3.20	4.36	0.77	2
O:13	2.17	3.11	7.34	0.64	5.22	10.33	4.30	3.90	1.04	2.10	0.23	1.86	0.98	0.22	0
O:14	3.22	1.95	6.82	0.79	4.69	9.96	3.59	4.93	0.99	2.72	0.54	1.37	1.20	0.65	0
O:15	1.79	1.79	33.20	0.51	3.91	10.06	2.17	6.79	0.63	1.83	0.28	1.91	2.22	0.47	1
O:16	20.82	11.97	125.40	5.32	8.94	23.39	8.52	19.62	3.89	8.92	1.83	3.58	7.95	0.79	1
O:17	18.94	11.08	129.80	4.98	9.76	27.42	11.68	17.84	2.98	9.23	3.18	4.82	8.24	0.63	1
O:18	16.79	9.87	99.33	7.32	8.39	22.83	8.95	30.27	2.92	9.84	3.30	4.40	6.68	0.74	2
O:19	17.74	8.22	90.04	5.13	6.19	19.02	6.22	17.21	2.63	7.54	1.96	2.23	3.49	0.60	3
O:20	14.19	18.64	239.93	2.84	9.12	268.31	7.93	65.29	3.85	8.70	0.89	2.93	3.07	0.98	3
O:21	24.76	27.60	235.80	9.40	11.80	219.62	18.26	73.25	7.48	72.42	2.43	7.11	8.33	0.76	3
O:22	19.54	18.35	193.70	8.96	10.94	263.82	15.33	143.26	9.83	73.25	1.73	8.77	8.52	0.94	3
O:23	11.13	17.93	218.05	4.23	18.48	232.80	4.26	114.35	2.79	7.99	0.28	1.42	2.21	1.08	2
O:24	3.30	10.24	20.98	1.02	2.18	42.85	2.42	62.25	0.80	2.16	0.29	1.21	3.29	1.43	0

Figure 3. Decision Table Imported with RSES

4.3 *Decision Table Discretization*

Experiment utilizes discretization method Rosetta[7] to do calculation on breaking point thus produces decision table document of discretization. Decision table of information entropy algorithm after discretization is shown in Figure 4.

Table: Network Behavior_ Time Quantum

24 / 15	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
O:1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O:2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O:3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O:4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O:5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O:6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O:7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O:8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
O:9	7	1	0	4	0	0	2	1	0	0	0	0	0	0	1
O:10	5	1	1	0	1	0	4	3	2	2	5	5	9	0	1
O:11	2	3	3	2	1	3	0	0	2	0	3	2	5	0	1
O:12	1	2	2	1	2	1	7	0	1	1	2	4	4	6	2
O:13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O:14	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
O:15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
O:16	7	1	3	6	3	5	4	3	4	6	5	5	7	7	1
O:17	5	1	3	4	5	5	6	3	2	6	7	7	7	2	1
O:18	3	0	2	7	2	4	5	4	1	7	8	6	6	4	2
O:19	4	0	0	5	0	2	1	2	0	3	6	1	3	1	3
O:20	0	5	6	0	4	9	3	6	3	5	1	3	1	8	3
O:21	8	5	6	8	6	7	8	6	5	8	6	8	8	5	3
O:22	6	5	4	8	6	9	8	8	5	8	4	8	10	8	3
O:23	0	4	5	3	7	8	0	7	1	4	0	0	0	9	2
O:24	0	0	0	0	0	6	0	5	0	0	0	0	2	10	0

Figure 4. Decision Table of Information Entropy Algorithm after Discretization

Table 1. describes the relevant discretization value of condition attribute "online music/television" after it is calculated by information entropy discretization algorithm.

Table 1. Interval Description of"Online Music/Television"after Information Entropy Algorithm Discretization

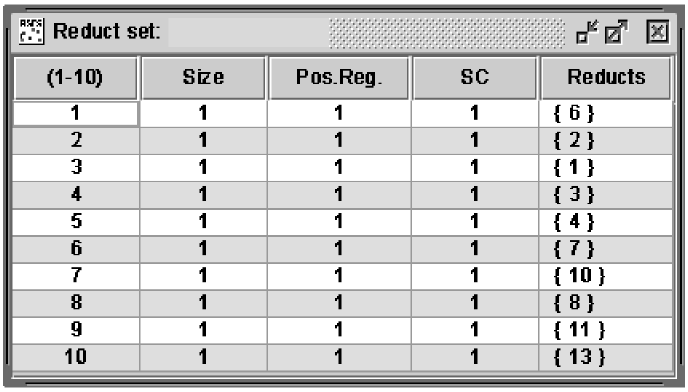
Interval Description	Description Value	Interval Description	Description Value
（0,17.266）	0	（23.111, 35.1345）	5
（17.266, 18.6735）	1	（35.1345, 131.235）	6
（18.6735, 19.171）	2	（131.235, 226.208）	7
（19.171, 21.077）	3	（226.208, 248.306）	8
（21.077, 23.111）	4	（248.306,*）	9

Notes：
a. Condition attribute "online music/television" produces 9 breaking points and 10 partition intervals after it is calculated by information entropy discretization algorithm;
b. (0,17.266) in the table stands for all the values whose intervals are less than 17.266 which is the first breaking point;
c. (248.306,*) in the table stands for all the values whose intervals are more than 248.306.

4.4 *Attribute Reduction*
According to three types of attribute reduction methods[8] including greedy algorithm, genetic algorithm and dynamic reduction, the experiment does comparison among results from different discretizations and calculates possible reduction values of each method which is shown in Table 2.

Table 2. Reduction Value Comparison of Three Attribute Reduction Methods

Reduction Algorithm	Greedy Algorithm					Genetic Algorithm					Dynamic Reduction Method				
Discretization Algorithm	equidistance	equifrequency	Naive	Semi Naive	information entropy	equidistance	equifrequency	Naive	Semi Naive	information entropy	equidistance	equifrequency	Naive	Semi Naive	information entropy
Possible Reduction Value	25	115	13	13	25	10	10	10	6	8	25	115	13	13	25


Reduct set:

(1-10)	Size	Pos.Reg.	SC	Reducts
1	1	1	1	{ 6 }
2	1	1	1	{ 2 }
3	1	1	1	{ 1 }
4	1	1	1	{ 3 }
5	1	1	1	{ 4 }
6	1	1	1	{ 7 }
7	1	1	1	{ 10 }
8	1	1	1	{ 8 }
9	1	1	1	{ 11 }
10	1	1	1	{ 13 }

Figure 5. Decision Table Results Acquired after Genetic Algorithm Reduction and Naïve Discretization

Figure 5. describes the decision table results acquired through genetic algorithm reduction which is operated after the Naïve Algorithm discretization is done[9]. Here Row"（1-10）"describes possible reduction values. Row "Size" describes the number of attribute in the reduction set. Row"Pos.Reg."describes the positive region value after reduction. Row "SC" describes stability coefficient of the reduction. Row "Reducts" describes specific attribute value condition of each reduction set.

4.5 *Rule Extraction*
Results of number of rules are acquired through different discretization algorithms and reduction methods which is shown in Table 3.

Table 3. Comparison of Number of Rules Acquired through Three Types of Attribute Reduction Methods

Reduction Algorithm	Greedy Algorithm					Genetic Algorithm					Dynamic Reduction Method				
Discretization Algorithm	equidistance	equifrequency	Naive	Semi Naive	information entropy	equidistance	equifrequency	Naive	Semi Naive	information entropy	equidistance	equifrequency	Naive	Semi Naive	information entropy
Number of Rules	435	2099	170	170	373	172	182	131	72	118	435	2099	170	170	373

Figure 6. describes the rule set results acquired through genetic algorithm reduction which is operated after the Naïve Algorithm discretization is done. Here Row"（1-131）"describes the size of rule set in which there exsits 131 pieces of rules. Row"Match"describes the object matching number of one certain piece of rule. Row"Decision Rules"describes the specific condition of a certain rule. The first piece of rule in this figure explains that supporting record number of rules in the state of "rest" is 7 when the value of network behavior "online music/television" is 0.

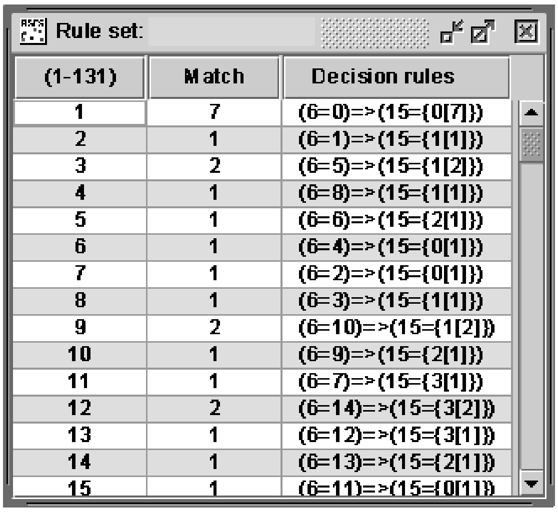


Figure 6. Rule Set Results Acquired after Genetic Algorithm Reduction and Naïve Discretization

4.6 *Data Analysis*

It is seen from the decision supporting condition of rule set acquired in the experiment that all the acquired supporting degree figures of rule set are similar with those in Fig.7 in appearance no matter which kind of method is used.

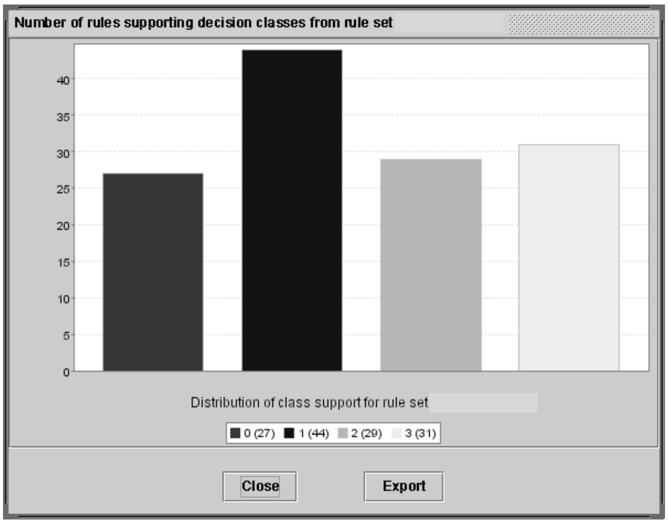


Figure 7. Supporting Degree Analysis of Rule Set Produced by Decision Table after Calculated through Genetic Algorithm Reduction and Naïve Discretization

Figure 7. shows that everyday network behavior of college students distributes relatively reasonable during the “having classes” time quantum. In this figure supporting rules of “1” are 44.

It is seen from the whole experiment that research results of college students’ everyday network behavior in this system are acquired through utilizing the system model of college students’ network behavior research based on rough sets and these results are applied to the management of college students’ network behavior. Through the final training or test verification, the validity and feasibility of this design are proved. There still exist many uncertain factors. For example students surf the Internet in bars, use agencies, there exists inconformity between decision attributes and timetable of part of students(some in class, others not.), etc. However, rules extracted keep certain generalization capacity and get favorable analyzing results facing with the influence of noises. Moreover it is seen from experiment that discretization algorithm and attribute reduction method of rough sets continuous attributes produce different effects in different applications. In this experiment, effects of information entropy discretization algorithm and genetic reduction algorithm are more obvious.

CONCLUSION

This paper establishes a type of application system model based on rough sets which only achieves a small number of targets for the analysis system of network behavior research and acquires some initial and periodical achievements. This provides a good foundation for the following decision support system for management of network behavior. There still needs more study and practice in the aspects of realizing algorithm and system optimization such as the improvement of parallel processing of rough analysis. As an emerging soft computing, rough sets has been respected by many fields for example computer, mathematics, artificial intelligence, controlling, etc. A great variety of research achievements are produced and widely applied in many fields. With further research and development of rough sets, more new research problems and directions would come into being.

REFERENCES

[1]. Yang G, He Q, Deng X. *International Journal of Advancements in Computing Technology*, **2011**, 3(10).pp. 266-273.
[2]. C. Nathalie Yuen, Michael J. Lavin, *CyberPsychology&Behavior*, **2004**,7(04),379-383.
[3]. Matthew S. Eastin, Teen Internet Use: Relating Social Perceptions and Cognitive Models to Behavior, *CyberPsychology&Behavior*,**2005**, 8(01),62-73.
[4]. Pawlak Z, Rough set, *International Journal of Computer and Information Sciences*,**1982**,11(05), 341-356.
[5]. YANG Gelan, Yue WU, and Huixia JIN. *Journal of Computational Information Systems*, **2012**, 8(10): 4315-4322.
[6]. Wang Guoyin, Rough Set Theory and Knowledge Acquisition,Xian: Xi’an Jiaotong University Press, **2001**,100-101.
[7]. Yang G., Le D., Jin Y., Cao S. Q. *International Journal of Distributed Sensor Networks*, **2014**, vol. 2014, Article ID 363584, doi:10.1155/2014/363584.
[8]. Dai Jianhua Li Yuanxiang and Liu Qun, *Computer Engineering and Applications*, **2003**,39(08),13-14.
[9]. CHEN Shu-zhen, *Journal of Wuhan Polytechnic University*,**2005**,24(03),122-124.