http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 11 (4): 508-511, 2012
ISSN 1812-5638 / DOI: 10.3923/itj.2012.508.511

An Analysis of Teaching Quality and Employment Status Based on Rough Set

Li Shunxin and Wang Changhong
College of Computer Science and Technology,
Wuhan University of Science and Technology, Wuhan, 430065, China

Abstract: With the popularization of higher education, the employment status of college graduates has become a social hot spot issue. This study extracts the information of teaching quality and employment status of colleges and universities by using data mining technology based on rough set theory. By analyzing the influence of various aspects in teaching upon employment status, this paper learns relevant rules and establishes an evaluation model.

Key words: Rough set, attribute reduction, rule extraction, importance degree

INTRODUCTION

Rough set theory was proposed by the Polish logician in 1982 which is a new mathematical tool to solve the problems with ambiguity and imprecision. Ever since it was published, it has become a new and the most important field of research with rapid development both in theory and application. Rough set theory provides a new and effective mathematical method for machine learning, knowledge acquisition, decision analysis, knowledge discovery in databases, expert systems, decision support systems, inductive reasoning, pattern recognition, fuzzy control and application in other aspects (Jensen and Shen, 2002).

Today, it is an important foundation that solving the problem of employment so as to build a harmonious society, and how to improve the employment rate and employment quality of schools has been a tough question. This study takes a series of indicators which impact on employment as the research objects, and analyzes the data by using the Rough Set-related methods to get the important degree of individual indicators and the corresponding decision-making rules. In the end, finding out the potential associated information among the various conditions, which will undoubtedly play a guiding role in reforming higher education and improving the teaching quality.

BASIC CONCEPT

Definition 1: The Confidence of Confidence Rule A→B (Confidence) is defined as: Cf (A→B) = | X∩Y | / | X |, in which x = {x∧A_x}, Y = {x∧B_x}, A_x means the condition attribute value of sample x satisfies with the formula A, B_x identifies the i decision attribute value of sample x satisfies with the formula B. In the other words, the set X is the set of samples whose condition attribute values meet the formula A , and the set Y is the set of samples whose decision attribute values meet the formula B (Jiang *et al*., 2010).

Definition 2: Let U be a domain, P and Q be two clusters (attribute set) of equivalence relation on U and the partitions of P and Q induced from U be X and Y respectively, X = {x_1, x_2, ..., x_n}, Y = {Y_1, Y_2, ..., Y_m}, The probability distribution of P and Q on the σ algebraic consisted by subset of U can be defined as:

$$(X:P)=\begin{bmatrix} X_1 & X_2 & \cdots & X_n \\ P(X_1) & P(X_2) & \cdots & [P(X]_n) \end{bmatrix}\ (Y:P)=\begin{bmatrix} Y_1 & Y_2 & \cdots & Y_m \\ P(Y_1) & P(Y_2) & \cdots & [P(Y]_m) \end{bmatrix} \quad (1)$$

Where:

$$P(X_i)=\frac{|X_i|}{|U|},\ i=1,\ldots n;\ \ P(Y_j)=\frac{|Y_j|}{|U|},\ j=1,\ldots m.$$

According to information theory, Information Entropy of knowledge P(H (P)) can be defined as:

$$H(P)=-\sum_{i=1}^{n} P(X_i)\log(P(X_i))$$

The relevant conditional entropy (H (Q|P)) of Knowledge Q and knowledge P can be defined:

Corresponding Author: Li Shunxin, College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, 430065, China

$$H(Q|P) = -\sum_{i=1}^{n} P(X_i) \sum_{j=1}^{m} P(Y_j|X_i) \log(P(Y_j|X_i))$$

where:

$$(P(Y_j|X_i) = \frac{|X_i \cap Y_j|}{|X_i|}$$

i = 1, ... n; j = 1, ... m (Chang *et al*., 1999).

Definition 3: Let T = (U, C ∪ D, V, f) be a decision table, where R⊆C. The importance (SGF (a, R, D)) of Any attribute (a, C-R) can be defined as (Wu and Gou, 2011):

$$SGF(a, R, D) = I(R \cup \{a\}; D] - I(R; D) = H(D|R \cup \{a\}) \quad (2)$$

ANALYSIS OF SAMPLES

Get data: Take the data of one class as sample and discrete the data (Table 1).

There are thirty data in the sample. In the property set V = {a, b, c, d, e, f, g, s} the condition attribute set is C = {a, b, c, d, e, f, g} and the decision attribute set is D = {s}. a→prize-winning status (1 stands for have got awards, 0 stands for no); b→be or not be student leaders (1 stands for yes, 0 stands no); c→scores of professional course (1 stands for weighted average score is good, 0 stands for bad); d→ideological situation (1 stands for good, 0 stands for bad active or healthy); e→scores of English (1 stands for weighted average score is good, 0 stands for no); f→normal graduation (1 stands for yes, 0 stands no); g→join in student unions (1 stands for yes, 0 stands for no); s→employment status (1 stands for the salary is high, 0 stands for low).

Make data consistency and divide the equivalence classes: Remove the inconsistent decisions items (Witlox and Tindemans, 2004) in the table. Divide the decision table into equivalence classes (Table 2).

A new decision table that is the same quality of classification as the original by deleting redundant rules are shown in the Table 3.

Calculate the importance degree of attribute (Amitava and Sankar, 2003):

- U|IND({s}) = {{x1, x5, x7, x10, x11},{x2, x3, x4, x6, x8, x9}}
- U|S = {{x1, x7, x10, x11},{x2, x3, x4, x6, x9}}
- U|IND({a,b,c,d,e,f,g})={x1,x2,x3,x4,{x5,x8},x6,x7,x9,x10,x11}
- U/IND({b,c,d,e,f,g})={x1,{x2,x11},{x3,x6},x4,{x5,x8},x7,x9,x10}

So the importance degree of attribute 'a' is:

- SGF (a,{b, c, d, e, g, f},{s}) = H ({s}|{b, c, d, e, g, f})-H ({ s } | { a , b , c , d , e , g , f }) = -2/11[(1/2)log(1/2)+(1/2)log(1/2)]-2/11[(1/2) log (1/2)+(1/2) log (1/2)]+2/11[(1/2) log (1/2)+(1/2) log (1/2)] = 0.1260

By the same token we can get the importance degree of other attributes:

- SGF(b,{a,c,d,e,f,g},{s})=0.1736 SGF(c,{a,b,d,e,f,g},{s})=0.2521 SGF(d,{a,b,c,e,f,g},{s})=0 SGF(e,{a,b,c,d,f,g},{s})=0SGF(f,{a,b,c,d,e,g},{s})=0 SGF(g,{a,b,c,d,e,f},{s}) = 0.1736

It can be learned that the importance of attribute 'c' is the maximum value of 0.2521, therefore 'c' is the most importance attribute in the attribute set {a,b,c,d,e,f,g}.The result shows that we should pay special attention to the teaching of professional courses, for the students' capabilities of professional courses will directly influence on their future employment and development.

REDUCTS

Attribute reduction: Remove the redundant attributes whose values are whole '0' or whole '1' (d , f). Remove the inconsistent decisions items (x8, x5).The new decision table are shown in Table 4.

Table 1: Sample and discrete the data

U	a	b	c	d	e	f	g	s	U	a	b	c	d	e	f	g	s	U	a	b	c	d	e	f	g	s
x1	1	1	1	1	0	1	1	1	x11	0	0	1	1	1	1	0	1	x21	0	0	0	1	1	1	1	0
x2	0	1	1	1	1	1	1	0	x12	0	0	0	1	1	1	1	0	x22	0	1	1	1	1	1	0	1
x3	1	1	0	1	0	1	1	0	x13	1	1	0	1	1	1	0	0	x23	1	1	1	1	1	1	1	1
x4	0	0	0	1	1	1	1	0	x14	0	0	1	1	1	1	0	1	x24	1	1	1	1	1	1	1	1
x5	1	1	1	1	0	1	1	1	x15	1	1	0	1	0	1	1	0	x25	0	0	0	1	1	1	1	0
x6	0	0	1	1	1	1	0	1	x16	0	0	1	1	1	1	0	0	x26	1	1	0	1	0	1	1	0
x7	0	1	0	1	0	1	1	0	x17	0	1	1	1	1	1	0	1	x27	0	1	0	1	0	1	1	0
x8	0	0	1	1	1	1	1	1	x18	1	1	1	1	0	1	1	1	x28	0	1	1	1	1	1	0	1
x9	1	1	0	1	0	1	1	0	x19	0	0	0	1	1	1	1	0	x29	0	1	1	1	1	1	0	1
x10	0	0	1	1	1	1	0	0	x20	1	1	0	1	0	1	1	0	x30	1	1	1	1	1	1	1	1

Table 2: Division of the decision table into equivalence classes

Equal class	a	b	c	d	e	f	g	s
x1,x5,x18	1	1	1	1	0	1	1	1
x2	0	1	1	1	1	1	1	0
x3,x9,x15,x20,x26	1	1	0	1	0	1	1	0
x4,x12,x19,x21,x25	0	0	0	1	1	1	1	0
x6,x11,x14	0	0	1	1	1	1	0	1
x7,x27	0	1	0	1	0	1	1	0
x8	0	0	1	1	1	1	1	1
x10,x16	0	0	1	1	1	1	0	0
x13	1	1	0	1	1	1	0	0
x17,x22,x28,x29	0	1	1	1	1	1	0	1
x23,x24,x30	1	1	1	1	1	1	1	1

Table 3: A new decision table having same quality of classification

U	a	b	c	d	e	f	g	s
x1	1	1	1	1	0	1	1	1
x2	0	1	1	1	1	1	1	0
x3	1	1	0	1	0	1	1	0
x4	0	0	0	1	1	1	1	0
x5	0	0	1	1	1	1	0	1
x6	0	1	0	1	0	1	1	0
x7	0	0	1	1	1	1	1	1
X8	0	0	1	1	1	1	0	0
x9	1	1	0	1	1	1	0	0
x10	0	1	1	1	1	1	0	1
x11	1	1	1	1	1	1	1	1

Table 4: The new decision table

E	a	b	c	e	g	s
x1	1	1	1	0	1	1
x2	0	1	1	1	1	0
x3	1	1	0	0	1	0
x4	0	0	0	1	1	0
x6	0	1	0	0	1	0
x7	0	0	1	1	1	1
x9	1	1	0	1	0	0
x10	0	1	1	1	0	1
x11	1	1	1	1	1	1

Table 5: New decision table same quality as the original based on the equivalence classes

E	a	b	c	g	s
x1	1	1	1	1	1
x2	0	1	1	1	0
x3	1	1	0	1	0
x4	0	0	0	1	0
x6	0	1	0	1	0
x7	0	0	1	1	1
x9	1	1	0	0	0
x10	0	1	1	0	1

Get the core properties of the decision table:

- U|R=U|{a,b,c,e,g}={x1,x2,x3,x4,x6,x7,x9,x10,x11}
- U|S={{x1,x7,x10,x11},{x2,x3,x4,x6,x9}}
- U|$R_{R\text{-}a}$=U|{ a,c, e,g }={x1,{x2,x11}, {x3,x6}, x4, x7, x9,x10 }
- $POS_{K\text{-}a}(S)$=$R_{R\text{-}a}$_S={x1,x3,x4,x6,x7,x9,x10} $r_{R\text{-}a}$=|$R_{R\text{-}a}$_S|/|U|=7/9

Here, we get $r_{R\text{-}a}$= 7/9 which means that after removal of condition attribute 'a' the decision table becomes inconsistent. So 'a' is core of attribute set {a,b,c,e,g}. By the same token we can get that: 'b','c' and 'g' are cores, while 'e' is redundant. Now reduce 'e',we can get new equivalence classes U/IND(a,b,c,g,s)={{x1,x11}, x2,x3,x4,x6,x7,x9,x10}. Then we get a new decision table same quality as the original based on the equivalence classes (Table 5).

Value reduction: Eliminate inaccurate values (Yan and Yang, 2007) of the property condition one by one, then check it whether it can keep the table consistent. First check (1)$a_1b_1c_1g_1 \rightarrow s_1$, $a_1b_1c_1 \rightarrow s_1$, $a_1c_1g_1 \rightarrow s_1$ they are all keep the table consistent while $b_1c_1g_1$ is oppose to (2) and $a_1b_1g_1 \rightarrow s_1$ is opposed to (3) (Xu *et al*., 2008). So the attribute (a,1) and (c,1) can not be reduced. By the same token we can get the reduced table as Table 6 (Only have the core value of the decision table).

$[1]_s$ = {1,7,10}; $[1_a]$ = {1,3,9}, $[1_b]$ = {1,2,3,6,9,10}, $[1_c]$ = {1,2,7,10}, $[1_g]$ = {1,2,3,4, 6,7}, obviously $[1]_a \cap [1]_c \subseteq [1]_s$ so we can get a reduced decision rule : from x1.

By the same token we can get x_2: $a_u b_1 g_1 \rightarrow s_0$ from x2; x_3: $c_0 \rightarrow s_0$ from x3; x_4: $c_0 \rightarrow s_0$ from x4; x_6: $a_0 b_1 g_1 \rightarrow s_0$ from x6;: $c_0 \rightarrow s_0$ from x7; x_9: a_1 g_0 and from x9';$c_0 \rightarrow s_0$: a_0 $g_0 \rightarrow s_1$ from x10. At the same time we get the new equivalence class: {{x3,x4,x6,x9'}, x1,x2,x6', x7 ,x9, x10'}. Remove the redundant rules and we get final decision rule table as follow (Table 7).

Table 6: Reduced table

E	a	b	c	g	s
x1	1	-	1	-	1
x2	0	1	-	1	0
x3	-	-	0	-	0
x4	-	-	0	-	0
x6	-	-	-	-	0
x7	-	0	1	-	1
x9	-	-	-	-	0
x10	-	-	-	0	1

Table 7: Final decision rule table

R	a	b	c	g	s
R1	1	-	1	-	1
R2	0	1	-	1	0
R3	-	-	0	-	0
R4	-	0	1	-	1
R5	1	-	-	0	0
R6	0	-	-	0	1
R7	-	-	1	0	1

CALCULATE CONFIDENCE

According to the definition 1, calculate the credibility based on the rules in Table 1 and 7:

- $CF(R_1 \rightarrow S_1)=|R_1 \cap S_1|/|R_1|=(6/30)/(6/30)=1$ 2
- $CF(R_2 \rightarrow S_0)=|R_2 \cap S_0|/|R_2|=(2/30)/(3/30)=0.67$

- $CF(R_3 \rightarrow S_0) = |R_3 \cap S_0|/|R_3| = (13/30)/(13/30) = 14$
- $CF(R_4 \rightarrow S_1) = |R_5 \cap S_1|/|R_5| = (4/30)/(6/30) = 0.67$
- $CF(R_5 \rightarrow S_0) = |R_6 \cap S_0|/|R_6| = (1/30)/(1/30) = 1\ 6$
- $CF(R_6 \rightarrow S_1) = |R_7 \cap S_1|/|R_7| = (7/30)/(9/30) = 0.78$
- $CF(R_7 \rightarrow S_1) = |R_8 \cap S_1|/|R_8| = (7/30)/(9/30) = 0.78$

COMPREHENSIVE EVALUATION

- 100% students who have gotten prizes from competitions and high scores in professional courses have a good employment status
- 67% students who have served as student leaders and joined in student unions without any prizes have a poor employment status
- 100% students who have poor performance in professional courses have a poor employment status
- 67% students who have served as student leader and gotten high score in specialization courses have a good employment status
- 100% students who have gotten prizes from competitions and joined in student unions have a poor employment status
- 78% students who have not gotten prizes from competitions or joined in student unions have a good employment status
- 78% students who have not joined in student unions but performed well in exams of professional courses have a good employment status

Based on attribute importance and the decision analysis above we can get: Professional courses teaching plays a very important role in university teaching. Therefore, professional courses teaching should be treated as the core in university education. Schools should encourage the students to participate in various competitions such as the innovation competitions or professional-related competitions to improve their ability. It is beneficial for students to join in student unions, but it is not advisable to spend too much time on joining in student unions.

In the study, the author just expounds the basic principles by computing a small amount of sample data, the results may not be entirely realistic situation. Only when the data reaches a certain degree of time can we see the actual situation. There are many issues we will confront to really put theory into practice such as the efficiency of the program we should to pay attention to as well as data acquisition and other aspects. Of course, there also has a lot accepts should to be improved, such as the complicated reduction method used in the property reduction and value reduction.

REFERENCES

Amitava, R. and K.P. Sankar, 2003. Fuzzy discretization of feature space for a rough set classifier. Pattern Recogn. Lett., 24: 895-902.

Chang, L.Y., G.Y. Wang and Y. Wu, 1999. An approach for attribute reduction and rule generation based on rough set theory. J. Software, 10: 1206-1211.

Jensen, R. and Q. Shen, 2002. Fuzzy-rough sets for descriptive dimensionality reduction. Proceedings of the IEEE International Conference on Fuzzy Systems, May 12-17, 2002, Honolulu, HI., USA., pp: 29-34.

Jiang, L.M., L. Junhui, L. Mengqiang and L. Junhui, 2010. Evaluation model of CMIS based on rough set. Distance Educ. China, 10: 66-69.

Witlox, F. and H. Tindemans, 2004. The application of rough sets analysis in activity-based modeling: Opportunities and constraints. Expert Syst. Applic., 27: 585-592.

Wu, S.Z. and P.Z. Gou, 2011. Attribute reduction algorithm on rough set and information entropy and its application. Comput. Eng., 37: 56-58.

Xu, F.F, D.Q. Miao, L. Wei, Q.R. Feng and Y.S. Bi, 2008. Algorithm for fuzzy-rough attribute reduction. J. Electron. Inform. Technol., 6: 1372-1375.

Yan, Y. and H.Z. Yang, 2007. Konwledge reduction algorithm based on mutual information. J. Tsinghua Univ. (Sci. Technol.), 47: 1903-1906.