Tamkang Journal of Science and Engineering, Vol. 10, No. 4, pp. 389–397 (2007)

Applying Data Mining Technologies to the Learning and Study Strategies Inventory

Chien-Chou Shih[1], Ding-An Chiang[2], Sheng-Wei Lai[2] and Yen-Wei Hu[3]

[1]*Department of Information and Communication, Tamkang University, Tamsui, Taiwan 251, R.O.C.*
[2]*Department of Computer Science and Information Engineering, Tamkang University, Tamsui, Taiwan 251, R.O.C.*
[3]*Center for General Education and Core Curriculum, Tamkang University, Tamsui, Taiwan 251, R.O.C.*

Abstract

To understand the thoughts and behavior of a large group of people, surveys are often used to obtain objective data. They are generally composed of dozens of questions that can be quite time-consuming, and a hassle to complete. To avoid this, the classification charts of the decision tree were used to search for critical and relevant questions, whereas the association rule reduced the number of questions within related categories. This paper applies data mining technology, which achieved the above-said outcome from performance evaluation results.

Key Words: Data Mining, Decision Tree, Association Rule, LASSI

1. Introduction

Whether a survey is successful or not is closely related to the number of survey questions included. More questions usually would mean greater accuracy in survey results. However, the number of survey questions is also one of the main factors influencing the willingness of the subject to participate. Through the technique of data mining, we can pinpoint the important characteristics of various indicators and their relationships to each other. As a result, we can also single out the problems of the counseled students and users in a minimum amount of time and number of questions. By reducing the number of questions and shortening the time required we also increase the willingness of subjects to fill out the questionnaires. This paper applies the results obtained through decision tree and association rules analysis to the Learning and Study Strategies Inventory (LASSI), and reduces number of questions needed dramatically.

This study will proceed in two phases. Phase one is that survey data collection and analysis. It is mainly to introduce the process of survey data collection and analysis. Phase two is to apply the results from phase one. In the first, we use questionnaire that concerning learning and study strategies to find out and analyze the characteristics and relations of questions. Then we utilize decision trees analysis to classify questions, and association rules analysis to relate with questions. Finally, we apply the results from analysis and related selection system to minimize the number of questions.

The remainder of this paper is organized into four sections. Section 2 describes the background knowledge about decision tree and association rule. Section 3 introduces the methods of our research. Section 4 shows the performances through two major steps. Section 5 concludes our study and discusses future directions.

2. Background Knowledge

2.1 Decision Tree Analysis

Decision tree [1–3] are based on the methodology of tree graphs and can be considered one of the more simple

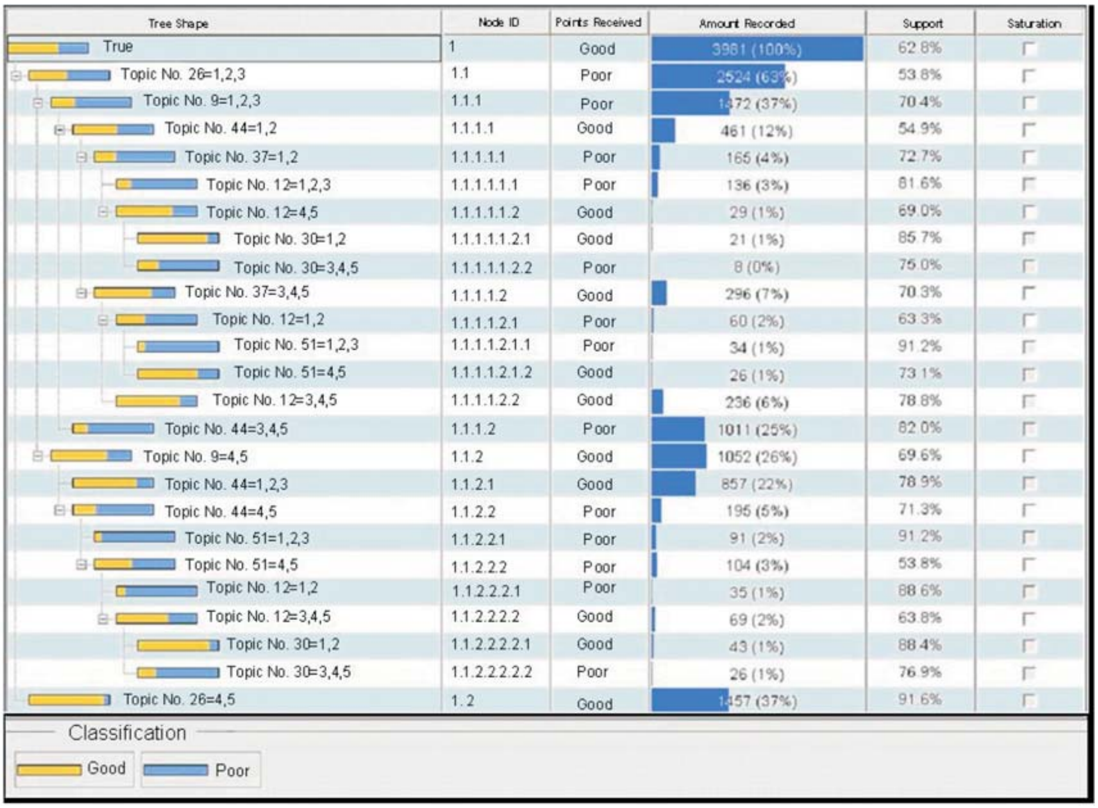
*Corresponding author. E-mail: ccs@mail.tku.edu.tw

inductive study methods. The decision tree is produced through a training set which can be constituted by known material and whose analysis results are based upon different sequences of variables. It can be used to predict unknown data and solve any problems faced. Decision tree analysis is a very suitable tool for users in data analysis.

The Figure 1 shows an example of results produced from a decision tree analysis for motivation scale (i.e. category) in study strategy. It can be transformed to storage in a database easily, which also can be applied for user inquiries. Therefore, this paper will use decision tree analysis to classify questions of survey. In Figure 1, column one is "Tree Shape" which shows all of related topics (i.e. questions) with motivation scale. The conditions connect with the color bar. For instance, "Topic No. 26=1,2,3" means that the node for answers 1, 2 or 3 on topic 26. Column two is "Node ID" which shows the visiting tree paths. Column three is "Points Received" which is a score for the topic. It shows poor's color bar longer than good's color bar in column one, so the "Points Received" is poor, and vice versa. Column four is "Amount Recorded" which is the proportion for the topic's answer. Each point has to inherit predecessor's conditions. For instance, the third node "Topic No. 9= 1,2,3" and "Topic No. 9=4,5" have to satisfy the conditions of "Topic No. 26=1,2,3". In other words, the value of "Topic No. 26=1,2,3" will equal "Topic No. 9=1,2,3" plus "Topic No. 9=4,5". Column five is "Support" which is the proportion for satisfy the conditions. For instance, "Topic No. 26=1,2,3" support is 53.8% which means 53.8% record's "points received" are poor in this conditions.

2.2 Association Rule Analysis

Association rule [4–11] analysis were first used to research the transaction records of market basket date [3] and then used to analyze the purchase behavior of customers. Relationships between different commodity groups were thus discovered and applied in considering the inventory and ornamentation of storage locations. The significance of an association rule is judged based on support and confidence, which must be greater than the lowest threshold. In addition to support and confidence, lift is another important factor to determine the effect of



Tree Shape	Node ID	Points Received	Amount Recorded	Support	Saturation
True	1	Good	3981 (100%)	62.8%	
Topic No. 26=1,2,3	1.1	Poor	2524 (63%)	53.8%	
Topic No. 9=1,2,3	1.1.1	Poor	1472 (37%)	70.4%	
Topic No. 44=1,2	1.1.1.1	Good	461 (12%)	54.9%	
Topic No. 37=1,2	1.1.1.1.1	Poor	165 (4%)	72.7%	
Topic No. 12=1,2,3	1.1.1.1.1.1	Poor	136 (3%)	81.6%	
Topic No. 12=4,5	1.1.1.1.1.2	Good	29 (1%)	69.0%	
Topic No. 30=1,2	1.1.1.1.1.2.1	Good	21 (1%)	85.7%	
Topic No. 30=3,4,5	1.1.1.1.1.2.2	Poor	8 (0%)	75.0%	
Topic No. 37=3,4,5	1.1.1.1.2	Good	296 (7%)	70.3%	
Topic No. 12=1,2	1.1.1.1.2.1	Poor	60 (2%)	63.3%	
Topic No. 51=1,2,3	1.1.1.1.2.1.1	Poor	34 (1%)	91.2%	
Topic No. 51=4,5	1.1.1.1.2.1.2	Good	26 (1%)	73.1%	
Topic No. 12=3,4,5	1.1.1.1.2.2	Good	236 (6%)	78.8%	
Topic No. 44=3,4,5	1.1.1.2	Poor	1011 (25%)	82.0%	
Topic No. 9=4,5	1.1.2	Good	1052 (26%)	69.6%	
Topic No. 44=1,2,3	1.1.2.1	Good	857 (22%)	78.9%	
Topic No. 44=4,5	1.1.2.2	Poor	195 (5%)	71.3%	
Topic No. 51=1,2,3	1.1.2.2.1	Poor	91 (2%)	91.2%	
Topic No. 51=4,5	1.1.2.2.2	Poor	104 (3%)	53.8%	
Topic No. 12=1,2	1.1.2.2.2.1	Poor	35 (1%)	88.6%	
Topic No. 12=3,4,5	1.1.2.2.2.2	Good	69 (2%)	63.8%	
Topic No. 30=1,2	1.1.2.2.2.2.1	Good	43 (1%)	88.4%	
Topic No. 30=3,4,5	1.1.2.2.2.2.2	Poor	26 (1%)	76.9%	
Topic No. 26=4,5	1.2	Good	1457 (37%)	91.6%	

Classification

Good Poor

Figure 1. The results from decision tree analysis for motivation study strategy.

association rule. It is used to strengthen the positive correlation. Lift is defined as follow:

$$\text{Lift}\ (\textit{condition} \rightarrow \textit{result}) = \frac{\text{Confidence}\ (\textit{condition} ==> \textit{result})}{\text{Support}\ (\textit{result})}$$

It is the ration of the density of the target to density of target overall. The bigger lift is more valuable. The Figure 2 shows a part of results produced from association rule analysis. It can be applied to find out the correlation of survey's questions. In Figure 2, column one is "rule" which found out from association rule analysis. For instance, the first rule is "[Poor Motivation] → [Poor Attitude]" which means that if someone's study strategy for motivation is poor, it will have effects on attitude which will also be poor. Column two is "Support" which means the proportion of records that satisfies the rule. For instance, the first rule's "Support" is 28.7365% which means 28.7365 records are "Poor Motivation" in any 100 records. Column three is "Confidence" which means the probability of that satisfied two conditions. For instance, the first rule's "Confidence" is 77.3% which means there are 77.3% [Poor Attitude] in all of the [Poor Motivation]. Column four is "Lift" which is defined above. The "Support" of [Poor Attitude] is 56.7%. According to the definition of lift, the first rule is 1.3635. When lift is greater than 1, then the resulting rule is better at predicting the result than guessing whether the resultant item is present based on item frequencies in the data. When lift is less than 1, which is the negative correlation. The rule is doing worse than informed guessing. So, all of them are useful rules except the third rule in Figure 2.

3. Research Methods

This paper utilized corpora from LASSI. To find out the characteristics and correlation for all of questions in study strategy through analysis of decision tree and association rule. The target of the research is to minimize questions used to find obstacles in study strategy.

3.1 LASSI for University Students

The resources of this survey are found in the "LASSI for University Freshmen" provided by the Consultation and Guidance Division of Student Affairs Office at Tamkang University. The LASSI is a 11-scale (i.e. category), 87-topic (i.e. question) assessment of students' awareness about and use of learning and study strategies as Table 1. It provides students with a diagnosis of their strengths and weaknesses, and what they can learn or enhance through educational interventions such as learning and study skills courses. The LASSI is scored using the Likert Scale's five-point scale. "Strongly Disagree" scored

Table 1. The topic number for LASSI

LASSI Sub-Scale	Included Topic Numbers
Attitude	13, 16, 34, 40, 46, 64, 81
Motivation	9, 12, 26, 30, 37, 44, 51
Time	3, 20, 32, 38, 43, 53, 61, 68, 87
Anxiety	1, 8, 23, 28, 49, 52, 58, 82
Concentration	5, 10, 35, 39, 41, 50, 56, 63
Information	11, 14, 21, 29, 36, 42, 62, 70, 83
Selection	2, 7, 55, 67, 71, 84
Studying	6, 17, 22, 45, 48, 57, 85
Self	4, 15, 19, 24, 27, 33, 60, 65, 86
Testing	18, 25, 31, 47, 54, 59, 66, 69
Solving	72, 73, 74, 75, 76, 77, 78, 79, 80

▲ Rule	Support	Confidence	Lift
[Poor Motivation]==>[Poor Attitude]	28.7365%	77.3000%	1.3635
[Poor Motivation]==>[Poor Time Management]	28.1085%	75.6100%	1.3757
[Poor Motivation]==>[Poor Stress Management]	28.4351%	76.4900%	0.9880
[Poor Motivation]+[Poor Attitude]==>[Poor Time Management]	23.1600%	80.5900%	1.4663
[Poor Motivation]+[Poor Attitude]==>[Poor Stress Management]	23.3861%	81.3800%	1.0512
[Poor Motivation]+[Poor Attitude]==>[Poor Test Strategies]	22.0799%	76.8400%	1.2788
[Poor Motivation]+[Poor Attitude]==>[Poor Ability to Select Importa	21.6780%	75.4400%	1.3336
[Poor Motivation]+[Poor Stress Management]==>[Poor Attitude]	23.3861%	82.2400%	1.4506
[Poor Motivation]+[Poor Stress Management]==>[Poor Time Mana	22.3311%	78.5300%	1.4288
[Poor Motivation]+[Poor Stress Management]==>[Poor Test Strate	23.0093%	80.9200%	1.3467
[Poor Motivation]+[Poor Stress Management]==>[Poor Ability to S	21.5775%	75.8800%	1.3414

Figure 2. A part of results from association rule analysis.

one mark. "Disagree" scored two marks. "Somewhat Agree" scored three marks. "Agree" scored four marks. "Strongly Agree" scored five marks. The reverse-scoring topic (i.e. question) reverses scored. Therefore, "Strongly Disagree" scored five marks, "Disagree" scored four marks and so on. This assessment takes about 20 to 25 minutes to complete. The LASSI Sub-Scales can be applied separately or totally.

Explanations for the two columns in Table 1 as follows:

- LASSI Sub-Scales: this survey is composed of 11 sub-scales (i.e. sub-categories), each of which represents one study strategy. Every one of the survey questions directly impacts the score of one of the components.
- Included Topic Numbers: lists the topic (i.e. question) numbers that are included under each sub-scale. Underlining represents a reverse-scoring topic.

The percentage rank norms for the average university student have been simplified into Table 2 for the sake of popularization and user convenience. The top row lists study strategies and the two outsides of the table represent the percentage rank. The scores for each study strategy corresponding to the percentage ranks can be found in the columns below each heading. Percentage ranks above 50 signify good study strategies and percentage ranks below 50 signify poor study strategies. According to Table 1 to calculate the assessment for all of the LASSI sub-scale scores, and to know the status of study strategies from corresponding score in Table 2.

3.2 Decision Tree Analysis for Various Study Strategies

Decision tree analysis is one of the most widely-used inductive reasoning algorithms. It uses classification method to classify all of the questions, filtering out critical questions in the process, greatly reducing the number of questions. The data of questionnaires were originally presented in an EXCEL format. Percentage rank and status of study strategies were not included here. Therefore, we had to recalculate related parameters according to Table 1 and Table 2, and converted into TABLE format for decision tree analysis. All of study strategies format following conversion can be seen as Table 3.

Table 2. The percentage rank norms for LASSI

Percentage	Att.	Mot.	Time	Anx.	Con.	Inf.	Sel.	Stu.	Self	Tes.	Sol.	Percentage
99	33	31	39	38	33	43	28	32	38	37	42	99
95	31	28	34	34	30	39	26	30	34	34	38	95
90	29	27	32	32	28	37	25	29	33	33	36	90
85	28	26	31	31	27	36	24	28	32	32	35	85
80	28	26	30	30	26	35	23	27	31	31	34	80
75	27	25	29	29	25	34	23	26	30	30	34	75
70	26	24	29	29	25	33	23	26	30	30	33	70
65	26	24	28	28	24	33	22	25	29	29	33	65
60	25	23	27	27	24	32	22	25	29	29	32	60
55	25	23	27	27	23	31	21	24	28	28	32	55
50	24	23	26	26	23	31	21	24	28	28	31	50
45	24	22	25	26	22	30	21	23	27	27	31	45
40	23	22	25	25	22	30	20	23	27	27	30	40
35	23	21	24	24	21	29	20	22	26	26	30	35
30	22	21	23	24	20	28	19	22	26	26	29	30
25	22	20	23	23	20	28	19	21	25	25	29	25
20	21	20	22	22	19	27	19	21	25	25	28	20
15	20	19	21	21	18	26	18	20	24	24	27	15
10	19	18	19	20	17	25	17	19	23	23	26	10
5	17	17	17	18	15	23	16	18	22	22	25	5
1	12	14	13	13	10	19	14	15	18	18	20	1

Table 3. Self-assessment converted table

Assigned Number	Sex	Topic #1	Topic #2	Topic #3	Topic #4	Topic #5	Topic #N	Topic #87	Att.	Mot.	…Strategy Analysis
1	F	4	4	3	4	3	…	5	Poor	Poor	…
2	M	4	4	2	3	3	…	3	Good	Poor	…
3	F	3	3	3	3	4	…	5	Good	Poor	…
4	F	4	4	3	4	2	…	4	Good	Good	…
5	M	3	3	3	3	3	…	2	Good	Good	…
6	M	3	4	5	3	4	…	2	Poor	Poor	…
…N	…	…	…	…	…	…	…	…	…	…	…
3980	F	5	3	4	3	3	…	5	Good	Good	…

The LASSI sub-scale (i.e. sub-category) will produce 11 predicted results for this survey, and therefore analyses aimed at 11 categories of study strategies must be carried out in decision tree analysis. Take for instance "Motivation" scale as a study strategy. The results of analysis are shown in Figure 1. The included topic numbers are 9,12,26,30,37,44 and 51 for "Motivation" scale in LASSI. The topic numbers were found from analysis that has an approximate effect on "Motivation" scale as LASSI. We can assess whether there are obstacles on "Motivation" scale from our analysis. According to the decision tree in Figure 1, we will inquire by topic number 26 in the first. When the answer is 1,2 or 3. We assume that there are obstacles on "Motivation" scale for 53.8% probability. Which can not be convinced due to "Support" is less than 75%. We can follow the decision tree again, and inquire by topic number 9. However, the "Support" of topic number 9 is still less than 75%. So, topic number 44 has to be inquired. If the students answer is 3,4 or 5, the "Support" is 82.0% more than 75% already. Therefore, we assume that there are obstacles on "Motivation" scale. Otherwise continue next topic for more detail prediction. On the other hand, if the students answer was 4 or 5 for topic number 26, we assume that there are not obstacles on "Motivation" scale for 91.6% probability. Finally, we carried out rule base processing for the results of Figure 1 and stored in a database.

3.3 Association Rule Analysis for Various Study Strategies

Association rule analysis can be applied to find whether the different topics (i.e. questions) are correlated or not. We use support, confidence and lift to pick out the necessary rules. This paper uses association rule to find the reference rules between different study strategies. The analyzed results are shown in Figure 2.2, which displays only a part of the study strategies and rule sorting for the reader's reference.

Some of them occur that the lift too low or conflict with one another. We omit any rules that have a lift less than 1.3, confidence less than 75% and those that conflict with one another. The processed rules from association rule analysis are shown in Table 4.

In this paper, the value of confidence was used as a base to judge whether the rules conform to prediction level. We found the applicable threshold of confidence as 75% from repeated experiments. Therefore, we considered the rules conform to prediction level as long as the confidence more than 75%. For instance, the second and third rules are "[Poor Self] $\rightarrow$ [Poor Information]" and "[Poor Self] $\rightarrow$ [Poor Select]" in Table 4. The two rules' confidences are higher than 75%. Students who scored poor in the self-test ability, from self-assessment, according to association rules, are also diagnosed to be poor in information processing and ability to select important points. So, if we considered who has obstacle in "Self" study strategy from self-assessment. In the same time, we can consider also who has obstacle in information processing and ability to select important points.

4. Performance Evaluations

In this paper, the goal was to find obstacles in study strategy with as few questions as possible. Two major steps were used to achieve the goal. First step is selecting candidate questions to reduce number of question from decision tree analysis. The other step is sequencing categories to reduce number of categories from association rule analysis. When selecting candidate questions from

Table 4. The processed rules from association rule analysis

Rule	Support	Confidence	Lift
Poor Test → Poor Select	0.459935	0.7655	1.3532
Poor Self → Poor Information	0.26727	0.7573	1.5397
Poor Self → Poor Select	0.273549	0.7751	1.3702
Poor Time → Poor Concentration	0.416729	0.7582	1.4013
Good Information → Good Solve	0.384828	0.7573	1.3399
Good Information → Good Study	0.381814	0.7514	1.3653
Poor Motivation → Poor Time	0.281085	0.7561	1.3757
Poor Motivation → Poor Attitude	0.287365	0.773	1.3635
Poor Concentration → Poor Time	0.416729	0.7702	1.4014
Poor Solve → Poor Select	0.327556	0.7533	1.3317
Poor Select → Poor Test	0.459935	0.8131	1.3532
Good Select → Good Solve	0.327054	0.753	1.3323

decision tree analysis, the most critical questions can be distinguished from the other questions according to the different study strategies. The questions selected in first step are nonsequence that may appear more than once. The sequence of categories was decided upon by using association rules to judge the level of relatedness of various study strategies in second step. Through the two steps, the total number of required questions was reduced.

4.1 Selecting Candidate Questions

We have introduced selecting candidate questions from decision tree analysis in Section 3, where we carried out decision tree analysis on 11 different study strategies. The results for 'Motivation' of study strategies are shown in Figure 1. In this step, we must focus on the 11 study strategies and choose critical question as candidate questions. The candidate of questions based on the length of decision tree, and which "Support" has to be more than 75%. In other words, the critical questions are located on the shortest path. So, the candidate of topics ID is the intermediate node and terminal for the shortest path in decision tree. The formula of candidate topics as follow:

Candidate topics = minimum (Node ID except root and "Support" >75%) – 1

For instance, the "Motivation" of study strategy's decision tree analysis shown in Figure 1. The minimum node except root and support more than 75% equals 4. Therefore, there are 3 candidate topics which are number 26,9 and 44 for "Motivation" of study strategy. Based on the answer given by the user, we then decide upon the following question by the above introduced formula. In the same way, to select question for various study strategies. The analysis results show in Table 5. We instanced

Table 5. Necessary question for all of study strategies table

Study strategy	Quantity of Question	Question Number	Quantity of Necessary Quality	Question number
Attitude	7	13,16,34,40,46,64,81	1	46
Motivation	7	9,12,26,30,37,44,51	3	26,9,44
Time	9	3,20,32,38,43,53,61,68,87	1	61
Anxiety	8	1,8,23,28,49,52,58,82	1	52
Concentration	8	5,10,35,39,41,50,56,63	1	41
Information	9	11,14,21,29,36,42,62,70,83	1	42
Selection	6	2,7,55,67,71,84	1	71
Studying	7	6,17,22,45,48,57,85	2	48,6
Self	9	4,15,19,24,27,33,60,65,86	3	19,65,24
Testing	8	18,25,31,47,54,59,66,69	1	31
Solving	9	72,73,74,75,76,77,78,79,80	1	78
Total	87		16	

"Motivation", if someone had obstacle in "Motivation" of study strategy who would be diagnosed through number 26,9 and 44. The syndrome was diagnosed by just using three questions.

Through the above mentioned procedures, the quantity of question to diagnose the obstacles of study strategies have been reduced from 87 to 16. In other words, it originally needs 87 questions to pinpoint the obstacles of study strategies. By applying the results from decision tree analysis, it now only requires 16 questions to accomplish the same process.

4.2 Sequencing Categories

We reduced category through sequencing categories. It bases on the level of relatedness. The candidate questions from decision trees analysis were separately selected with the various study strategies. Therefore, at least 16 questions are needed to understand any of the obstacles in all of the study strategies. By applying association rule analysis, the quantities of category and questions needed in the assessment were reduced. These selected category must be sequenced based on their level of relatedness to find out all of reach the rule suspected to be the standard (confidence is greater than 75%), and then associate the related study strategies again and again, until it is no more correlation. The relatedness is defined as follows:

Let $A \rightarrow B$ be an association rule, where A is a study strategy, B is related to A.

$R(A) = \sum B - Q(A)$, where R is relatedness, Q is necessary questions.

Therefore, the relatedness of A is the summation of all related study strategies with A which confidence more than 75% subtract the number of necessary questions about A. The $\sum B$ can be achieved by follow recursive algorithm.

Let $A := \{A\}$;
Repeat
    old $A^+ := A^+$;
    For each association rule $B \rightarrow C$ **do**
        If B related with A   **Then** $A^+ := A^+ \cup C$;
Until(old $A^+ = A^+$);
Return $|A^+|$;

In above algorithm, when old $A^+ = A^+$ means that all rules related with A have been calculated. So, it will terminate. If the calculated results of relatedness are equal, then the average confidence values for all related scales should be calculated and the one with a greater average value should be given a higher position in the sequence. Table 6 shows the relatedness of the 11 scales (i.e. categories) of study strategies.

The goal of self-assessment is to diagnose obstacle for student's study strategies. In Table 6, the first column lists all of study strategies and then connects the quantity of necessary questions. It means the quantity of the questions found satisfied confidence more than 75% from decision tree analysis, and we hypothesized that all questions for found out poor study strategy. The "Relation" columns show the correlative study strategies for each study strategy, and connect which confidence to deter-

Table 6. The relatedness of all study strategies

Study Strategy	Quantity of Necessary Questions	Relation 1	Confidence	Relation 2	Confidence	Relatedness	Average Confidence Value
Solving	1	Selection	0.7533			2	0.7533
Selection	1	Testing	0.8131			1	0.8131
Concentration	1	Time	0.7702			1	0.7702
Self	3	Information	0.7573	Selection	0.7751	1	0.7662
Testing	1	Selection	0.7655			1	0.7655
Motivation	3	Attitude	0.773	Time	0.7561	1	0.76455
Time	1	Concentration	0.7582			1	0.7582
Anxiety	1		0			0	0
Attitude	1		0			0	0
Information	1		0			0	0
Studying	2		0			-1	0

mine whether it conform to prediction level. When the confidence value is more than 75%, we determined the relation rule as useful.

In Table 6, the "Motivation" of study strategy is related with "Attitude" and "Time", and both confidences are more than 75%. So, we assume that when someone is diagnosed poor on "Motivation", will also be poor on "Attitude" and "Time". According to the definition of relatedness, the "Motivation" of study strategy includes itself, "Attitude" and "Time". There are 3 related study strategies. By the way, the "Time" of study strategy is related with "Concentration" and the "Concentration" of study strategy is related with "Time", then no more related. Therefore, there are 4 study strategies related with "Motivation", and the quantity of necessary questions is 3. So, the relatedness of "Motivation" is all of related study strategies subtract the quantity of necessary questions that equals 1.

We can sequence scales (i.e. categories) based on the statistical results shown in Table 6 by comparing levels of relatedness and average confidence values. The scales of study strategy sequence as follow: Solving > Selection > Concentration > Self > Testing > Motivation > Time > Anxiety > Attitude > Information > Study.

First of all, the "Solving" of study strategy will be selected according to the sequence. The "Selection" and "Testing" will be omitted due to related with "Solving". The other 8 study strategies have to recalculate for the second selection. The relatedness of "Self" will change due to the "Selection" be omitted. The new sequence as follow: Concentration > Motivation > Time > Anxiety > Attitude > Information > Self > Study. Therefore, the "Concentration" of study strategy will be selected according to the new sequence. The "Time" will be omitted due to related with "Concentration". The other 6 study strategies have to recalculate again. No other study strategy will be omitted as no other correlation is found.

The new sequence as follow: Anxiety > Attitude > Information > Motivation > Self > Study. The "Attitude", "Anxiety" and "Information" will be selected sequentially. By the way, the relatedness of "Motivation" and "Self" will change due to "Attitude" and "Information" was selected. So, the last 3 study strategies sequence as follow: Study > Motivation > Self. They will be selected sequentially. The quantity of study strategies has been reduced from11 to 8. Therefore, the quantity of question to diagnose the obstacle of study strategies has been reduced from 16 to 13. In other words, it originally needs 87 questions to diagnose the obstacle of study strategies. Applying the results from decision tree analysis, only 16 questions are required to diagnose the obstacle of study strategies. Following the total questions formula, the summation of necessary question of all study strategies subtracts the summation necessary questions of related with "Solving" and "Concentration" study strategies equal 13. Thus, applying the results of association rule analysis can greatly reduce essential questions.

5. Conclusions and Future Directions

This paper applied data mining technologies to self-assessment for LASSI. From our performance evaluations, it is possible to carry out a complete self-assessment with only 13 questions. The survey was originally composed of 87 questions that required 20–30 minutes to complete, but in our study, the whole process only takes 2–3 minutes. The question and time required for completion have been drastically reduced, at the same time increasing the willingness of the user to participate. The integration of analysis results in this paper made the application of data mining have diversification even more.

Currently, we utilize static method to select question. It follows the results from sequencing categories. So, the questions sequence is the same order completely. In the future, that will decide on the next question according to user's selection. Inaddition, we will research multilevel association that allows more than one rule at the same time. Therefore, the sequencing categories will be variable that depend on user's selection. The necessary questions which diagnose the obstacle of study strategies will be less than now. Besides, our corpora only focus on university freshman. In the future, we will survey and analyze all students to improve the accuracy and practicability for our researches.

References

[1] Quinlan, J. R, "Induction of Decision Tree," *Machine Learning,* Vol. 1, pp. 81–106 (1996).

[2] Quinlan, J. R, *C4.5: Programs for Machine Learning,* San Mateo, Calif.: Morgan Kaufmann (1993).

[3] Stuart Russell, Peter Norving, *Artifical Intelligence a Modern Approach,* Prentice-Hall Internation Editions.

[4] Agrawal, R., Imielinski, T. and Swami, A., "Mining Association Rules between Sets of Items in Large Databases," *Proceedings of the 1993 ACM SIGMOD Conferencen on Management of Data,* Washington D.C., pp. 207–216 (1993).

[5] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. and Verkamo, A. I., *Fast Discovery of Association Rules, Advances in Knowledge Discovery and Data Mining,* U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, eds., AAAI/MIT Press, pp. 307–328 (1996).

[6] Brin, S., Motwani, R., Ullman, J. and Tsur, S., "Dynamic Itemset Counting and Implication Rule for Market Basket Data," *Proceedings of the 1997 SIGMOD Conference on Management of Data,* pp. 255–264 (1997).

[7] Han, J., Pei, J. and Yin, Y., "Mining Frequent Patterns without Candidate Generation," *Proceedings of the 2000 ACM SIGMOD Conference on Management of Data,* Dallas, Texas, USA, pp. 1–12 (2000).

[8] Hipp, J., Güntzer, U. and Nakhaeizadeh, G., "Algorithms for Association Rule Mining – A General Survey and Comparison," *SIGKDD Explorations,* Vol. 2, pp. 58–64 (2000).

[9] Park, J. S., Chen, M. S. and Yu, P. S., "Using a Hash-Based Method with Transaction Trimming for Mining Association Rules," *IEEE Transactions on Knowledge and Data Engineering,* Vol. 9, pp. 813–825 (1997).

[10] Agrawal, R. and Srikant, R., "Mining Sequential Patterns," *Proceedings of the 11th International Conference on Data Engineering,* Taipei, Taiwan, pp. 3–14 (1995).

[11] Srikant, R. and Agrawal, R., "Mining Sequential Patterns: Generalizations and Performance Improvements," *Proceedings of the 5th International Conference on Extending Database Technology,* Avignon, France, pp. 3–17 (1996). (An extended version is the IBM Research Report RJ 9994)

Manuscript Received: Dec. 29, 2006
Accepted: Jun. 8, 2007