Hirashima,T. et al. (Eds.) (2010). Workshop Proceedings of the 18th International Conference on Computers in Education. Putrajaya, Malaysia: Asia-Pacific Society for Computers in Education.

Analysis of Learners' Activity in a Question-Posing Learning Support System by Association Rule Mining

Yuki HIRAI[a] & Tomoo INOUE[a]
[a]*Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan*
s0930522@u.tsukuba.ac.jp, inoue@slis.tsukuba.ac.jp

Abstract: Many support systems for learner-centered question-posing learning have been developed. Some effectiveness in using such systems have been reported. However, most researches have indicated a single factor or combined two factors as the effectiveness of question-posing learning. In this paper, association rule mining was used for analyzing the effectiveness in using a question-posing learning support system. The effectiveness of the combination of more than two factors was investigated. From the analysis, it was revealed that making high-quality question-items related to increasing comprehension of learning materials even though learners posed, answered and/or assessed not many questions.

Keywords: Question-posing learning, Learning support system, Data mining, Association rule mining

1. Introduction

Question-posing is higher cognitive activity than question-answering and has positive effect on the learners' problem solving skills [1]. Recently, question-posing learning activity includes question-posing, question-answering, question-assessing and some discussion. These activities improve the learners' comprehension [2]. Additionally, many Web-based learning support systems have been developed and collaborative learning in a distributed asynchronous environment has been supported [3]. According to abovementioned background, many question-posing learning support systems, such as "MONSAKUN [4]", "QPPA [2]" and "Concerto [5][6]", have been developed.

In this study, the learners' activities in Concerto III were analyzed and the learning effectiveness in using the system was identified. The effectiveness of posing questions has been discussed in many precedent studies. Effects of other learning activities, though they are considered to be important, have not been reported enough. The positive effectiveness of making high-quality question-items have been reported [7]. The question-items consist of question-stem, the answer and the explanation. The high-quality question-items define as "the question-item where there are few mistakes, such as wrong answer and misspelling". Also, the positive effectiveness of assessing many questions have been published [8]. Thus, making questions, making high-quality question-items and/or assessing questions have positive effect on learning. However, most precedent researches have indicated a single factor or combined two factors as the effectiveness of question-posing learning activities. Akahori [3] has suggested that the factor of effectiveness is not only one, but complicated especially when the subject of an experiment is a human beings.

In this paper, the learning effectiveness in using a question-posing learning system was analyzed with "association rule mining". The effectiveness of the combination of more than two factors could be found by this analysis. The results indicated that making high-quality question-items related to increasing comprehension of learning materials even though learners posed, answered and/or assessed not many questions.

2. Related Work

a. The Learners' Activity Analysis in a Question-posing Learning Support System

In the early research about a question-posing learning support system, the qualitative analysis, such as one of the questionnaire to which the learners who used the system responded, has been conducted. One example of the results is as follows:

- Learners' interactive skills were improved by creating questions and group-review of posed questions. Learners' comprehension of learning materials was also improved. Learners were motivated to learn. Therefore, the posing and review activity had significantly positive effect on learning. [9]

There are some reports about the positive effectiveness of question-posing activity by analyzing the questionnaire. (e.g. [2][10][11][12][13])

Recently, the quantitative analysis, such as one of logs stored in the system, has been carried out. The pre- and post-test score have been used in many researches. These scores have been used to show that learners' comprehension was increased in the comparison of "before" and "after" using the system. One instance of the analysis is as follows:

- Students were divided into two groups by the pre-test score, and students in each group were also divided into two groups by the number of posed questions. Then, the pre- and post-test score were compared between the groups. The results indicated that our system improved the problem solving ability of the students which the pre-test score was low and the number of posed questions was high. [4]

There are some researches conducted the similar analysis. (e.g. [9][14][15][16][17])
The effectiveness of question-posing activity varies according to the condition of experiment. Abovementioned researches have indicated at most combined two factors (e.g. the students which the pre-test score was low and the number of posed questions was high) as the effectiveness of question-posing activity.

b. Data Mining Methods

Generally, learners' activity logs are stored in the database for LMS (Learning Management System) included a question-posing learning support system [18]. Data mining techniques are frequently used to analyze learners' activity, learning effectiveness and/or other perspective. There are various mining techniques, such as association rule mining [19], decision tree, SVM (Support Vector Machine), clustering and many more. Many insights are revealed by these techniques. However, it may be difficult for the analysts to interpret if the findings are too much. Therefore, we used "association rule mining" because the number of findings can be controlled by the definition of threshold level.

3. Outline of Association Rule Mining

Association rule mining has received a lot of attention in recent years [18]. An association rule [19] is an implication of the form "A => B", where A and B disjoint subsets of U (a universal set). The sets of items A and B called LHS (left-hand-side) and RHS (right-hand-side). The quality of an association rule is defined by its *confidence* and *support*. *Confidence* of a rule is the conditional probability of B given A. *Support* of a rule is an estimate of the probability of itemsets in the union of A and B. The correlation between A and B in the rule

may be strong if the value of *confidence* is high. However, the rule may not be persuasive if the value of *support* is low. Thus, evaluation of rules needs to consider both these values.

4. Experiment of using Concerto III

a. Outline of Concerto III

In this study, the data stored in Concerto III [7] was analyzed. Concerto III is a Web-based learning support system and provides the following functions:

- Question-posing: learners can make a question in the form of a multiple-choice or story question and pose it to the system. They can use figures on question-making and revise the posed question if necessary.
- Question-answering and Question-assessing: learners can answer the posed question and assess it. They can assess it in terms of its originality, difficulty and/or quality on the five point Lickert scale (1-5) and additional comments. The assessment is done at the same time when the learners answer the question. As for the assessment of the quality, the scale is as follows: (1) Contradiction exists in the question. He/She can not answer the question (e.g., no answers for the question). (2) He/She can answer the question, but there exist some errors in the answer or explanation. (3) There exist no errors in the question, answer and explanation, but I want the question-poser to explain the question and answer in more details. (4) There exist no errors in the question, answer and explanation. The question-poser explains the question very well. However, there are incorrect characters in the question, answer or explanation. (5) The question is perfect.
- Discussion: Learners can discuss the posed question with others. The system provides BBS (Bulletin Board System) for each question.

b. Outline of the Experiment

Concerto III has been applied to a university course “Introduction to computer systems”. The course was offered by the department of information education at Tokyo Gakugei University in the 2008 and 2009 academic year. A pre-test has been carried out before the application of Concerto III. The content of pre-test is what learners have already learned in the course. A post-test has also carried out after the application. The content of post-test is what learners learned in the course. Table 1 shows the usage results of Concerto III. Table 2 shows an example of question-item posed by a learner.

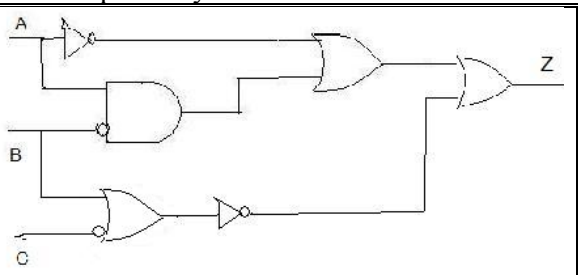
Table 1. Usage results of Concerto III.

Year	2008	2009
Duration of application	6 weeks	7 weeks
No. of user registration	94	77
Total No. of login	1384	611
Total No. of question-posing	273	94
Total No. of question-answering	3815	1340
Total No. of question-assessing	2561	697
Total No. of submitted the comments in the BBS	23	9

5. Extraction of Association Rules

a. Data Sets

Table 2. Example of question-item posed by a learner.

Question: Given that the logical circuit using MIL symbol, express this same circuit using Boolean algebra. Then, simplify the Boolean expression.	
Answer: Z = not A + not B	
Explanation: Z = notA + A * notB + not(B + notC) = (notA + A) (notA + notB) + notB * C = notA + notB + notB * C = notA + notB (1 + C) = notA + notB	

In this analysis, learners' activity logs were stored in Concerto III and the pre- and post-test score were used. Table 3 shows the recorded activities and their explanations. There were some learners who were absent from the pre- or post-test. The data of these learners were not used in the following analysis. The assessed data were collected through the Web page of the system. It might be desirable that the quality was evaluated by the instructors for the accuracy, but was done by the learners in this practice. This instead realized the collaborative learning. The evaluation by the instructors is one of the future works.

Table 3. The system logs and its explanation.

Log	Explanation
Login	A learner logged in to the system
Pose	A learner posed a question
Answer	A learner answered a pose question
Assess	A learner assessed a posed question
Submit	A learner submitted a comment to the BBS
Quality	The quality of a learner's posed question

b. Methods

The analysis began with calculation of the number of each learner's question-posing, question-answering and so on. Table 4 shows part of the calculated data. For example, learner L_1 posed 4 questions, answered 30 questions and his/her score of pre-test was 6 and the post-test is 62.

Table 4. The summary of users' information.

Learner	No. of question-posing	No. of question-answering	…	Pre-test score	Post-test score
L_1	4	30	…	6	62
L_2	6	218	…	5	73
…	…	…	…	…	…
L_n	3	65	…	7	48
Average	2.93	41	…	-	-

Then, we classified the calculated data as H (High-group) and L (Low-group). Table 5 shows part of the classified data. For instance, the learner which the number of question-posing was higher than average value belongs to the high-group. As for the pre- and post-test score, each learner's T-score was calculated and classified as "increased" or "decreased". The learner which his/her T-score of post-test was higher than one of pre-test belongs to the

"increased" group. "Increased" defines that learner's comprehension of learning materials was increased in the comparison of pre-test and post-test. As for the classified and normalized data shown in Table 5, the data in 2008 and 2009 was combined. There were 119 learner's data in the combined data.

After completing the classification process, mining association rules was carried out. We used the software "R 2.9.1 [20]" (for statistical processing), the package "arules [21]" (for mining association rules) and the apriori algorithm [22]. The minimum value of *support* and *confidence* was set up to control the number of extractable rule. The minimum *support* was 0.20. This means that there are 24 or more learners who satisfy the condition LHS of a rule. The minimum *confidence* was the proportion which satisfy the condition RHS unconditionally. This is because the rule which the value of *confidence* is higher than the minimum *confidence* is significant in the principle of this mining. "T-score = increased" or "T-score = decreased" was selected as the condition RHS because our goal of this study was to identify the learning effectiveness.

Table 5. The summary of the classified groups.
("1" means that the learner belongs to the group)

Learner	No. of question-posing		No. of question-answering		…	T-score	
	H	L	H	L		increased	decreased
L_1	1	0	0	1	…	0	1
L_2	1	0	1	0	…	1	0
…	…	…	…	…	…	…	…
L_n	0	1	1	0	…	0	1

c. Results

Table 6 and Table 7 contain the extracted rules in this mining. 24 rules related to "T-score = increased" and 20 rules related to "T-score = decreased" were extracted. The rules were numbered in order of *confidence* and *support*. For example, Rule I_1 means that T-score of the learner was increased and the learner satisfied the following three conditions:

- The number of logging into the system was lower than the average value.

Table 6. Rules related to increase of T-score.

Rule No.	LHS	RHS	*Support*	*Confidence*
I_1	Login=L, Submit=L, Quality=H	T-score=increased	0.210	0.595
I_2	Login=L, Quality=H	T-score=increased	0.218	0.577
I_3	Pose=L, Submit=L, Quality=H	T-score=increased	0.260	0.574
I_4	Submit=L, Quality=H	T-score=increased	0.327	0.573
I_5	Pose=L, Quality=H	T-score=increased	0.268	0.542
I_6	Answer=L, Submit=L, Quality=H	T-score=increased	0.226	0.540
I_7	Pose=L, Assess=L, Quality=H	T-score=increased	0.201	0.533
I_8	Pose=L, Answer=L, Quality=H	T-score=increased	0.210	0.531
I_9	Assess=L, Submit=L, Quality=H	T-score=increased	0.218	0.530
I_{10}	Quality=H	T-score=increased	0.344	0.512
I_{11}	Submit=L	T-score=increased	0.428	0.490
I_{12}	Login=L, Answer=L, Submit=L	T-score=increased	0.243	0.483
I_{13}	Answer=L, Quality=H	T-score=increased	0.226	0.482
I_{14}	Pose=L, Answer=L, Submit=L	T-score=increased	0.260	0.476
I_{15}	Login=L, Answer=L	T-score=increased	0.243	0.475
I_{16}	Login=L	T-score=increased	0.294	0.472
I_{17}	Assess=L, Quality=H	T-score=increased	0.218	0.472
I_{18}	Answer=L, Submit=L	T-score=increased	0.277	0.4714

I_{19}	Login=L, Submit=L	T-score=increased	0.277	0.4714
I_{20}	Pose=L, Assess=L, Submit=L	T-score=increased	0.344	0.4712
I_{21}	Pose=L, Submit=L	T-score=increased	0.201	0.470
I_{22}	Answer=L, Assess=L, Quality=H	T-score=increased	0.302	0.467
I_{23}	Assess=L, Submit=L	T-score=increased	0.302	0.467
I_{24}		T-score=increased	0.462	0.462

Table 7. Rules related to decrease of T-score.

Rule No.	LHS	RHS	*Support*	*Confidence*
D_1	Pose=L, Quality=L	T-score=decreased	0.201	0.685
D_2	Submit=L, Quality=L	T-score=decreased	0.201	0.666
D_3	Quality=L	T-score=decreased	0.210	0.641
D_4	Answer=L, Assess=L	T-score=decreased	0.352	0.575
D_5	Login=L, Pose=L, Assess=L	T-score=decreased	0.285	0.5666
D_6	Assess=L	T-score=decreased	0.394	0.5662
D_7	Answer=L	T-score=decreased	0.369	0.564
D_8	Pose=L, Answer=L, Assess=L	T-score=decreased	0.294	0.555
D_9	Login=H	T-score=decreased	0.210	0.555
D_{10}	Login=L, Assess=L	T-score=decreased	0.294	0.546
D_{11}	Login=L, Pose=L, Submit=L	T-score=decreased	0.294	0.546
D_{12}	Login=L, Pose=L	T-score=decreased	0.310	0.544
D_{13}	Pose=L, Answer=L	T-score=decreased	0.310	0.544
D_{14}	Answer=L, Assess=L, Submit=L	T-score=decreased	0.310	0.544
D_{15}	Login=L, Pose=L, Answer=L	T-score=decreased	0.260	0.543
D_{16}	Pose=L	T-score=decreased	0.428	0.542
D_{17}	Login=L, Answer=L, Assess=L	T-score=decreased	0.268	0.542
D_{18}	Pose=L, Assess=L	T-score=decreased	0.327	0.541
D_{19}	Login=L, Assess=L, Submit=L	T-score=decreased	0.285	0.539
D_{20}		T-score=decreased	0.537	0.537

- The number of submitted comments to the BBS was lower than the average value.
- The quality of his/her posed question was higher than the average value.

Then, *Support* of the rule was 0.210 and *Confidence* was 0.595.

6. Discussion

a. Interpretation of Extracted Rules

The rules I_1 to I_{10} include the condition "Quality = H". According to this, we can guess that posing high-quality question-items relates to the increase of T-score. Moreover, "Quality = H" and other conditions, such as "Login = L", "Submit = L", "Pose = L", "Answer = L" or "Assess = L", were combined. The *Confidence* of these combined rules are higher than one of the single condition "Quality = H". On the other hand, the rules D_1 to D_3 include the condition "Quality = L". According to this, we can guess that posing low-quality question-items relates to the decrease of T-score. There were some rules consisted of the single condition, such as "Pose = L", "Answer = L" or "Assess = L".

b. Comparison of This Study and the Precedent Researches

In this section, the results of this research are compared to the results of the precedent researches. In all of these researches, the learners were undergraduate students. The subjects were "Introduction to computer systems" [7][8] and "Information processing" [9]. Web-

based learning support systems were provided and the learners' comprehension of learning materials have been investigated.

(1) The effectiveness of question-posing

Takagi et al. [9] has suggested that there were some effectiveness of question-posing by investigating the questionnaire. They also reported that the number of question-posing did not influence on the examination score. Another study [7] has conducted that the examination score of learners who posed many questions were more increased than those who not. We can not identify the influence because "Pose = H" is not listed in Table 6 and 7. However, we can guess that one factor which the number of question-posing is lower than the average value related to another factor which the examination score is decreased.

(2) The effectiveness of posing a high-quality question-item

Hirai et al. [7] have reported that the examination score of learners who posed high-quality question-items were more increased than those who not. As for this, we confirm after the fact because the rule satisfied the condition "if Quality = H then T-score = increased" was extracted in Table 6.

(3) The effectiveness of question-assessing

Hirai et al. [8] have indicated that the examination score of learners who assessed many questions were more increased than those who not. We can not identify whether the number of question-assessing influence on the examination score because "Assess = H" is not listed in Table 6 and 7. However, we can guess that one factor which the number of question-assessing is lower than the average value related to another factor which the examination score is decreased.

c. The Effectiveness in Using a Question-posing Learning Support System

In section 6.1, we described that there were rules consisted of the single condition, such as "Pose = L", "Answer = L" or "Assess = L" in LHS of Table 7. In other words, these rules related to the decrease of T-score. However, when these factors and one factor "Quality = H" are combined, the rule relates to the increase of T-score. The following are some examples of this phenomenon:

- (Pose = L) and (Quality = H) => (T-score = increase) …rule I_7
- (Answer = L) and (Quality = H) => (T-score = increase) …rule I_{13}
- (Assess = L) and (Quality = H) => (T-score = increase) …rule I_{17}

In short, the result suggested that making high-quality question-items had positive effect on learning even though learners posed, answered and/or assessed not many questions.

d. Hypothesis on increasing comprehension by making high-quality question-items

Learners may check question-items to upgrade when making the question. As for checking questions, Yu et al. [2] have reported that peer-assessing learning activities encourage participants to gauge objectively and critically the adequacy and correctness of posted questions. Then, learners presumably add details to their existing cognition, explore and correct their misconceptions, and/or re-organize their current knowledge structures by online interaction and open communication with peers pertaining to the question and the correct answer. This report describes peer-assessing activities. We hypothesized that learners may do these activities during making the question. Therefore, we considered that making high-quality question-items caused increasing comprehension.

7. Conclusion

In this paper, the learners' activities have been analyzed by association rule mining. The combined factors as the effectiveness of question-posing learning have been clearly identified. The interpretation of extracted rules have indicated that posing high-quality question-items related to increasing comprehension of learning materials even though learners posed, answered, and/or assessed not many questions. Some support depending on the interpretation of extracted rules will be developed in the future.

Acknowledgement

We are grateful to Yusuke TAKAGAWA for his contribution. Also, we would like to thank Atsuo HAZEYAMA from Tokyo Gakugei University in Japan for kindly providing us with data used in this study.

References

[1] Yokoyama, T., Hirashima, T., Okamoto, M. & Takeuchi, A. (2005). Development of an Environment for Learning by Problem-posing as Integration of Sentences, *Proc. of the 19th Annual Conference of the Japanese Society for Artificial Intelligence*, 2A1-03. (in Japanese)

[2] Yu, F.-Y., Liu, Y.-H., & Chan, T.-W. (2005). A Web-based learning system for question posing and peer assessment, *Innovations in Education and Teaching International*, 42(4), 337-348.

[3] Akahori, K. (2003). Invitation to educational technology: Methodology for problem solving in education, *JUSTSYSTEM Press*. (in Japanese)

[4] Yokoyama, T., Hirashima, T., Okamoto, M. & Takeuchi, A. (2007). Long-term Usage and Evaluation of Learning Environment for Problem-posing as Sentence-integration, *Japan journal of education technology*, 30(4), 333-341. (in Japanese)

[5] Hirai, Y. & Hazeyama, A. (2007). A Learning Support System based on Question-posing and Its Evaluation, *Proc. of C5*, 180-185.

[6] Hirai, Y. & Hazeyama, A. (2008). Question-posing based Collaborative Learning Support System and Its Application to a Distributed Asynchronous Learning Environment, *Journal of Information Processing Society of Japan*, 49(10), 3341-3353. (in Japanese)

[7] Hirai, Y., Takagawa, Y., Hazeyama, A. & Inoue, T. (2009). Analysis of the Learners' Assessment Activity in a Collaborative Learning Support System based on Question-posing, *Proc. of ICCE'09 Workshop*, 25-29.

[8] Hirai, Y., Hazeyama, A. & Inoue, T. (2010). A Learning Support System based on Question-posing by Learners and Effectiveness of Its Application to a Distributed Asynchronous Environment, *Transaction of Japanese Society for Information and Systems in Education*, 27(1), 62-73. (in Japanese)

[9] Takagi, M., Tanaka, M. & Teshigawara, Y. (2007). A Collaborative WBT System Enabling Students to Create Quizzes and to Review Them Interactively, *Journal of Information Processing Society of Japan*, 48(3), 1532-1545. (in Japanese)

[10] Toumoto, T., Ichi, M., Hirashima, T. & Takeuchi, A. (2007). Interactive Environment for Learning by Problem-Posing in Multi-Digit Subtraction, *Japan journal of education technology*, 31(1), 61-68. (in Japanese)

[11] Sugawara, N., Oda, K., Akaike, H. & Kakuda, H. (2007). SHoes: Realtime e-learning System for Classroom Education and Organizational Learning, *Journal of Information Processing Society of Japan*, 48(8), 2791-2801. (in Japanese)

[12] Hayashi, T., Mizuno, T., Tominaga, H., Tarumi, H. & Yamasaki, T. (2008). A Learner-centered e-Learning Environment by Contributing and Sharing Problems Created by Learners, *Proc. of CollabTech*, 94-98.

[13] Watanabe, H. (2005). Group Activities using Discussion Function in Information Education, *Proc. of the 3rd Conference of Japan WebCT User*, 83-88. (In Japanese)

[14] Denny, P., Hamer, J., L.-Reily, A. & Purchase, H. (2008). PeerWise: Students Sharing their Multiple Choice Question, *Proc. of ICER*.

[15] Okamoto, M. (1996). Effectiveness of Problem-Posing in Acquirement of Problem Solving Schema, *Proc. of Annual convention of Japanese Association of Educational Psychology*, 38, 368. (in Japanese)

[16] Hirashima, T., Yokoyama, T., Okamoto, M. & Takeuchi, A. (2008). Long-term Use of Learning Environment for Problem-Posing in Arithmetical Word Problems, *Proc. of ICCE*, 817-824.
[17] Nakano, A., Hirashima, T. & Takeuchi, A. (2002). Intelligent Learning Environment for Understanding Basic Operations of Arithmetic by Problem Posing, *Journal of Japanese Society for Artificial Intelligence*, 598-607. (in Japanese)
[18] Ueno, M. (2007). Data Mining in e-Learning, *Japan journal of education technology*, 598-607. (in Japanese)
[19] Agraval, R., Imielinski, T. & Swami, A. N. (1993). Mining association rules between sets of items in large database, *Proc. of SIGMOD*, 207-216.
[20] arules: Mining Association Rules and Frequent Itemsets, The Comprehensive R Archive Network, http://cran.r-project.org/web/packages/arules (refer: 2010/06/02)
[21] The R Project for Statistical Computing, http://www.r-project.org (refer: 2010/06/02)
[22] Agrawal, R. & Srikant, R. (1994). Fast Algorithms for Mining Association Rules, *Proc. of VLDB*, 487-499.