AN EVALUATION OF A MODEL FOR AUTOMATED ASSESSMENT OF INFORMATION SYSTEM STUDENTS' CONTRIBUTIONS TO INTERNET FORUM

Mee Chin Wee and Zaitun A. B.
Faculty of Computer Science & Information Technology
University of Malaya, 50603 Kuala Lumpur, Malaysia
mcwee@yahoo.com
zab@um.edu.my

ABSTRACT

Students collaboration via discussion in an Internet forum when working on a project for Information Systems (IS) courses is common in Malaysia. Assessing students' contributions to Internet forums has become important. However, manually assessing students' contributions to Internet forums is a time consuming task. A model was proposed to help IS educators in assessing students' contributions to an Internet forum. This paper reports on the evaluation of the proposed model incorporated into an Internet forums to generate performance indicator scores for students' contributions. The evaluation was carried out in Malaysia. The evaluation results show that the assessment model can be used to predict IS students' marks for their contributions to Internet forums.

1. INTRODUCTION

In order to obtain degrees especially in IS courses, project or assignment is one of the essential component that students have to complete. Students discussed their learning issues in Internet forums when working on a project or assignment is common for IS courses. Student discussion in an Internet forum is important as it is difficult to schedule for regular project meeting and discussion shall not be limit to project meeting. In an Internet forum, each student can view another student's contributions and providing feedback to one another online. The process of reflection and articulation of content, writing about what they have learned, engages students in an activity-based learning experience. Permanent storage of messages in an Internet forum provides support for reflection. Moreover, educators do not usually provide much feedback for their students to complete their project (Helic et al., 2005). Assessing students' contributions to Internetforums has become important. The two main reasons for assessing students' contributions are that it encourages students' participation, and it allows students to focus on the given topics. Student participation is a key to effective collaborative learning (Hardless et al, 2001). This finding indicates that students need to be active participants in order to succeed. Besides, assessment criteria can served as a clear guide to students for learning outcomes and the expected quality of thinking and work, and as a means of aligning teaching and learning behaviours and goals (Ho, 2002; Jones et al., 2000). There is a number of assessment criteria stated in the literature such as assess students' performance based on total number of students' postings, total message length, timeliness of message, describing and categorizing postings using SCAFFOLD (Scale for Forums/Online Discussion Assessment) (Dringus and Ellis, 2004), content analysis using Henri's Analytical Model (Henri, 1992), and Garrison and Anderson's Practical Inquiry Model of Cognitive Presence (Garrison and Anderson, 2003). However, manually assessing students' contributions to Internet forums is a time consuming task as reported in literature. To reduce the workload of IS educators, a computer generated performance indicator (PI) is proposed.

2. OBJECTIVE

This paper reports on the results of the evaluation of a model for automated assessment of students' contributions to Internet forums for IS courses. The paper also presents the elements of the assessment models, the evaluation procedure and followed by the evaluation results.

3. AN ASSESSMENT MODEL

The objective of PI is to predict students' marks by analyzing the class messages posted by IS students in Internet forums. The PI is

generated from four aspects: the quality of their work, the quantity of their efforts, the timeliness and the activeness of their participation. Four measures – message category, message length, message date and number of messages – are derived from the class messages to measure each assessment aspect respectively. The authors assume that quality of learning in Internet forums is revealed by the quality of the messages generated by a student. The category of a message is analyzed along SCAFFOLD to reflect the depth of knowledge of the author, so the message category could be an indicator for the learning quality. SCAFFOLD was adopted (as summarized in Table 1) since it is comprehensive and contains elements of higher order thinking skills, that is, 'analysis', 'synthesis', and 'evaluative', which are also the high levels of the IS knowledge metric. These high level of knowledge is important as IS academic community emphasized the importance of developing students' problem solving and critical thinking abilities as the exit characteristics of its IS. SCAFFOLD is build upon the theoretical foundation established in the literature. Dringus and Ellis summarised a range of participation indicators identified in the literature (including models for analyzing the process of learning in Internet forums such as Garrison, Anderson, and Archer's Practical Inquiry Model of Cognitive Presence, and Jeong's sequential analysis of group interaction and critical thinking in online discussion) and develop a list of 19 participation indicators (Dringus and Ellis, 2004). SCAFFOLD was used by faculty and students to rate the 13 postings contained in a discrete segment of a masters-level discussion forum in a multimedia systems course (Dringus and Ellis, 2005). The results of evaluation shows that SCAFFOLD could be used for developing and conveying feedback on Internet forums as the students and faculty members had a measure of commonality in interpreting the meaning of the 19 indicators and using the SCAFFOLD to describe postings in Internet forums.

Table 1 SCAFFOLD

The Contribution
Acknowledging: responded to another contribution
Analysis: provided analysis of the problem being discussed
Broadened: increased the scope of the discussion
Clarification: supplied or sought clarification as needed in responses
Closure: helped lead to a conclusion on a topic
Comprehensive: was complete, but not overly lengthy
Error Free: contained accurate information
Evaluative: was evaluative, assessing the meaningfulness or validity of ideas being shared
Originality: contained new ideas or approaches to the topic
Problem: identified a worthy problem related to the topic
Questioning: raised thoughtful questions about the
Reflective: interjected personal commentary or experiences
Resolution: promoted cooperation to resolve issues of debate or disagreement
Resources: exchanged useful resources with others such as links or citations
Social: conversational or social in nature
Solutions: suggested meaningful solutions
Summarizing: summarized the topic discussion overall
Synthesis: contained well formed, clear, connected, and synthesized ideas
Topical: was on topic

An approach to resolve the challenges of collecting and coding large data sets might be to directly involve students in a process of categorizing their own discussion in such context. Knowlton (2001) argues that "For the benefits of online discussion to be realized, students must have formal opportunities for self evaluation". Students must practice evaluating their own contributions to an online discussion against a clearly articulated set of criteria. Knowlton (2001) emphasized on the important of giving a minimum length for postings since it takes some length to construct perspectives that can become the basis of knowledge. Therefore, students' effort in the virtual dialogue could be reflected by the amount of words they post to the system. Message length measures a student's effort in the class and is found by counting all the words, no matter duplicated or not, in the student's messages. In terms of timeliness, due dates were best for stimulating the discussion online. It is

important that messages are posted or reply on time (Knowlton, 2001; Pendergast, 2006). Student participation is a key to effective collaborative learning (Hardless et al, 2001). If posting a message is considered as one class activity, activeness of participation can be measured by message count, which is the number of messages posted by a student. Combining the assessments from multiple aspects has been proven useful for increasing the forecast accuracy (Winkler and Clemen, 2004). The authors apply the idea of weighting to assign weights to assessment criteria. The four measures are combined to compute a PI score, which is

PI score = a*Total message + b*Total message length + Message category + Timeliness;

Each message has a date. For timeliness, a message's date that fall before a given date (deadline for the discussion as provided by the educator) would be included for the calculation of the PI for each student. Then, the formula is

PI score = a*Tot_Mess + b*Tot_Length + SCAFFOLD;

'Closure', "error free", 'topical', 'solutions', 'comprehensive', 'originality', 'problem', 'reflective' were eliminated from the SCAFFOLD list as these items can be logically grouped and represented by other items in the SCAFFOLD list after the authors take into consideration the result of findings reported in Dringus and Ellis (2005). As a result, after replacing SCAFFOLD with it's elements,

PI score = a*Tot_Mess + b*Tot_Length + c* Count_ Acknowledging + d*Count_ Analysis + e*Count_Broadened + f*Count_Evaluative + g*Count_Clarification + h * Count_Questioning + j*Count_Resolution + k* Count_Resources + m*Count_Social + n*Count_Summarzing + p*Count_Synthesis;

Where
a, b, c, d, e, f, g, h, j, k, m, n, p are coefficients,
Tot_Mess – Total message posted by a member;
Tot_Length – Total message lengths posted by a member;Count_ Acknowledging – total count of 'acknowledging'messages posted by a member;
Count_ Analysis - total count of 'analysis' messages posted by a member;
Count_Broadened - total count of 'broadened' messages posted by a member;
Count_Evaluative - total count of 'evaluative' messages posted by a member;
Count_Clarification - total count of 'clarification' messages posted by a member;
Count_Questioning - total count of 'questioning' messages posted by a member;

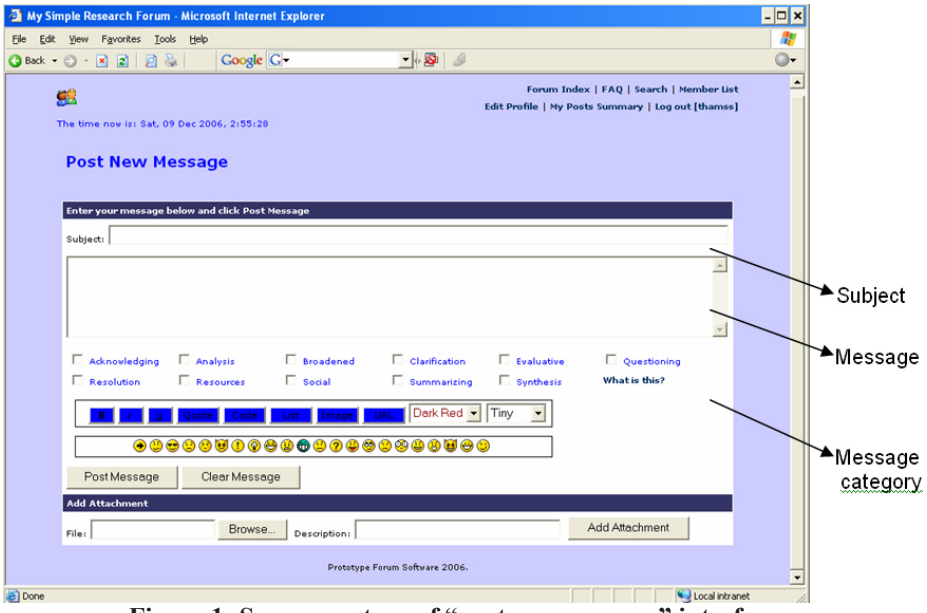

Figure 1: Screen capture of "post new message" interface

Count_Resolution - total count of 'resolution' messages posted by a member;
Count_Resources - total count of 'resources' messages posted by a member;
Count_Social - total count of 'social' messages posted by a member;
Count_Summarzing - total count of 'summarizing' messages posted by a member;
Count_Synthesis - total count of 'synthesis' messages posted by a member;

In order to implement a computer generated PI, open source forum software was adopted. After reviewing the list of forum software (Woolley, 2006), class-1 Forum Software is adopted in this research. class-1 Forum Software is written and distributed under the GNU General Public License which means that its source is freely-distributed and available to the general public. Using OSS approach, the authors do not need to redevelop the basic features available in existing Internet forum. The new feature that added to the forum software was self categorize posts. To implement the proposed features, SCAFFOLD checkbox is developed to allow members to categorize their message before posting as shown in figure 1. The new features that added to administrator interface were set performance indicators, and group performance statistics. To implement the proposed features, the forum software should enable administrator to set the coefficient for the criteria of the PI for each forum. Coefficients that can be set in a PI are total message, total message length, and each category of SCAFFOLD. Each criterion is given a coefficient to be filled in by an administrator. This feature allows the IS educator to select the preferred grading criterion as the criteria in the PI is and will remain, in a large way, a subjective option of the IS respondents. In mathematics, a coefficient is a constant multiplicative factor of a certain object (variable). A zero coefficient for a criterion indicates that the IS educator does not used it as a grading criterion. Criteria that have similar coefficient value means the criteria are equally important. A criterion that is preferred to use as grading criterion by the IS educator could be given a higher coefficient value. Students (members) have to categorize their own message before posting. Students can not edit or delete message after posting. IS educator can perform edition or deletion of messages. The IS educator can change the category of a message if found incorrect.

4. EVALUATION OF THE ASSESSMENT MODEL

The main purpose of the evaluation is to determine the accuracy of the assessment model in predicting student mark for their contributions to Internet forum. To measure the accuracy of the assessment model, Pearson product-moment correlations between the PI scores and the actual grades were calculated. The evalu ation of the forum software was conducted in the second semester of an academic year. Two IS educators from Faculty of Computer Science and Information Technology of the oldest university in Malaysia were agreed to participate in the evaluation of the forum software. Two IS courses (identify using ID = W1 and W2) with a total of sixty four students were involved in the evaluation. The students were taking IS courses that required them to complete a project; hence they had a suitable background for the evaluation. At the end of the project duration, all the IS students' discussions were compiled into tables. Each table contains posts detail such as posts' subjects, time and date of posts, and aggregate contribution of a student in the forum software. The compiled data files were sent to the two IS educators involved for assessment purpose. This is a common approach for manual grading. Three other IS assessors were contacted independently to assess students' contributions for both IS courses. All the assessors have more than six years teaching experience of IS courses. The assessors felt comfortable reviewing the discussion. The IS assessors who participated in the evaluation were considered as a representative sample of IS educators who may potentially use forum software for PBL in IS education. The forum software was available for the IS assessors (a total of five assessors) to view the learning context even though the student discussion was over. The projects' title and description were sent to the three assessors as well. The authors set the coefficients (c, d, e, f, g, h, j, k, m, n, p) of SCAFFOLD to 1, a = 1 (which

means 1 post is assessed as 1 point) and b = 0.01(which means 1 word is assessed as 0.01 point). This is because the IS assessors grading preferences were unknown. The criteria in the PI score are equally important in this case since they are given the same coefficient except for total message length. However, when the grad ing preferences are known, it is easy to adjust the coefficients to reflect the grading preferences. The same coefficients were set for the two IS courses throughout the evaluation. The Pearson product-moment correlations between the PI and the actual grades were calculated. Correlations between individual measures (except for timeliness) and the actual grades were also calculated as shown in Table 2 for W1 and Table 3 for W2. The results in the second column of Table 1 and Table 2 demonstrate that there is a high correlation between the PI and the actual grades (0.827 – 0.996). The results in each row of Table 2 and Table 3 demonstrate the correlation between the PI and the actual grades given by different IS assessors. According to a report in the es say grading literature, agreement between computer graders and human judges varies from 0.4 to 0.9 approximately, and that is comparable to or even better than agreement between two human graders. The results also show that, in most cases, PI performs slightly better than any of the three measures, that are total message count, total message length and SCAFFOLD.

Table 2 Correlations for IS course W1

	R(PI-G)	R(TM-G)	R(TL-G)	R(S-G)
Assessor 1	0.988	0.979	0.887	0.980
Assessor 2	0.852	0.806	0.905	0.857
Assessor 3	0.936	0.928	0.758	0.935
Assessor 4	0.885	0.900	0.797	0.863

Table 3 Correlations for IS course W2

	R(PI-G)	R(TM-G)	R(TL-G)	R(S-G)
Assessor 1	0.996	0.993	0.853	0.991
Assessor 2	0.930	0.923	0.905	0.926
Assessor 3	0.942	0.937	0.835	0.939
Assessor 4	0.827	0.789	0.901	0.848

Where
R(PI-G): correlation between the PI and the actual gradesR(TM-G): correlation between the total message count (TM) score and the actual grades
R(TL-G): correlation between the total message length and the actual grades
R(S-G): correlation between the SCAFFOLD and the actual grade

5. LIMITATIONS

During the evaluation period, the network in the university involved was not reliable towards the end of evaluation. This has caused redundant data (message) appear in the forum and data lost. The forum software evaluated the redundant data more than once.

The evaluation only focused on two final year degree-level discussion forum in IS courses and five IS assessors were involved. It would not be prudent to over generalize the evaluation results.

6. CONCLUSION

In the evaluation of the accuracy of the assessment model, the authors found that the PI score generated from the model were highly correlated with the actual grades assigned by the IS assessors. The difference between the judgments of different IS assessors grading the same class independently is also low. It is reasonable to assume that such correlation is comparable to what has been reported in the automatic essay grading literature (Williams, 2001). Thus, the evaluation results suggest that the performance of the assessment model is comparable to, if not better than, that of a human instructor. Therefore, the computer generated PI can be implemented as a teaching tool to help IS educators obtain a reference to students' performance without reading through the huge amount of class messages, which is a tedious and intensive procedure. The tool could be employed as a supplementary grader to help IS educators make better judgments with reduced workload.

REFERENCES

Helic, D., Krottmaier, H., Maurer, H., Scerbakov, N. Enabling Project-Based Learning in WBT Systems, In International Journal on E-Learning (IJEL), Vol. 4, Issue 4, 2005. pp 445-461.

Hardless, C., Lundin, J., and Nulden, U. Mandatory Participation in Asynchronous Leanring Networks. Proceedings of the 34th Hawaii International Conference on System Sciences (HICSS). 2001.Maui, USA.

Ho, S., Evaluating students' participation in on-line discussions. The Eighth Australian World Wide Web Conference (AusWeb 2002),2002. Sunshine Coast, Queensland.

Jones, A., Scanlon, E. and Blake, C. , Conferencing in communities of learners: examples from social history and science communication. Educational Technology & Society 3(3). 2000.

Dringus L. P. and Ellis T. J. Building the SCAFFOLD for Evaluating Threaded Discussion Forum Activity: Describing and Categorizing Contributions, ASEE/IEEE Frontiers in Education Conference. 20-23 October 2004. Savannah, GA.

Dringus L. P. and Ellis T. J. Evaluating Threaded Discussion Forum Activity: Faculty and Student Perspectives on Categories of Activity. ASEE/IEEE Frontiers in Education Conference. 19-22 October 2005. Indianapolis.

Henri, F. Computer Conferencing and Content Analysis. In A. R. Kaye (Ed), Collaborative learning through computer conferencing: The Najaden Papers, Berlin: Sprnger-Verlag. 1992. pp117-136.

Garrison, D.R., Anderson, T. E-Learning in the 21st Century. London, RoutledgeFalmer.2003.

Knowlton, D. S. Promoting durable knowledge construction through online discussion. Sixth Annual Instructional Technology Conference : Today's successes/tomorrow's horizons. 8 – 10 April 2001. Middle Tennessee State University.

Pendergast, M. An Analysis Tool for the Assessment of Student Participation and Implementation Dynamics in Online Discussion Forums. ACM SIGITE Newsletter, 3(2), 2006 .

Winkler, R. L. and Clemen, R. T. Multiple Experts vs. Multiple Methods: Combining Correlation Assessments. Decision Analysis. 1(3). 2004. Pp 167-176

Williams, R. Automated essay grading: An evaluation of four conceptual models. In A. Herrmann and M.M. Kulski(Eds), Expanding Horizons in Teaching and Learning. Proceedings of the 10th Annual Teaching Learning Forum.2001

Woolley, D. R. Forum Software for the Web. [Internet] Available from: http://thinkofit.com/webconf/forumsoft.htm [Cited 5 August 2006]