Kong, S.C., Ogata, H., Arnseth, H.C., Chan, C.K.K., Hirashima, T., Klett, F., Lee, J.H.M., Liu, C.C., Looi, C.K., Milrad, M., Mitrovic, A., Nakabayashi, K., Wong, S.L., Yang, S.J.H. (eds.) (2009). *Proceedings of the 17th International Conference on Computers in Education [CDROM]*. Hong Kong: Asia-Pacific Society for Computers in Education.

Inferring Appropriate Feedback Timing from Answering Styles for Open-Ended Questions

Chun-Hsin LIAO[a], Tsai-Yen LI[a], Chun-Yen CHANG[b]
[a]*Computer Science Department, National Chengchi University, Taiwan*
[b]*Earth Science Department, National Taiwan Normal University, Taiwan*
g9430@cs.nccu.edu.tw

1. Introduction and System Overview

Among the possible forms of test, open-ended question has a non-replaceable role in assessing high-level thinking in science education. However, teachers usually hesitate to give this type of tests because of the grading efforts involved unless the grading can be processed automatically by the computer. Recently, there has been much research on automatic text processing in the area of language learning and writing [1] and science education [2]. The system in [2] was further customized into an intelligent tutoring system, called VIBRANT, that can provide feedback to the students according to the answers that have been inputted so far [3]. In this system, the students were asked to answer an open-ended question in the form of ideation and explanation. When an idea or explanation is entered, appropriate comments or feedbacks are provided automatically according to the established user model. However, this system requires a model of domain expert to compute the score for the students. In addition, only the overall collection of answers counts without regard to the course of how they were inputted.

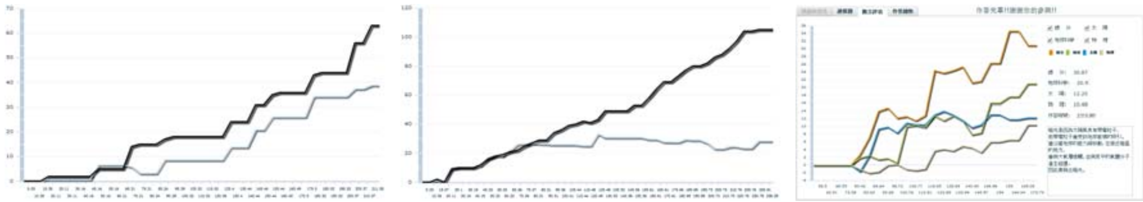
In this paper, we proposed a system aiming to infer good timings for system feedbacks from the answering style of a student. Unlike the previous work on automatic grading for open-ended questions [2], no explicit expert model is constructed. The expected answers are free texts instead of in the semi-structured format of ideation and explanation. For this type of open-ended questions, it is a challenge to know the knowledge status of the student for the purpose of providing timely feedbacks.

First, like in [4], we use a regression method in machine learning to train the model for automatic scoring in an off-line step. The learned grading model can then be used to provide automatic grading on-line. In addition, we hope to understand their status from the traits of how they enter the answer as well as the scores they have obtained. For example, some students tend to think of the questions thoroughly before inputting answers while other may quickly write down what occurs to their minds and then revise it later. Can these behaviors be categorized and even quantitatively measured? Are these behaviors related to the scores that they have accumulated over time? We will report some preliminary observations from some experiments in this paper.

2. Implementation and Experiments

In order to understand the answering styles, we have developed a web-based system to collect preliminary data and perform analysis. At the client side, the system consists of web pages implemented in FLEX to provide text input for open-ended questions and send the answers back to the server periodically (every 5 sec. in our experiments). When the answers are received at the server side, Java-based servlets are used to compare the progress of the text inputs and perform automatic scoring. For each transaction, the system

Kong, S.C., Ogata, H., Arnseth, H.C., Chan, C.K.K., Hirashima, T., Klett, F., Lee, J.H.M., Liu, C.C., Looi, C.K., Milrad, M., Mitrovic, A., Nakabayashi, K., Wong, S.L., Yang, S.J.H. (eds.) (2009). *Proceedings of the 17th International Conference on Computers in Education [CDROM]*. Hong Kong: Asia-Pacific Society for Computers in Education.



(a) (b) (c)

Fig. 1. Trending of input words and obtained scores. (a) and (b) are two typical user inputs. (c) is the grading results for each individual concepts contributing to the final overall answer.

saves the current answer, the time, the overall score, and the individual score for each constituent concept. After the test is completed, the trending of the users' input performance and score is plotted for further analysis.

We have invited 25 high school students to participate in the experiment. They were asked to answer the open-ended question of "Why does polar light occur?" by entering their answers in a text box. The number of entered texts and the scores for partial answers over time for two typical students are shown in Fig. 1. The horizontal axis represents the time while the vertical axis is for the numbers of words (dark line) and the scores (light line). For the case of Fig. 1(a), 63 words were entered, and the overall score at the end is 3.9 out of 6. One can observe that his/her score grew consistently with the number of entered texts. On the other hand, in the case of Fig. 1(b), the score of the student stopped growing at one third of the course although he/she continued to input new texts for the whole test. By taking a closer look at the answer inputted by the students, we found that indeed the student could not provide more meaningful explanation after that point, which implied that appropriate feedbacks to the student from the system might be desirable at this time.

However, in order to construct useful feedbacks to the students, we have to analyze the entered texts further. In the off-line learning step, we asked human graders to give a grade to each of the three concepts involved in answering the question. The learned model is then used to assess students' individual concepts from their answers. For example, in the case of Fig. 1(c), the differences of the scores for the three concepts implied that the student is weak in one of the three concepts (the lowest curve) especially in the beginning. Appropriate feedbacks such as hints or additional multiple-choice questions for this concept can be provided according to the application context.

3. Conclusion and Future Work

By observing how a student answers, a good tutor should be able to tell the knowledge status of the student and the missing concepts critical for constructing a correct answer. In this paper, we have proposed a prototype system aiming to collect the answering behaviors of a student and perform on-line diagnosis by automatic scoring on open-ended questions. We are currently working on developing appropriate measures for categorizing these answering behaviors as well as the mechanism for providing on-line feedbacks.

References

[1] Shermis, M. D., & Burstein, J. C. (2003). *Automated essay scoring: A cross-disciplinary perspective*. NJ: LEA.

[2] Wang, H.C., Li, T.Y. & Chang, C.Y. (2005). A User Modeling Framework for Exploring Creative Problem-Solving Ability. In *Proc. of the 12th Intl. Conf. in Artificial Intelligence in Education*.

[3] Wang, H.C., Li, T.Y., Huang, C.C. & Chang, C.Y. (2006) VIBRANT: A Brainstorming Agent for Computer Supported Creative Problem Solving. In *ITS, LNCS 4053*, (pp. 787 – 789).

[4] Wang, H.-C., Chang, C.-Y. & Li, T.-Y. (2008) Assessing Creative Problem-Solving with Automated Text Grading. *Computer and Education*, 51, pp.1450-1466.