Towards the Prediction of User Actions on Exercises with Hints Based on Survey Results

Pedro J. Muñoz-Merino, Abelardo Pardo, Mario Muñoz-Organero, and Carlos Delgado Kloos

Universidad Carlos III de Madrid, Department of Telematics Engineering,
Avda de la Universidad, 30, E-28911 Leganés (Madrid), Spain
{pedmume,abel,munozm,cdk}@it.uc3m.es

Abstract. The actions a user performs on exercises depending on the different hinting techniques applied, can be used to adapt future exercises. In this paper, we propose a survey for users in order to know their different actions depending on different conditions. The analysis of preliminary results for some questions of the model shows that there is a correlation between some survey questions and the real student actions, but there is a case in which there is not such correlation. For the cases where that correlation exists, this correlation leads to think that some prediction of users actions based on survey results is possible.

Keywords: student behavior, adaptation, exercises, hints, survey.

1   Introduction and Related Work

The generation of adaptive hints for tutoring systems based on exercises has been addressed (e.g. [1], [2]). The paper [1] takes into account the student knowledge level and their mental state so that Bayesian networks can be applied to adapt hints in the Andes system. While article [2] applies the Item Response Theory to calibrate the hints over the SIETTE system based on the student knowledge about the topics. In addition, paper [3] also proposes an adaptive solution for the delivering of hints.

All these works dealt with the adaptation of hint contents but they did not adapt the different hinting techniques of an exercise without changing the knowledge contents. An example of hinting technique adaptation would be that a system can decide if there is some penalty for viewing hints or not depending on the user. The use of different hinting techniques can imply different learning gains in some situations [4].

Monitoring systems can be implemented in order to retrieve these user interactions, and a post-processing of the data must be done in order to have the desired student actions. Different works [3], [5], [6], [7], [8] give excellent ideas about the type of user interactions that can be retrieved into hinting systems, including the number of tries and selected hints, or the incorrect answers. From previous students actions we can have an estimation of their future behavior [1], or the prediction of their test scorings [5], as well as the correlation of different student behaviors with respect to their learning gains [8], or an evaluation of the instructional effectiveness of different exercises [8]. Therefore, user interactions in the hint context are valuable information.

In addition, these user interactions can be written into suitable formats called Contextualized Attention Metadata [9] that can bring together different feeds and store different user events. Some tools have been created for monitoring and processing user interactions such as CAMERA [10] with very interesting applications.

Instead of monitoring the different student actions, we propose giving a survey to the users. Although decisions based on the monitoring can be better e.g. for adaptation than those based on survey results, nevertheless sometimes this method can be worth, because the costs of implementation of monitoring and processing systems can be saved, as well as the problem of the initial training for the system is solved. In order to make this approach scalable for many students, a web survey form has been designed so that students can submit their survey answers through the Web.

There are user questionnaires for adaptation purposes such as [11] for feedback taking into account a model about learning styles [12]. There are also other prediction works not applied in the hinting context, such as to try to predict a user model from user behavior in a shopping context [13] or the interests of a museum visitor [14].

2   Student Survey Regarding Exercises with Hints

The survey included three questions to determine students' hint abuse. The questions and the hint abuse concept are in the direction of the work presented in [7]:

1.- If I am quite sure to know the solution of an exercise:
   a) I answer the solution of the problem immediately
   b) I like to view all the hints previously
2.- If I have no idea of the solution of an exercise:
   a) I like to make several attempts before viewing the hints
   b) I like to view the hints previously before making several attempts
3.- If I am not sure to know the answer to an exercise:
   a) I like to make several attempts before viewing the hints
   b) I like to view the hints previously before making several attempts

Next, the survey includes some questions about different hinting techniques without changing the hint contents. Tables 1 and 2 show the rest of the survey questions that include aspects about the importance, fairness, preferences and student behavior. Each question is rated between 1 and 7 by students. Seven categories were selected based on studies that suggest the optimal number of categories (e.g. [15]).

Table 1. Survey questions for rating the importance of different hinting techniques

Question
- Hints that are only available when a user answers a problem incorrectly
- Hint sequence (a hint that is composed by several steps. Each step is a problem)
- Selection of a maximum number of hints out of several that are proposed
- Penalties on the scoring for viewing hints
- Penalties on the scoring for incorrect attempts
- Rewards on the scoring for answering hints correctly
- Penalties on the scoring for answering hints incorrectly

Table 2. Survey questions about the fairness, preferences and the own perceived behavior

Question
- I think it is fair that each incorrect attempt implies penalties on the scoring
- I think it is fair that viewing hints implies penalties on the scoring
- I think it is fair that answering hints incorrectly implies penalties on the scoring
- I think it is fair that answering hints correctly implies rewards on the scoring
- I think it is fair that hints can only be activated as a result of a student incorrect attempt
- When the system shows several hints and the user can select a maximum of them, I think that it is fair such limitation in the number of hints that can be selected
- I prefer hints that are only text instead of hints that are other problems to solve
- I prefer hints that are available directly at the beginning, instead of hints that are only available after an incorrect answer to the initial problem
- I prefer a huge number of hierarchical hint levels (concept of hints about hints)
- I prefer a hint that is divided into a sequence of many problems (hints that are several problems in different steps) instead of a hint that is only a problem
- I prefer questions with score in order to evaluate my knowledge
- The fact of penalizing on the scoring for incorrect attempts lead me not to answer
- The fact of penalizing on the scoring for viewing hints lead me not to select them
- The fact of penalizing on the scoring for answering hints incorrectly lead me not to answer
- The fact of increasing the scoring for answering a hint correctly, lead me to answer

3 Research Question and Description of the Experience

The main research question of this work is if some questions of a specific survey to students are correlated with their future user actions on exercises with hints. There are several factors that can make a difference between the different student survey answers with respect to their real actions, such as the following:

- While some survey questions are directly related to the future students behaviors, there are other questions that do not, e.g. about the user preferences.
- A student can misunderstand a question so that she does not reflect her real thoughts and feelings. Or as the survey is filled in before the user interactions, students cannot understand well the different proposed situations.
- Students can lie or they cannot be aware of their own future actions.

An experience took place in the classroom in 2010 in the context of a Computer Architecture Laboratory Course (http://www.it.uc3m.es/pedmume/LAO/index.html). Students received a survey with most of the questions of Section 2. The survey was filled in by students before their interaction with the system. Next, students performed a pre-test. Next, they interacted with 26 exercises with hints in XTutor during two sessions: the first one of 90 minutes, and the second one of 110 minutes. Each student was provided with different exercises with associated hints, the most relevant for this work were: six exercises with penalties for viewing hints, and other six without any penalties; three exercises with hints directly available, and other three without them. The number of samples was 29, as this is the number of students that filled in the survey correctly and attended both sessions.

4 Preliminary Results

Table 3 shows some students actions information that were used for this research:

Table 3. Retreived information from the tutor regarding the experience

Acronym	Description
T	The pre-test results.The pre-test was composed by 8 questions, and each student obtained an scaled grade between [0,1]
PHS	The percentage of selected hints considering the total number of exercises that a student interacted with. It is considered that a student selected a hint if she selected at least one of the hints before answering the exercise correctly
PW	The percentage of exercises for which a student selected a hint when there were no penalties for doing it
PP	The percentage of exercises for which a student selected a hint when there were penalties for doing it
PD	The percentage of exercises for which a student selected a hint when the hint is directly available
PA	The percentage of exercises for which a student selected a hint when the hint is only available as a result of an incorrect attempt
Diff_HS1	PW – PP: The difference in the percentage of selected hints when there is not any effect on the scoring with respect to having a penalty for selecting hints
Diff_HS2	PD – PA : The difference in the percentage of selected hints when the hint is directly available with respect to not having it directly available

We present some preliminary results that compare three types of student actions with respect to some questions of the survey and the pre-test grades. The selected student actions to be predicted were: *PHS, Diff_HS1*, and *Diff_HS2*.The dedicated time for the sessions gave only chance to have enough data for these three actions.

The *Diff_HS1* would be tried to be predicted based on the student answers to the survey question of "The fact of penalizing in the scoring for viewing hints led me not to select many" (*Survey_1*), while the *Diff_HS2* based on the survey question: "I prefer hints that are available directly at the beginning, instead of hints that are only available after an incorrect answer to the initial problem" (*Survey_2*).

Finally, the *PHS* would be tried to be predicted based on the three initial survey questions about hint abuse, but also taking into account the student grade on the pre-test because a student with more knowledge does not need so many hints. For the three initial questions, a student is given a value from [0,3], this is the hint abuse rate (*AR*), depending on the times she prefers to view hints previously. The proposed estimator for trying to predict *PHS* would be the multiplication of *AR* by *T*.

Table 4 shows the Pearson linear correlation coefficient between the commented parameters (where N=29 for all the cases). We can observe that two correlations were statistically significant ($p<0.05$) while one was not. Therefore, students have some perception about their own behavior depending on if there are penalties or not for viewing hints. And their preferences about the hint being available at the beginning or only as a result of an incorrect attempt are according to their real behavior. But in both cases, with the monitored data, a linear model with statistically significant coefficients cannot be provided to predict the user actions based on the survey results.

Table 4. Pearson correlation coefficient between different parameters

Parameters	Correlation	p
AR*T and PHS	0.112	0.282
Survey_1 and Diff_HS1	0.322	0.044
Survey_2 and Diff_HS2	0.365	0.031

The result between *AR*T* and *PHS* can mean that the real student actions about the percentage of hints selected is not according to what they answered in the survey. Indeed, analyzing the data, we can observe that there are three students that answered that they like to view all the hints previously in all the cases but they did not really view many hints. The reasons of this can be the same as commented in section 3.

In addition, we can think that neither *AR*T* is a proper parameter nor *PHS*, and more proper estimators and parameters should be selected to be taken into account in a prediction model. The *AR*T* and *PHS* parameters are high level ones, since *AR*T* considers the hint abuse rate as the result of joining three different survey questions and the total pre-test grade, and *PHS* does not take into account an analysis of each specific question but joins all in a group without a difference between the different initial student knowledge levels for each question. But more specific estimators can be determined, considering for each question the initial student knowledge level in that question and determining if a student really needs a hint or not for it, and if it is according to what she answered in the specific question of the survey.

5 Conclusions

We have proposed a new survey to retrieve the features of a user model regarding different hinting techniques that can be applied to exercises, and about user preferences for viewing hints depending on their knowledge in an exercise.

In one case, there was not a statistically significant correlation between the real student actions and the selected survey questions for that aspect combined with the pre-test grade. On the other hand, there were statistically significant correlations between two student actions and two survey questions. This leads to think that for some cases we can rely on the survey results to try to predict the real user actions.

These are only preliminary results since from this experience we can only analyze three different types of student actions. Furthermore, there are only 29 student samples, and future experiences might complement this study. In addition, other variables might also be included in future studies such as the specific course topics.

Another interesting topic for future research would be to study students' actions to gain the system (e.g. answer the survey questions lying in a way that they know that they will receive more hints) and possible measures to prevent from it.

A challenge is to find the best model that can predict the real student actions based on the different questions of the survey, when this is possible. In order to achieve it, more student interactions with exercises and more student samples would be needed.

References

1. Gertner, A.S., Conati, C., Vanlehn, K.: Procedural help in Andes: Generating hints using a Bayesian network student model. In: 15th National Conference on Artificial Intelligence, pp. 106–111. AAAI Press, Menlo Park (1998)
2. Conejo, R., Guzmán, E., de-la Cruz, J.-L.P., Millán, E.: An Empirical Study About Calibration of Adaptive Hints in Web-Based Adaptive Testing Environments. In: Wade, V.P., Ashman, H., Smyth, B. (eds.) AH 2006. LNCS, vol. 4018, pp. 71–80. Springer, Heidelberg (2006)
3. Zhoul, Y., Freedman, R., Glass, M., Michael, J.A., Rovick, A., Evens, M.W.: Delivering hints in a dialogue-based intelligent tutoring system. In: 16th National Conference on Artificial intelligence, pp. 128–134. AAII (1999)
4. Muñoz-Merino, P.J., Delgado Kloos, C., Muñoz-Organero, M.: Deciding on Different Hinting Techniques in Assessments for Intelligent Tutoring Systems. International Journal of Innovative Computing Information and Control 7(2), 841–858 (2011)
5. Feng, M., Heffernan, N.T., Koedinger, K.R.: Predicting state test scores better with intelligent tutoring systems: developing metrics to measure assistance required. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 31–40. Springer, Heidelberg (2006)
6. Baker, R., Corbett, A.T., Roll, I., Koedinger, K.R.: Developing a generalizable detector of when students game the system. User Mod. and User-Adapted Int. 18, 287–314 (2008)
7. Aleven, V., Mclaren, B., Roll, I., Koedinger, K.R.: Toward Meta-cognitive Tutoring: A Model of Help Seeking with a Cognitive Tutor. Int. J. Artif. Intell. 16, 101–128 (2006)
8. Feng, M., Heffernan, N.T., Beck, J.: Using learning decomposition to analyze instructional effectiveness in the ASSISTment system. In: 14th International Conference on Artificial Intelligence in Education (AIED 2009), pp. 523–530. IOS Press, Amsterdam (2009)
9. Wolpers, M., Najjar, J., Verbert, K., Duval, E.: Tracking Actual Usage: the Attention Metadata Approach. Educational Technology & Society 10(3), 106–121 (2007)
10. Schmitz, H., Scheffel, M., Friedrich, M., Jahn, M., Niemann, K., Wolpers, M.: CAMera for PLE. In: Cress, U., Dimitrova, V., Specht, M. (eds.) EC-TEL 2009. LNCS, vol. 5794, pp. 507–520. Springer, Heidelberg (2009)
11. Vasilyeva, E., Pechenizkiy, M., De Bra, P.M.E.: Adaptation of Elaborated Feedback in e-Learning. In: Nejdl, W., Kay, J., Pu, P., Herder, E. (eds.) AH 2008. LNCS, vol. 5149, pp. 235–244. Springer, Heidelberg (2008)
12. Felder, R.M., Silverman, L.K.: Learning and Teaching Styles in Engineering Education. Journal of Engineering Education 78, 674–681 (1988)
13. Mori, J., Matsuo, Y., Koshiba, H., Aihara, K., Takeda, H.: Predicting Customer Models Using Behavior-Based Features in Shops. In: Houben, G.-J., McCalla, G., Pianesi, F., Zancanaro, M. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 126–137. Springer, Heidelberg (2009)
14. Bohnert, F., Zukerman, I., Berkovsky, S., Baldwin, T., Sonenberg, L.: Using Interest and Transition Models to Predict Visitor Locations in Museums. AI Communications 21 (AICom) – Special Issue on Recommender Systems, 195–202 (2008)
15. Malhotra, N.K.: Questionnaire design and scale development. In: The Handbook of Marketing Research Uses Misuses and Future Advances, pp. 83–94. Sage Publications, Inc., Thousand Oaks (2006)