Modeling the Affective States of Students Using an Intelligent Tutoring System for Algebra

Thor Collin S. Andallaza[1], Ma. Mercedes T. Rodrigo[1], Maria Carminda V. Lagud[2], Rina Joy M. Jimenez[1], and Jessica O. Sugay[1]

[1]Ateneo Laboratory for the Learning Sciences, Department of Information Systems and Computer Science, Ateneo de Manila University, Quezon City, Philippines
[2]Education Department, Ateneo de Manila University, Quezon City, Philippines
{tandallaza, mrodrigo, jsugay}@ateneo.edu, minminvl@gmail.com, jimenezrjm@yahoo.com

Abstract. The medium-term goal of this project is to implement an emotionally intelligent learning companion that will provide algebra students with encouragement and support. To this end, this study sought to improve upon existing models of affect of students using Aplusix, an intelligent tutor for algebra. Continuing earlier work, this model was a refined analysis of student interaction logs with the ITS using linear regression. Unfortunately, the models produced had low correlations with the data.

Keywords: affect, Aplusix, embodied conversational agent, feature selection, student modeling

1 Introduction

Intelligent tutoring systems are computer applications that are capable of providing individualized instruction to learners through the use of artificial intelligence, thereby supporting the learner and facilitating the learning process [12], [16]. ITSs are able to support this level of interaction through the maintenance of a cognitive model of the student, where the model is an assessment of the student's mastery of the skills being taught. Examples of cognitive-based modeling include rule-based models, which allow the ITS to determine the knowledge used in a generation of step-by-step solutions [1]. Cognitive models allow ITSs to comprehend student actions, such that they are able to map out these actions to know solution paths, and thus generate appropriate content that adapt to the capabilities of individual students [1].

In addition to this, however, literature (e.g. [3], [13]) has shown that there is a progression towards student modeling beyond the cognitive aspect. Examples of such non-cognitive information include the ability of a student to be challenge-seeking [12], as well as patterns in behavior that are indicative of "gaming the system", i.e., an abuse of system functionality through systematic guessing [3]. The use of non-cognitive models allows ITSs or other software, such as agents (e.g. [2], [3], [13]), to provide motivational and emotional support and to suppress undesirable or non-learning behavior.

One example of these systems is Aplusix, an intelligent tutor for algebra [5], [6], [10]. Aplusix features an advanced editor that allows for step-by-step solutions to algebraic and arithmetic problem sets. The ITS keeps track of student progress through logs generated during a session, and provides visual feedback on the student's progress through the editor and interface [6], [10].

Previous work with Aplusix (e.g., [2], [10]) attempted to generate various non-cognitive models. Lagud and Rodrigo's study [10], in particular, has been able to map out students' learning profiles with their affective profiles, thus revealing features that are indicative of particular affective states. Lagud defined a student's affective profile as a vector with seven percentages. Each percentage represented the proportion of time during which the student exhibited an affective state during the observation period. A student's learning profile, on the other hand, was another vector with four terms: the number of correct items solved, the average number of steps taken to solve each problem, the average time to solve each problem, and the highest difficulty attempted [10].

In a subsequent study, Andallaza and Jimenez [2] used these findings to construct student models per problem type instead of per student or per session basis. The refinement used standard deviation and terciles to generate threshold values that can be used to detect and evaluate current student affect on a real time basis. They then created an embodied conversational agent (ECA) with the models, thereby enabling the ECA to detect and respond to student affect in the hopes of directing and sustaining motivation in learning. Upon testing the agent, however, recommendations from the study revealed the need for more refined and more accurate models that will enable the ECA to provide more timely and more appropriate motivational responses. Such is the objective of this study, where another form of analysis through linear regression will be done on models determined in previous work [10]. It is hoped that the analysis will be able to establish better baseline values for the affective states of engagement, confusion, and boredom. These, in turn, can then set the foundation for this study's medium-term goal: the development of an emotionally intelligent agent for Aplusix.

2 Methods

In this section, we discuss the testbed for the study, the data collection methods, and our approach to data analysis.

2.1 Aplusix

Aplusix (Figure 1) covers a variety of topics, from factorization to solving equations and inequalities. It allows users to use a step-by-step method in arriving towards solutions [6], [10]. Students can choose to tackle these topics in problem sets, where each set is of a certain level of difficulty. The ITS provides visual feedback on student progress through the use of two parallel lines. The lines are black when the connected steps are equivalent (e.g. gravitating towards a known solution), and the lines are red with an additional X when the connected steps are not equivalent. Apart from the advanced editor that allows for step-by-step calculations, Aplusix also generates reports on current student progress in the attempt to resolve the problem, as well as domain-based agents in the form of Chloe, Julien, and Olivia that students may interact with to get hints or the final solution to their current problem.

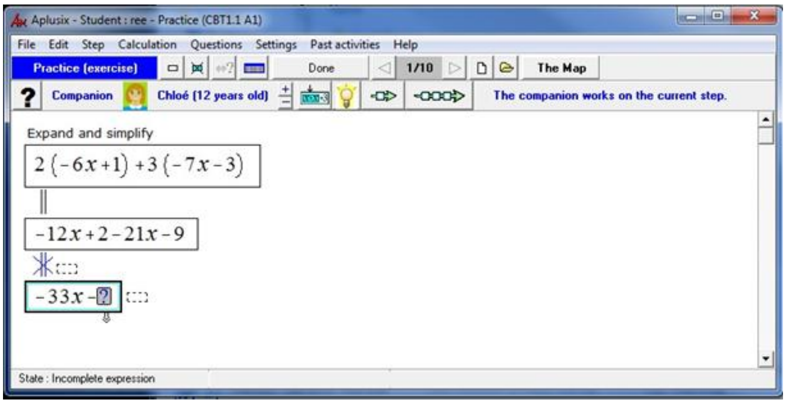

Fig. 1. The Aplusix environment.

2.2 Data Collection

The student-tutor interaction data and affect observation data were collected in the study of Lagud and Rodrigo [10]. The experiment was done with high school students whose ages ranged from 12 to 15, with 13.5 as the average age and 14 as the modal age. The one hundred and forty students, of which 83 are male and 57 are female, came from five different high schools – four of which are located in Metro Manila, and one located in Cavite. In addition, all of the students were computer-literate, but none have used nor are familiar with Aplusix [10].

The methods for collecting the data of the students' affective states were adapted from the study of Baker et al. [3], which made use of human observers rather than cameras or any specialized equipment. Observations were done by three pairs of observers, all of whom were briefed and trained through a series of pre-observation discussions, as well as practices regarding coding strategies for these affective states. The list of affective states to be observed for the study were taken from a study of Rodrigo et al. [14], the list including engaged concentration [4] (a subset of flow [8]), boredom, confusion, delight, surprise, frustration, and the neutral state. The students used Aplusix for 42 minutes, and a total of thirteen pairs of observations were carried out, each lasting twenty seconds, with each student being observed once every 180 seconds. The inter-rater reliability, which is a measure of agreement between observers, was then computed using Cohen's Kappa, whose resulting value turned out to be acceptably high (K = 0.63) [10]. This meant that the observers' agreements were not by chance.

2.3 Preliminary affect modeling

In [10], Lagud and Rodrigo arrived at relationships between learner and affective profiles of students using Aplusix. Each student's learner profile was defined as the number of problems correctly solved, the highest difficulty level attempted, the average time to solve a problem, and the average number of steps used to solve the problem. A student's affective profile, on the other hand, was the percentage of the student was observed exhibiting each of the seven affective states of interest.

An ANOVA performed on the learner and affective profiles showed that students with the highest number of correct answers exhibited the most engaged concentration while students with the lowest number of correct answers experienced confusion and boredom the most. Students who attempted the most difficult problems exhibited the most engaged concentration while students who tried the lowest levels experienced more boredom and confusion. Students who took the longest time in solving the algebra problems exhibited the most confusion while students who took the shortest time exhibited confusion least. Students who used the most number of steps to solve a problem exhibited confusion and boredom. Students who used the least number of steps exhibited the most engaged concentration.

The limitation of [10] was that the models were coarse-grained. They found relationships between affect and learning indicators from an entire session's worth of data. If the goal was to be able to provide students with motivational support in real-time, the models had to be finer grained.

2.4 Earlier attempts at model refinement and ECA development

Andallaza and Jimenez [2] attempted to arrive at a finer-grained model. The authors collapsed every student's attempt to solve a problem into a single vector and then sorted these vectors by problem type. For each problem type they computed the standard deviations and terciles for the number of steps taken, and time spent solving problems of that level.

For each of these criteria, both the standard deviations and tercile analyses divided students into three groups—an average group, an above average group, and a below average group (See Table 2). The major difference between these methods of division is that each of the terciles had more or less the same number of rows. The standard deviations method on the other hand had a variable number of rows per division.

Table 1. Sample student model for problem type/level A1, number of steps using standard deviations.

Problem Type	Standard Deviation Method Values		
	Mean	SD	% Popl'n above or below 1 SD
A1	18	23	9.62

Table 2. Sample student model for problem type/level A1, number of steps using terciles.

Terciles Method Values				
Group	Tercile Size	Min	Max	Mean
Above average	19	3	9	6
Average	16	10	13	11
Below Average	17	15	143	37

Of the resulting values from both methods of analysis, only the threshold values from the terciles method were usable. The standard deviation values themselves turned out to be larger than their corresponding mean values, indicating that the variance in the original log data.

[2] generated a script hat mapped the agent responses to particular affective states and to specific subconditions within these states. For instance, a student who is engaged and takes less time than most engaged students to solve a problem would receive the message, “That was fast! Good job!” On the other hand, a student experiencing boredom, yet is near the correct answer receive the message, “Don’t give up just yet.” This was an attempt to ensure that the agent would respond appropriately, given an observed state.

The responses were delivered through an animated agent called Grimace [17] (Figure 4). Grimace’s face showed an appropriate expression well as the text form of the message. In addition, there was a text-to-speech module that allowed students to hear the text message [2].

Fig. 2. The initial ECA version with Aplusix shown at the back.

The application design, especially its multimodal output, enabled the agent to closely resemble a human tutor-student environment, giving it the potential to fulfill its design goals of providing motivation in the study of algebra. Unfortunately, the agent was not very effective. The agent delivered its responses too quickly and too frequently, causing some students to feel irritated at the agent. Some even elected to mute the sound because of the unwanted intervention. These observations called for the second iteration of the agent, which is the medium-term goal of this study. Consequently, this also called for the construction of better student models, which meant revisiting the original Aplusix log data taken from previous work [10] and viewing it in another perspective.

2.5 Description of the Data

We once again revisited the learner and affective profiles from [10] and combined them in a worksheet show in Figure 5.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	school	run	student	flow	bor	con	del	fru	neu	sur	raw score	# steps	duration	level	levelnum
2	alphonsus	1	7	0.8462	0	0.1154	0	0	0.0385	0	3	48.67	2.19	B1	1
3	alphonsus	1	8	0.9231	0	0	0.0385	0	0	0.0385	16	76.44	1.72	B2	2
4	alphonsus	1	9	0.8077	0	0.1154	0.0385	0	0.0385	0	21	65.24	1.33	B3	4
5	alphonsus	1	10	0.9231	0.0385	0.0385	0	0	0	0	9	92.89	2.28	B1	1

Fig. 3. Screenshot of the data table used for the analysis.

The first three columns contained the student code, with the student's school name, run number, and student number. The next seven columns were the affective profile of the student: The percentage of time a student was observed to be in a specific affective state [10]. Finally, the remaining columns the make up the student's learning profile, derived values features logged by Aplusix [10].

2.6 Linear Regression

We used Weka [9] to produce linear regression models of all seven affective states. In a linear regression, each of the observed affective states is defined as a formula of the form

$$x = c_1f_1 + c_2f_2 + \ldots + c_nf_n + \text{constant} \quad (1)$$

A variable x (in this case, an observed affective state) is expressed as the sum of the coefficient values (c1, c2, …, cn, constant) multiplied by respective features associated with x, given by f1, f2, …, fn.

Once these models were generated, we computed for their correlation coefficients using 10-fold cross-validation. The coefficients are in the range [-1, 1], where -1 indicates a perfectly inverse relationship between the model and the value being predicted, 1 indicates that the model always predicts the said value, and 0 indicates that the model is not predictive at all.

3 Results and Analysis

The results of the linear regression yielded the following models as shown in Table 3. Although none of the models were good, most of the results followed the observed patterns discussed in [10].

The models imply that when students take more steps to solve a problem, they are less likely to be engaged. On the other hand, students who take more steps are more likely to be bored. In addition, students who have less number of correct answers, who take more steps, and who take more time in solving each problem are said to be confused. The interesting observations include surprise and neutral, where students are said to be more surprised if they have more correct answers, and are more neutral when they less attempt problems of higher difficulty.

Table 3. Linear regression results

Affective state	Model	Correlation Coefficient
Engaged concentration	- 0.001 * *average no. of steps*+ 0.8117	0.1843
Boredom	0.0003 * *average no. of steps* + 0.0097	-0.0046
Confusion	- 0.0021 * *no. of correct answers* + 0.0004 * *average no. of steps* + 0.006 * *average time to solve each problem* + 0.1235	0.2334
Delight	+ 0.061	-0.1712
Surprise	0.0002 * *no. of correct answers* + 0 * *average no. of steps* + -0.0041	-0.1486
Frustration	+ 0.0275	-0.3805
Neutral	-0.0025 * *highest level attempted* + 0.0189	0.1591

Although the results made some intuitive sense, none of them were good according to the computed coefficients for each model. In fact, none of the features appeared in the models for delight and frustration. Out of the five remaining models, four of them made use of the feature *average no. of steps*, which again made intuitive sense -- the actual number of steps it takes for a student to complete a problem has implications on what the student is feeling. In particular, a student who is engaged takes fewer steps on average, while a student who is bored or confused takes more steps to finish.

There are several possible reasons that might explain why the models produced were weak. It is possible that the features selected were not telling enough. In the current table, only four features were used as inputs to the model. Additional feature engineering is necessary to arrive at other student behaviors that might be indicative of affective states.

Some affective states occurred rarely: Boredom only occurred 3% of time [15]. Frustration only occurred 2% of the time. Finally, delight occurred less than 10% of the time. There might not have been enough data to arrive at models of these states.

Hence, other modeling approaches might have to be used to flush them out. Alternatively, we can focus on modeling the states that are prevalent yet are still interesting, specifically, confusion.

4 Future Work

The creation of these student models, which are essentially iterations over the previous student models used in the previous agent, is the first, yet one of the most crucial steps in the development of the ECA for Aplusix. The goal of this study was to produce a new model that should be able to provide a more accurate evaluation of a student's current affective state, the detection of any possible changes to it, and detection of opportunities for agent intervention. Unfortunately, the new approach did not lead us to good models. They did confirm the importance of the average number of steps as a feature that is indicative of student affect. Further study is needed, though, in order to arrive at more accurate models. Ways forward include the generation of other features from the raw data and focusing on more frequently-occurring affective states such as confusion.

An additional way forward is to reduce the grain size of the data in future analysis. The present analysis represents each student with a single vector. This vector is a summary of over 40 minutes of interaction time. Thus, any model produced will only be able to predict student affect based on 40 minutes of data. However, for an ECA to be useful, it has to be able to respond to a student in a timely manner. Hence, future research will have to find a way to split the raw data into several time windows.

We will try applying the Bayesian Knowledge Tracing (BKT) [7] framework to analyze the data. Bayesian Knowledge Tracing a way of estimating student knowledge as well as predict future performance based on prior performance. It assumes that students either know or do not know a concept. A concept can transition from the unknown to known (or learned) state through opportunities to practice. It is not automatic, though, that an opportunity to practice leads to learning. It is possible that a student knows a concept but slips, leading to an error. It is also possible that a student guesses and answers correctly. Hence, the probability that a student has learned a concept is updated after each opportunity to practice, based on whether the answer of the student was correct or not. Part of the challenge to taking this approach is mapping Aplusix's log features to the features needed for Bayesian Knowledge Tracing. Upon generating the BKT values, these can then be included among existing features to general a new model.

Once an acceptable model is reached, the next step to be taken is the integration of these models to the working iteration of the agent and the inclusion of additional responses, as well as revisions to the appearance of the agent. Finally, we will conduct an actual field test of the agent, where it will be used by introductory algebra high school students in order to determine what behaviors, reactions, and overall effect the agent will have on math learning.

Acknowledgements. We thank the Ateneo Laboratory for the Learning Sciences for unique contributions and support for the research. We thank Department of Science and Technology Philippine Council for Industry, Energy, and Emerging Technology Research and Development (PCIEERD) for making this research a reality through the grant entitled, "Development of Affect-Sensitive Interfaces".

References

1. Alven, V.: Rule-Based Cognitive Modeling for Intelligent Tutoring Systems. In: Nkambou, R. et al. (eds.) Advances in Intelligent Tutoring Systems. SCI, vol. 308, pp. 33-62. Springer, Heidelberg (2010)
2. Andallaza, T.C.S., Jimenez, R.J.M.: Design of an Affective Agent for Aplusix. Undergraduate thesis, Ateneo de Manila University, Quezon City (2012)
3. Baker, R.S.J.d., Corbett, A.T., Koedinger, K.R., Wagner, A.Z.: Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System". Proceedings of ACM CHI 2004: Computer-Human Interaction, 383-390 (2004)
4. Baker, R.S.J.d., D'Mello, S.K., Rodrigo, M.M.T., Graesser, A.C.: Better to be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States during Interactions with Three Different Computer-Based Learning Environments. Intl. Journal of Human-Computer Studies, vol. 68 (4), pp. 223-241 (2010)
5. Bouhineau, D., Nicaud, J.F., Chaachoua, H., Huguet, T., Bittar, M., Bronner, A.: Two years of use of the Aplusix System. In: Proceeding of the 8th IFIP World Conference on Computer in Education, p. 23. University of Stellenbosch, Cape Town South Africa (2005)
6. Chaachoua, H., Nicaud, J., Bronner, and Bouhineau, D.: APLUSIX, a Learning Environment for Algebra, Actual Use and Benefits. In: 10th International Congress on Mathematics Education (2004)
7. Corbett, A.T., Anderson, J.R.: Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. User Modeling and User-Adapted Interaction 4, 253–278 (1995)
8. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper & Row, New York (1990)
9. Holmes, G., Donkin, A., Witten, I.H.: Weka: A Machine Learning Workbench. In: Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Tutoring Systems, pp. 357-361 (1994)
10. Lagud, M.C.V., Rodrigo, M.M.T.: The affective and learning profiles of students while using an intelligent tutoring system for algebra. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010. LNCS, vol. 6094, pp. 255-263. Springer, Heidelberg (2010)
11. Nicaud, J., Bouhineau, D, and Huguet T.: The Aplusix-editor: A new kind of software for the learning of algebra. In: Cerri, S., Gouardères, G., Paraguaçu, F. (eds.) ITS 2002. LNCS, vol. 2363, pp. 178-187. Springer, Heidelberg (2002)
12. Nwana, H.: Intelligent tutoring systems: An overview. Artificial Intelligence Review vol. 4, pp. 252-277 (1990)
13. Rebolledo-Mendez, G., Du Boulay, B., Luckin, R.: Motivating the Learner: An empirical evaluation. In: 8th International Conference on Intelligent Tutoring Systems, pp. 545-554 (2006)
14. Rodrigo, M.M.T., Baker, R.S.J.d.: Comparing learners' affect while using an intelligent tutor and an educational game. Research and Practice in Technology Enhanced Learning vol. 6, pp. 43-66 (2011)

15. Rodrigo, M. M.T., Baker, R.S.J.d., D'Mello, S.K., Gonzalez, M.C.T., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sugay, J.O., Tep, S., Viehland, N.J.B.: Comparing Learners' Affect While Using an Intelligent Tutoring System and a Simulation Problem Solving Game. In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) ITS 2008. LNCS, vol. 5091, pp. 40–49. Springer, Heidelberg (2008)
16. Rosi, M., Stankov S., Glavinic, V.: Intelligent tutoring systems for asynchronous distance education: In: 10th Mediterranean Electrotechincal Conference (2000)
17. Spindler, O., Fadrus. T. Grimace Project Doumentation. Vienna University of Technology (2009)