Modeling Affect in Student-driven Learning Scenarios

Paul Salvador Inventado* , Roberto Legaspi, Rafael Cabredo*, and Masayuki Numao
The Institute of Scientific and Industrial Research, Osaka University
8-1 Mihogaoka, Ibaraki, Osaka, Japan, 567-0047
{inventado,cabredo,roberto}@ai.sanken.osaka-u.ac.jp, numao@sanken.osaka-u.ac.jp

ABSTRACT
Much research has been done on affect detection in learning environments because it has been reported to provide better interventions to support student learning. However, students' actions inside these environments are limited by the system's interface and the domain it was designed for. In this research, we investigated a learning environment wherein students had full control over their activities and they had to manage their own goals, tasks and affective states. We identified features that would describe students' learning behavior in this kind of environment and used them for building affect models. Our results showed that although a general affect model with acceptable performance could be created, user-specific affect models seemed to perform better.

Keywords
affect modeling, educational data mining, student-driven learning

1. INTRODUCTION
The current society and workplace is dynamic and requires people to continuously learn new skills and adapt to what is needed. In order to prepare students for this kind of environment, they need to learn how to manage their learning goals, their time, their motivation and their affective states in environments wherein they receive little or no support and they have complete control over their learning.

Self-regulated learners are likely to be capable of adapting to such environments because they can effectively manage the different aspects of a learning scenario. One of the most important yet difficult skills to learn in self-regulation is monitoring one's cognitive and affective states. Knowledge of one's thoughts and affective states helps students evaluate the current situation and identify if it is better to continue with the current activity or change it. When students learn on their own, it is likely that they will spend time doing non-learning related activities. These do not always serve as distractions because they have also been shown to help regulate emotions [12]. Self-monitoring becomes essential in this case because students need to identify when and how much time spent in non-learning related activities is acceptable so that they can still achieve their learning goals. Self-monitoring is not easy because it is a complex meta-cognitive activity requiring much attention and sophisticated reasoning [13]. Learning in complex domains increases cognitive load and makes self-monitoring even more difficult.

In this research we are moving towards the creation of systems that can help students self-monitor by automatically detecting their affective states. It will be helpful for students to be informed about their affective states when they experience high cognitive load so that they can change their behavior accordingly. Such systems can also suggest activities to help them learn better (e.g., refer to notes, seek help) when certain affective states are detected. Another goal of the research is to use a data collection methodology that does not disrupt students' usual learning behavior. The succeeding sections discuss our methodology, the data we used, our affect model creation process and our results.

2. RELATED WORK
Many researchers have tried improving existing learning systems by incorporating affect detection for better feedback. D'Mello et al. [7] for example developed affect models using data from students' interactions with a conversational agent in the domain of computer literacy. The features they used for building these models were based on students' responses, the correctness of the students' answers, their progress and the type of feedback provided by the system. The model they built to distinguish each affective state from each other did not perform very well (i.e., Kappa = 0.163), however the models they built for distinguishing affective states from a neutral state performed better (i.e., Kappa = 0.207 - 0.390).

Baker et al. [1] also developed affect models for students using Cognitive Tutor Algebra. They used features that described students' actions, the correctness of their actions and their previous actions. They built affect models which distinguished one affective state from another (e.g., bored vs. not bored, frustrated vs. not frustrated) whose resulting Kappa values ranged from 0.230 to 0.400.

*also affiliated with: Center for Empathic Human-Computer Interactions, College of Computer Studies, De La Salle University, Manila, Philippines

Our work differs from previous research because we built affect models using data from students who were not limited to learning in a certain domain, who controlled their own learning, who did not receive feedback and there was no information regarding their learning progress.

3. LEARNING BEHAVIOR DATA

In our previous work, we collected data from one male undergraduate student, one male master's student and two female doctoral students who engaged in research activities as part of their academic requirements [9]. The students were aged between 17 and 30 years old wherein three of them were taking Information Science while one doctoral student was taking Physics. During the data collection period, two of the students were writing conference papers and two made power point presentations about their research. Students had control over how they conducted their learning activities and did not receive any direct support from their supervisor. These conditions required all students to manage their own cognitive and affective states as they learned which satisfied our target learning scenario.

Data was collected in five separate two-hour learning episodes from each student over a span of one week. Students freely decided on the time, location and type of activities they did but were required to learn in front of a computer that recorded their learning behavior. All students used a computer in doing their research so the setup was naturalistic and they did not have to change the ways in which they usually learned.

Data about the students' learning behavior was collected by asking them to annotate their behavior after each learning episode using a behavior recording and annotation tool we developed called Sidekick Retrospect [9]. At the beginning of a learning episode, students inputted their learning goals. The system then began logging the applications they used, taking screenshots of their desktop and capturing image stills from their webcam with corresponding timestamps. After a learning episode, students were presented a timeline which showed the desktop screenshots and image stills depending on the position of their mouse on the timeline. This helped students recall what happened during the learning episode so they could annotate it.

Students made annotations by selecting a time range and inputting their *intentions*, *activities* and *affective state*. Intentions can either be goal related or non-goal related relative to the goals set at the beginning of the learning episode. Activities referred to any activity done while learning which could either be done on the computer (e.g., using a browser) or out of the computer (e.g., reading a book). Two sets of affect labels were used for annotating affective states wherein goal-related activities were annotated as delighted, engaged, confused, frustrated, bored, surprised or neutral and non-goal related activities were annotated as delighted, sad, angry, disgusted, surprised, afraid or neutral. Academic emotions [4] were used for annotating goal related intentions because they gave more contextual information about the learning activity. However, academic emotions might not have captured other emotions outside of the learning context so Ekman's basic emotions [8] were used to annotate non-goal related intentions.

Students would inherently recall what happened during a learning episode when they made annotations so it would be easier for them to identify the appropriate labels. Going through the entire learning episode sequentially would also help them annotate more accurately because they would see how and why their activities changed as well as its outcomes. It is possible that students might not annotate the data correctly for fear of judgment or getting lower scores. However, in our experiment we made it clear to the students that their learning behavior would not affect their grades in any way and assured them that these would not be shown or discussed with their supervisors.

The students' annotations were processed and cleaned so that contiguous annotations had a different intention, activity or affective state. Those that were exactly the same were merged. The resulting data consisted of 1,081 annotations from all students with an average of 54.05 annotations (N=20; σ=27.18) in each learning episode.

4. FEATURE ENGINEERING

The data consisted of only three features (i.e., timestamp, intention and activity) and the affective state label for creating affect models. Models built using these initial features performed poorly so new features had to be designed.

Although the students worked on different topics and used different applications, all of them processed and performed experiments on previously collected data, searched for related literature and created a report or document about it. Although students performed many different activities, analyzing the data showed that these activities can be categorized into six general types – information search (e.g., using a search engine), view information source (e.g., reading a book, viewing a website), write notes, seek help from peers (e.g., talking to a friend), knowledge application (e.g., paper writing, presentation creation, data processing) and off-task (e.g., playing a game). This was used as a new feature which we called task. We also added as features the duration of the task and its position in the learning episode. A task's position in the learning episode was expressed as a normalized time value ranging from zero to 100, wherein zero indicated the start of the episode, 50 indicated the middle of the episode and 100 indicated the end of the episode.

Previous research has shown that the occurrence and duration of previous cognitive or affective states influenced the student's current affective state [2, 9]. However, there is no study that describes how long their influences last. For this study, we only considered the effects of cognitive and affective states in the last five minutes which was based on the average duration of tasks in our data. Similarly, there was no study indicating how many elements in a sequence of previous tasks influenced the current task. However, the data showed that students performed only a maximum of five tasks within a five minute interval (i.e., when students quickly shifted from one task to another). So, we considered the past five tasks relative to the current task as features.

To express the relationship between the previous and current tasks, we used *task frequency* (i.e., the number of times a certain type of task was performed in the last five minutes), *task duration* (i.e., the number of seconds each type of task

Table 1: Affect model performance

Classifier	Kappa	F-measure	Accuracy
Naïve Bayes	**0.345**	**0.349**	**63.10%**
J48 (C4.5)	0.333	0.290	62.57%
JRip	0.326	0.331	61.22%
SVM	0.286	0.351	59.91%
Rep-tree	0.284	0.362	59.06%
Bayesian Network	0.181	0.317	39.43%

was performed in the last five minutes), *most frequent task* (i.e., the most frequent type of task in the last five minutes) and *dominant task* (i.e., the type of task that was performed for the longest time in the last five minutes).

The new set of 22 features was used with the affect label for affect modeling.

5. AFFECT MODELING

Rapidminer 5.3 [11] was used in running different machine learning algorithms to build the affect models. Batch cross-validation was used for evaluating the models such that all the data from one student was held out for testing each time. This was used to test if the resulting model would generalize over students.

RapidMiner's genetic algorithm feature selector was used to identify the most relevant features for the classification task. The fitness of each feature subset was calculated by running a given machine learning algorithm on that subset and then using the resulting model's batch cross-validated kappa value. Cohen's Kappa [3] was used because it considers misclassifications of multiple class labels which cannot be handled by other measures like accuracy. Kappa has also been used frequently in classifying educational data [1, 7].

Table 1 shows the results of the evaluation where the Naïve Bayes model gave the highest kappa value of 0.345 using the features selected by the feature selector. This indicates that the model can perform around 34% better than chance.

The feature selector used 12 out of the 22 features for building the affect models. Three of these features were related to the student's current state (i.e., position in the learning episode, duration and task). Five of the features were related to the past tasks employed by the student (i.e., task$_{n-1}$ ... task$_{n-5}$). Two of the features were related to the amount of time spent performing a previous task in the last five minutes (i.e., information search duration and write notes duration). Finally, two features were related to the frequency of performing tasks in the last five minutes (i.e., apply knowledge frequency and off-task frequency).

Just like other research, features related to previous actions were also found to correspond with the current affective state [2, 6, 9, 12] and is probably the reason why these were selected. The task feature was probably selected because some affective states occurred more frequently while performing a particular task. For example, confusion and engagement were commonly associated with knowledge application and viewing information sources most likely because these activities require utilizing current knowledge and understanding

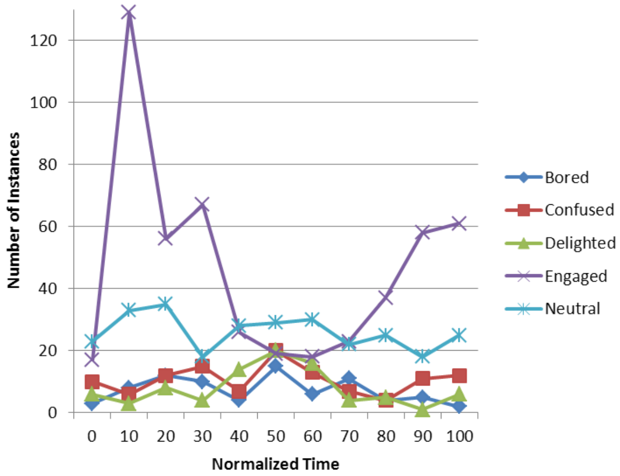

Figure 1: Occurrence of affective states over time

Table 2: Kappa values of user-specific affect models

	J48	JRip	Rep-Tree	SVM	BN	NB
1	**0.675**	0.665	0.628	0.625	0.604	0.310
2	**0.270**	0.147	0.166	0.164	0.229	0.206
3	**0.498**	0.490	0.400	0.425	0.414	0.375
4	**0.532**	0.445	0.484	0.529	0.300	0.256

new information. However, boredom and neutral affective states were commonly experienced only when viewing information sources probably because unlike knowledge application, some information sources may not have been relevant to the student. Delight was usually experienced when students performed off-task activities probably because students engaged in activities they enjoyed during this period.

Certain affective states were experienced more frequently at certain positions in the learning episode which might have caused it to be selected as a feature. Figure 1 shows the total number of times an affective state was experienced by students at certain points of the learning episode. Engagement was experienced more frequently at the start and at the end of the session while the occurrence of confusion and delight increased in the middle of the episode. This is indicative of the phases of flow wherein a learner starts in an equilibrium state of understanding, which is challenged by new knowledge usually exhibited by feelings of confusion and is later assimilated leading back to an equilibrium state [5, 10].

We also investigated the performance of user-specific affect models by using each student's data separately. The models were evaluated with batch cross-validation using the session number to see if it generalized over learning sessions.

Table 2 shows that the kappa values of the user-specific affect models were higher compared to the general affect model. Among all machine learning algorithms, J48 performed best with a Kappa value of 0.675. The feature selector selected features similar to those in the general affect model with subtle differences in the features related to the frequency and duration of previous tasks. For example, in one student's affect model, the frequency of information searches in the past five minutes was selected as a feature while the frequency of writing notes in the past five minutes was selected instead in another student's model. These are indicative of students' affective states being influenced differently by certain tasks. This also shows that individual differences play a part in the affective states experienced by a student making them behave differently in similar contexts.

The features selected in both the general and user-specific models described the frequency, duration and type of previous actions performed by the students as well as the students' current learning state. These are contextual information about the students' learning state which seems to be good predictors of students' affective states as shown by the performance of the resulting affect models.

6. CONCLUSION

In this paper, we have presented the development of affect models that are capable of identifying students' affective states. The features used described the context in which the student learned such as the previous and current tasks they performed and are currently doing. The novelty of our work is that the affect models we built could identify affective states in a learning environment wherein students were not bound by a particular domain or learning system and the students had complete control over their activities. Even though information regarding the students' progress was unavailable, the performance of the models was still acceptable. Our results could not directly be compared to previous work because the affective states predicted by our models were in the context of a particular task unlike the works of Baker et al. [1] and D'Mello et al. [7] that predicted affective states in particular time intervals. However, the approach seems promising because the performance of the model was almost as good as the results in these works.

We acknowledge that the general affect model was created using only a few participants. However, the important observation we got was that user-specific models had better results indicating the importance of individual differences in building affect models. Evaluating the performance of affect models using data from more students would help confirm such findings.

There is a need to find features that could increase the performance of these affect models and experiment on different thresholds (e.g., task frequencies and durations in either less than or more than five minutes prior to the current task)

Affect models built with our methodology can be used by other systems for monitoring affective states. Students can then be made aware of their affect through prompts so they can adapt their activities accordingly. These systems could also suggest changes to particular tasks when certain affect is detected. Enabling systems to help students self-monitor can help them self-regulate and thus learn better.

Acknowledgements

This work was supported in part by the Management Expenses Grants for National Universities Corporations from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) and JSPS KAKENHI Grant Number 23300059. We would also like to thank all the students who participated in our data collection.

7. REFERENCES

[1] R. S. Baker, S. M. Gowda, M. Wixon, J. Kalka, A. Z. Wagner, A. Salvi, V. Aleven, G. W. Kusbit, J. Ocumpaugh, and L. Rossi. Towards Sensor-Free affect detection in cognitive tutor algebra. In *Proceedings of the International Conference on Educational Data Mining*, pages 126–133, 2012.

[2] R. S. Baker, M. M. Rodrigo, and U. E. Xolocotzin. The dynamics of affective transitions in simulation Problem-Solving environments. In *Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction*, ACII '07, pages 666–677, 2007.

[3] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

[4] S. D. Craig, A. C. Graesser, J. Sullins, and B. Gholson. Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29(3):241–250, 2004.

[5] M. Csikszentmihalyi. *Flow : the psychology of optimal experience.* Harper & Row, 1990.

[6] S. D'Mello and A. Graesser. Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2):145–157, 2012.

[7] S. K. D'Mello, S. D. Craig, A. Witherspoon, B. Mcdaniel, and A. Graesser. Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1):45–80, 2008.

[8] P. Ekman. Are there basic emotions? *Psychological Review*, 99(3):550–553, 1992.

[9] P. S. Inventado, R. Legaspi, R. Cabredo, and M. Numao. Student learning behavior in an unsupervised learning environment. In *Proceedings of the 20th International Conference on Computers in Education*, pages 730–737, 2012.

[10] B. Kort, R. Reilly, and R. W. Picard. External representation of learning process anddomain knowledge: affective state as a determinate of its structure and function. In *Artificial Intelligence in Education*, pages 64–69, 2001.

[11] I. Mierswa, M. Wurst, R. Klinkenberg, M. Scholz, and T. Euler. YALE: Rapid prototyping for complex data mining tasks. In L. Ungar, M. Craven, D. Gunopulos, and T. Eliassi-Rad, editors, *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 935–940, 2006.

[12] J. Sabourin, J. Rowe, B. Mott, and J. Lester. When Off-Task is On-Task: The affective role of Off-Task behavior in Narrative-Centered learning environments. In G. Biswas, S. Bull, J. Kay, and A. Mitrovic, editors, *Artificial Intelligence in Education*, volume 6738 of *Lecture Notes in Computer Science*, pages 534–536. 2011.

[13] B. J. Zimmerman. Self-regulated learning and academic achievement: An overview. *Educational psychologist*, 25(1):3–17, 1990.