Sensor-Free Predictive Models of Affect in an Online Learning Environment

Avery Harrison
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA
aeharrison@wpi.edu

Naomi Wixon
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA
nbwixon@wpi.edu

Anthony Botelho
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA
abotelho@wpi.edu

Ivon Arroyo
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA
iarroyo@wpi.edu

ABSTRACT
A significant amount of research has illustrated the impact of student emotional and affective state on learning outcomes. Just as human teachers and tutors often adapt instruction to accommodate changes in student affect, the ability for computer-based systems to similarly become affect-aware, detecting and personalizing instruction in response to student affective state, could significantly improve student learning. Personalized and affective interventions in tutoring systems can be realized through affect-aware learning technologies to deter students from practicing poor learning behaviors in response to negative affective states and to optimize the amount of learning that occurs over time. In this paper, we build off previous work in affect detection within intelligent tutoring systems (ITS) by applying two methodologies to develop sensor-free models of student affect with only data recorded from middle-school students interacting with an ITS. We develop models of four affective states to evaluate and determine significant predictors of affect. Namely, we develop a model which discerns students' reported interest significantly better than majority class.

Keywords
Intelligent tutoring systems; Affect detection.

1. INTRODUCTION
The ability to identify and deliver personalized interventions that are effective for individual students can greatly benefit the learning process by recognizing and addressing specific student needs. However, it's often unfeasible to realize personalized instruction and support in traditional classrooms with large numbers of students per teacher. With growing access to technology in classrooms, online learning platforms such as MathSpring have provided personalized learning opportunities that have shown positive achievement outcomes for student users [3]. Recently, such work has shifted focus to acknowledge and leverage the impact that emotion has on learning. Affect-aware learning technologies, including online learning platforms, can be developed and deployed to monitor and predict affect to provide appropriate interventions to maximize student learning.

Modeled after the control-value theory of emotion in education [16] and previous work in affect detection, we aim to develop sensor-free predictive models of affect from user behavior and performance within MathSpring. We will use students' self-reported levels of confidence, interest, excitement, and frustration during four user sessions to create predictive models and detect how student behaviors in the system relate to levels of affective states. Doing so will further efforts to build, study, and deploy affective interventions within MathSpring to optimize learning with feasible means for educational settings.

2. LITERATURE REVIEW
2.1 Affect and Control-Value Theory
A growing body of research has investigated emotion and affect in the context of education [12, 13]. To differentiate emotions and affect, consider emotions to be intuitive feelings, such as joy and anger, while affect broadly captures the manifestations of those feelings, such as pleasure and frustration, particularly in educational settings [17]. From perspectives in psychology, education, and computer science, a large amount of evidence suggests that student affect influences learning and deeper comprehension, both positively and negatively [5, 10, 12]. This research highlights the importance of affect in learning to provide content that effectively challenges students.

Our framework is based on the control-value (CV) theory of emotion [16]. Pekrun's CV theory of achievement emotions posits that student beliefs of their control over success in a subject and their value in understanding said subject will most influence their affect and, consequently, overall learning. For example, a student might feel enjoyment during an activity for which that student feels greater confidence in learning the content. The CV theory attributes student affect to feelings of control and subject value within a learning environment, underscoring the necessity of providing students with appropriately challenging tasks and adaptive content to maintain emotions that will positively influence learning in a given activity.

2.2 Sensor-Free Affect Detection
Efforts have been made to develop sensor-free affect detectors with tutoring systems for educational settings, particularly by

pairing student log files with human observations to detect behavior that might be representative of affect. For instance, Baker and colleagues developed BROMP [15] for observers to code student affect over short intervals and then match the observed affect with student activity logs [8]. Researchers have also observed facial expressions and body movement to create a framework for mapping affect onto student behavior [10, 17]. However, it is difficult to implement student affect detectors with physiological sensors or observational data in a permanent school setting [7] due to cost and the potential threat to validity introduced by such methods caused by alterations of the learning environment. Researchers have previously tried to predict self-reported affect with log data and questionnaires [9] but less work has been done with solely log data.

3. CURRENT STUDY

MathSpring is an intelligent tutoring system that covers Common Core mathematics curriculum for students in 6th-10th grade to prepare for standardized tests [1]. The system adapts to provide content that will likely keep the student in the zone of proximal development [3] while providing scaffolding and fostering growth mindsets through personalized, pedagogical support. Affective support is realized through text, audio, and images from an animated learning companion as students solve problems [3]. Studies have found that using MathSpring leads to significant performance gains on standardized math tests as opposed to students who do not practice with MathSpring [1].

We currently aim to utilize student data from MathSpring build affect predictors from only student logs. We intend to respectively construct predictive models for confidence, interest, excitement, and frustration levels reported over brief intervals in user sessions. Based on previous affect detection work, and gaps in the literature, we hypothesize that students changing topics and viewing progress in MathSpring will contribute to affect predictions [1]. We predict that topic changes may indicate frustration and be negatively related to positive affective states.

This study was conducted with 85 eighth-grade students at a middle school in Massachusetts. Between December 2016 and May 2017, students participated in four, hour-long sessions with MathSpring. In each session, students worked on assigned problem sets corresponding with class material. Students typically completed the assigned problem set in that time or completed the set in the next session. Throughout each session, users saw a learning companion that delivered messages to remind students of the hint button or provide encouragement. Previous work has looked at the effects of interventions with different affective messages, such as empathetic, growth mindset, and success/failure messages [1]. Growth mindset messages were used in this study because they are the default for MathSpring. If a user selected an incorrect answer, the hint button would flash. After a second wrong attempt, the learning companion delivered a growth mindset message. Students could skip problems or return to the "My Progress" page any time where they could view topic mastery and choose to continue, change topics, challenge themselves, or review content.

Drawing from previous work on affect detection in learning technologies [1, 2, 4, 6], we inquired about levels of excitement, interest, confidence, and frustration during user sessions. Roughly every five minutes between problems, students received a prompt to self-report affect on a 5-point Likert scale ranging from 1=*Not at all* to 5=*Extremely,* with the option to skip the self-report. Prompts randomly alternated between the affective states but contained the same wording. For example, the prompt for Confidence would read, "*Please tell us how you are feeling. Based on the last few problems tell us about your level of Confidence in solving math problems*."

4. MODELS AND ANALYSES

We first reconstructed data from student log files. Affect self-reports were randomized throughout the user sessions so the order and summation of self-reports for each affect varied by user. For example, a student could have reported on confidence followed by interest level while another student could have been prompted for frustration and then excitement level. Due to this variation between and within students, we chose to use the "mini-sessions" of activity between each affect report. This is supported by previous findings that recently completed problems are more predictive of affect than an entire user session [6] and alleviates the possible effect of elapsed time on affect reports.

Table 1. Descriptive statistics on affect self-reports.

	Excitement	Frustration	Interest	Confidence
N Affect Reports	138	129	133	154
Mean (*SD*)	1.78 (1.17)	2.36 (1.67)	1.98 (1.29)	3.23 (1.53)

With mini-sessions of self-reported affect (N=554; Table 1), we aggregated behavior variables, such as the number of problems seen between reports, that corresponded with a given affect report then separated mini-sessions by affect. Observations without a reported emotion level (N=196) were culled. A PCA with "mini-session" level variables revealed four factors. We selected one variable per factor for the models. *Topic changes* refer to a student changing problem sets due to completion or topic mastery, prolonged poor performance, or self-electing to return to the progress page and choose a different topic. The *average number of hints* refers to the average seen per problem. The *percentage of problems answered correctly* is calculated within the "mini-session". Lastly, the *number of interventions* sums hint button flashes and messages from the learning companion during problem solving.

Based on past MathSpring work [1], we tried two methods of building predictive models of affect by constructing logistic regressions with five-fold cross-validations at the student level. The "at least somewhat" models attempt to predict whether students would report "Not at all" to "A little" (1-2 on the self-report scale) or "Somewhat" to "Extremely" (3-5) of a given affect. Then, the "at least a little" models attempt to predict whether students reported any degree of a given affect (2-5) or not at all (1).

Table 2 summarizes the performance of models. Notably, both models of interest perform comparably to other predictive models of affect (kappa > 0.20) [14]. While there is variation across affect and discretization, with both Confidence models and the "at least a little" model for Frustration performing below chance (AUC < 0.50; kappa < 0), five of the models appeared to be performing above chance with disagreeing AUC and kappa values. Unlike AUC, accuracy, F_1, and kappa values are sensitive to the choice of rounding threshold of model estimates, particularly with unbalanced labels. This incongruence between AUC and kappa has been seen in other work on sensor-free affect detection using deep learning [8]. Given the imbalance of labels within each affective state, we calculated an optimized

Table 2. Logistic regression model performance.

Model	AUC	Kappa	F_1	Optimized Accuracy	Optimized F_1	Optimized Kappa
At Least Somewhat						
Interest	**0.75**	**0.24**	**42.41**	**72.29**	**58.54**	**0.38**
Confidence	**0.70**	**0.02**	**73.76**	**68.41**	**70.33**	**0.32**
Excitement	**0.68**	**0.10**	**25.00**	**63.17†**	**37.50**	**0.10**
Frustration	0.53	-0.04	18.46	59.36	18.46	-0.04
At Least A Little						
Interest	**0.73**	**0.32**	**60.34**	**67.14***	**61.57**	**0.35**
Confidence	0.69	-0.04	84.18	64.76†	70.94	0.22
Excitement	**0.62**	**0.06**	**42.11**	**65.42**	**58.18**	**0.20**
Frustration	0.39	-0.17	32.03	36.33†	32.03	-0.17

Note: Bolded rows indicate model performance above chance ($0.50 < AUC \leq 1$; $0 < kappa \leq 1$). Optimized accuracies significantly better ($p<.05$) than a base rate model are denoted with (*), while optimized accuracies significantly worse than base rate are denoted with (†).

metric by learning a reasonable rounding threshold of model estimates using the training set of each fold. We also compared each model's optimized accuracy to the respective base rate, majority class model to determine significance. It is found that only the model for "at least a little" Interest has a significantly higher accuracy than the base rate. Table 3 details standardized coefficients for each model. Number of interventions was the most frequent predictor across affects and discretization levels. Percentage of correct problems was also a strong predictor of interest ($p < 0.01$). Topic changes positively predicted interest and excitement and negatively predicted frustration.

Table 3. Standardized coefficients (β) of predictors by model.

Model	Topic Changes	Avg. Hints	Correct Problems (%)	Number of Interventions
At Least Somewhat				
Interest	**0.69**	-0.10	**0.95**	**-0.80**
Confidence	0.39	0.40	0.04	**-0.48**
Excitement	0.48	0.12	-0.23	**-0.80**
Frustration	**-0.83**	0.17	-0.37	**0.75**
At Least A Little				
Interest	0.41	-0.13	**0.67**	**-0.53**
Confidence	0.37	0.16	0.01	**-0.56**
Excitement	**0.55**	-0.03	0.10	**-1.04**
Frustration	-0.26	0.10	0.05	0.26

5. DISCUSSION

We presented predictive models of affect within MathSpring with a model of "at least a little" interest that performs significantly well. In general, the "at least somewhat" models perform better, suggesting that this discretization split should be used in future projects to predict student affect. While some of the models do not perform well, this is not surprising given that sensor-free affective models are more difficult to build than models profiting from detectors or pre- and post-study data.

However, it is surprising that the number of topic changes, contrary to our hypothesis, was positively related to interest and excitement levels and negatively related to frustration. This implies that higher frequencies of topic changes between affect reports indicate positive affective states. Conversely, a student who does not change topics between affect reports is more likely to report a higher level of frustration. We assumed that students would change topics if they performed poorly (indicating that the content is too challenging to be productive) or were bored. However, students could also change topics if they mastered or completed a topic (indicating the content is too easy). Considering the positive relationship between interest and topic change, and excitement and topic change, perhaps students were more likely to change topics because of completion or mastery. This suggests that students might conflate the concepts of interest and excitement with feelings of achievement.

The other predictor to note, number of interventions, was the most common, statistically significant predictor of affect level across models. Number of interventions was negatively related to positive affect which suggests that fewer interventions led to higher reports of positive affective states. Conversely, the number of interventions positively predicted frustration, suggesting that more interventions predicted a higher level of frustration. Assuming that interventions increased as student attempts increased, it is unsurprising that higher numbers of interventions precede higher reports of frustration and lower reports of positive affective states. The number of topic changes and interventions between affect reports were the main predictors of affect across models, while percent of problems answered correctly only positively related to interest. The lack of strength in the four predictive attributes suggests that we should consider other variables from the four PCA components.

There are other caveats to consider. Namely, self-report from middle school students might not be accurate and prompting students to self-report throughout user sessions might disrupt natural affect. Also, the type of intervention might influence affect rather than the quantity of interventions. For instance, students who saw affirmative messages after answering a problem correctly might have felt differently towards the learning companion and MathSpring than a student who saw growth mindset messages after attempting a problem multiple times. That said, using interventions as a variable in the sensor-free predictive models is only beneficial to data from MathSpring until we better comprehend the underpinnings of how interventions influence student affect more broadly.

This work poses future directions to give better consideration to these questions. It might be worthwhile to construct ordinal regression to predict the level of affect reported rather than a binary classification. We also intend to create a feature that indicates the previous self-report level of the affect in question. This feature was not included for the initial round of analyses due to the randomization of affect report ordering. A student might only report on a given affect once or twice towards the beginning of the session, rendering the information less useful than if the same affect were reported on twice in a row across a shorter span. Even with potential irregularity of affect reporting, previously-reported same-affect level could be suggestive of the dynamics of affect throughout user sessions. Pursuing these directions will help us better understand the dynamic between student affect and behavior in tutoring systems.

6. CONCLUSION

We presented a high-performing predictive model of interest, as well as predictive models of excitement and confidence that perform above chance, demonstrating the ability to build sensor-free detectors of affect in MathSpring. Given the limitations of the current models and future plans with the data from this study, we consider this to be a first effort. We intend to utilize the data to improve sensor-free affect detection so that socio-emotional interventions in MathSpring can be better realized to optimize student support and learning. Progress in sensor-free affect detection research has positive implications for classroom implementation of affect-aware learning technologies and sustainable data collection through student activity files.

7. REFERENCES

[1] Arroyo, I., Wixon, N., Muldner, K., Karumbaiah, S., Lizarralde, R., Allessio, D., Burleson, W., and Woolf, B. P. 2017. *Addressing student emotion in personalized digital learning environments*. Manuscript submitted for publication.

[2] Arroyo, I., Shanabrook, D. H., Burleson, W., & Woolf, B. P. 2012. Analyzing affective constructs: Emotions 'n attitudes. *Proceedings of the 11th International Conference on Intelligent Tutoring Systems, Greece*, 714-715.

[3] Arroyo, I., Woolf, B. P., Burleson, W., Muldner, K., Rai, D., and Tai, M. 2014. A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition, and affect. *International Journal on Artificial Intelligence in Education, Special Issue on "Landmark AIED Systems for STEM Learning",* 24, 387-426.

[4] Arroyo, I. 2014. Analyzing affective constructs: Emotions, attitudes, and motivation. Retrieved from: http://digitalcommons.wpi.edu/ssps-papers/2

[5] Arroyo, I., Cooper, D.G., Burleson, W., and Woolf, B.P. 2010a. Bayesian networks and linear regression models of students' goals, moods, and emotions. In *Handbook of Educational Data Mining,* C. Romero, S. Ventura, M. Pechenizkiy, and R. Baker (Eds.). CRC Press, Boca Raton, 323-338.

[6] Baker, R.S.J.d., D'Mello, S.K., Rodrigo, M.M.T., and Graesser, A.C. 2010. 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. *International Journal of Human-Computer Studies*, 68, 4 (2010), 223-241.

[7] Baker, R.D., Gowda, S., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit,, G., Ocumpaugh, J., and Rossi, L. 2012. Towards sensor-free affect detection in cognitive tutor algebra. *Proceedings of the 5th International Conference on Educational Data Mining*, 126-133.

[8] Botelho, A., Baker, R. S., and Heffernan, N. T. 2017. Improving sensor-free affect detection using deep learning. In *International Conference on Artificial Intelligence in Education*, 40-51, Springer, Cham.

[9] Conati, C., and Maclaren, H. 2009. Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, 19, 267-303.

[10] D'Mello, S.K., and Graesser, A.C. 2010. Multimodal semi-automated affect detection from conversational cures, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, 20, 2, 147-187.

[11] D'Mello, S., Person, N., and Lehman, B. 2009. Antecedent-consequent relationships and cyclical patterns between affective states and problem solving outcomes. In *Artificial Intelligence in Education. Building Learning Systems that Care: from Knowledge Representation to Affective Modelling,* V. Dimitrova, R. Mizoguchi, B. du Boulay, and A. Grasser (Eds.). IOS Press, 57-64.

[12] D'Mello, S.K., Strain, A.C., Olney, A., and Graesser, A. 2013. Affect, meta-affect, and affect regulation during complex learning. In *International Handbook of Metacognition and Learning Technologies*, R. Azevedo and V. Aleven (Eds.). Springer International Handbooks of Education 26.

[13] Ekman, P. 1999. Basic emotions. In *Handbook of Cognition and Emotion,* T. Dalgleish and M. Power (Eds.). John Wiley & Sons Ltd., 45-60.

[14] Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., and Heffernan, C. 2014. Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology*, 45, 3, 487-501.

[15] Ocumpaugh, J., Baker, R.S., and Rodrigo, M.M.T. 2015. Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual. *New York, NY and Manila, Philippines: Teachers College, Columbia University and Ateneo Laboratory for the Learning Sciences.*

[16] Pekrun, R. 2006. The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18, 315-341.

[17] Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., and Picard, R. 2009. Affect-aware tutors: Recognising and responding to student affect. *International Journal of Learning Technology*, 4, 3/4 (2009), 129-164.