Emotion Sensors Go To School

Ivon ARROYO,[a,1] David G. COOPER,[a] Winslow BURLESON[b]
Beverly Park WOOLF, [a] Kasia MULDNER, [b] Robert CHRISTOPHERSON [b]
[a] *Department of Computer Science, University of Massachusetts Amherst*
[b] *School of Computer Science and Informatics / Arts, Media and Engineering Arizona State University*

Abstract. This paper describes the use of sensors in intelligent tutors to detect students' affective states and to embed emotional support. Using four sensors in two classroom experiments the tutor dynamically collected data streams of physiological activity and students' self-reports of emotions. Evidence indicates that state-based fluctuating student emotions are related to larger, longer-term affective variables such as self-concept in mathematics. Students produced self-reports of emotions and models were created to automatically infer these emotions from physiological data from the sensors. Summaries of student physiological activity, in particular data streams from facial detection software, helped to predict more than 60% of the variance of students emotional states, which is much better than predicting emotions from other contextual variables from the tutor, when these sensors are absent. This research also provides evidence that by modifying the "context" of the tutoring system we may well be able to optimize students' emotion reports and in turn improve math attitudes.

Keywords. Intelligent tutor, student emotion, sensors, physiological activity, infer cognition, infer student affect

Introduction

If computers are to interact naturally with humans, they must recognize affect and express social competencies. However, the role of affect in instruction is at best in its infancy. The possibility of intelligent tutoring systems that trace students' emotions is a attractive concept. Tutors provide individualized teaching in multiple domains [1], however, much previous research has tended to privilege the cognitive over the affective in which learning theories view thinking as information processing, marginalizing or ignoring affect [2].

Our research goals are two-fold: 1) to systematically examine the relationship(s) between student affective state and desired outcomes, i.e., to identify whether a dependency exists between students' reported emotions and their learning of, motivation for, and attitudes toward mathematics, and 2) to evaluate the importance and feasibility of tracing students' emotional states within real-world classroom settings. Thus, this paper explores how the students' experience with a tutoring system shapes their feelings while learning, and motivation and attitudes towards the subject. We describe the first steps in creating data-driven models for emotion, via the use of physiological sensors (camera, mouse, chair, and wrist) in concert with online context and learners' feedback.

[1] Corresponding Author. Ivon Arroyo, ivon@cs.umass.edu

1. Background and Related Work

Characterization of Emotion. No comprehensive, validated, theory of emotion exists that addresses learning, explains which emotions are most important in learning, or identifies how emotion influences learning [2]. Most studies of emotion do not include the phenomena observed in natural learning situations, such as interest, boredom, or surprise. We identified a subset of emotions based on Ekman's analyses of facial expressions that includes joy, anger, surprise, fear, disgust/contempt, and interest [3], with the intention of recognizing these emotions in student behavior and then provide interventions. We added a cognitive component to ground this categorization of emotion in an educational setting, resulting in four orthogonal bipolar axes of cognitive-affect [4]. For example, given Ekman's *fear* category, the proposed scale is: "I feel anxious . . . I feel very confident."

Prior Work in Affective Support. Research shows that learning is enhanced when empathy or support is present [5, 6]. Various studies have linked interpersonal relationships between teachers and students to increased student motivation over the long term [7, 22]. Thus great interest exists to embed affective support into tutoring applications. Since affect recognition is a key aspect of tailored affective support, research has focused on automated detection of affective states in a variety of learning contexts [e.g., 8, 9, 10]. This work has shown promising results and some affective states (e.g., frustration or boredom) can be detected within one intelligent tutoring system [e.g., 10, 11, 12].

To date, much of the existing work has focused on inferring students' affective states with software (e.g., machine learning) [e.g., 13]. Hardware sensors have the potential to provide information on students' physiological responses that have been linked to various affective states [e.g., 3]. Most of these past research efforts, however, have been conducted in laboratory experiment settings. In addition, we believe these models have to be recreated for each individual learning environment, population of students and tutoring software, as the ways people interact with them are very likely to vary across populations, domains and subject matters being taught. Our research explores various sensors' potential for affect recognition in real public schools educational settings. Addressing students emotions while attitudes toward STEM fields are being developed is simply invaluable. We build on work started by Burleson [14], who developed a learning companion that depended on a sensor framework (incorporating a mouse, posture chair, video camera, skin conductance bracelet) to recognize and respond to student affect. Dialog and posture features were used to discriminate among affective states of boredom, confusion, flow and frustration [9].

Our research questions include: What is the value of tracing students' emotions in real educational settings? Are students' reported emotions associated to learning and/or attitudes-motivation for learning math? Do real students (within real classroom settings) express these self-reported emotions physically? We now describe our first experiments related to answering these questions.

2. Methodology and experiment design.

We conducted two studies during Fall 2008 involving the use of sensors with Wayang Outpost, a multimedia adaptive tutoring system for geometry [15, 16]. The tutor without sensors has been used with thousands of students in the past and has demonstrated improved learning gains in state standard exams. We implemented gendered learning companions that emphasize the importance of effort and perseverance and encourage the students to use the Help function. Students used the software as part of their regular math class for 4-5 days and took mathematics tests and a survey questioning their perceptions of math, before and after using the tutor software.

The first study involved 38 high school students (HS), 15-17 year olds from a public school, enrolled in three different math classes and the second study involved 29 female undergraduate students (UG) taking a mathematics class for elementary school teachers at UMass. The software was used as part of the regular mathematics class for 4-5 days and covered topics in the class, though HS students had not, and UG students had, seen most of the topics in the class, while UG students had. Nearly 100% of students in both classrooms were outfitted with all four sensors, though we received data from only about 50% of the students as discussed in Section 4. Students took a mathematics pretest and a survey to assess motivation [17] and assess self-concept in math and math value [18]. Post-test surveys also included questions that measured student perceptions of the software and the help provided by the tutor. All items had a six-point scale, except for the two learning vs. performance orientation items [17]. The system iterated through different topics, and problems were chosen adaptively depending on students' ongoing math performance. In addition, every 5 minutes, and after students finished a problem, a screen queried about emotions: "How [interested/excited/confident/frustrated] do you feel right now?" Students choose one of 5 points on a continuum, where the ends were labeled (e.g. I feel anxious… Very confident) and where 3 corresponded to a neutral value. The emotion queried was randomly chosen (obtaining a report per student per emotion for most subjects) and no students complained about the frequent tutor requests for self-reports.

Overall results. General outcomes of these studies can be summarized as follows: High school students (HS) had less math incoming ability than the undergraduate students group (UG). Stepwise linear regression was used to identify good predictors of each emotion with student self-reports as the independent variable, see Table 1. Students in the HS study were more "pessimistic" than the UG study, both in pretest surveys and self-reports of emotions, while UG students were not generally frustrated, HS students reported more frustration, and less interest, confidence and excitement. The combination of both populations provided an interesting mix of students with different feelings and math abilities. Both populations learned based on post-pre test tests; they improved an average 10% in math performance (25% proportional learning gain).

3. Students report their emotions.

Our studies show that students' self-report of emotion depends on events that occurred in the previous problem and not on their incoming beliefs. Thus if a student self-reports " I feel frustrated" it is likely that he had several incomplete attempts in the previous problem. If he reported "high confidence," he likely just solved a problem. We analyzed the relationship between the sample mean interest, excitement, confidence and frustration reported by each student and their corresponding incoming math knowledge, self-concept, math value [18] and learning orientation [17] by analyzing statistical dependencies between those variables for the full set of data. Reported emotions showed only a marginal statistical significant correlation with most of these variables regarding incoming attitudes, feelings and knowledge of math. One significant correlation was found between Pretest%Correct and Interest self-reports (N=55, R=0.29, p=0.03), suggesting that students who had higher math knowledge to begin with also reported being more interested within the tutor than students with low incoming math ability.

	Tutor context only	Camera + Tutor	Seat + Tutor	Wrist + Tutor	Mouse + Tutor	All Sensors + Tutor
Confident	R=0.49, N=62	**R=0.72, N=20 -- a**	R=0.35, N=32		R=0.55, N=28	R=0.82, N=17
Frustrated	R=0.53, N=69	R=0.63, N=25	**R=0.68, N=25 -- d**	R=0.56, N=45	R=0.54, N=44	R=0.72, N=37
Excited	R=0.43, N=66	**R=0.83, N=21 -- b**	R=0.65, N=39	R=0.42, N=37	R=0.57, N=37	R=0.70, N=15
Interested	R=0.37, N=94	**R=0.54, N=36 -- c**	R=0.28, N=51		R=0.33, N=51	

a --(Solved? ConcentratingMax); b -- (concentratingMax; unsureMin; HintsSeen; LearningCompanion?); c -- (InterestedMin LearningCompanion?); d -- (IncAttempts; SecsToFirstAtt; TimeInSession; sitForwardStdev)

Table 1. R values for Linear Regression Models *(best fit models for each emotion in bold)*. Empty cells mean that no fit model was found for that data set. N values vary because each case corresponds to one emotion report x the data for each sensor –mean, minimum value and maximum value corresponding to each sensor for the last problem before the report. Full data for all sensors is limited to a subset of students.

Instead, fluctuating student emotions are apparently related to what has just happened in the tutor, i.e. to the "context," see Figure 3, e.g., did students solve the problem *immediately before* correctly, or did they need much help. We analyzed each emotion in relation to the following contextual variables regarding that last problem: number of incorrect attempts (IncAttempts), whether the problem was solved correctly in the first attempt (Solved?), time elapsed in the session (TimeInSession), time so far in the tutor (TimeInTutor), number of hints seen in the last problem (NumHints), seconds until the first attempt to answer (SecsFirstAttempt), seconds until the problem was solved correctly (SecsToSolve), and the gender of the learning companion that gave some feedback. We then used stepwise regression to identify good predictors of each emotion, see Table 1. The results suggest that emotions *can* be predicted from these contextual variables, as they depend significantly on what has just happened in the previous problem. Interest (14% of the variance) can be predicted from IncAttempts and the gender of the learning companion. Excitement (18% of the variance) can be predicted from Solved?. Frustration (28% of the variance) can be predicted from NumHints, IncAttempts and TimeInSession. Confidence (24% of the variance) can be predicted from NumHints and whether the previous problem was solved correctly. All these statistically significant dependencies indicate that students' emotion self-reports depend on what has just happened in the tutoring session, while they marginally depend on their incoming beliefs.

Table 1 describes variables that were entered into the model with the stepwise regression method. For instance, in the first cell (top, left) there are 62 reports of students' confidence. The regression model generated has a fit of R=0.49, accounting for 24% of the variance. The variables found to predict confidence are included in Table 1. For instance, when considering the confidence reports have camera data available (N=20), the stepwise regression method found that Solved? and ConcentratingMax (the maximum probability of the student being concentrated during the last problem, according to the MindReader software [20]) can predict 52% of the variance of confidence (R=0.72), which is more than double of the variance predicted without the camera.

4. Students express their emotions physically

As mentioned above, a set of non-invasive hardware sensors recorded students' physiological behavior as described in [19]. The hardware (with the exception of the camera developed at MIT) was manufactured at Arizona State University from validated instruments first developed by the Affective Computing group at MIT.

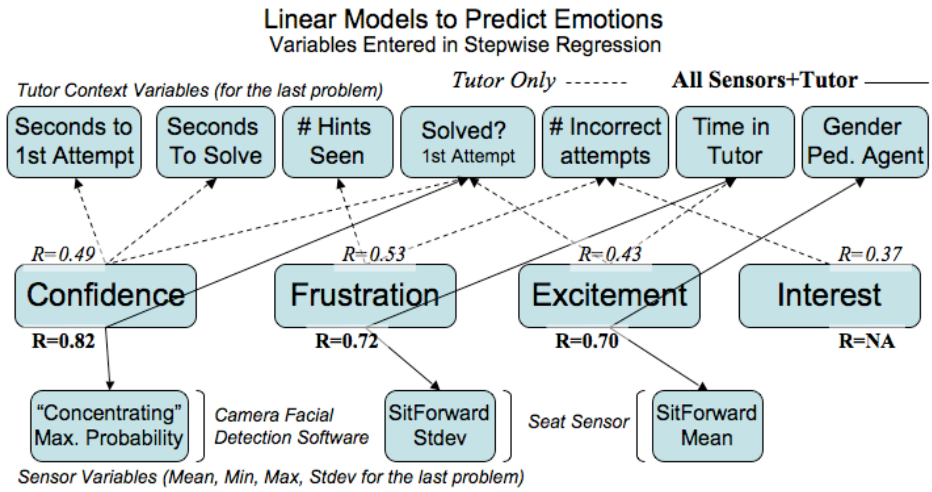

Figure 1. Models resulting from linear regression studies of students' self-report and related variables. These results suggest that emotion can be predicted from what just happened in the previous problem (dashed arrows, R values in italics). However, student physical activity as determined by sensors improve predictive power (solid arrows, R values in bold).

Twenty-five sets of each sensor were manufactured for simultaneous use in classrooms in Massachusetts and Arizona. The four sensors, shown in Figure 2, include: a facial expression recognition sensor that incorporates a computational framework to infer a user's state of mind [20] called the MindReader software (provides a probability of a students being concentrated, interested and a few others, at any given time); a wireless conductance bracelet based on an earlier glove that sensed skin conductance, developed at the MIT Media Lab; a pressure mouse to detect the increasing amounts of pressure that students place on their mice related to increased levels of frustration [21]; and low-cost/low resolution pressure sensitive seat cushions and back pads with an incorporated accelerometer to measure elements of a student's posture and activity.

Figure 2. Sensors used in the classroom (clockwise): Facial Expression Sensor; Conductance Bracelet, Pressure Mouse and Posture Analysis Seat.

One research issue was to determine the extent of the benefits of using sensor data to detect students' emotions, instead of making inferences from contextual variables (e.g., student time on problems, number of hints or attempts) as discussed in Section 3. This issue was addressed by analyzing the improved emotion predictions when sensor data was available compared to when inferences were limited to information about student behavior in the tutor context. One caveat is that regression works well with a full set of data and not all sensors were available at all times for all students, because of several real-life practical problems. As a result, we have full data for each emotion for approximately half of the students. Figure 1 shows the generated models with reduced (but complete) data sets that include all sensors. However, in order to be more precise about the potential contribution of each sensor, we created another set of models showing the contribution of each individual sensor separately, shown in Table 1. The column towards the left after the *Tutor Context Only* column shows the added contribution of the camera, e.g., we can create a linear model of R=0.72, accounting for 52% of the variance (more than double that with tutor context only, without sensors). The variables that were found to predict some of the emotions are shown below the table, e.g., the variables used to predict confidence after the camera data was added are shown in table footnote "***a***" and were Solved? and concentratingMax (the maximum probability that the student was "concentrating," a value given by the facial expression software, for the last problem before the student confidence report). When we consider only those emotion reports for students who also have seat posture data, the seat features (SitForwardMax, Min, Mean, and Stdev) generate a worse model.

Figure 3 shows data analyzed from the camera alone. The graphs on the left show facial expression software predictions that students' who reported low confidence were concentrating minutes before those self report were given, whereas students who reported high confidence did not seem to be concentrating. That is, students who were working hard to figure out a problem felt unsure/anxious and not confident about their ability to solve it. The graphs on the right show that students who self-reported low frustration were predicted by the facial expression software to be thinking seconds before completing the problem and reporting their frustration level. The small letters (O, X, ?, F) indicate actions taken by students in the tutor (correct, incorrect, Hint or sit forward).

5. Discussion and Future Work

This article makes several important contributions to the field of sensor recognition of student affect in intelligent tutors. We showed that students' self-reports of emotion can be automatically inferred from physiological data that is streamed to the tutoring software for students in real educational settings. Summaries of this physiological activity, in particular data streams from facial detection software, can help tutors predict more than 60% of the variance of students emotional states, which is much better than when these sensors are not used.

Using sensors in two classrooms we analyzed how students feel and behave while solving mathematics problems in a public school setting. We identified state-based fluctuating student emotions through student's self-reports that were highly dependent on the tutoring scenario, particularly on indicators of effort in the last problem seen. These fluctuating student reports were related to longer-term affective variables (e.g., value mathematics and self-concept) and these latter variables, in turn, are known to predict long-term success in mathematics, e.g., students who value mathematics and have a positive self-concept of their mathematics ability perform better in mathematics classes [22]. In fact, the subjective value of mathematics and student's self-concept of mathematics ability drops as students' transition from elementary school to junior high school in the USA. In addition, there is a sharp rise in the perceived difficulty of math beginning around the 7th grade and persisting through to 12th grade. An opportunity exists for tutoring systems to optimize not only learning, but also long-term attitudes related to students' emotions while using the software. By modifying the "context" of the tutoring system including students' perceived emotion around mathematics, a tutor might optimize and improve their mathematics attitudes.

Future work consists of refining emotion models to predict desirable and undesirable learning states and attitudes. In the short term, we intend to deploy more sensors in classrooms, gather more data, and refine our emotion models. We will continue to enhance learning agents that work in concert with emotion detectors to help students cope with their negative emotion, supporting them to return to on-task behavior [4].

In the long term we intend to develop detectors that evaluate which are the available input streams (e.g. cameras that are already incorporated into the computer) and use the best possible model available given the sensor input that the tutoring system has available. Thus, the tutoring system would still make a reasonable prediction of emotion, even if a subset of sensors are available, or none at all. This would allow to scale up the use of the system to other schools in the country that may not have all the sensors available to them.

Acknowledgement

This research was funded by two awards (1) National Science Foundation, #0705554, IIS/HCC *Affective Learning Companions: Modeling and supporting emotion during teaching*, Woolf and Burleson (PIs) with Arroyo, Barto, and Fisher; and (2) U.S. Department of Education to Woolf, B. P. (PI) with Arroyo, Maloy and the Center for Applied Special Technology (CAST), *Teaching Every Student: Using Intelligent Tutoring and Universal Design To Customize The Mathematics Curriculum*. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

We acknowledge contributions to the system development from Rana el Kaliouby, Ashish Kapoor, Selene Mota and Carson Reynolds. We also thank Joshua Richman, Roopesh Konda, and Assegid Kidane at ASU for their work on sensor manufacturing.

References

[1] Koedinger, K. R., Anderson, J. R., Hadley, W. H., Mark, M.A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education 8*(1), 30-43.
[2] Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D., Strohecker, C. (2004). Affective Learning--A Manifesto. *BT Technical Journal, 2*(4), 253-269.
[3] Ekman, P. (1999). *Facial Expressions*. New York: John Wiley & Sons Ltd.

[4] Arroyo, I., Woolf, B. P., Royer, J.M. and Tai, M., (2009) Affective Gendered Learning Companions, *14th International conference on Artificial Intelligence and Education (AIED 2009)* V. Dimitrova and R. Mizoguchi (Eds), IOS Press.
[5] Graham, S., & Weiner, B. (1996). Theories and principles of motivation. In Berliner, D. & Calfee, R (Eds), *Handbook of Educational Psychology*. 63-84. New York: Macmillan.
[6] Zimmerman, B. J. (2000). Self-Efficacy: An Essential Motive to Learn. *Contemporary Educational Psychology, 25*, 82-91.
[7] Wentzel K and Asher S R, 1995, Academic lives of neglected, rejected, popular, and controversial children', Child Development, 66, pp 754—763 (1995).
[8] Conati C. & Mclaren H. (2004). Evaluating A Probabilistic Model of Student Affect. *Proceedings of ITS 2004, 7th International Conference on Intelligent Tutoring Systems*, Lecture Notes in Computer Science, Volume 3220/2004, Springer Berlin / Heidelberg p. 55-66.
[9] D'Mello, S. & Graesser, A. (2007). *Mind and Body: Dialogue and Posture for Affect Detection in Learning Environments*. Paper presented at the Frontiers in Artificial Intelligence and Applications.
[10] McQuiggan, S. & Lester, J. (2006). *Diagnosing Self-Efficacy in Intelligent Tutoring Systems: An Empirical Study*. Eighth International Conference on Intelligent Tutoring Systems, M. Ikeda, K. Ashley, T.W. Chan (Eds), Jhongli, Taiwan
[11] Graesser, A. C. , Chipman, P., King, B., McDaniel, B., and D'Mello, S (2007). Emotions and Learning with AutoTutor. *13th International Conference on Artificial Intelligence in Education (AIED 2007).* R. Luckin, K. Koedinger, and J. Greer (Eds), (pp 569-571). IOS Press.
[12] D'Mello, S. K., Picard, R. W., and Graesser, A. C. (2007). Towards an Affect-Sensitive AutoTutor. S*pecial issue on Intelligent Educational Systems IEEE Intelligent Systems, 22* (4), 53-61.
[13] Conati, C., Gertner, A. and VanLehn, K. (2002). Using Bayesian Networks to Manage Uncertainty in Student Modeling. *User Modeling and User-Adapted Interaction, 12*(4), 371-417.
[14] Burleson, W. (2006). *Affective Learning Companions: Strategies for Empathetic Agents with Real-Time Multimodal Affective Sensing to Foster Meta-Cognitive Approaches to Learning, Motivation, and Perseverance.* MIT PhD thesis, Available at http://affect.media.mit.edu/pdfs/06.burleson-phd.pdf.
[15] Arroyo, I., Beal, C.R., Murray, T., Walles, R., Woolf, B.P. (2004). Web-Based Intelligent Multimedia Tutoring for High Stakes Achievement Tests. In J. C. Lester, R. M. Vicari & F. Paraguaçu (Eds.), *Intelligent Tutoring Systems, 7th International Conference, ITS 2004,* (pp. 468-477). Maceiò, Alagoas, Brazil, Proceedings. Lecture Notes in Computer Science 3220: Springer 2004.
[16] Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Mehranian, H., Fisher, D., Barto, A., Mahadevan, S., Woolf, B. (2007). Repairing Disengagement With Non Invasive Interventions. International Conference on Artificial Intelligence in Education, Marina del Rey, CA.
[17] Mueller, C.M., Dweck, C.S. (1998). Praise for intelligence can undermine children's and performance. Journal of Personality and Social Psychology , 75 (1), 33-52.
[18] Wigfield, A., & Karpathian, M. (1991). Who am I and what can I do? Children's self-concepts and motivation in achievement solutions. *Educational Psychologist, 26, 233–261.*
[19] Dragon, T., Arroyo, I., Woolf, B.P., Burleson, W., El Kaliouby, R., Eydgahi, H. (2008) Viewing Student Affect and Learning through Classroom Observation and Physical Sensors. International Conference on Intelligent Tutoring Systems 2008: 29-39
[20] El Kaliouby, R. (2005). *Mind-Reading Machines: Automated Inference of Complex Mental States.* Unpublished Ph. D. thesis, University of Cambridge.
[21] Dennerlein, J., Becker, T., Johnson, P., Reynolds, C. and Picard, R. (2003) Frustrating Computer Users Increases Exposure to Physical Factors, Proceedings of the International Ergonomics Association, pp. 24--29
[22] Royer, J. M., & Walles, R. (2007). Influences of gender, motivation and socioeconomic status on mathematics performance. In D. B. Berch and M. M. M. Mazzocco, (Eds), *Why is math so hard for some children.* Baltimore, MD: Paul H. Brookes Publishing Co. (pp. 349-368).