Semi-Supervised Classification of Realtime Physiological Sensor Datastreams for Student Affect Assessment in Intelligent Tutoring

Keith Brawner[1], Robert Sottilare[1], Avelino Gonzalez[2]

[1] United States Army Research Laboratory
Human Research and Engineering Directorate
Learning in Intelligent Tutoring Environments Laboratory
Simulation and Training Technology Center, Orlando, FL 32826
{Keith.W.Brawner, Robert.Sottilare} @us.army.mil

[2] Univercity of Central Florida
Department of Electrical Engineering and Computer Science
Orlando, FL 32826
Gonzalez@ucf.edu

Abstract. Famously, individual expert tutoring holds the promise of two standard deviations of improvement over classroom-based instruction. Current content-scaling techniques have been able to prove one standard deviation of improvement. However, just as expert tutors take the motivation and emotional state of the student into account for instruction, so too must computer instructors. Differences between individuals and individual baselines make this difficult, but this information is known across one training session. The construction of assessing modules in realtime, from the available performance and sensor datastreams, skirts these problems, but is technically difficult. This research investigates automated student model construction in realtime from datastreams as a solution from which to base pedagogical strategy recommendations.

Keywords: Intelligent Tutoring, Affective Computing, Datastream Mining

adfa, p. 1, 2011.

1 Background, Research, and Direction

Artificial Intelligence is a collection of methods that are used to solve problems. The most frequent problem solved is the automation of decision making, based upon the classification of inputs. The classification problem can be separated into two categories: unsupervised and supervised. Supervised classification problems have training data with provided 'answers', known as 'labels', and testing data. Unsupervised artificial intelligence problems attempt to classify data without knowing the true class of the observation.

Physiological data presents a unique problem to the realm of classification. One of the overwhelming trends in the field of psychology is that all people are different, known as individual differences. As such, the observed behavior of individuals varies widely. This trend represents itself well among physiological sensors as well[1]. Psychology studies relating to physiological measurements frequently involve the 'baseline' of an individual in order to correct for this problem. This is, inherently, an unsupervised learning problem. For example, galvanic skin responses (GSR), which are specific to the individual, must be learned without explicit second-by-second updates on the person's emotions, due to impracticality.

While there have been many studies that use physiological data in order to establish meaning among individuals or groups[2], the problem of individual differences forces the researcher to evaluate each individual individually. While this approach is helpful to psychology researchers, a different approach must be taken for an intelligent tutoring system. If an engineered system was to respond to the needs of its user, this data would have to be parsed, interpreted, and recommended for action in realtime. Because of individual differences, day-to-day variations, inter-day variations, sensor placements, and a host of other issues, baseline measurements cannot be stored for the individual[3]. Establishment of the meaning of these sensors measurements must be made as close to instantaneously as possible. This presents its own problems, starting with the ideas that the data can be of potentially infinite length, and all points and trends on a new individual are unknown.

Intelligent Tutoring comes in many forms. It can be a virtual world where the student can play and practice skills, a computer-led classroom presentation, a computer-human mixed-discussion activity, or other teaching methods. The two fundamental inputs to the human tutor are the assessments of knowledge and the assessments of the affect of the student[4]. Expert human tutors achieve learning gains of two sigma, or roughly two letter grades[5]. Web-based computer tutors, which perform only one of these assessments, have been shown to produce one sigma of learning gain[6]. In order to increase the effectiveness of computer-based learning activities, the intelligent tutor should mirror the approach of human tutoring, and account for the affect of the person being trained[7].

All of the above describes the effort of the author to solve part of a problem which is not only important, but novel. Intelligent tutoring systems should respond to the needs of their students, by assessing their affect, from sensor data taken from the student in realtime, and classified along with self assessments and performance measures. This research addresses this issue through the comparison of supervised against unsupervised methods of machine learning on a dataset of wide-ranging sensors.

This research will develop realtime, unsupervised or semi-supervised methods of affect detection. These models will be directly compared against the supervised linear regression tree models built from validated benchmarks collected in another experiment using low-cost sensors as measurement and high-cost EEG as a moment-by-moment ground truth[8]. The three main thrusts of this research are:

- Group classification models of sensor data are impractical or nonexistent
  - Individual classification models must be built
  - Shown via literature
- Offline individual models of sensor-based affect are not reusable
  - Models must be built in realtime
  - Shown via literature
- Realtime-constructed models are comparable to their offline counterparts
  - Making them usable in Intelligent Tutoring Systems
  - Shown via experiments and artificial intelligence datastream development[9]

References.

1. Baker, R.S.J.d.(2010). Mining data for student models. In R. Nkmabou, R. Mizoguchi, & J. Bourdeau (Eds.) *Advances in Intelligent Tutoring Systems, Studies in Computational Intelligence* (Vol. 308, pp. 323-337*)*. Heidelberg: Springer Verlag.
2. D'Mello, S.K., Graesser, A.C.: Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. User Model. User-Adapt. Interact. 20(2), 147–187 (2010)
3. Bersak, D., McDarby, G., Augenblick, N., McDarby, P., McDonnell, D., McDonal, B. and Karkun, R. (2001) "Biofeedback using an Immersive Competitive Environment". Online Proceedings for the Designing Ubiquitous Computing Games Workshop, Ubicomp 2001.
4. Scandura, J.. (2011). What TutorIT Can Do Better Than a Human and Why: Now and in the Future. In Tech., Inst., Cognition and Learning, Vol. 8, pp. 175–22. Old City Publishing, Inc.
5. Bloom, B. S. (1984). The 2 sigma problem. The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4-16.
6. Verdú, E., Regueras, L.M., Verdú, M. J., De Castro, J.P., & Pérez, M.A. (2008). Is Adaptive Learning Effective? A Review of the Research. In L. Qing, S. Y. Chen, A. Xu, & M. Li (Eds.), Proceedings of the 7th WSEAS International Conference on Applied Computer & Applied Computational Science (ACACOS '08) (pp. 710-715). Stevens Point, Wisconsin: WSEAS Press.
7. Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., Picard, R.: Affect-Aware Tutors: Recognizing and Responding to Student Affect. International Journal of Learning Technology. 4, 129--164 (2009)
8. Carroll, M., Kokini, C., Champney, R., Sottilare, R., & Goldberg, B. (2011). Modeling Trainee Affective and Cognitive State Using Low Cost Sensors. In Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL, November 2011.
9. Brawner, K., & Gonzalez A. (2011). Realtime Clustering of Unlabeled Sensory Data for User State Assessment. In Proceedings of International Defense & Homeland Security Simulation Workshop of the I3M Conference. Rome, Italy, September 2011.