Opportunities and Challenges in Generalizable Sensor-Based Affect Recognition for Learning

Jonathan P. Rowe[1], Bradford W. Mott[1], and James C. Lester[1]

[1]Center for Educational Informatics, North Carolina State University, Raleigh, NC 27695
{jprowe, bwmott, lester}@ncsu.edu

Abstract. Recent years have witnessed major research advances in sensor-based affect recognition. Alongside these advances, there are many open questions about how effectively current affective recognition techniques generalize to new populations and domains. We conducted a study of learner affect with a population of cadets from the U.S. Military Academy using a serious game about tactical combat casualty care. Using the study data, we sought to reproduce prior affect recognition findings by inducing models that leveraged posture-based predictor features that had previously been found to predict affect in other populations and learning environments. Our findings suggest that features and techniques, drawn from the literature but adapted to our setting, did not yield comparably effective models of affect recognition. Several of our affect recognition models performed only marginally better than chance, and one model actually performed worse than chance, despite using principled features and methods. We discuss the challenges of devising generalizable models of affect recognition using sensor data, as well as opportunities for improving the accuracy and generalizability of posture-based affect recognition.

Keywords: Affect Recognition, Posture, Microsoft Kinect, GIFT

1 Introduction

Affect is instrumental to learning. Students' affective experiences shape their learning behaviors and outcomes, and vice versa. Growing recognition of this relationship has led to the emergence of work on affect-enabled learning technologies, which endow educational software with the ability to recognize, understand, and express affect. Several affect-enabled learning technologies have been developed in recent years, spanning a broad range of domains, including computer science education [1], reading comprehension [2], mathematics [3], and computer literacy [4]. Although these bespoke affect-sensitive systems have yielded promising results, there are many open questions about whether existing affect recognition techniques generalize to new domains, populations, and settings.

Recent work on sensor-based affect recognition holds promise for yielding generalizable models. Because sensor-based models typically do not rely on features that are specific to particular learning environments, in principle, they should port across domains and settings. Sensor-based affect recognition models have been devised for a range of modalities, including facial recognition, gaze tracking, speech analysis, physiological signals (e.g., heart rate, electrodermal activity), hand gesture, and posture [5]. In this work, we focus on posture-based affect recognition, which has shown promise for its capacity to predict student affect [1, 3, 4]. Motion sensors, such as Microsoft Kinect, can be used to gather rich data streams about posture, they are relatively low-cost, and they are increasingly getting integrated into mainstream computers [6]. By modeling these rich data streams with machine learning techniques, posture-based affect recognition models have been induced that can effectively predict participants' affective self-reports, as well as expert judgments of affect gleaned from freeze-frame video analyses [1, 3, 4].

In this paper, we summarize our work on posture-based affect recognition with the Generalized Intelligent Framework for Tutoring (GIFT). In collaboration with Teachers College Columbia University and the U.S. Army Research Laboratory, we conducted a study of learner affect with cadets from the U.S. Military Academy (USMA) using a serious game for learning tactical combat casualty care skills. Using this study data, we sought to reproduce prior affect recognition findings, leveraging posture-based predictor features that had previously been found to predict affect in other populations and learning environments. However, our results indicated that the same features and techniques, adapted to our setting, did not yield comparably effective models. Our affect recognition models performed only marginally better than chance, and in fact, one model actually performed worse than chance. We discuss the challenges of devising generalizable models of affect recognition using sensor data, and describe opportunities for improving the predictive accuracy of posture-based affect recognition models.

2 Posture Sensor-Based Affect Recognition

Several research labs have investigated multimodal affect recognition in learning environments over the past decade. Our research on generalizable sensor-based affect recognition is strongly influenced by this work. To date, posture-based affect recognition models have been induced with data from pressure-sensitive chairs [3, 4], as well as motion sensors, such as Microsoft Kinect [1]. These two data streams, drawing from distinct types of sensors, are superficially different, but can be distilled into analogous predictor features that have similar relationships with affective states such as engagement, boredom, frustration, and confusion. Features can be distilled from both types of data to indicate leaning forward, leaning backward, sitting upright, and fidgeting. We summarize several representative studies that have utilized these types of features to recognize learner affect, and that have influenced our own work.

D'Mello and Graesser utilized posture data from the Body Pressure Measurement System (BPMS) to predict judgments of student affect during learning with AutoTutor [4]. The BPMS is a pressure-sensitive system that is comprised of a grid of sensing elements placed across a chair's seat and back. In their study, participants were video recorded, and several judges analyzed the video using freeze frame analysis in order to code participants' affective states retrospectively. Using this data, D'Mello and Graesser induced a series of emotion-specific binary logistic regression models, each distinguishing a particular affective state from neutral, using 16 posturebased features as predictors. Their findings indicated that the models, averaged across judges, explained approximately 11% of the variance in affective state, with findings in line with an attentive-arousal theoretical framework. Specifically, affect such as delight and flow coincided with forward leaning, boredom coincided with a tendency to lean back, and states such as confusion and frustration coincided with an upright posture.

Cooper et al. used a suite of sensors to collect data on student affect in Wayang Outpost, an ITS for high school geometry [3]. The sensors included a skin conductance bracelet, pressure sensitive mouse, pressure sensitive chair, and mental state camera, which provided data on student posture, movement, grip tension, arousal, and facial expression. The pressure sensitive chair was a simplified version of the sensing system utilized by D'Mello & Graesser [4], utilizing a series of six forcesensitive resistors distributed across the seat and back of a seat cover cushion. Data from these channels was distilled into predictor features to predict students' emotion self-reports, which were queried every five minutes throughout the learning interaction. The posture-based features included net change in seat and back pressure between the current timestep and previous timestep, and a feature indicating whether the student was leaning forward or not. Step-wise linear regression models were induced to predict students' emotion self-reports. Results indicated that posture-based features were significantly predictive of self-reported excitement during learning, although they were not part of the best-performing models for other emotional states.

Grafsgaard et al. have investigated postured-based affect prediction using Microsoft Kinect sensors with an intelligent tutoring system for introductory programming [1]. Posture features were distilled from depth image recordings by tracking the distance between the depth camera and the participant's head, upper torso, and lower torso. The features included discretized distance indicators, such as near, mid, and far head positions, each determined by whether the tracked head point was closer or farther from the median head position by one standard deviation. In addition, a postural movement feature was distilled to label occasions where the average amount of acceleration of the head tracking point was greater than the population average over a one-second window. The posture-based predictor features were combined with features distilled from other multimodal streams to induce multiple regression models for predicting students' retrospective self-reports of engagement and frustration. Findings indicated that posture features were predictive of both self-reported affective states: leaning forward was predictive of both higher engagement and higher frustration, and postural movement was associated with increased frustration and reduced learning.

Building upon this foundation, we set out to distill similar predictor features from the data collected at USMA, and apply similar machine learning methods, to produce affect recognition models for predicting field observations of affect.

3 Kinect-Driven Affect Recognition in GIFT

We collected learning and affect data from 119 USMA cadets as they used the vMedic serious game environment for learning tactical combat casualty care skills. In vMedic, the learner adopts the role of a combat medic who must properly treat and evacuate one (or several) of her injured fellow soldiers by following standard medical procedures within the game environment. All participants completed the same training module, which was managed by GIFT. The training module consisted of a pretest, a brief PowerPoint on tactical combat casualty care, four training scenarios in vMedic, and a post-test.

Each participant was assigned to a research station that consisted of an Alienware laptop, a Microsoft Kinect for Windows sensor, an Affectiva Q Sensor, and a mouse and pair of headphones. As participants completed the study materials, a pair of field observers regularly recorded participants' physical displays of emotion. The field observers followed an observation protocol, BROMP, developed by Baker et al. [7], in which observers walked around the perimeter of the study room, discreetly recording observations of each participant's affect in a round robin sequence. The field observers coded for seven affective states: concentration, confusion, boredom, surprise, frustration, contempt, and other.

The study produced several parallel data streams, including vMedic trace data, Kinect position tracking data, electrodermal activity data, pre- and post-test response data, and field observation data. In this work, we focus on analysis of the Kinect and field observation data, which were fused into a single time-synchronized dataset. The dataset was cleaned and filtered in order to remove any Kinect-tracking glitches, as well as non-essential vertex data. Afterward, 73 predictor features were distilled, which characterized participants' postural positions and dynamics, inspired by similar features from the research literature on posture-based affect recognition. The features included summary statistics for three points tracked by the Kinect: head, top_skull, and center_shoulder. Specifically, we computed features for the current distance and depth of each vertex; the minimum, maximum, median, and variance in distance of each vertex observed thus far; the same statistics for 5, 10, and 20-second windows; several features that characterized net changes in vertex distance, analogous to the net_change features reported in [3, 4]; and sit_forward, sit_back, and sit_mid features analogous to those reported in [1, 3].

Using this feature data, we induced separate affect detectors for each emotional state using a range of machine learning techniques in RapidMiner 5.3, inclu ing J48 decision trees, naïve Bayes, support vector machines, logistic regression, and JRip [8]. The detectors were cross-validated using 10-fold participant-level cross validation. Oversampling was used to balance class frequency by cloning minority class instances in the training sets. Forward feature selection was performed to reduce the number of predictor features used in the models. We calculated Kappa and A' to assess the models' performance.

Across all of the emotions, our posture-based affect recognition models achieved an average Kappa of 0.064, and 0.521 for A' [8]. The best performing model was for boredom, which showed Kappa=0.109, A'=0.528 using logistic regression. Overall, the models performed slightly better than chance, with the exception of the surprise detector, which actually performed worse than chance, Kappa=-0.001, A'=0.493.

These results were surprisingly modest, despite our best efforts to run a carefully designed study and reproduce previously reported methods. There are several possible explanations. It is possible that BROMP labels, which are based on holistic judgments of affect over 20-second windows, are ill matched for methods that leverage low-level postural features as predictors. Previous work utilized self-reports and freeze frame video analysis, which have different tradeoffs than BROMP. Additionally, much of the work on posture-based affect recognition has taken place in laboratory settings with a single participant at a time. In our study, up to 10 participants were present, with each research station having a slightly different sensor position and orientation. This variation may have introduced additional noise to the data, which could have been problematic for the methods reported here. Further, the population of learners we used in the study, USMA cadets, showed considerable restraint in their physical expressions of affect. As such, the displays of affect via body language may have been different than those encountered in prior work, making them ill matched for the predictor features that we engineered. These findings underscore the challenges to be overcome in efforts to devise generalizable models of affect recognition.

We draw several lessons for our continued work on sensor-based affect recognition with GIFT. First, orienting Kinect sensors' position and orientation to track points on participants' lower torso could prove important for posture detection. In the present study, our sensor configuration enabled us to track only vertices on participants' upper torso and head, which may have limited the features we could distill.

Second, it would be useful to validate the Kinect vertex data recorded by GIFT against the sensor's raw depth video data. Prior work on Kinect-based posture detection directly leveraged raw depth channel data, but this method is memoryintensive and requires custom implementation of posture tracking algorithms [1]. While vertex data produced by Kinect should in principle provide the same information about posture as raw depth data, validating this fact would ensure that our findings relate to the generalizability of affect recognition techniques, and not assumptions about underlying data sources.

Third, investigating alternate machine learning techniques could prove useful for enhancing the predictive ability of posture-based predictor features. It is possible that temporal models, such as dynamic Bayesian networks, which explicitly model shifts in posture and affect, could yield improved results. Furthermore, recent work on deep learning techniques may show promise, given their capacity to perform automated representation learning. Although additional work is merited to manually engineer high-level features to match the holistic encodings of affect provided by BROMP, it would be ideal to automate this manual feature engineering process, as is one of the promises of representation learning techniques such as deep learning.

4 Conclusions

We have described work investigating the generalizability of posture sensor-based affect recognition. We collected a multimodal dataset on affect and learning with a group of USMA cadets using a serious game for tactical combat casualty care. Leveraging techniques from the affective computing research literature, we distilled a range of posture-based predictor features for modeling participants' affective states with machine learning. Our results indicated that posture-based features and models, which had previously been found to yield effective affect recognition systems, did not work as effectively on our data as had been found with other populations and learning environments. In fact, most of our affect recognition models performed only marginally better than chance, despite the use of principled features and models. Although there are several directions to investigate for enhancing our posture-based affect recognition models, the failure of existing techniques to generalize to our data is notable. These findings underscore the challenges, and opportunities, in research on affect recognition and generalizable approaches to intelligent tutoring.

Acknowledgments. The authors wish to thank the U.S. Army Research Laboratory for supporting this research. We are grateful to Vasiliki Georgoulas, Michael Matthews, and James Ness for facilitating the study at the United States Military Academy. Additionally, we wish to thank our collaborators, Ryan Baker, Jeanine DeFalco, and Luc Paquette at Teacher's College Columbia University, and Keith Brawner, Benjamin Goldberg, and Bob Sottilare from ARL.

References

1. Grafsgaard, J. F., Wiggins, J. B., Vail, A. K., Boyer, K. E., Wiebe, E. N., Lester, J. C.: The Additive Value of Multimodal Features for Predicting Engagement, Frustration, and Learning during Tutoring. In: Proceedings of the 16th ACM International Conference on Multimodal Interaction, pp. 42–49. (2014)
2. Mills, C., Bosch, N., Graesser, A., Mello, S. D.: To Quit or Not to Quit: Predicting Future Behavioral Disengagement from Reading Patterns. In: Proceedings of the 12th International Conference on Intelligent Tutoring Systems, pp. 19–28. (2014)
3. Cooper, D. G., Arroyo, I., Woolf, B. P., Muldner, K., Burleson, W., Christopherson, R.: Sensors model student self concept in the classroom. In: Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization, pp. 30–41. (2009)
4. Mello, S. D., Graesser, A.: Mining Bodily Patterns of Affective Experience During Learning. In: Proceedings of the 3rd International Conference on Educational Data Mining, pp. 31–40. (2010)
5. D'Mello, S. K., Kory, J.: A Review and Meta-Analysis of Multimodal Affect Detection Systems. ACM Computing Surveys, 47(3), 43 (2015)
6. Intel. (2015, March 20). Intel RealSense. Retrieved from http://www.intel.com/realsense.
7. 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. International Journal of Human-Computer Studies, 68(4), 223–241 (2010)

8. Paquette, L., Rowe, J., Baker, R., Mott, B., Lester, J., DeFalco, J., Brawner, K., Sottilare, R., Georgoulas, V.: Sensor-Free or Sensor-Full: A Comparison of Data Modalities in Multi-Channel Affect Detection. Under review for the 8th International Conference on Educational Data Mining, (under review)