The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Relating Children's Automatically Detected Facial Expressions To Their Behavior in RoboTutor

Mayank Saxena
Delhi Technological University
New Delhi, India
mayank26saxena@gmail.com
+91-9910002161
http://mayank26saxena.github.io/

Rohith Krishnan Pillai
Carnegie Mellon University
Doha, Qatar
rohithkp@andrew.cmu.edu

Jack Mostow
Carnegie Mellon University
Pittsburgh, PA, USA
jmostow@andrew.cmu.edu

Abstract

Can student behavior be anticipated in real-time so that an intelligent tutor system can adapt its content to keep the student engaged? Current methods detect affective states of students during learning session to determine their engagement levels, but apply the learning in next session in the form of intervention policies and tutor responses. However, if students' imminent behavioral action could be anticipated from their affective states in real-time, this could lead to much more responsive intervention policies by the tutor and assist in keeping the student engaged in an activity, thereby increasing tutor efficacy as well as student engagement levels. In this paper we explore if there exist any links between a student's affective states and his/her imminent behavior action in RoboTutor, an intelligent tutor system for children to learn math, reading and writing. We then exploit our findings to develop a real-time student behavior prediction module.

Introduction

Recently there has been a significant amount of research in trying to make intelligent tutor systems reactive to student's responses, to allow for a more intuitive interface with the tutor and also more importantly to be able to keep the student engaged in the activity for longer periods. Current tutor systems use computer vision and other sensors to detect the affective states of the users (happy, sad, content, disgust, fear, etc.) to gauge user engagement levels at end of learning activity (Whitehill et al. 2014) (D'mello and Graesser 2010). However, these systems are deployed mostly on adults and in controlled environments. Also such systems stop at the level of discerning the students engagement levels to determine the tutor's intervention policy, which means that the policy only has a limited type of input namely the engagement levels. It might be more fruitful to base the policy on the anticipation of a behavior that the student might exhibit in the near future, such as when the student might choose to exit an activity prematurely. Such inferences about the student behavior/actions provide more inputs for an intervention policy to make more nuanced, responsive reactions.

In this research we discuss two main questions, namely, given the additional information of the students affective state, what can we infer about the student behavioral action? Is it possible to predict the students next action based on their affective states in a real-time system? The analyses was conducted in the context of RoboTutor, an open source Android application to teach children aged between 6-12, who have little or no prior schooling nor access to technology, math, reading and writing tutor. We discuss our findings in the next sections.

Data Collection and Processing

RoboTutor consists of multiple educative activities for children such as writing, Akira (a car racing game), bubble pop, and myriad read-along illustrated stories with speech recognition. As RoboTutor runs on an Android tablet, the front facing camera is used for affective state detection. The video feed of children using RoboTutor, along with the log files containing all the in-app user actions by the child is stored on the cloud, using Google Drive. Then the log files are parsed and added to respective tables in a database for use for querying, and analyses.

We use OpenFace (Baltrušaitis, Robinson, and Morency 2016), an open-source facial behavior analysis toolkit to extract facial action units (FAUs) from the video feed. OpenFace was trained on diverse data with respect to age, gender, ethnicity. We identified a set of pedagogically relevant emotions such as boredom, confusion, frustration, surprise, delight, neutral. Next, we created a mapping from FAUs returned by OpenFace to the above mentioned emotions, all using existing literature on the subject (Craig et al. 2008; McDaniel et al. 2007). Comprehensive literature on this subject could not be cited here due to lack of space. The affective states from the video were then joined with its respective parsed log data file, containing behaviors such as back button presses, completion of activity, correctness, etc. to create a file conducive for analysis. We used a total of 17 different video of length ranging between 20 to 30 minutes log session pairs to conduct the analysis (N = 17). The analysis included visualization of the data, and finding statistically significant correlations between emotions and the students behavior. The validation of detection accuracy for the emotions were done by spot-checking samples of data, and testing external validity of detected emotions to see if their correlations with logged tutor events, e.g. response correctness vs. delight, are statistically reliable and psychologically plausible.

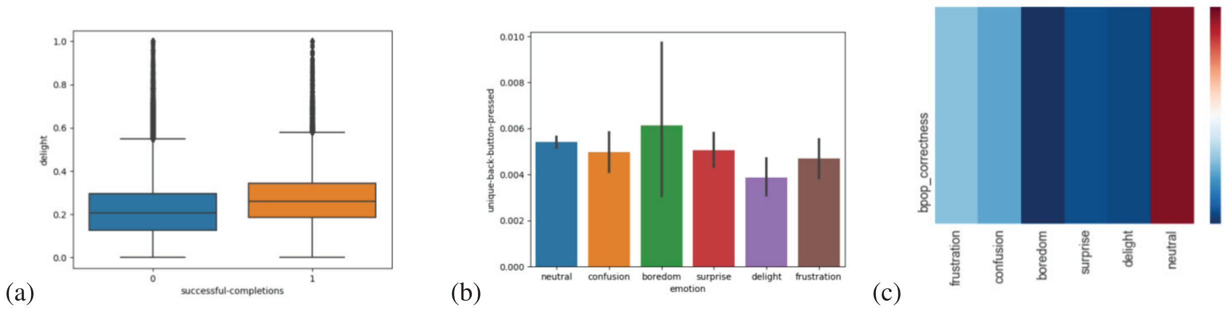

Figure 1: Statistical correlations between emotions and the students behavior in RoboTutor

Results and Discussion

Some of the key results from the analyses are shown in Figures 1(a) to 1(c). Figure 1(a) shows that the median delight that students expressed across all activities was much higher than when the student did not complete the activity. The difference in delight between the two groups is statistically significant (p-value < 0.05 at a confidence value of 95%). Similarly, Figure 1(b), illustrates the distribution of the affective states exhibited by the students when students hit the back button while using tutor. Although boredom is more frequent, only neutral, surprise and delight were statistically significant. Figure 1(c) the only positive correlation of correctness in the bubble pop activity in the RoboTutor application is the emotion, neutral/flow. This is in stark comparison to all other emotions which negatively correlated with correctness and performance in the bubble pop activity.

An important observation is that affective state delight could be a predictor of completion of activity. Similarly, the second observation is that affective state surprise is good indicator of the students pressing the back button during an activity. Although neutral value is higher, it is also more uniformly distributed emotion across the activities, but surprise seems to be high enough to distinguish itself from the emotion delight, and clearly shows the links between itself and the quitting an activity. Also from Figure 3 note that correctness in different activities can also be predicted using a certain affective state, such as neutral/flow. This makes sense as neutral/flow is most notably seen in students when they are actively learning, and processing the information presented, and so students would be more likely to get tasks right in this context.

These results show that we can definitely see links and possible predictors for different behaviors on tutors by carrying out similar analyses as described in this paper. This is good evidence that we could possibly create a real-time system for finding optimal timings for the intervention policy to keep the student engaged in the app. We believe that this methodology and technology can be extended to more platforms and that this will help in making tutors more personalized. It can be seen that once the application will be able to predict that the user is about to quit a session or activity, it could change its strategy and display content which engages the user more, which could lead to a possible decrease in the session dropout rate in intelligent tutors systems.

Conclusion

Our results show that there do exist links between student behavior and their affective states. This indicates that indeed a real-time system that can detect affective states of the students can possibly be used to predict the behavior of students given ample data. This finding led us to develop such a real-time system and integrate it in RoboTutor (an intelligent tutor system). This system running real-time alongside the tutor helps the tutor to produce more complex and nuanced intervention policies due to the larger number of inputs that the system provides over a traditional system solely providing levels of engagement.

Acknowledgements

Special thanks to Vishnu Rajan Tejus for integrating the emotion detector into RoboTutor. This research was supported by Robotics Institute Summer Scholars (RISS)

References

Baltrušaitis, T.; Robinson, P.; and Morency, L.-P. 2016. Openface: an open source facial behavior analysis toolkit. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, 1–10. IEEE.

Craig, S. D.; D'Mello, S.; Witherspoon, A.; and Graesser, A. 2008. Emote aloud during learning with autotutor: Applying the facial action coding system to cognitive–affective states during learning. *Cognition and Emotion* 22(5):777–788.

D'mello, S. K., and Graesser, A. 2010. Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction* 20(2):147–187.

McDaniel, B.; D'Mello, S.; King, B.; Chipman, P.; Tapp, K.; and Graesser, A. 2007. Facial features for affective state detection in learning environments. In *Proceedings of the Cognitive Science Society*, volume 29.

Whitehill, J.; Serpell, Z.; Lin, Y.-C.; Foster, A.; and Movellan, J. R. 2014. The faces of engagement: Automatic recognition of student engagementfrom facial expressions. *IEEE Transactions on Affective Computing* 5(1):86–98.