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 at end of learning session to determine their engagement levels, but this provides a one-dimensional input for 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 complex intervention policies by the tutor. This in turn would assist in keeping the student engaged in an activity, thereby increasing tutor efficiency as well as student engagement levels. In this paper we first explore if there exist any links between a student's affective states and his/her imminent behaviour 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

Intelligent tutor systems are being more commonly used in education over the last few years. Many such intelligent tutors use a variety of activities to teach students. Keeping students engaged in these activities is essential for the effectiveness of a tutor. Recently there has been a significant amount of research in trying to make such 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.). These affective states are analyzed 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. Many times this might be insufficient to base a policy over, and 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, complex reactions. For example, if a module provides the tutor with inferences about the next most probable behavior being the student clicking out of an activity prematurely with a very high confidence value, the tutor could possibly change the current approach of teaching the content in an attempt to prevent disengagement.

In this research we discuss the analysis conducted to study if there are any interesting links between student behavior and their affective states in an intelligent tutor system and whether affective states can be used to predict students behavior in real-time. 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 had two main questions to answer, namely,

(i) Given the additional information of the students affective state, what can we infer about the student, the different activities and the subject categories?

(ii) Is it possible to predict the students next action based on their affective states in a real-time system?

We discuss our findings in the next sections.

Data Collection and Processing

RoboTutor has 2 testing facilities in Tanzania, where the children use and interact with the application. The tutor in the application consists of multiple activities such as bubble pop, writing, Akira (the car racing game), 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 camera video 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 them. We identified a set of pedagogically relevant emotions and a mapping from the FAU returned by OpenFace and these emotions using existing literature on the subject (Craig et al. 2008; McDaniel et al. 2007). 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.

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 the 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.

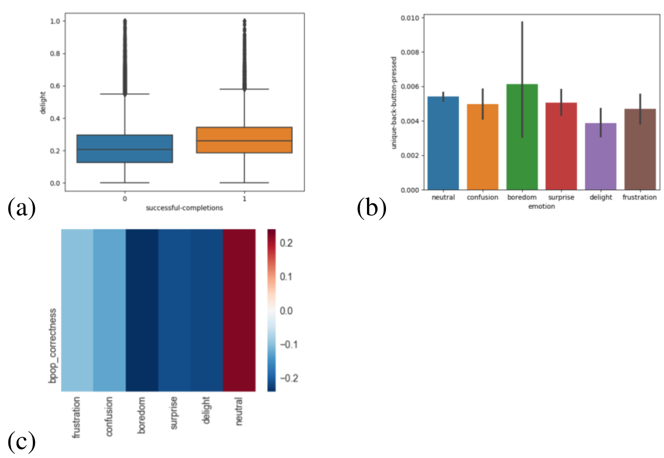

Figure 1: Statistical correlations between emotions and the students behavior in RoboTutor

An important observation is the fact 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 a 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.

All these showcase the fact the we can definitely see links and possible predictors for different behaviors on RoboTutor by carrying out similar analyses described to those described in this paper. This shows good evidence that we could possibly create a real-time system for finding optimal timings for the intervention policy to be used 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 all tutors become more personalized. Other effect of such a student behavior anticipation module include a possible decrease in the session dropout rate in intelligent tutors systems. It can be seen that once the application will be able to predict that user is about to quit a session or activity, it could change its strategy and display content which engages the user more.

Conclusion

Our results show that there do exists 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. Development of such an real-time system is the next logical step for intelligent tutor systems like RoboTutor. Currently work on designing this system for RoboTutor as an android service is underway. Such a system running real-time alongside the tutor would be able to help the tutor to produce more complex and nuanced intervention policies due to the larger number of inputs that such a system provide over a traditional system providing solely levels of engagement, and thereby help create more effective tutor systems.

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.

SUPPLEMENTARY MATERIAL

MAYANK SAXENA [1] & ROHITH KRISHNAN PILLAI [2] & JACK MOSTOW [3]

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 WHAT IS ROBOTUTOR?

RoboTutor is one of five $ 1M Finalists in the $ 15M Global Learning XPRIZE Competition (http://learning.xprize.org). This international competition seeks to address the acute shortage of teachers in developing countries. The goal is to create an open-source Android tablet app that enables children ages 7-10 with little or no access to schools to learn basic reading, writing, and arithmetic without adult assistance.

2 MOTIVATION FOR WORK

For the purpose of making Intelligent Tutor Systems adaptive, current models take into consideration the affective states of happiness, sadness, anger, disgust, contempt and neutral. However, prior research has informed us that such emotions are not very relevant to an educational environment where learning is being measured. Instead the affective states of delight, surprise, confusion, frustration, boredom and neutral (flow) are more commonly seen and their relations to learning also more widely studied and accepted. We fix this gap by developing a model which takes

[1] *Department of Mathematics, Delhi Technological University, New Delhi, India*
[2] *Department of Computer Science, Carnegie Mellon University, Doha, Qatar*
[3] *School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA*

into account pedagogically relevant affective states as well as the student's predicted behaviour action to adapt the content of the tutor dynamically to improve learning and engagement of the student.

3 DATA COLLECTION PIPELINE

There were 2 kinds of data that we used for the analysis. The first one being the video files of screen recorded video of children interacting with RoboTutor. The second being the log files which are automatically generated while the child interacted with the RoboTutor application. In our case, the log files provided information regarding the students actions, performance during the sessions, and overall in-app behavior of the students. In total, we had 17 log files in .json format and their corresponding video files. The length of the video files varied from 20 to 30 minutes.

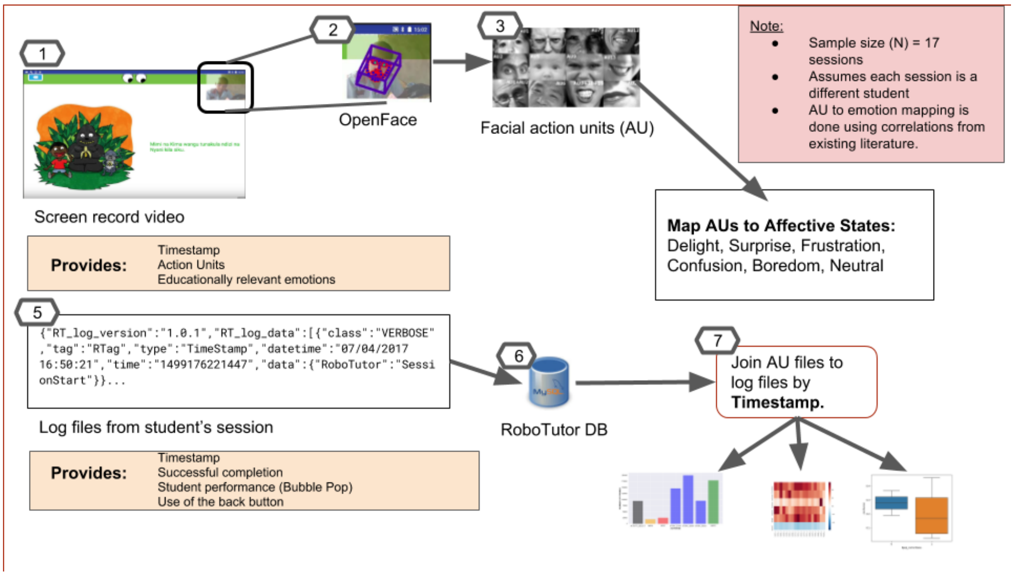

Figure 1: Data Collection Pipeline

As we can see in Figure 1, we first obtained the facial data from the video files. The next step involved running the OpenFace tool to extract information regarding the Facial Action Units (FAU). Using this information about the FAUs, we found out which affective state i.e. which emotion is present and with what probability. Next, we obtained information about the student's actions from the .json files which we got from the RoboTutor database. The final step involved joining the FAU files and the log files by the time-stamps.

4 DATA PROCESSING

The mapping of the FAUs to their corresponding affective states was done by looking up existing literature to query the correlation between the different FAUs and our required educationally relevant affective states. We use the correlation factors to guide us in weighting the FAUs to create a linear weighted average of the FAUs for each of these 6 affective states. This creates the numerical values for the emotions that we then normalized, to create the normalized affective state value and the maximum likelihood affective state for that particular time stamp.

Emotion	Pattern	Action Units Description
Frustration	1,2 1->2	Inner and Outer Brow Raised Together Presence of an inner brow raise will trigger an outer brow raise and vice versa
Confusion	4,7 4->7	Brow lowered with tighened lids Tightened lids will lead to lowered brows
Surprise	1 2 5 26	Inner brow raise, Outer brow raise Upper lid raiser Jaw drop
Delight	7 12 25 26	Lid tightener Lip corner puller Lips part Jaw drop
Neutral	-	Absence of all action units

Table 1: Mapping for FAUs to Pedagogically Relevant Affective States

5 EXAMPLE

Our model when given an image as input, gives as output the different probabilities for the presence of the emotion. In Figure 2, the "delight" emotion has the maximum probability.

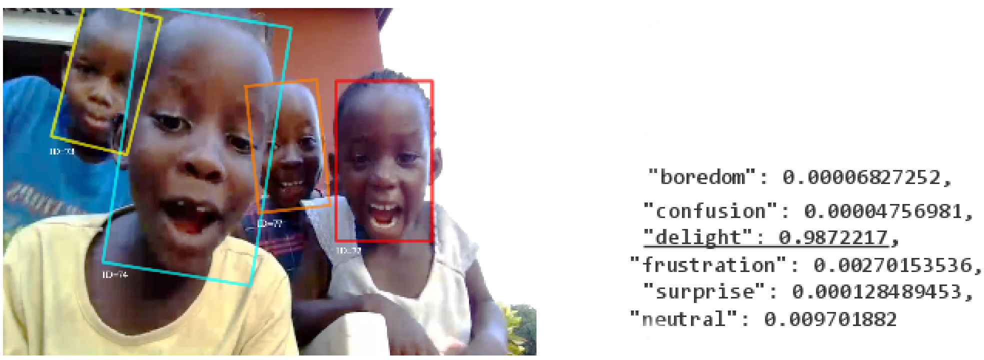

Figure 2: Statistical Probe of Tutoring

6 FUTURE WORK

There is a lot of scope for future work in this research. We propose a framework which could detect affective states such as frustration, confusion etc in real time using computer vision techniques and could prompt the tutor to change its strategy in case of such a situation. This could help improve the experience of the student tremendously and will also lead to improved learning and increased engagement levels. This framework could be implemented in the form of an API or a background service, which could be work along with every tutor and communicate with the tutor so that the tutor can dynamically change its strategy based on inputs about the learner's emotional affective states.

We have already developed a framework for the Android platform which is currently being used by RoboTutor to determine the imminent behaviour action of the child in real time. This information is being used to dynamically adapt the content of the tutor.