

RPE Proposal

Rohit Murali

Supervisor: Cristina Conati

1 Background and Motivation

Emotions play a large part in learning. Studies (Baker et al. 2010; Wortha et al. 2019) show that emotions can affect learning in Intelligent Tutoring Systems (ITSs). This encourages educators to design ITSs which can adapt to a student's emotions during their interaction (Grawemeyer et al. 2016; Woolf et al. 2009). These systems require detecting the user's affective state to an extent during learning, and there is research in this area that tries to predict emotion valence through interaction data (Lallé et al. 2018), single emotion using eye-tracking data (Jaques et al. 2014; Sims et al. 2020) and co-occurring emotions (Lallé et al. 2021). This proposal outlines a research idea focused on predicting co-occurring emotions during learning with the ITS, MetaTutor (Azevedo et al. 2013). Prediction tasks with MetaTutor are usually done by collecting students' eye-tracking and interaction data. (Lallé et al. 2021) predicts pairs of co-occurring emotions (Boredom and Frustration; Curiosity and Anxiety) using eye-tracking and interaction data with MetaTutor. This paper shows that eye-tracking and interaction data used in isolation can predict the co-occurrence of the two emotion pairs significantly better than a baseline. This work was done as part of an RA project that started in Winter Term 1 2020.

2 Proposed Project

We will extend the task of predicting co-occurring emotions using eye-tracking and interaction data in (Lallé et al. 2021) by combining two datasets of different user studies with MetaTutor; the study used in (Lallé et al. 2018), and the study used in (Lallé et al. 2021) and (Jaques et al. 2014) with . Both studies involve university students who learn about the circulatory system through MetaTutor. The students' gazes are tracked with eye-trackers and asked to self-report their emotions at different stages of the experiment in both studies. The dataset in (Lallé et al. 2021) has valid eye-tracking data for 270 self-reports and valid interaction data for 325 self-reports. The dataset in (Lallé et al. 2018) has valid eye-tracking for 126 self-reports and valid interaction data for 176 self-reports. The two studies are structurally similar, so it makes sense to attempt to merge the two datasets, however, there are variations among the studies that makes this task non-trivial. These include structural differences, such as the time intervals between emotion reports and number of emotion

reports asked per student, and measurement differences, such as the type of eye-tracker used. We restrict our prediction tasks to pairs of co-occurring emotions as higher-order classification tasks might be too hard considering the limited size of the dataset. The first part of this RPE project involves identifying these differences in detail, coming up with a way to combine the two datasets, and choosing the right pair(s) of co-occurring emotions for prediction tasks, which may be different from those in (Lallé et al. 2021).

With the larger dataset (*MetaTutor dataset* onward), the next step of the RPE project is to deploy machine-learning models for the 4-way classification task related to the emotion pairs we decide are suitable. We will use the random forest (RF) and logistic regression (LR) classifiers used in (Lallé et al. 2021) on the MetaTutor dataset as a baseline since they are the current existing standard in literature for the task of predicting co-occurring emotions. In (Lallé et al. 2021), we found that simple feature fusion of the two types of data did not improve performance. So we plan on investigating an ensemble classifier involving the RF and LR models trained on the different data sources, eye-tracking and interaction data. Moreover, with a larger dataset, deep-learning models may work well and so we will evaluate the performance of a fully-connected neural network and later VTnet (Sims et al. 2020) compared to the ensemble and baseline classifiers.

The last part of the project involves investigating the feasibility of the VTnet architecture (Sims et al. 2020), used in a different context of users interacting with ValueChart (Carenini et al. 2004), on the MetaTutor dataset. The VTnet architecture used in (Sims et al. 2020) uses raw eye-tracking data, including scan-paths, to classify user confusion by connecting the outputs of an Recurrent Neural Network (RNN) and a Convolutional Neural Network (CNN) and sending them to a fully-connected neural network. This part of the project will involve a novel upgrade to the VTnet architecture by including both eye-tracking and interaction data as inputs to the network, and assessing its performance against the other models.

3 Timeline

A proposed timeline of the project will be as follows.

- **May to Mid-June:** Identify differences between the two datasets and understand how to combine them.
- **Mid-June to Mid-July:** Evaluate performance of the ensemble model and the fully-connected neural network to the current models in literature, namely the RF and LR models in (Lallé et al. 2021).
- **Mid-July to End-August:** Upgrade VTnet architecture to include interaction data and assess its performance.

References

- Azevedo, Roger, Jason Harley, Gregory Trevors, Melissa Duffy, Reza Feyzi-Behnagh, François Bouchet, and Ronald Landis (2013). “Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems”. In: *International handbook of metacognition and learning technologies*. Springer, pp. 427–449.
- Baker, Ryan SJd, Sidney K D’Mello, Ma Mercedes T Rodrigo, and Arthur C Graesser (2010). “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”. In: *International Journal of Human-Computer Studies* 68.4, pp. 223–241.
- Carenini, Giuseppe and John Loyd (2004). “Valuecharts: analyzing linear models expressing preferences and evaluations”. In: *Proceedings of the working conference on Advanced visual interfaces*, pp. 150–157.
- Grawemeyer, Beate, Manolis Mavrikis, Wayne Holmes, Sergio Gutierrez-Santos, Michael Wiedmann, and Nikol Rummel (2016). “Affecting off-task behaviour: how affect-aware feedback can improve student learning”. In: *Proceedings of the sixth international conference on learning analytics & knowledge*, pp. 104–113.
- Jaques, Natasha, Cristina Conati, Jason M Harley, and Roger Azevedo (2014). “Predicting affect from gaze data during interaction with an intelligent tutoring system”. In: *International conference on intelligent tutoring systems*. Springer, pp. 29–38.
- Lallé, Sébastien, Cristina Conati, and Roger Azevedo (2018). “Prediction of student achievement goals and emotion valence during interaction with pedagogical agents”. In: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 1222–1231.
- Lallé, Sébastien, Rohit Murali, Cristina Conati, and Roger Azevedo (2021). “Predicting Co-Occurring Emotions from Eye-Tracking and Interaction Data in MetaTutor”. In: *Proceedings of the 22nd Conference on Artificial Intelligence in Education*.
- Sims, Shane D and Cristina Conati (2020). “A neural architecture for detecting user confusion in eye-tracking data”. In: *Proceedings of the 2020 International Conference on Multimodal Interaction*, pp. 15–23.
- Woolf, Beverly, Winslow Burleson, Ivon Arroyo, Toby Dragon, David Cooper, and Rosalind Picard (2009). “Affect-aware tutors: recognising and responding to student affect”. In: *International Journal of Learning Technology* 4.3-4, pp. 129–164.
- Wortha, Franz, Roger Azevedo, Michelle Taub, and Susanne Narciss (2019). “Multiple negative emotions during learning with digital learning environments–Evidence on their detrimental effect on learning from two methodological approaches”. In: *Frontiers in psychology* 10, p. 2678.