

# 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 (Lallé et al. 2018; Salmeron-Majadas et al. 2014), single emotion (Jaques et al. 2014; Lallé et al. 2016; Paquette et al. 2014; Sabourin et al. 2011; Sims et al. 2020; Woolf et al. 2009) and co-occurring emotions (Lallé et al. 2021). We look specifically at the ITS, MetaTutor (Azevedo et al. 2013), that delivers content about the circulatory system via text and diagrams, and includes mechanisms to support Self-Regulated Learning (SRL). Prediction tasks with MetaTutor are done by collecting students’ eye-tracking and interaction data. During the interaction with Metatutor, users’ gaze data are tracked with an eye-tracker. Users are asked regularly to report if they felt any of the Pekrun’s emotions (Pekrun et al. 2014), by completing an Emotions and Value (EV) Questionnaire (Azevedo et al. 2013).

Jaques et al. 2014 works on predicting boredom and curiosity in a MetaTutor study using eye-tracking data. The eye-tracker used is the Tobii T60. The dataset (dataset A onward) has valid eye-tracking data for 270 EV reports and valid interaction data for 325 EV reports. Lallé et al. 2021 looks at predicting pairs of co-occurring emotions (boredom and frustration; curiosity and anxiety) using eye-tracking and interaction data with MetaTutor. This work was done as part of my RA project that started in Winter Term 1 2020. The dataset used (dataset A from now on) was the same as in Jaques et al. 2014. Our results show that eye-tracking and interaction data used in isolation can predict the co-occurrence of the two emotion pairs significantly better than a baseline classifier that predicts based on the probability distribution it sees in the training set. Further, classifiers trained on eye-tracking and interaction data individually complemented each-other depending on the target class. We also found that simple feature fusion of the two types of data did not improve performance over models trained on individual data sources.

Lallé et al. 2018 works on predicting emotion-valence using eye-tracking data on a different MetaTutor study. The eye-tracker used in this study is the SMI RED 250. This dataset (dataset B onward) has valid eye-tracking for 126 EV reports and valid interaction data for 176 EV reports. For this RPE project, we will first extend the task of predicting co-occurring emotions using eye-tracking and interaction data in Lallé et al. 2021 by combining dataset A and dataset B. We will then evaluate the efficacy of deep-learning and ensemble models for the task of predicting co-occurring emotions.

## 2 Proposed Project

The two user studies for dataset A and dataset B are structurally similar, so it makes sense to attempt to combine the two datasets, however, there are variations among the studies that make this task non-trivial. These include structural differences, such as the time intervals between EV reports and the total number of EVs asked per student, and measurement differences, such as the type of eye-tracker used. The first part of this RPE project involves identifying these differences in detail and coming up with a way to combine the two datasets. Combining the two datasets would give us valid eye-tracking data for 396 EV reports and valid interaction data for 501 EV reports.

With the larger dataset (*MetaTutor dataset* onward), the next step of the RPE project is to deploy machine-learning models to predict co-occurring emotions. Choosing the right pair(s) of co-occurring emotions for prediction tasks depends on the distribution of the emotions for the MetaTutor dataset, which may be

different from dataset A. We restrict our prediction tasks to pairs of co-occurring emotions as higher-order classification tasks might still be difficult considering the limited size of the dataset. Thus, we will look at 4-way classifications of an emotion pair just as in Lallé et al. 2021. We will use standard machine learning models such as random forest (RF), support-vector machine (SVM) and logistic regression (LR) classifiers as a baseline since they have been extensively used for affect detection (Zeng et al. 2008) and have been used for the task of predicting co-occurring emotions (Lallé et al. 2021). We will implement and evaluate the performance of a fully-connected neural network against these baseline classifiers. This would be a first look at deep-learning models to target the prediction of co-occurring emotions in ITS.

Using interaction data and eye-tracking data has shown promise in affect detection (Lallé et al. 2016, 2021). A limitation of (Lallé et al. 2021) was that even though classifiers trained on either eye-tracking and interaction data worked well individually and complemented each other depending on the target-class, combining the two datasets through feature-fusion did not improve performance. So we plan on investigating ensemble classifiers involving the standard machine-learning models trained on the different data sources, eye-tracking and interaction data. An ensemble model could leverage the fact that models trained on the two sources complement each other and have high class accuracies for different classes.

In Sims et al. 2020, user confusion is predicted using deep-learning models trained on raw sequences of eye-tracking data. Predictions are done in a different context of users interacting with ValueChart (Carenini et al. 2004). The work features a Recurrent Neural-Network (RNN) trained on raw eye-tracking sequences, a Convolutional Neural-Network (CNN) trained on scan-paths, and an ensemble model, VTnet, combining the CNN and the RNN. The VTnet architecture in Sims et al. 2020 works by connecting the outputs of the RNN and the CNN and forwarding them to a fully-connected neural network. They found that VTnet combines the strength of CNNs in spatial reasoning with the strength of RNNs in temporal reasoning and outperformed an existing RF model in literature (Lallé et al. 2016) in terms of accuracy. The last part of the project involves investigating the feasibility of the VTnet architecture used in Sims et al. 2020 on the MetaTutor dataset. This involves 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. This would be the first work that combines eye-tracking and interaction data in a deep-learning model for affect prediction.

### 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 fully-connected neural network and the ensemble models and to the baseline models.
- **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, Cristina Conati, and Giuseppe Carenini (2016). “Predicting Confusion in Information Visualization from Eye Tracking and Interaction Data.” In: *IJCAI*, pp. 2529–2535.
- 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*.
- Paquette, Luc, Ryan SJD Baker, Michael A Sao Pedro, Janice D Gobert, Lisa Rossi, Adam Nakama, and Zakka Kauffman-Rogoff (2014). “Sensor-free affect detection for a simulation-based science inquiry learning environment”. In: *International conference on intelligent tutoring systems*. Springer, pp. 1–10.
- Pekrun, Reinhard and Markus Bühner (2014). “Self-report measures of academic emotions”. In: *International handbook of emotions in education*. Routledge, pp. 571–589.
- Sabourin, Jennifer, Bradford Mott, and James C Lester (2011). “Modeling learner affect with theoretically grounded dynamic Bayesian networks”. In: *International Conference on Affective Computing and Intelligent Interaction*. Springer, pp. 286–295.
- Salmeron-Majadas, Sergio, Olga C Santos, and Jesus G Boticario (2014). “An evaluation of mouse and keyboard interaction indicators towards non-intrusive and low cost affective modeling in an educational context”. In: *Procedia Computer Science* 35, pp. 691–700.
- 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.
- Zeng, Zhihong, Maja Pantic, Glenn I Roisman, and Thomas S Huang (2008). “A survey of affect recognition methods: Audio, visual, and spontaneous expressions”. In: *IEEE transactions on pattern analysis and machine intelligence* 31.1, pp. 39–58.