Student programmer emotion detection using gaze, code, and face data collected via off-the-shelf equipment

Name: Mario Carreon
Laboratory's name: Ubiquitous Computing Systems Laboratory
Supervisor's name: Prof. Keiichi Yasumoto

Abstract

Teachers of Introduction to Programming (CS1) courses can detect student emotions through observation of their overall demeanor, allowing timely interventions for those who are struggling. However, such observation is challenging in large classes and even more so in online environments, where teachers have limited visual access to students. While emotion detection is possible using biometric sensors, such equipment is generally inaccessible to the average student. This dissertation addresses the following challenge: To what extent can off-the-shelf sensing equipment be used to detect programmer emotions? To explore this question, we developed an online integrated development environment (IDE) called MarioWindow++, which utilizes a standard webcam, ubiquitous in modern laptops, to capture video data for emotion detection. While student emotions can be inferred from their facial expressions, additional data streams can be extracted from the current setup. Gaze tracking can be extrapolated from the video feed. This type of software-based gaze tracking is not as accurate as dedicated hardware-based gaze tracking, it provides a low-cost means of providing more data from which to infer emotions. Furthermore, while not directly related to emotions, existing research has attempted to extrapolate student difficulty through how they write their code. As such, we now describe our second research challenge: does combining face, gaze, and code data in a multimodal approach enhance emotion recognition? From accuracies of 56.0%, 60.0%, and 59.3% for considering code, face, and gaze individually, multimodal model training combining code, face, and gaze datasets increased emotion detection accuracy to 66.4% with a precision of 72.7%, recall of 77.3%, and an F1-score of 74.9%, using an XGBoost classifier trained on 8 hours and 20 minutes of data from 26 participants. These results, gathered using only a webcam, are statistically similar to those obtained using dedicated hardware-based gaze tracking. Overall, the system proved generally effective in detecting student emotions during programming tasks. Its deployment may enable instructors to identify struggling students more accurately, offer timely support, and ultimately enhance learning outcomes.