Students’ Emotional Self-Labels for Personalized Models

Sinem Aslan[1], Eda Okur[1], Nese Alyuz[1], Sinem Emine Mete[2], Ece Oktay[3], Utku Genc[1], Asli Arslan Esme[1]

[1]Intel Corporation, USA
{sinem.aslan, eda.okur, nese.alyuz.civitci, utku.genc, asli.arslan.esme}@intel.com

[2]Bahcesehir University, Turkey
e.sinemmete@gmail.com

[3]Bogazici University, Turkey
oktayece@gmail.com

ABSTRACT

Although there are some implementations towards understanding students’ emotional states through automated systems with machine learning models, one of the key challenges still remain unaddressed: Generic detectors of emotions lack enough accuracy to autonomously and meaningfully trigger any interventions to infuse positive change in students. Collecting self-labels from students as they assess their internal states can be a way to collect labeled subject specific data necessary to obtain personalized emotional engagement models. In this paper, we outline preliminary analysis on emotional self-labels collected from students while using a 1:1 math learning platform.

CCS Concepts

• Human-centered computing~Empirical studies in HCI • Applied computing~Learning management systems

Keywords

Personalized emotional engagement; personalized learning; self-report; affective computing; Intelligent Tutoring Systems (ITS).

1. INTRODUCTION

One of the major goals of teachers is to create a nurturing environment facilitating positive emotions in learning. Leveraging this relationship, digital learning environments with artificial intelligence capacity (e.g., Intelligent Tutoring Systems - ITSs) have been studied for enabling personalized learning experiences by leveraging students’ emotions [1], [2]. Unfortunately, use of ITSs has generally been limited to cognitive goals of learning process [4]. Considering the important role of emotions in learning, ITSs need emotion-awareness capability [3], [5].

Despite efforts in emotion-aware ITSs, one major challenge is still unaddressed: Generic AI models of emotions lack enough accuracy to autonomously and meaningfully trigger any interventions for infusing positive change in students [6]. In [6], we show that models personalized to each individual using the corresponding labeled subject-specific data have high performance for emotional engagement detection. However, for online usage, these models require incoming subject-specific data to be labeled. To address this problem, we investigate the use of self-labels as self-reported measures of students’ emotional states.

2. METHODOLOGY OVERVIEW

2.1. Research Questions

There are three major research questions to address: (1) What is the distribution of emotional states as labeled by the human experts (i.e., ground-truth labels)? (2) What is the distribution of emotional self-labels as reported by the students? (3) What are the overlap ratios of emotional states between emotional self-labels as reported by the students and ground-truth labels?

2.2. Data Collection and Labeling

The data collection took place in 13 sessions (40 minutes each) of a Math Course with 17 students in 9th grade. The students used an online math platform: They watched instructional videos and solved related questions. Our data collection application running in the background, recorded the videos of the individual students through a camera (i.e., Intel® RealSense™ Camera F200) and captured students’ desktop screens. We had around 113 hours of student data to be labeled with respect to emotional states: *Satisfied*, *Bored*, and *Confused*. We employed the Human Expert Labeling Process (HELP) [7] to have the data labeled by five expert labelers with an undergraduate/postgraduate degree in Educational Psychology/Psychology. This process resulted in around 845 hours of total data labeling and about 169 hours of labeling per labeler.

2.3. Emotional Self-Labels

The data collection application collected real-time emotional self-labels from the students as self-reported measures of their emotional states. To set the groundwork and enable student cooperation on self-labels, we created a scenario for students (See Figure 1).

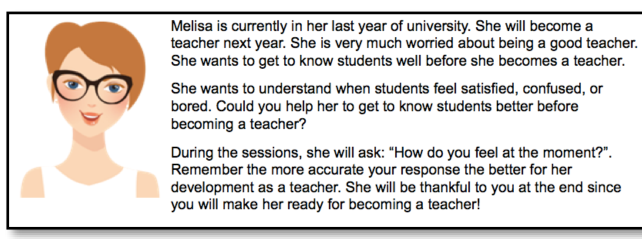

Figure 1. Scenario given to enable cooperative self-labeling.

After introduction of this scenario, we elaborated on the meaning of the three emotional states [7] with the help of the course teacher. As suggested by the course teacher, in the self-labeling, we used “Fine” as a replacement for the word: “Satisfied”. There were two methods we implemented to collect these self-labels: (1) Voluntary emotional self-labels: The students were able to provide an emotional self-label at any time using the window that stayed at the top right corner of the page (See Figure 2) and (2) mandatory emotional self-labels: The system asked the students to enter an emotional self-label at random intervals via a pop-up window.

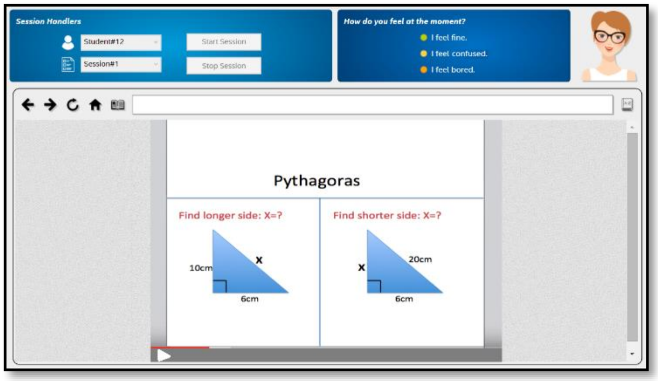

Figure 2. Visualization of the self-labeling interface.

3. PRELIMINARY ANALYSIS & RESULTS

RQ1: Distribution of Ground Truth Labels

The data were preprocessed to construct instances with a length of 8-seconds and an overlap of 4-seconds. Final instance-wise ground truth labels were then assigned by applying majority voting together with validity filtering. If there was no majority, “Can’t Decide” was assigned as the label (See [7]). The overall distributions for the final ground truth labels are given in Table 1.

Table 1. Distribution of the final ground truth labels

CLASS	INSTRUCTIONAL		ASSESSMENT		TOTAL	
	#instances	Perc (%)	#instances	Perc (%)	#instances	Perc (%)
Can’t Decide	6008	20.47	8109	27.94	14117	24.18
Satisfied	3599	12.26	8694	29.95	12293	21.06
Bored	19731	67.22	9387	32.34	29118	49.88
Confused	15	0.05	2838	9.78	2853	4.89
TOTAL	29353		29028		58381	

RQ2: Distribution of Self-Labels

The analysis of the self-labels showed that there is a major difference between the mandatory (See Figure 3(a)) and voluntary (See Figure 3(b)) self-labels in terms of emotional state distributions. When mandatory, “Satisfied” state was selected as often as “Bored” state (39%), and “Confused” state was selected less frequently (22%). However, when voluntary, students mostly selected “Bored” state (68%). Moreover, the overall state distribution for the ground truth labels is provided in Figure 3(c). The distribution of the overall ground truth labels is similar to the distribution of the emotional states for the voluntary self-labels.

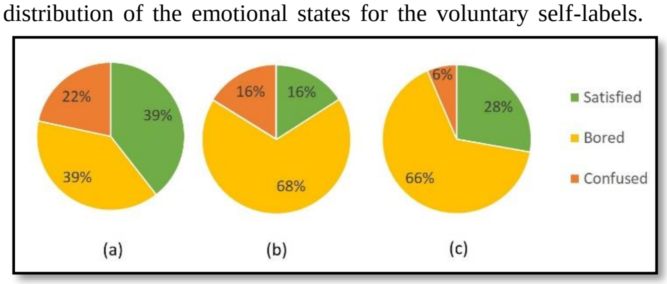

Figure 3. Emotional state distributions for (a) mandatory and (b) voluntary self-labels and (c) overall ground truths.

RQ3: Overlap Ratios

We compared the agreement between self-labels and the final ground truth labels using overlap ratios as the agreement measure: To calculate the overlap ratios, we compared self-labels assigned per instance (for previous N seconds) with the final expert labels assigned per instance (again for previous N seconds). This previous N seconds is the self-label span to be investigated. For our initial experiments, we considered 20-second-label span for self-labels (i.e., a given self-label is valid for the instances of the previous 20 seconds), an overall overlap ratio of 0.58 was obtained: For voluntary and mandatory self-labels, 0.65 and 0.46 ratios were achieved, respectively.

4. CONCLUSIONS AND FUTURE WORK

The preliminary results of this study showed that the collection approach of the self-labels impacted the emotional state distribution. This study also indicated that there was a relatively higher overlap ratio between voluntary self-labels and ground-truths. As a future work, we will conduct further statistical analysis on self-labels (e.g., different label spans, inter-rater agreement), and experiment on personalizing engagement models using these self-labels.

5. REFERENCES

[1] N. Bosch, S. D'Mello, R. Baker, J. Ocumpaugh, V. Shute, M. Ventura and W. Zhao, "Automatic detection of learning-centered afective states in the wild," in *Int. Conf. on Intelligent User Interfaces*, 2015.

[2] B. Woolf, W. Burleson, I. Arroyo, T. Dragon, D. Cooper and R. Picard, "Affect-aware tutors: recognising and responding to student affect," *Int. Journal of Learning Technology*, vol. 4, no. 3, pp. 129-164, 2009.

[3] R. S. Baker, S. K. D'Mello, M. M. T. Rodrigo, & A. C. Graesser, A. C. (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. International Journal of Human-Computer Studies, 68(4), 223-241.

[4] R. W. Picard, S. Papert, W. Bender, B. Blumberg, C. Breazeal, D. Cavallo, T. Machover, M. Resnick, D. Roy, and C. Strohecker. "Affective learning—a manifesto." *BT Technology Journal* 22, no. 4 (2004): 253.

[5] O. C. Santos, "Emotions and personality in adaptive e-learning systems: an affective computing perspective." In *Emotions and Personality in Personalized Services*, pp. 263-285. Springer International Publishing, 2016.

[6] N. Alyuz, E. Okur, E. Oktay, U. Genc, S. Aslan, S. E. Mete, D. Stanhill, B. Arnrich and A. A. Esme, "Towards an emotional engagement model: Can affective states of a learner be automatically detected in a 1:1 learning scenario?," in *ACM Conf. on User Modeling, Adaptation and Personalization (UMAP) - Workshops*, 2016.

[7] S. Aslan, S. E. Mete, E. Okur, E. Oktay, N. Alyuz, U. Genc, D. Stanhill and A. A. Esme, "Human Expert Labeling Process (HELP): Towards a reliable higher-order user state labeling by human experts," in *Int. Conf. on Intelligent Tutoring Systems (ITS) - Workshops*, 2016.