



# Towards Better Affect Detectors: Detecting Changes Rather Than States

Varun Mandalapu and Jiaqi Gong<sup>(✉)</sup>

University of Maryland, Baltimore County, Baltimore, MD 21050, USA  
{varunml, jgong}@umbc.edu

**Abstract.** Affect detection in educational systems has a promising future to help develop intervention strategies for improving student engagement. To improve the scalability, sensor-free affect detection that assesses students' affective states solely based on the interaction data between students and computer-based learning platforms has gained more and more attention. In this paper, we present our efforts to build our affect detectors to assess the affect changes instead of affect states. First, we developed an affect-change model to represent the transitions between the four affect states; boredom, frustration, confusion and engagement concentration with ASSISTments dataset. We then reorganized and relabeled the dataset to develop the affect-change detector. The data science platform (e.g., RapidMiner) was adopted to train and evaluate the detectors. The result showed significant improvements over previously reported models.

**Keywords:** Affect change · Affect states · Sensor-Free  
Educational data mining

## 1 Introduction

The intelligent tutoring systems have come to force in recent years, especially with the rise of affective computing that deals with the possibility of making computers to recognize human affective states in diverse ways [1, 2]. The relationship between the students' emotions and affective states and their academic performance, the college enrollment rate, and their choices of whether majored in STEM field has been revealed through learning data captured by these tutoring systems [3–6].

In this paper, we attempt to enhance sensor-free affect detection derived from the existing psychological studies. Previous affect detectors have focused on single affective states and used a clip slice of learning data captured by the learning system as a data sample for model training and testing. We focus on the changes of affective states, reorganize the dataset to emphasize the transitions among the affective states and verify whether the models of affect changes can produce a better predictive accuracy than those prior algorithms.

## 2 Dataset

This work adopts the dataset drawn from the ASSISTments learning platform to evaluate our proposed approach to detecting affective states. Most of the previous papers have mentioned the concern of the imbalance issue of the dataset and developed resampling methods to solve this problem [7, 8]. However, none of them illustrated the imbalance issue explicitly and did not provide much detail regarding the resampling methods. However, the resampling methods have a significant influence on the performance of machine learning algorithms adopted by previous affect detectors. Therefore, transparent detail regarding resampling methods is needed to increase the confidence of the affect detectors.

Since this paper focuses on the transitions among the four types of affective states, we also conducted statistical analysis of these transitions. Figure 1(a) represents the transition between affective states of students. Table 1 illustrates the statistical analysis of the students who did/not experienced affect changes. We examined the students who were always confused or bored, none of their number of the clips are longer than 3.

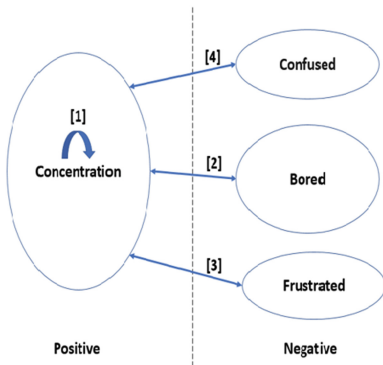


Figure 1 (a)

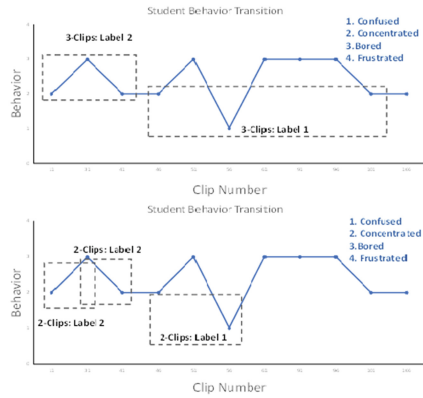


Figure 1 (b)

**Fig. 1.** (a) The affect-change model. (b) Illustration of the organizing and labeling process of 3-clip (up) and 2-clip (down) dataset. (student number: 4)

The analysis in Table 1 has two implications. First, we can simplify the models to four types of transitions; (Always Concentrated), (Concentration  $\diamond$  Confusion), (Concentration  $\diamond$  Bored), and (Concentration  $\diamond$  Frustration), which is equivalent to the trained models in previous work [7]. Second, down sampling the clips of the students who were always concentrated should be an excellent solution to solve the imbalance issue, because the clips without affect change should have reliable feature distribution that is not influenced significantly by the down sampling process. Based on these implications, we developed our method to build the affect detectors for detecting affect changes rather than states.

**Table 1.** The students who did/not experienced affect changes.

Affect change	Number of students	Percentage of students
Always concentrated	511	67.60%
Always bored	7	0.93%
Always confusion	1	0.13%
Always frustration	0	0%
Concentration $\diamond$ Confusion	58	7.67%
Concentration $\diamond$ Bored	126	16.66%
Concentration $\diamond$ Frustration	53	7.01%
Bored $\diamond$ Confusion	15	1.98%
Confusion $\diamond$ Frustration	7	0.93%

### 3 Methodology

According to our affect-change model, we reorganize and relabel the dataset, and then generate two types of the dataset with new format and labels. We adopt the RapidMiner [9] as the model training and testing platform, and test six training models; Logistic Regression, Decision Tree, Random Forest, SVM, Neural Nets, and AutoMLP. To conduct a fair comparison with previous work, we keep the settings for trained models as the same as previous work [7].

Based on the affect-change model, we developed two types of organizational strategies for the dataset, called “3-clips” and “2-clips” data format as shown in Fig. 1(b). It is noteworthy that we conduct a down sampling process on the clips that are always concentrated. The data organization with down sampling process solved the imbalance issue of the original dataset.

To facilitate the comparable experiments with previous work, we adopted the same data science platform, RapidMiner, as the tool for model training and testing. The selected models include Logistic Regression, Decision Tree, Random Forest, SVM, Neural Nets, and AutoMLP. All models are evaluated using 5-fold cross-validation, split at the student level to determine how the models perform for unseen students. These training and testing strategies were set up as the same as previous work [7].

### 4 Results

The evaluation measures for the results of each of our models include two statistics, AUC ROC/A' and Cohen's Kappa. Each Kappa uses a 0.5 rounding threshold. The best detector of each kind of affect change is identified through a trade-off between AUC and Kappa. The performance of efficient model is compared in Tables 2 and 3. In all the detectors, the models trained by SVM performed better than others both AUC and Kappa wise. There is no much difference between the raw data and average data. Most of the evaluation measures are close to each other. The only difference is that for 3-clip data, AutoMLP performs slightly better than neural nets. To save the computation complexity, we prefer to choose the average data to reduce the data dimension.

**Table 2.** Model performance for each individual affect label using the 2-clip dataset.

Affect change	Raw data			Average data		
	Models	AUC	Kappa	Models	AUC	Kappa
Always concentration	SVM	0.818	0.470	SVM	0.839	0.530
Concentration $\diamond$ Confusion	Neural Nets	0.674	0.162	Neural Nets	0.727	0.308
Concentration $\diamond$ Bored	SVM	0.826	0.482	SVM	0.836	0.498
Concentration $\diamond$ Frustration	Neural Nets	0.678	0.179	Neural Nets	0.687	0.137
Average		0.749	0.323		0.772	0.368
Wang et al. [7]		0.659	0.247			
Botelho et al. [8]		0.77	0.21			

**Table 3.** Model performance for each individual affect label using the 3-clip dataset.

Affect change	Raw data			Average data		
	Models	AUC	Kappa	Models	AUC	Kappa
Always concentration	SVM	0.817	0.435	AutoMLP	0.828	0.457
Concentration $\diamond$ Confusion	Neural Nets	0.651	0.100	Neural Nets	0.668	0.123
Concentration $\diamond$ Bored	SVM	0.789	0.402	SVM	0.761	0.409
Concentration $\diamond$ Frustration	Neural Nets	0.757	0.344	AutoMLP	0.708	0.258
Average		0.753	0.320		0.741	0.311
Wang et al. [7]		0.659	0.247			
Botelho et al. [8]		0.77	0.21			

## 5 Conclusion and Future Work

In this paper, we attempt to develop an affect-change model to build the relationship between the domain knowledge and the learning dataset previously studied using traditional feature engineering and machine learning algorithms. Our future work will include (1) developing models integrating semantic context to identify the affect states, (2) verifying and validating the trained models in population studies, such as Blackboard System that has collected plenty of interaction data at University of Maryland, Baltimore County, (3) combining sensor-free models and sensor-based models to develop more robust and flexible systems.

## References

1. Ma, W., Adesope, O.O., Nesbit, J.C., Liu, Q.: Intelligent tutoring systems and learning outcomes: a meta-analysis. *J. Educ. Psychol.* **106**(4), 901 (2014)
2. Steenbergen-Hu, S., Cooper, H.: A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *J. Educ. Psychol.* **106**(2), 331 (2014)

3. San Pedro, M.O.Z., Baker, R.S.J.d., Gowda, S.M., Heffernan, N.T.: Towards an understanding of affect and knowledge from student interaction with an intelligent tutoring system. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 41–50. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-39112-5\\_5](https://doi.org/10.1007/978-3-642-39112-5_5)
4. Pardos, Z.A., Baker, R.S.J.d., San Pedro, M.O.C.Z, Gowda, S.M., Gowda, S.M.: Affective states and state tests: investigating how affect throughout the school year predicts end of year learning outcomes. In: Proceedings of the Third International Conference on Learning Analytics and Knowledge, pp. 117–124. ACM (2013)
5. Pedro, M.O., Baker, R., Bowers, A., Heffernan, N.: Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In: Educational Data Mining 2013 (2013)
6. Pedro, S., Ofelia, M., Ocumpaugh, J., Baker, R.S., Heffernan, N.T.: Predicting STEM and Non-STEM college major enrollment from middle school interaction with mathematics educational software. In: EDM, pp. 276–279 (2014)
7. Wang, Y., Heffernan, N.T., Heffernan, C.: Towards better affect detectors: effect of missing skills, class features and common wrong answers. In: Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, pp. 31–35. ACM (2015)
8. Botelho, A.F., Baker, R.S., Heffernan, N.T.: Improving sensor-free affect detection using deep learning. In: André, E., Baker, R., Hu, X., Rodrigo, Ma.Mercedes T., du Boulay, B. (eds.) AIED 2017. LNCS (LNAI), vol. 10331, pp. 40–51. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-61425-0\\_4](https://doi.org/10.1007/978-3-319-61425-0_4)
9. Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T.: Yale: rapid prototyping for complex data mining tasks. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 935–940. ACM (2006)