Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Exploring Relationships Between Temporal Patterns of Affect and Student Learning

Anthony F. BOTELHO[a], Seth A. ADJEI[b], Vedant BAHEL[c], & Ryan S. BAKER[d]
[a]*University of Florida, USA*
[b]*Northern Kentucky University, USA*
[c]*University of British Columbia, Canada*
[d]*University of Pennsylvania, USA*
*a.botelho@ufl.edu

Abstract: Numerous prior articles have studied the relationships between student affect and various outcomes of learning. Prior research has found these relationships are complex; shifts in student affective experiences and the duration in which students remain in particular states form temporal patterns that are often difficult to interpret. Much of the existing research in this area focuses on the correlation between the overall prevalence of particular affective states and learning outcomes, ignoring the temporal and sequential characteristics of student affect. In this work, we leverage temporal clustering methods to identify emerging patterns of affect while students work within a computer-based learning platform to explore how sequence patterns of student concentration, boredom, confusion, and frustration correlate with student performance on a delayed assessment test. Similar to prior work, we find strong relationships between affect and student learning, even accounting for a measure of prior knowledge. Additionally, we identify that the directionality of these correlations in regard to specific affective states differs across clusters. While some affective states such as boredom are identified to exhibit negative relationships with learning within some patterns, a strong positive relationship between frustration and learning is found within one of the emerging pattern clusters.

Keywords: Affect, Temporal Clustering, Student Knowledge, Affect Dynamics

1. Introduction

There is a complex relationship between students' affective experiences and their learning. Affective transitions are highly complex and often do not seem to follow a specific interpretable sequence (Karumbaiah et al., 2021a), and there is also variation in the duration of individual affective experiences. Despite these known complexities, much of the research on the relationship between affect and learning looks for linear relationships between the prevalence of affective states with learning gains. This line of work has shown some clear relationships (see research synthesis in Karumbaiah et al., 2022). For instance, boredom is associated with poorer learning outcomes in several studies (e.g., Rodrigo et al., 2009; Bosch & D'Mello, 2017; Gong et al., 2019)) and engaged concentration is replicably associated with better learning outcomes in several studies (e.g., Rodrigo et al., 2009; Brom et al., 2014; Fancsali, 2014). However, confusion has inconsistent relationships with learning, varying considerably between studies (Karumbaiah et al., 2022). One empirical study suggested that the key to the inconsistent findings around learning's relationship to confusion may be the duration of these affective experiences. Liu and colleagues (Liu et al., 2013) found that extended confusion and frustration were associated with poor learning outcomes but that brief confusion and frustration were associated with positive learning outcomes.

While previous works have observed trends within and across affective states, the current work attempts to observe how patterns of affect relate to student learning as measured by a delayed measure of learning (described further in a later section). It is therefore important to further define what these patterns refer to and how this work may be distinguished from other related research on student affective patterns (Andres et al., 2019) as introduced in the previous section. To help illustrate this, consider the example sequence in Figure 1. Each of the lines represents a probabilistic estimate of one of four affective states (engaged concentration, boredom, confusion, and frustration) over the span of the assignment. In describing a sequence such as this one, prior works including D'Mello & Graesser (2011)

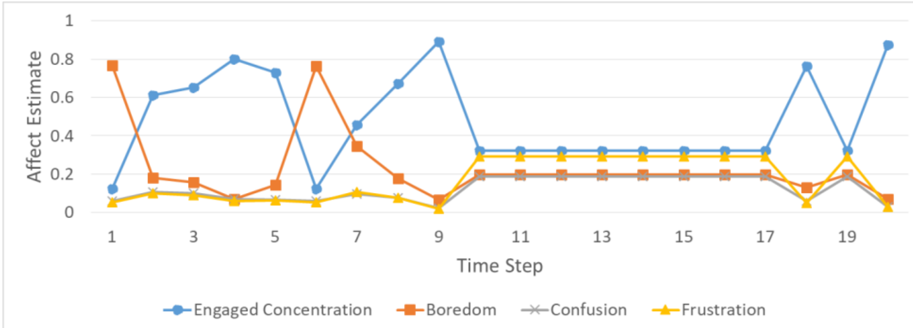

Figure 1. Example sequence of affect estimates for a student working on an assignment.

and Botelho et al. (2018) would have described affective chronometry, or periods of this sequence where the student remained in a single affective state for an extended time period. Alternatively, other works such as D'Mello & Graesser (2012) would have described the sequence's affect dynamics, characterized by the shifting between one affective state and another. Observing this sequence, however, it can be noted that there are few consistent trends; shifts and periods of affect appear to occur in bursts, with subsequent periods of uncertainty where there is arguably no clear singular state exhibited. The patterns exhibited within and across these affective states may provide novel insights into the relationships between affect dynamics and student learning.

Work on affect dynamics has not been extensively connected to the work on affect chronometry. Typically, chronometry is only considered in affect dynamics in terms of comparing brief (20 second or less) and extended periods (60-80 seconds) of the same affective state (Liu et al., 2014; Andres et al., 2019). Another limitation in the affect dynamics literature is the focus on relatively simple one-step models, which only look at single affective transitions (i.e. from state 1 to state 2). Affect appears to manifest in complex patterns over time, and our understanding of these is limited.

Temporal or sequence clustering is now a well-studied practice in research pertaining to pattern recognition (Liao, 2005). While specific implementations of time-series clustering methods vary, many rely on the generation of correlation-, spectral-, and wavelet coefficient-based features (Montero & Vilar, 2014) in combination with "warping" or scaling procedures (Berndt & Clifford, 1994) to describe the sequence holistically, before clustering based on these generated features.

There have been major advances in the detection of affect as well. Early work in affect detection within learning relied upon a complex array of sensors (Conati et al., 2003; D'Mello et al., 2007), utilized in laboratory settings, making it difficult to scale across long time periods and large selections of learners. Although more scalable sensor-free affect detection soon emerged, early models were only moderately better than chance (D'Mello et al., 2008; D'Mello & Graesser, 2012). Recent work leveraging deep learning algorithms has increased the quality of sensor-free affective detectors to the point where they can plausibly be used for fine-grained work (Botelho et al., 2017; Hutt et al., 2019), achieving AUC values approaching or equaling 0.8 for many affective states. Furthermore, even early affect detectors were more likely to obtain statistically significant findings about rare affective states than self-report or observational methods (Karumbaiah et al., 2022), possibly because they provided the opportunity to collect a much larger number of measurements, by 3 or more orders of magnitude.

In this paper, we leverage contemporary affect detectors to study affect dynamics and learning, specifically applying affect detectors developed using deep learning for the ASSISTments platform (Botelho et al., 2017). Specifically, we apply temporal clustering algorithms to investigate the patterns of student affect that emerge within student work on an assignment and examine whether affective states correlate differently with learning outcomes between the emerging patterns.

2. Methodology

The dataset used in this work is composed of interaction logs from students who used the ASSISTments learning platform (Heffernan & Heffernan, 2014) to complete mastery-based "skill builder" assignments during the 2016-17 through 2018-19 academic years. A skill builder is a problem set from which the system continues to present a series of problems to a student until they answer 3 consecutive

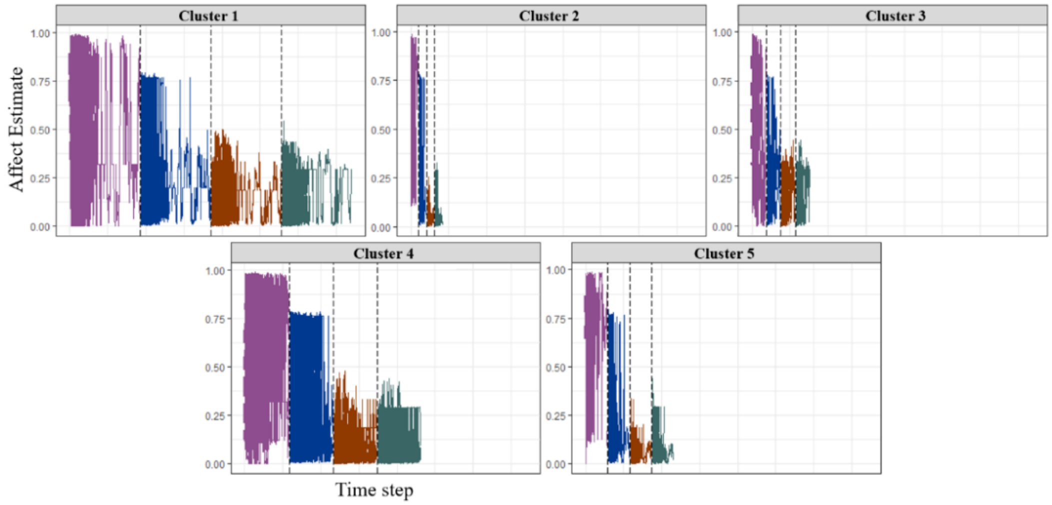

Figure 2. The sequences of affect estimates within each of the five affect pattern clusters.

problems correctly. The dataset used here contains 3065 skill builder assignments from 867 students; this data did not include students from rural settings, following prior works which have identified that the affect detectors used in this work have difficulty extrapolating to this population (Ocumpaugh et al., 2014; Karumbaiah et al., 2021b). We applied a deep-learning-based affect detector, pre-trained on historic ASSISTments data, developed and reported in Botelho et al. (2017), based on the collection of affect labels collected using BROMP classroom observations (Baker et al., 2020). The Botelho et al. (2017) model utilized in this work inherently assumes that students exhibit only one affective state at a time and produces a probability distribution over the 4 affective states.

To examine the relationships between patterns of student affect and learning, we utilize another feature of ASSISTments which provides a delayed learning outcome measure. The Automatic Reassessment and Relearning System (ARRS) is a feature of the ASSISTments ecosystem which can be coupled with skill builder assignments where teachers opt to have students reassessed on the knowledge component they had learned during the initial skill builder assignment (Xiong et al., 2013). This assessment has previously been used as a measure of student learning (Wang & Heffernan, 2015).

2.1 Patterns of Affect in Relation to Learning

To explore the relationship between patterns of affect and this measure of student learning, we apply a two-phase methodology. First, we cluster sequences of affect estimates and then we use those clusters within a regression analysis. Traditionally, clustering methods are used to group similar samples based on their proximity within a defined feature space. However, in this work we are observing sequences of student affect, where each time step in a sequence is defined by four probabilistic affect estimates.

In this work, we utilize the TSClust R package (Montero & Vilar, 2014) to generate the temporal features for our observed sequences of affect estimates before then applying a K-means clustering method based on a dynamic time warping distance measure (Ratanamahatana & Keogh, 2004; Lemire, 2009). This process involves first creating a matrix of samples as described by a vector of features before then using a simple elbow method to examine the ratio of within- and cross-sum of squared distances of samples at increasing numbers of clusters (i.e. the range of 2 to 15 clusters in our analysis). From this, we select K=5, and use a coding package-supplied method to plot these 5 emerging clusters as seen in Figure 2. These clusters represent different temporal patterns of student affect.

We build 5 separate logistic regression models, one for each cluster, to examine how each affective pattern correlates with our learning outcome. For each cluster, we generate a set of features to help describe the dynamics of each affective state within the cluster; these features include the natural log of the length of each sequence in addition to the mean and variance of detector estimates of each of three observed affective states (boredom, confusion, and frustration) composing each sequence; the fourth affective state, engaged concentration, is omitted due to a high inverse correlation and collinearity between these estimates and the other three affective states. Due to the large differences in

Table 1. *Coefficients for each cluster's regression predicting the one-week assessment performance. Boldface indicates statistical significance while marginally-significant values are denoted with* †.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Intercept	-0.214	0.143	-0.320	0.522	3.078
Prior Percent Correct	**2.521**	**3.140**	**3.221**	0.352	1.520
Log Sequence Length	**-0.457**	-0.236	**-0.428**	**-0.355**	-0.870
Mean Boredom	0.189	-0.538	**-0.559**	-0.242	-0.721
Boredom Variance	0.265	0.824†	-0.301	-0.310	-1.474
Mean Confusion	0.275	0.286	0.149	-1.002	-0.710
Confusion Variance	0.168	-0.390	0.005	-0.113	-1.117
Mean Frustration	-0.234	-0.837	0.299	**1.827**	0.110
Frustration Variance	0.221	-0.899	0.048	0.687	0.312
High Boredom X Low Boredom Var.	0.336	0.203	0.422	0.583	2.562†
High Confusion X Low Confusion Var.	-0.051	0.555	0.154	-0.199	1.307
High Frustration X Low Frustration Var.	-0.261	0.046	-0.430	-0.979	-0.541

scales of estimates across the different affective states, likely caused by the rarity of confusion and frustration, we use a median split to dichotomize the mean and variance values to represent high/low mean and high/low variance.

3. Results

Figure 2 shows the results of the temporal clustering analysis. Within each cluster, the chart displays 4 sections denoted by a vertical dotted line, ordered by engaged concentration, boredom, frustration, and confusion. Within each cell, the x-axis refers to the time step of the contained sequences while the y-axis represents the estimates of each affective state; the lines within each section of each cell represent an overlay of these estimates for each of the affective sequences found within the cluster. Clusters 1 and 4 show the longest sequences; this denotes that the students who spent the most time and attempted the most problems are contained within these clusters, but does not necessarily indicate that students remained in any particular affective state for a longer period compared to other clusters.

The 5 regression analyses are shown in Table 1. Across these regressions, there are several observations worth highlighting. First, the measure of prior knowledge (prior percent correct), reliably predicted assessment performance in Clusters 1, 2, and 3, but not in Clusters 4 and 5. The length of the sequences was a reliable (negative) predictor in Clusters 1, 3, 4, and 5, likely as an inverse measure of both initial knowledge and speed of learning within this mastery-based assignment.

4. Discussion and Future Works

Analyzing the relationship between affect and learning across five distinct temporal clusters, we find that boredom and frustration exhibit reliable relationships with learning, but only in some clusters. Within several of the clusters, boredom was negatively correlated with learning (cf. Karumbaiah et al., 2021a). Curiously, a relatively strong positive relationship between high frustration and student learning was found for Cluster 4, aligning to previous findings by Pardos et al. (2014); from our analysis, students with high frustration were twice as likely to answer the assessment item correctly. Distinctive from Cluster 1 where we see similarly-long sequence lengths, frustration emerges as a positive predictor of learning; in other clusters it is associated negatively or weakly with learning. The findings across clusters align with other prior work (Baker et al., 2010). The temporal clusters we have generated provide means to identify this "productive" frustration previously theorized in (D'Mello & Graesser 2012). It is also worth noting that this positive relationship between frustration and our learning outcome is found in the cluster with the lowest overall performance on the assessment. Cluster 4 appears to characterize students who are struggling to learn the material, where frustration is an indicator of the most successful of these students; perhaps for a struggling student, the only way to succeed is to persist through that struggle and the frustration it generates (e.g. D'Mello & Graesser 2011).

As the clusters were generated using multivariate, continuous-valued estimates of affect over time, it is worth highlighting that we see at most one affective state emerging as a reliable predictor of learning within each cluster. These significant states were rarely the highest estimate among the detector outputs; in other words, if one were to convert the estimates to discrete labels by simply taking the state corresponding with the highest estimate, subtle affective trends would not have been possible to identify. However, the lack of a discrete label makes it more difficult to interpret the patterns that emerged (cf. (Andres et al., 2019). Retaining full probability information capturing more complex relationships within and across affective states, but sacrifices a level of interpretability as a result.

We have, then, a trade-off between characterizing complex affective patterns and interpretability. Future work may overcome this by exploring better ways of describing the patterns within each cluster. We approached this initially by dichotomizing the means and variances of each affective state, but future work could consider additional measures of affective dynamics and chronometry, to better represent the properties being captured by the temporal clustering.

Another potential limitation of the current work is in the use of sensor-free detectors to measure affect. While the detectors have been shown to exhibit high performance in predicting human-recorded observation labels (Botelho et al., 2017), (further) improvements to these detectors could provide more accurate insights into patterns of student affect. It is also important to emphasize that the detector model used in this work was trained on a dataset collected between 2010 and 2012, as reported in Botelho et al. (2017). It is uncertain whether later changes to the system or user population might lead to lower detector reliability (cf. Levin et al., 2022). Future work should examine longitudinal and contextual factors that may affect detector performance, to identify and mitigate potential threats to validity.

Overall the findings of our analyses provide new insights into the relationship between the dynamics and chronometry of affect and student learning. The subtle complexities of affect, in a temporal sense, are not currently well-understood, but the methods presented in this work provide a means to better explore this. This method may be particularly useful for better understanding affective and behavioral patterns exhibited by students and eventually we may be able to use these clusters to identify what recommendations to give students (for instance, who should persist and who should seek help). In this fashion, these analyses may be a step towards building better supports for students that avoid negative interactions and experiences that may lead to poorer learning outcomes.

Acknowledgements

We would like to thank multiple agencies for supporting this work including the National Science Foundation (#1903304) and Schmidt Futures.

References

Andres, J. M. A. L., Ocumpaugh, J., Baker, R. S., Slater, S., Paquette, L., Jiang, Y., ... & Biswas, G. (2019, March). Affect sequences and learning in Betty's Brain. In *Proc. of the 9th Int'l Conf. on Learning Analytics & Knowledge* (pp. 383-390).

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & 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. *Int'l Journal of Human-Computer Studies, 68*(4), 223-241.

Baker, R. S., Ocumpaugh, J. L., & Andres, J. M. A. L. (2020). BROMP quantitative field observations: A review. Learning Science: Theory, Research, and Practice. New York, NY: McGraw-Hill.

Berndt, D. J., & Clifford, J. (1994, July). Using dynamic time warping to find patterns in time series. In *KDD workshop* (Vol. 10, No. 16, pp. 359-370).

Bosch, N., & D'Mello, S. (2017). The affective experience of novice computer programmers. *Int'l Journal of Artificial Intelligence in Education, 27*(1), 181-206.

Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017, June). Improving sensor-free affect detection using deep learning. In *Proc. of the 2017Int'l Conf. on Artificial Intelligence in Education* (pp. 40-51).

Botelho, A. F., Baker, R. S., Ocumpaugh, J., & Heffernan, N. T. (2018). Studying Affect Dynamics and Chronometry Using Sensor-Free Detectors. *In Proc. of the 11th Int'l Conf. on Educational Data Mining,* 157-166.

Brom, C., Buchtová, M., Šisler, V., Děchtěrenko, F., Palme, R., & Glenk, L. M. (2014). Flow, social interaction anxiety and salivary cortisol responses in serious games: A quasi-experimental study. *Computers & Education*, *79*, 69-100.

Conati, C., Chabbal, R., & Maclaren, H. (2003, June). A study on using biometric sensors for monitoring user emotions in educational games. In *Workshop on assessing and adapting to user attitudes and affect: Why, when and how*.

Baker, R. S., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A., ... & Rossi, L. (2012). Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. *In Proc. of the 5th Int'l Conf. on Educational Data Mining*, 126-133.

D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion*, *25*(7), 1299-1308.

D'Mello, S., Chipman, P., & Graesser, A. (2007). Posture as a predictor of learner's affective engagement. In *Proc. of the Annual Meeting of the Cognitive Science Society* (Vol. 29, No. 29).

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145-157.

D'mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., & Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User modeling and user-adapted interaction*, *18*(1), 45-80.

Fancsali, S. (2014, July). Causal discovery with models: behavior, affect, and learning in cognitive tutor algebra. In *Proc. of the 7th Int'l Conf. on Educational Data Mining*, 28-35.

Gong, L., Liu, Y., & Zhao, W. (2019, March). Dynamics of Emotional States and Their Relationship with Learning Outcomes during Learning Python with MOOC. In *Proc. of the 2019 7th Int'l Conf. on Information and Education Technology* (pp. 71-76).

Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *Int'l Journal of Artificial Intelligence in Education*, *24*(4), 470-497.

Hutt, S., Grafsgaard, J. F., & D'Mello, S. K. (2019, May). Time to scale: Generalizable affect detection for tens of thousands of students across an entire school year. In *Proc. of the 2019 CHI Conf. on human factors in computing systems* (pp. 1-14).

Judkins, D. R., & Porter, K. E. (2016). Robustness of ordinary least squares in randomized clinical trials. *Statistics in Medicine*, 35(11), 1763-1773.

Karumbaiah, S., Baker, R., Tao, Y., & Liu, Z. (2022, March). How does Students' Affect in Virtual Learning Relate to Their Outcomes? A Systematic Review Challenging the Positive-Negative Dichotomy. In *LAK22: 12th Int'l Learning Analytics and Knowledge Conf.* (pp. 24-33).

Karumbaiah, S., Baker, R. B., Ocumpaugh, J., & Andres, A. (2021a). A re-analysis and synthesis of data on affect dynamics in learning. *In IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2021.3086118.*

Karumbaiah, S., Lan, A., Nagpal, S., Baker, R. S., Botelho, A., & Heffernan, N. (2021b). Using Past Data to Warm Start Active Machine Learning: Does Context Matter?. In *LAK21: 11th Int'l Learning Analytics and Knowledge Conf.*,151-160.

Lemire, D. (2009). Faster retrieval with a two-pass dynamic-time-warping lower bound. *Pattern recognition*, *42*(9), 2169-2180.

Levin, N., Baker, R.S., Nasiar, N., Fancsali, S., Hutt, S. (2022) Evaluating Gaming Detector Models for Robustness Over Time. *In Proceedings of the 15th Int'l Conf. on Educational Data Mining*.

Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern recognition*, *38*(11), 1857-1874.

Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R. (2013, July). Sequences of frustration and confusion, and learning. *In Proc. of the 6th Int'l Conf. on Educational Data Mining*, *59-66*.

Montero, P., & Vilar, J. A. (2015). TSclust: An R package for time series clustering. *Journal of Statistical Software*, *62*, 1-43.

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for Educational Data Mining models: A case study in affect detection. *British Journal of Ed. Technology*, *45*(3), 487-501.

Pardos, Z.A., Baker, R.S., San Pedro, M.O.C.Z., Gowda, S.M., Gowda, S.M. (2014) Affective states and state tests: Investigating how affect and engagement during the school year predict end of year learning outcomes. *Journal of Learning Analytics, 1* (1), 107-128.

Ratanamahatana, C. A., & Keogh, E. (2004, April). Making time-series classification more accurate using learned constraints. In *Proc. of the 2004 SIAM Int'l Conf. on data mining* (pp. 11-22).

Rodrigo, M. M. T., Baker, R. S., Jadud, M. C., Amarra, A. C. M., Dy, T., Espejo-Lahoz, M. B. V., ... & Tabanao, E. S. (2009, July). Affective and behavioral predictors of novice programmer achievement. In *Proc. of the 14th Conf. on Innovation and technology in computer science education* (pp. 156-160).

Wang, Y., & Heffernan, N. T. (2014, June). The effect of automatic reassessment and relearning on assessing student long-term knowledge in mathematics. In *Int'l Conf. on Intelligent Tutoring Systems* (pp. 490-495).

Xiong, X., Li, S., & Beck, J. E. (2013, May). Will you get it right next week: Predict delayed performance in enhanced ITS mastery cycle. In *The 26th Int'l FLAIRS Conference.*