Student Engagement Profiling Using Automated Sensing and Unsupervised Clustering

Yu Fang

April 30, 2024

Abstract

The development of gaze tracking has been shown to facilitate students' learning [1–4] in controlled environments. There has been little evidence to support that gaze tracking can help identify distracted students in a traditional classroom setting. By utilizing Edusense's [5] gaze features throughout an entire course in 2019, we found there are 7 gaze-type clusters. While there are no clusters that correspond to distraction, we found gaze cluster outliers with respect to their in-class locations. These outliers happen sparsely throughout a class, and are perhaps a proxy for distracted students. Using this metric, we explored the two temporal factors that contributed to drop in average students' attention. First, average students' attention dropped steeply towards the end of class. Second, average students' attention dropped mid-way through the semester.

Introduction

The advent of machine learning technologies has ushered in a new era in education, particularly in classroom analytics. Our prior work showed that automated sensing output, including, but not limited to, student gaze movement in 3D collected via EduSense [5] correctly classified classroom activity (COPUS [6]). To further expand on the how students' gaze reflect their learning outcome, we decided to identify motifs of students gaze and to explore the possibility of using this as the proxy for student engagement.
Students' gaze has been utilized to evaluate learning [1–4]. However, most studies have focused on identifying where each student's gaze land on a screen (allocentric) [1, 2]. In the event where the study analyzes egocentric gaze direction [3], the analyses can only be done on one individual at a time (personal laptop webcam). Here, we utilize unsupervised clustering [1] on students' gaze features collected by Edusense [5] and hope to utilize additional classroom analytics such as student location to provide insight into student engagement.
By analyzing video data from 05391A course throughout a semester in 2019, we will reveal the amount of student engagement within and across sessions. Moreover, we seek to explain the features that comprises each engagement mode, detailing the composite gaze dynamics in attentiveness. Altogether, the two main research questions we are hoping to answer: How do students' different gaze directions show whether they pay attention in class? How does time impact students' attention within a class and from week to week.

Methods

Unsupervised K-means clustering validation

There are 7 total gaze features (SB_gaze_left, SB_gaze_right, $SB\_longest\_left$, $SB\_longest\_right$, SB_facing_back, SB_facing_front, SB_gazing_down) obtained from our sensing system. To prevent creating unnecessary combinations of gaze modes as well as protecting against occasional noise isolated in single features, we ran unsupervised k-means clustering to consolidate the gaze motifs [1]. We systematically explored different initial cluster values and evaluated both the euclidean distances and Silhouette scores (Fig. 1A-B).

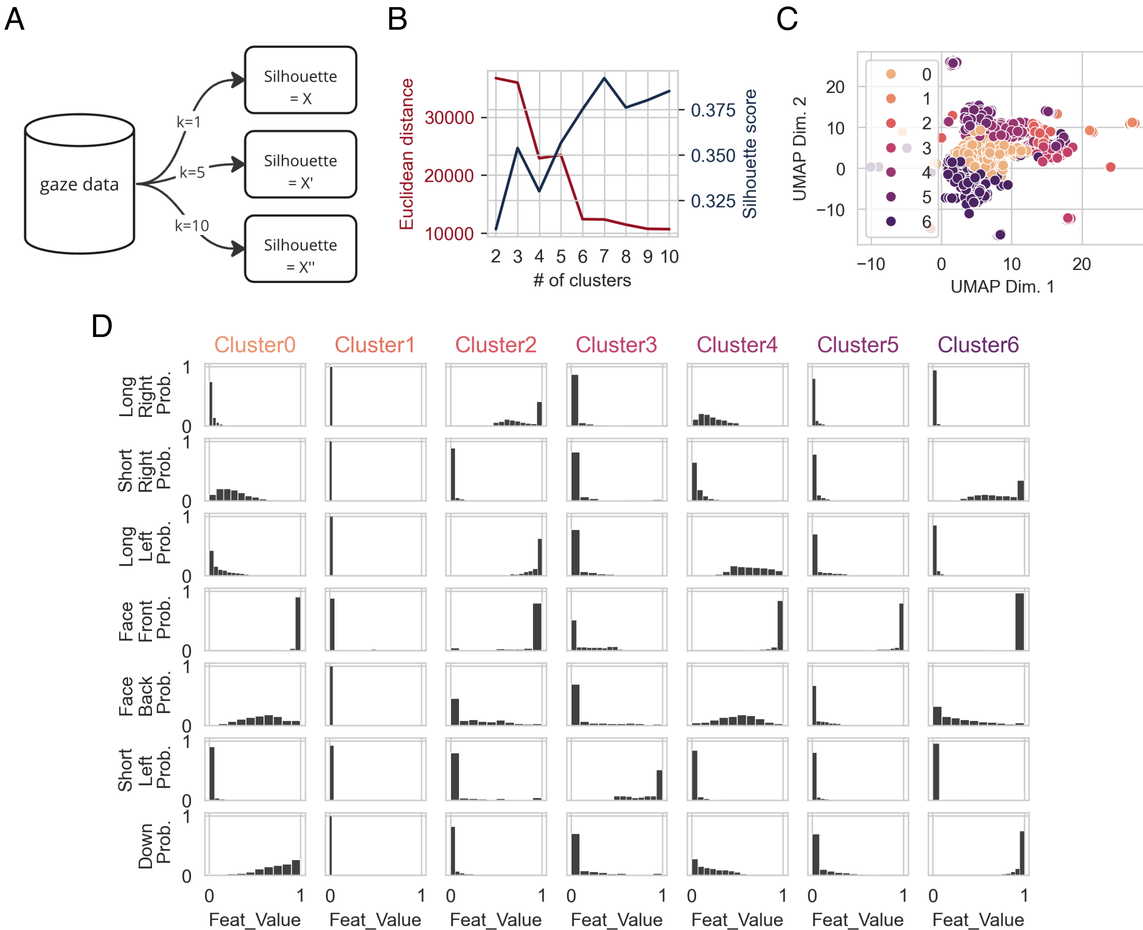

Figure 1: Unsupervised K-means clustering revealed 7 gaze motifs. A-B) Silhouette score versus the number of 'k' clusters. C) Low-dimensional representation of gaze features, colored by cluster identity. D) Feature distribution histograms for each cluster identity (columns). Note the cluster name and color matches the one in C).

In addition, to better visualize the unsupervised k-means clustering, we projected all 7 features down to a lower, 2-dimensional plane using Uniform Manifold Approximation and Projection (UMAP). This method would help us visually inspect the similarity within and between k-means clusters (Fig. 1C).

Feature weights defined by reproducible machine learning

Although the gaze motifs can be clearly defined based on the composite gaze features, it may be difficult to understand the main feature differences between clusters. Using a Random Forest Classifier, we can reliably predict each gaze motif using the 7 gaze features. To differentiate between clusters, the main feature differences are defined as the classifier feature importances.

Nearest neighbor outlier detection algorithm

For each student's (x, y) location, we identify the 5 nearest students and their gaze motifs. If none of the neighboring students consist of the reference student's gaze motif, then we would prescribe this student as an outlier, a proxy to disengaged students. In the rare instance (1 total) where the 5 closest neighbors are all outliers, we remove such reference student as an observation.

Results

Unsupervised clustering of student gaze

In our dataset, we have students' locations (x, y coordinate in the video frame) and their corresponding gaze features (SB_gaze_left, SB_gaze_right, $SB\_longest\_left$, $SB\_longest\_right$, SB_facing_back, SB_facing_front, SB_gazing_down). Although one could prescribe exactly 7 motifs based on the direction described in the feature names, there may also exist combinatorial motifs. Since we do not know if these exist, we decided to run unsupervised k-means clustering on different initial conditions (number of clusters). Using two metrics, euclidean distances and Silhouette scores, we found that, coincidentally, 7 clusters provided the best fit (Fig. 1A-B). In addition, when visually inspected, the cluster identities are quite compact and non-overlapping in a 2-dimensional representation UMAP 1C). To define each of the 7 identified clusters, we examined the each feature's distribution. Some of these clusters are only defined by one feature type, i.e. Cluster3 and Cluster 5 (Fig. 1D), while others appear to be a heterogeneous mix of features. This result show that some features are perhaps highly correlated. Moreover, we also identified a group of pure noise (Cluster1 - where all features' likelihood values are at 0, Fig. 1D).

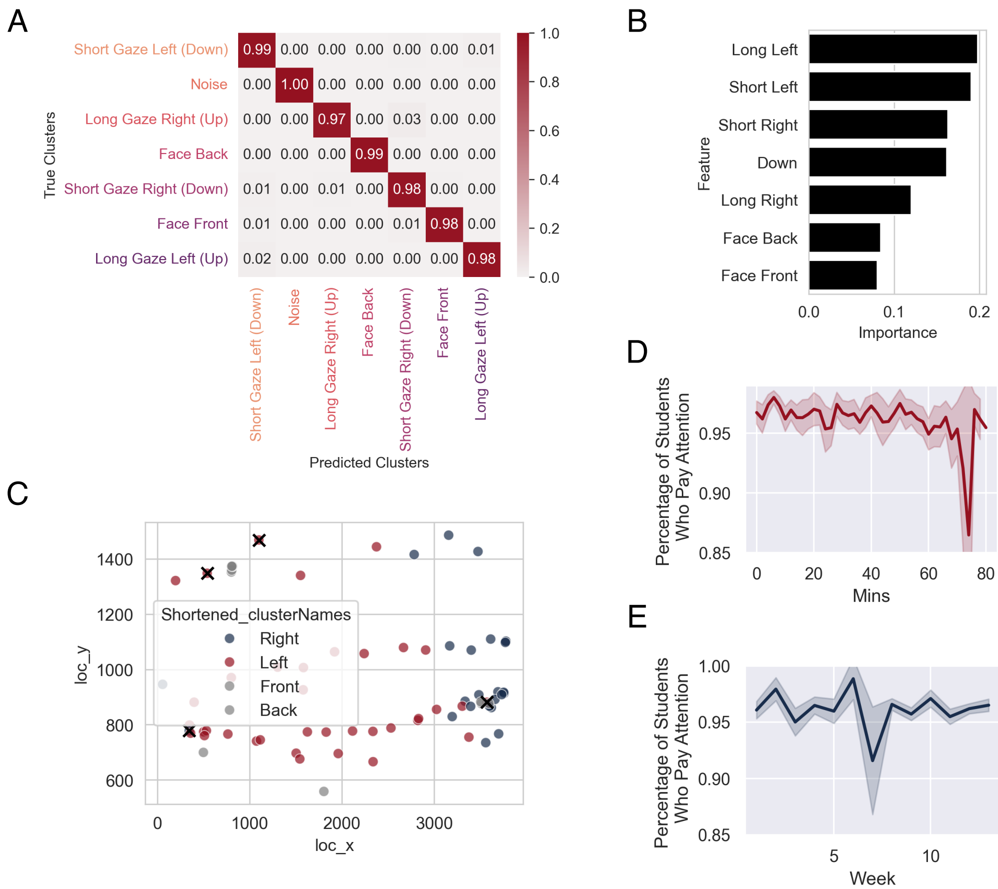

Figure 2: Student engagement profiling and the influence of in-class and semester time. A) Confusion matrix demonstrating 30% held-out test dataset that were successfully predicted using the remaining 70% training. B) Feature importance bar graph listing the most differentiable features, from top to bottom. C) Location scatterplot at 2-4 mins mark of an example session colored by cluster; 'X' marks outliers. D) Percentage of students who pay attention (mean with CI) over an 80-minute session in 05391A. E) Weekly percentage of students who pay attention (mean with CI) in 05391A.

Using classifier to reproduce clustering

Although visually inspecting and defining clusters may work, we leveraged supervised machine learning classifiers (Random Forest in our case) to identify the features that best separates these cluster identities. Upon splitting the gaze features and their corresponding k-means cluster identity into a train (70%) and a test (30%) set, we evaluated the reproducibility of such clustering with a classifier. The impeccable performance shown in the confusion matrix demonstrated there are clear rules that exist for each cluster identity (Fig. 2A). In other words, each cluster is distinct from the others. These rules are heavily impacted by 4 gaze features (SB_gaze_left, SB_gaze_right, $SB\_longest\_left$, SB_gazing_down, Fig. 2B). As shown in the top 4 gaze features, the 'left' and 'right' attributes are pivotal for distinguishing between different clusters. Therefore, we specifically focused on using 'left' and 'right' directions for student engagement (Fig. 2C).

The use of outliers to calculate mean students' attention

Now, we hope to explore the relationship between students' gaze patterns and their attention levels in the classroom. We hypothesize that a student who is attentive will exhibit gaze directions that align with those of at least one of their five nearest peers. Conversely, a deviation from this pattern may suggest that the student is not paying attention. After deploying such algorithm, we can visually mark the outliers who deviate from their neighbors with an 'X' (Fig. 2C). We then tally the number of students that are marked as an outlier and compute the ratio of outliers to all students as the percentage of attentive students. To explore the impact of in-class time, we plotted percentage of attentive students (mean with CI, n=19 classes) over the 80-minute class period and found a steep decrease after the one-hour mark (Fig. 2D). To evaluate the influence on number of semester weeks, we plotted percentage of attentive students (mean with CI, n=40 blocks) over the a 13-week semester and found a decrease midway through the semester, but came back up (Fig. 2E).

Conclusion

To conclude, our analysis shows that the employment of unsupervised clustering techniques can delineated seven distinct motifs of students' gaze. Using students' in-class locations and their respective gaze motifs, we can identify whether a student is paying attention by contrasting their gaze direction to their nearest peers. Furthermore, our study indicates a decline in students attention starting at an hour mark in an 80 minute session class in general. Across sessions, students' attention tend to drop in the midway of the semester. We only calculate the mean attention levels rather than assessing individual students to avoid potential misuse of the data. Specifically, we aim to prevent situations where results might be used to assign participation scores or to reprimand students based on perceived attentiveness. Instead, instructors or professors will receive only aggregated data, such as the average attention levels of students per session or across multiple sessions. This approach ensures that the focus remains on general trends rather than on individual performance, thereby safeguarding student privacy and promoting a more equitable educational environment. Altogether, these insights shows the complexity of students engagement during a class period which shows the potential of applying machine learning methods in educational settings to improve the learning environment for students and the teaching experiences.

Limitations

There are a couple limitations in the design of our study. First, the 5 closest neighbor parameter may require adjustment if this is wanted to be applied for another class (potentially with a percentage of students, rather than a set number). Second, this analysis did not need a workaround for situations where all 5 nearest students are outliers. We can imagine such a case becomes more prevalent than what we have seen. Therefore, further investigation is needed to ensure comprehensive data integrity and the robustness of the findings.

References

1. Gomes, J. S., Yassine, M., Worsley, M. & Blikstein, P. Analysing Engineering Expertise of High School Students Using Eye Tracking and Multimodal Learning Analytics. *Educational Data Mining* (2013).
2. Rahman, Y. *et al.* Exploring Eye Gaze Visualization Techniques for Identifying Distracted Students in Educational VR, 868–877 (June 2020).
3. Sharma, P. *et al.* Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning. *Communications in Computer and Information Science* **1720 CCIS,** 52–68. ISSN: 18650937. https://arxiv.org/abs/1909.12913v5 (Sept. 2019).
4. Wang, Y., Lu, S. & Harter, D. Multi-Sensor Eye-Tracking Systems and Tools for Capturing Student Attention and Understanding Engagement in Learning: A Review. *IEEE Sensors Journal* **21,** 22402–22413. ISSN: 15581748 (Oct. 2021).
5. Ahuja, K. *et al.* EduSense: Practical Classroom Sensing at Scale. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol* **3,** 71. https://doi.org/10.1145/3351229 (2019).
6. Smith, M. K., Jones, F. H., Gilbert, S. L. & Wieman, C. E. The classroom observation protocol for undergraduate stem (COPUS): A new instrument to characterize university STEM classroom practices. *CBE Life Sciences Education* **12,** 618–627. ISSN: 19317913. https://www.lifescied.org/doi/10.1187/cbe.13-08-0154 (Dec. 2013).

Signature: ______________________ E-Mail: ______________________