Teaching style differences explained by a machine learning model

Yu Fang

December 4, 2023

Abstract

With the recent development of automated classroom analytics using machine learning - Edusense [1], there is still little evidence on the impact of how this type of education technology can be useful to improve personalized learning. There are, however, studies supporting the benefits of student-centered learning over traditional lecturing [2–4], albeit that the students' perception of learning were anti-correlated with their learning outcome [4]. Beyond their self-evaluation, students tend to negatively review instructors of harder courses on RateMyProfessor [5], subsequently discouraging fellow students from taking those classes. Although prior studies have shown the benefit from providing more in-class participation, not all students could be willing to participate, due to personality differences [3].

To address these nuances, our study employs a classifier trained to predict Classroom Observation Protocol for Undergraduate STEM (COPUS) [6]. Our goal is to move beyond the simple dichotomy of active versus passive learning and gain a deeper understanding of classroom dynamics. First, we improved the model generalizability by incorporating a few sessions from the new classroom. Second, we found that the synchrony across COPUS activities were different in the new classroom 05391A. Lastly, based on the 4 main categories: Presenting, Guiding, Administration and Other [6], we easily identify the 3 sessions that contained an increased ratio of guiding. Through this analysis, we described teaching style differences through the lens of a machine learning model.

Introduction

Traditional lecturing is increasingly being replaced by student-centered methods due to its effectiveness in engaging students to regurgitate information. Additionally, there exist multiple forms of student engagement [7], and the ability to capture these will better characterize learning. The advent of machine learning technologies has ushered in a new era in education, particularly in classroom analytics. Therefore, we propose to utilize the Classroom Observation Protocol for Undergraduate STEM (COPUS) to objectively characterize classroom activity. The research question we want to ask is how learning can be defined using a more comprehensive description of classroom activity, i.e. time in each COPUS.

Preliminary analyses showed that classroom sensing data collected via EduSense [1] correctly classified classroom activity. Although each activity could have an unbalanced ground truth, we synthetically resampled the training data using Synthetic Minority Oversampling Technique [8] to solve the issue. By transforming sensing data into classroom activity predictions, we can begin to characterize teaching styles. Moreover, a longitudinal classroom study could help reveal teaching trends that evolve from beginning to the end of a semester. Overall, we have explored the utility of using a machine learning model to monitor/summarize teaching styles both across courses and within a course. The question we seek to answer in this study: Can we describe the teaching styles for different classes and/or different sessions of the same class?

Results

Machine learning model learned a different classroom design

The primary objective was to ascertain the model's effectiveness in analyzing classroom videos and accurately deducing teaching practices. The model was originally trained using data from courses that predominantly followed a lecture-based format, such as courses numbered 36-200 Reasoning with Data, 15-251 Great Theoretical Ideas in Computer Science, 21-122 Integration and Approximation, and 73-102 Principles of Microeconomics. This approach, however, differed from the teaching style of Professor Harrison in the course '05-391 Designing Human-Centered Software,' which is known for its interactive and dynamic classroom environment. To adapt the model to this varied teaching context, we integrated data from five previously recorded sessions of Professor Harrison's course into the training dataset. We annotated these sessions to ensure the accuracy and relevance of the data.

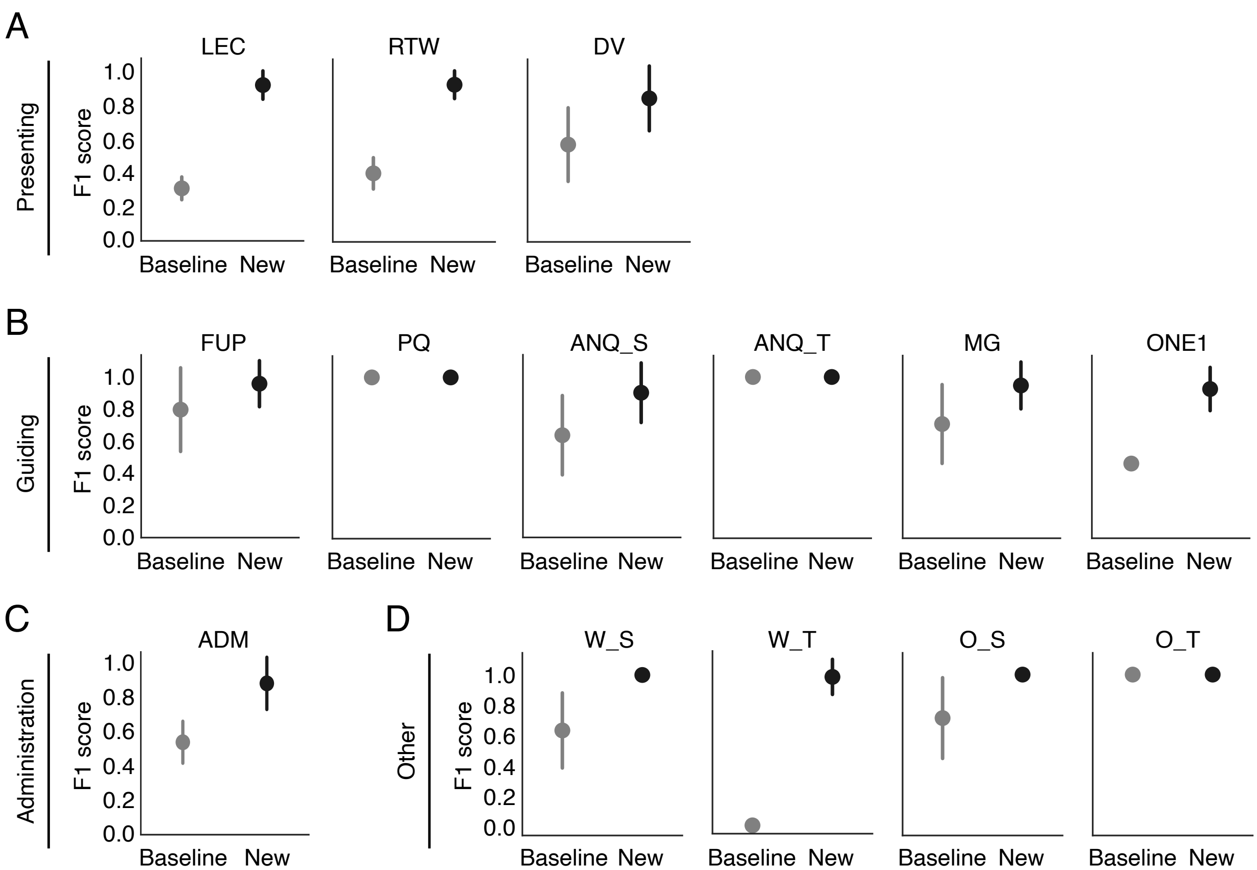

Figure 1: COPUSense performance on a new classroom. Mean ± standard deviation of performance on 20 randomly partitioned held-out new classroom data for each of the four main categories (A: Presenting; B: Guiding; C: Administration; D: Other). New is after incorporating 5 sessions from the new classroom - 407, whereas baseline is before.

To evaluate the improvement in performance, we randomly isolated 20% of data from annotated 05391A for held-out test set. We then compared the prediction performance for each activity using either the original model (containing only lecture-based formats) versus the updated model (one that incorporated the remaining 80% of data from 05391A). Upon cross-validation over 20 random seeds, we found that all activity improved significantly (Fig. 1). This improvement shows the potential of the model to be continuously generalizable to other classrooms.

COPUS synchrony revealed teaching style differences

After running the updated model on lecture-based and the new classroom (05391A), we found interesting differences in pairs of COPUS activities. By definition, most of lecture time (LEC) points will coincide with instructor writing on board (RTW) (Fig. 2A), but almost never when the instructor is waiting (W_T) (Fig. 2B). We can compute such pairwise correlation with all the COPUS activity to characterize overall styles.

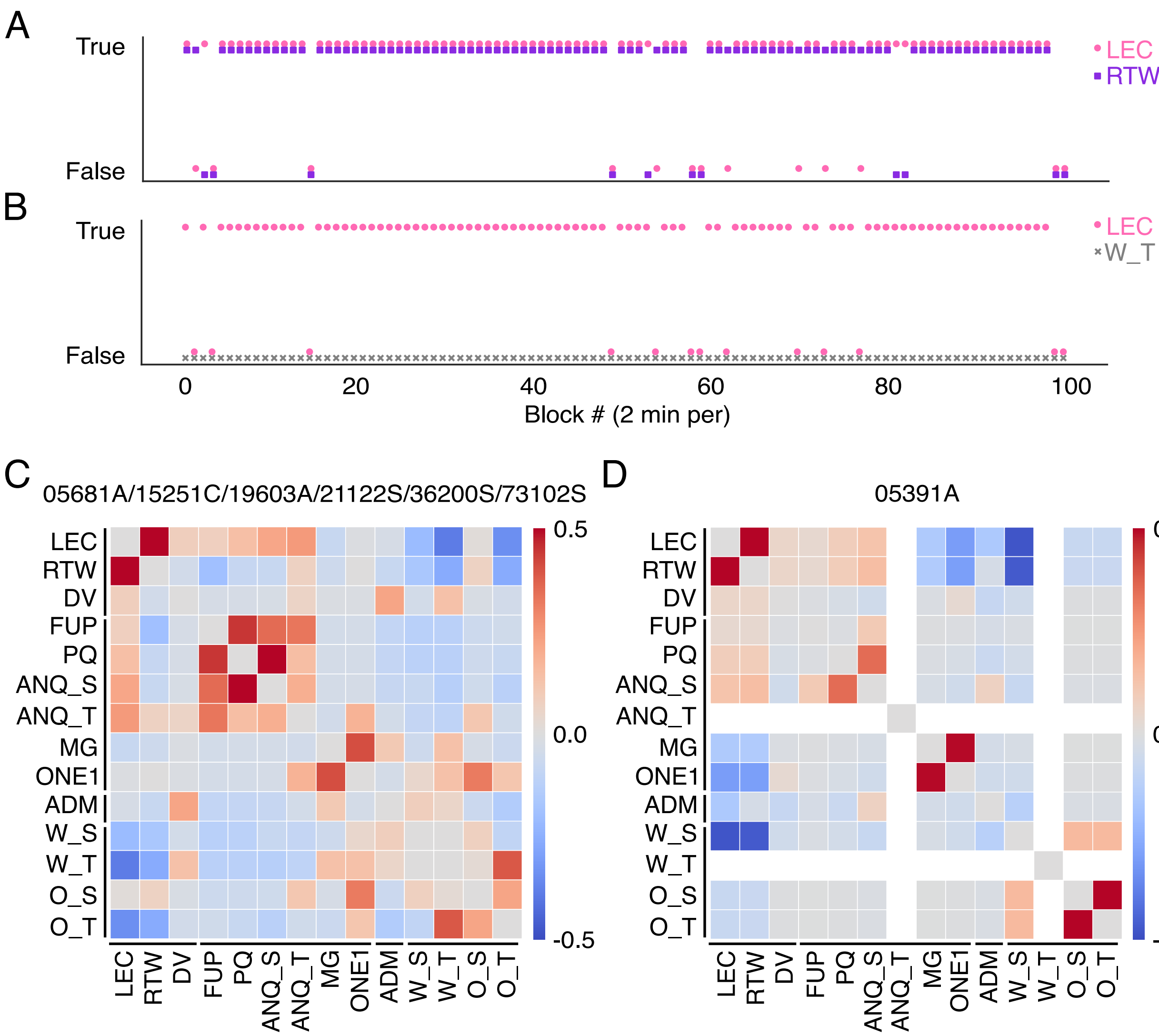

Figure 2: Teaching style differences demonstrated by inter-COPUS activity correlation. A) An 100-block example of predicted occurrences of LEC (lecturing) and RTW (real-time writing). B) The same 100-block example of predicted occurences of LEC (lecturing) and W_T (intructor waiting). C) Heatmap of pair-wise inter-COPUS activity correlation for classrooms 05681A, 15251C, 19603A, 21122S, 36200S, and 73102S. D) Heatmap of pairwise inter-COPUS activity correlation for classroom 05391A.

We found that in lecture-based classes (05681A, 15251C, 19603A, 21122S, 36200S, and 73102S), there appears to be higher correlation amongst individual guiding exercises, the likes of follow-up questions (FUP), posting clicker questions (PQ), students asking questions (ANQ_S), and instructor asking questions (ANQ_T) (Fig. 2C). On the other hand, that correlation structure is minimal and only appears between PQ and ANQ_S (Fig. 2D) in 05391A.

Moreover, in 05391A the presenting COPUSes (lecturing (LEC) and instructor writing on board (RTW)) were negatively correlated with the occurrences of a couple guiding exercises (moving through class guiding students during active tasks (MG) and one-on-one discussions (ONE1)). However, that does not be the case for the other classes, aligning with the lecture-based format.

Teaching style evolves throughout 05391A

Our technology can also study how classes that focus on different concepts may represent on a COPUS level. It is believed that each session will include a varying amount of presenting versus guiding simply due to the nature of the materials. We examined 05391A throughout the entire 2019 Spring semester. For each of the four main COPUS categories (presenting, guiding, administration, and other), we compute the percentage of occurrence averaged across all activity that belonged to that category (LEC, RTW, DV belong to presenting, for example). To better understand the relative amount, we divide each category's mean percentage of occurrence to the total (summation of all four categories). As expected, the ratio between these four main COPUS categories are dynamic, and easily identified sessions 5, 16, and 17 contain more guiding activities than the rest (Fig. 3A).

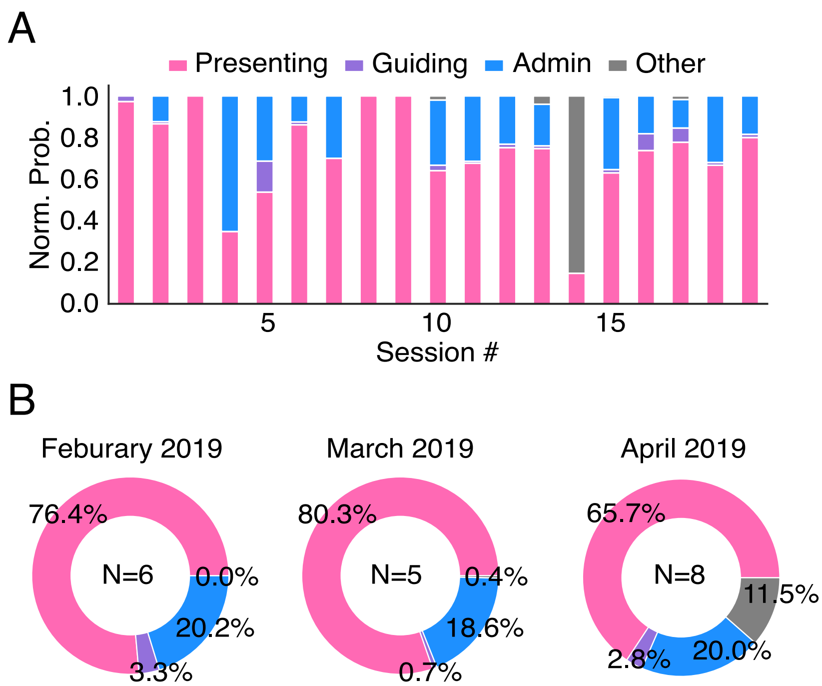

Figure 3: Utility of COPUSense in a longitudinal 05391A classroom study. A) Stacked bar plot with normalized activity ratio for each session. B) Pie chart of normalized activity ratio binned by each month.

Additionally, when we grouped by beginning, middle, and late semester (February 2019, March 2019, and April 2019), we found an interesting progression. During the middle part of the semester, the instructor seems to spend more time presenting, and less time guiding the students through active exercises. It can be due to a plethora of reasons, including, but not limited to, instructor becoming less engaged, students attendance dropped, or simply due to course material being less practical to be guided. Overall, we gained valuable insights into instructor teaching styles through our machine learning model.

Discussion

Before we updated our model, we observed some inconsistencies in its output, which could be attributed to a variety of factors, including classroom design. To solve that issue, we annotated a fraction of the data from the new classroom and found a significant improvement in model performance on the new classroom.

Additionally, when annotating, we observed that some COPUS activity seems to always coincide with another. This observation was particularly prominent in Instructor Lecturing (Lec), Instructor Real-time writing on board (RTW). To that end, we performed pairwise correlation for all COPUS variables and discovered different correlation structures between lecture-based classes and the new classroom that teaches 05391A. We found that the negative correlation between one-on-one extended conversation with a few students (ONE1) and LEC could reveal that this course (05391A) had more of these ONE1 activity than the other lecture-based courses, which aligned with what was known already. However, what was surprising was there was minimal correlation between clicker questions and follow-up question in 05391A compared to the other classes, potentially revealing the nature of the clicker questions.

Lastly, a longitudinal study on 05391A revealed a drop off in guiding exercises in March 2019. If guiding exercises do indeed improve student's class performance, this could be an avenue of improvement, selectively focusing on including more in-class guiding exercises in March.

Recommendations

In future developments of our study on smart classroom technology, it is essential to adapt the interface to better suit different user groups. Currently, the interface is highly detailed, catering effectively to instructors but potentially posing complexity challenges for students. A streamlined version of the interface for students would enhance their ease of use and interaction. Additionally, for administrative staff, the interface can be optimized to facilitate tasks such as assigning classrooms to professors. This could include analyzing factors like class timing and its impact on faculty course evaluations — noting, for instance, that classes scheduled at 8 am might result in lower evaluations. The interface could also incorporate analysis of classroom occupancy, comparing the actual number of students present with those registered, and examining how teaching methods might vary in settings where small groups are placed in large rooms. These insights would be instrumental in aiding registrars in making more objective and efficient room assignments. By tailoring the interface to meet the specific needs of each user group, the utility and applicability of smart classroom technologies can be significantly enhanced.

Limitations

One of the primary limitations in our study is the outdated nature of the training and testing sessions. The most recent data we have at our disposal is from 2019. This presents a significant challenge, as there have likely been considerable changes in teaching methods, classroom dynamics, and technological advancements in the field of education since then. The absence of up-to-date data could hinder the model's ability to accurately mirror the current state of classroom environments and emerging educational trends. Furthermore, the quantity of training data available for the model raises concerns. In the case of the original EduSense model, the training dataset comprises only 12 sessions. Additionally, for the specific analysis of Professor Harrison's class, we have a mere 5 sessions to work with. This limited amount of data may not sufficiently cover the diverse range of classroom interactions and scenarios, making it challenging to train the model effectively. Such a constraint in the dataset can adversely affect the model's learning capabilities, potentially leading to inaccuracies and limiting the generalizability of its findings across various classroom contexts and teaching styles. Addressing these limitations is crucial for enhancing the model's performance and applicability in contemporary educational settings.

References

1. 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).
2. Freeman, S. *et al.* Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 8410–8415. ISSN: 10916490. `https://www.pnas.org/doi/abs/10.1073/pnas.1319030111` (June 2014).
3. Rocca, K. A. Student Participation in the College Classroom: An Extended Multidisciplinary Literature Review. **59,** 185–213. ISSN: 1479-5795 (2010).

4. Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K. & Kestin, G. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences of the United States of America* **116,** 19251–19257. ISSN: 10916490. https://www.pnas.org/doi/abs/10.1073/pnas.1821936116 (Sept. 2019).

5. Gao, B. J. & Katrompas, A. A Preliminary Experimental Analysis on RateMyProfessors. *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020,* 5679–5681 (Dec. 2020).

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).

7. Fredricks, J. A., Blumenfeld, P. C. & Paris, A. H. School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research* **74,** 59–109. ISSN: 00346543 (2004).

8. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority Oversampling Technique. *Journal of Artificial Intelligence Research* **16,** 321–357. ISSN: 1076-9757. https://www.jair.org/index.php/jair/article/view/10302 (June 2002).

Signature: Amy Ogan E-Mail: aeo@cs.cmu.edu