Shih, JL. et al. (Eds.) (2023). Proceedings of the 31st International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Evaluating the Assessment of Comment Quality in Learning Communication Skills using Active Video Watching

Raul Vincent LUMAPAS[a*], Antonija MITROVIC[a], Matthias GALSTER[a], Sanna MALINEN[a], Jay HOLLAND[a], & Negar MOHAMMADHASSAN[a]
[a]*University of Canterbury, New Zealand*
*raulvincent.lumapas@pg.canterbury.ac.nz

Abstract: Supporting student engagement remains one of the key challenges in video-based learning. This challenge is addressed by active video watching (AVW), a learning approach that supports engagement through different interventions, such as note-taking in the form of comments that learners submit while watching videos. One platform to support AVW is AVW-Space. Previous studies on AVW-Space detail improvements in the system, such as the integration of Artificial Intelligence and Machine Learning (ML) models in the comment feature of the system. This study investigates two machine learning models used to automatically assess the quality of comments when learning communication skills via AVW. One model is generated based on a large set of comments created by students when engaging with videos about presentation skills. For this study, a new model is developed from comments that students submitted when engaging with videos about communication skills. Results show that the new model, which was created from data on communication skills, performed better when assessing comments for communication skills compared to the model generated from comments for another skill. This has been demonstrated by the higher value of inter-rater agreement with the comment quality assessment made by human coders.

Keywords: Video-based Learning, AVW-Space, Comment Quality

1. Introduction

Video-based learning increases students' motivation to learn, provides flexibility and opportunities for self-regulated learning (Chatti et al., 2016; Dimitrova & Mitrovic, 2022). However, there are still challenges in video-based learning, such as the lack of interaction between learners and teachers, resulting in passive learning (Yousef et al., 2014) and a low level of engagement (Dimitrova & Mitrovic, 2022). This need to support student engagement is addressed in active video watching (AVW). A platform to support AVW is AVW-Space developed at the University of Canterbury (Mitrovic et al., 2016; Mitrovic, Dimitrova, et al., 2017). In AVW-Space, teachers create spaces based on topics containing videos and activities. Student watch videos and write comments on the videos uploaded in the platform. AVW-Space provides a multitude of features that help in learning, such as personalized prompts called Nudges and Visualizations.

To increase engagement, Mohammadhassan et al. (2020) developed nudges as personalized reminders to students who write no comments or write only low-quality comments. This particular improvement uses Artificial Intelligence (AI) in assessing the quality of comments. Comments were classified into five categories: (1) Affirmative, negative, or off-topic, (2) Repeating, (3) Critical and analytical, (4) Self-reflective and (5) Self-regulating comments. Category 1 and 2 comments are pedagogically undesirable and are classified as low-quality comments as these comments do not convey deep thinking. On the other hand, categories 3 – 5 are considered high-quality comments as learners elaborate on the video, reflect, and relate the video content to previous experience. In the same study by

Mohammadhassan et al., a random-forest classifier was trained to assess comment quality. After selecting the best machine learning model for assessing the comment quality, the best-performing combination of the comment categories was identified, retaining categories 1 and 2, and combining categories 3-5 into category 3. This improvement enhanced student engagement, resulting in an increase in comments and the number of high-quality comments. Despite these improvements, Mohammadhassan et al. suggested that there is still room for improving the automatic assessment of comment quality. This suggestion is then explored in our study.

2. Methodology and Results

The above-mentioned machine learning model was trained using comments submitted by students in ENGR101, a first-year general introductory engineering course at the University of Canterbury (UC). This course used the online resources in AVW-Space for training presentation skills. The model was trained using the comments collected from the 2017-2019 ENGR101 cohorts. The same model was then also used for the automatic assessment of comment quality in another course and for another skill: SENG202, a course in the second year for Software Engineering students, that uses AVW to teach communication skills.

As the current ML model was generated using comments submitted when learning presentation skills, there is a need to assess its performance when assessing comments submitted when leaning communication skills. To evaluate this, we analyzed comments from the most recent instances of SENG202 (from 2020 to 2022) and the classifications generated by the current model. Using the comments collected during this period (159, 691, and 688 comments for 2020, 2021, and 2022 respectively), we trained a new machine learning model specifically for communication skills. We used the testing set to analyze the performance of the current model used in AVW-Space and the new model trained using the data set generated from comments on communication skills. Two human coders have also manually assessed the quality of the comments in the testing dataset.

Table 1. Distribution of Individual Ratings by the models and the human coders

Comment Quality	Rater 1	Rater 2	Current Model	New Model
Category 1	3	5	1	2
Category 2	148	118	240	176
Category 3 – 5	157	185	67	130

Of the total of 1,529 comments, 80% (1,231 comments) were used to train the model. The remaining 20% (308 comments) were used as the test set. The number of comments used for training and testing the new model is smaller than what was used in the current model (Total: 2,343. Training: 1796 & Testing: 447). As shown in Table 1, it is noticeable that both models classified most comments as Repeating comments (Category 2), a form of low-quality comment. Both human coders had a higher rate for Category 3 as compared to the machine learning models.

Table 2. Results of Analyzing Inter-rater Agreement between the models and the human coders

Results	Current Model	New Model
Krippendorff's Alpha	0.442	**0.623**
Average Pairwise Cohen's Kappa	0.478	0.627
Pairwise Cohen's Kappa - System & Rater 1	0.295	0.557
Pairwise Cohen's Kappa - System & Rater 2	0.407	0.593
Pairwise Cohen's Kappa - Rater 1 & Rater 2	0.732	0.732

Table 2 shows the inter-rater agreement between the human coders and the ML models. It can be observed that human coders agree more often, having a Cohen's Kappa of 0.732, a substantial level of agreement. The results show that the human coders do not often agree with the current model. However, there is an increase in the agreement between the human coders and the new model. Despite falling short of the acceptable minimum Krippendorff's Alpha value ($\alpha>0.66$) (Krippendorff, 2010), the new model still significantly outperforms the current model in classifying comments, even though that the new model was trained using a smaller number of comments.

3. Conclusions

In this research, we assessed two machine learning models to assess the quality of comments submitted when engaging with videos to learn communication skills. The new model was trained with comments from SENG202 classes from 2020 to 2022. The classifications generated by the current and new models were compared to those produced by two human coders. The result shows higher agreement between the human coders and the new model.

Despite the improvement in the performance of the model, there is room for further improvement, particularly in classifying lower-quality comments. With more comments, the new model generated in this study can be improved further, possibly increasing the result in the inter-rater agreement with the human coders. Creating a more generalizable model for assessing comment quality regardless of the concept or skill can also be explored to make the model applicable beyond presentation and communication skills.

4. References

Chatti, M. A., Marinov, M., Sabov, O., Laksono, R., Sofyan, Z., Fahmy Yousef, A. M., & Schroeder, U. (2016). Video annotation and analytics in CourseMapper. Smart Learning Environments, 3(1), 10. https://doi.org/10.1186/s40561-016-0035-1

Dimitrova, V., & Mitrovic, A. (2022). Choice Architecture for Nudges to Support Constructive Learning in Active Video Watching. Artificial Intelligence in Education, 32(4), 892–930.

Krippendorff's Alpha. (2010). In N. Salkind, Encyclopedia of Research Design (p. 670). SAGE Publications, Inc. https://doi.org/10.4135/9781412961288.n206

Mitrovic, A., Dimitrova, V., Weerasinghe, A., Lau, L. (2016) Reflexive experiential learning: using active video watching for soft skills training. In: Chen, W. et al. (Eds.) Proc. 24th Int. Conf. Computers in Education, pp. 192-201. Mumbai, India, Nov 28 – Dec 2, 2016. Asia-Pacific Society for Computers in Education (ASPCE)

Mitrovic, A., Dimitrova, V., Lau, L., Weerasinghe, A., & Mathews, M. (2017). Supporting Constructive Video-Based Learning: Requirements Elicitation from Exploratory Studies. In E. André, R. Baker, X. Hu, Ma. M. T. Rodrigo, & B. du Boulay (Eds.), Artificial Intelligence in Education (Vol. 10331, pp. 224–237). Springer International Publishing. https://doi.org/10.1007/978-3-319-61425-0_19

Mohammadhassan, N., Mitrovic, A., Neshatian, K., Dunn, J. (2020) Automatic quality assessment of comments in active video watching using machine learning techniques. In: So, H.J. et al. (Eds.) Proc. 28th International Conference on Computers in Education, pp. 1-10. Asia-Pacific Society for Computers in Education. ISBN978-986-97214-5-5.

Yousef, A. M. F., Chatti, M., & Schroeder, U. (2014). The State of Video-Based Learning: A Review and Future Perspectives. International Journal on Advances in Life Sciences, 6, 122–135.