MBOX Lightweight Voice Analysis Sensors for MMLA

Daniel Spikol[a], Zaibei Li[a], Sergio Serrano-Iglesias[b], Hamza Ouhaichi[c] and Bahtijar Vogel[c]

[a]University of Copenhagen, Department of Computer Science, Denmark
[b]GSIC-EMIC Research Group, Universidad de Valladolid, Spain
[c]Malmö University, Department of Computer Science, Sweden

Abstract
This abstracts presents MBOX, a scalable and lightweight system that integrates data collection, data analysis and instructive feedback to evaluate participants' engagement levels of learning activities.

Keywords
Audio Analysis, Machine Learning, Multimodal Learning Analytics, IOT

1. Introduction

Conversation, defined as the communication between two or more people, is an indispensable part of interpersonal communication and teamwork in our daily lives. People can exchange their thoughts and ideas during conversations by hearing each other while observing body language. As conversations usually accompany human activities, it is plausible and meaningful to evaluate the activity's effectiveness by analysing the conversations. Through Multimodal Learning Analytics (MMLA) we can provide opportunities for understanding and supporting collaborative problem-solving. In the following abstract, we focus on capturing a conversation focusing on speaker diarization and patterns of conversation.

We are interested in creating a scalable and lightweight system that integrates data collection, data analysis and instructive feedback to evaluate participants' engagement levels of learning activities. This prototype concentrates on the data pipeline's feature extractor, which accepts audio data collected from sensors and outputs several conversational characteristics in segments. This project concentrates on the data pipeline's feature extractor, which accepts audio data collected from sensors and outputs several conversational characteristics in segments. The audio analysis is part of the larger MBOX (Multimodal Box) which aims to connect to different smart learning environments (SCARLETT) and enables a better scaling of computational resources used within the learning context [1, 2]

CROSSMMLA SLE @ LAK'22: Learning Analytics for Smart Learning Environments Crossing Physical and Virtual Learning Spaces, March 22, 2022, Online

2. Background

Over the past years, several studies have investigated how sensor-based analytics can be used to identify patterns of speech in the classrooms [3, 4, 5] on teacher discourse in the classroom. Recent work also illustrates the benefits of acoustic models and machine learning [6] to analyse teaching practices with mobile devices. In addition, different approaches for speaker identification and conversational pattern recognition have provided approaches that can be scaled to smaller computer devices allowing for the possibility of developing and deploying the MBOX system [7].

3. Scenario

The proposed scenario is a collaborative activity between 3-4 students to solve an open-ended design activity. The task is to design and program a robotic arm to respond to a different input. The students need to design and assemble the robot, program the appropriate behaviour, and create a shared document that is guided by inquiry-based activities. We use the Scarlett system to provide the learning activity and the MBOX system to capture and process the audio add the context for capturing the learning environment.

4. Challenges

The major challenges for the MBOX audio analysis is several-fold, first the design of the scenario that allows the system to capture the training data of the students, second the the environment, how will such systems function in real-world environments with multiple teams of students working in the same room, thirdly the design of the data pipeline and performance of the models, and lastly the integration into the SCARLETT system. However, the more significant, more relevant issue is the perceived benefits to the learners and teachers for such tools and how to improve educational activities without reducing them to performance data.

References

[1] B. Tabuenca, S. Serrano-Iglesias, A. Carruana-Martin, C. Villa-Torrano, Y. A. Dimitriadis, J. I. Asensio-Perez, C. Alario-Hoyos, E. Gomez-Sanchez, M. L. Bote-Lorenzo, A. Martinez-Mones, et al., Affordances and core functions of smart learning environments: A systematic literature review, IEEE Transactions on Learning Technologies (2021).

[2] H. Ouhaichi, D. Spikol, B. Vogel, Mbox: Designing a flexible iot multimodal learning analytics system, in: 2021 International Conference on Advanced Learning Technologies (ICALT), IEEE, 2021, pp. 122–126.

[3] Z. Wang, K. Miller, K. Cortina, Using the LENA in Teacher Training: Promoting Student Involement through automated feedback, volume 4, na, 2013.

[4] M. Ford, C. T. Baer, D. Xu, U. Yapanel, S. Gray, The lenatm language environment analysis system (2008).

[5] E. Jensen, M. Dale, P. J. Donnelly, C. Stone, S. Kelly, A. Godley, S. K. D'Mello, Toward Automated Feedback on Teacher Discourse to Enhance Teacher Learning, Association for Computing Machinery, New York, NY, USA, 2020, p. 1–13. URL: https://doi.org/10.1145/3313831.3376418.

[6] D. Schlotterbeck, P. Uribe, R. Araya, A. Jimenez, D. Caballero, What classroom audio tells about teaching: A cost-effective approach for detection of teaching practices using spectral audio features, in: LAK21: 11th International Learning Analytics and Knowledge Conference, LAK21, Association for Computing Machinery, New York, NY, USA, 2021, p. 132–140. URL: https://doi.org/10.1145/3448139.3448152. doi:10.1145/3448139.3448152.

[7] M. Yousefi, J. H. Hansen, Frame-based overlapping speech detection using convolutional neural networks, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 6744–6748.