Majumdar, R. et al. (Eds.) (2025). Proceedings of the 1st International Conference on Learning Evidence and Analytics. Asia-Pacific Society for Computers in Education

Automatic Role Estimation of Students' Utterance Using Generative AI

Yasuhisa TAMURA[a*], John AUGERI[b], Yuuki ITOHARA[a] & Risa OTSUKA[a]
[a]*Faculty of Science and Engineering, Sophia University, Japan*
[b]*Ile-de-France Digital University, France*
*ytamura@sophia.ac.jp

Abstract: This study explores the use of generative AI to estimate speaking roles in educational group discussions. By analyzing Zoom-recorded conversations transcribed into text, a modified ChatGPT model classified utterances into roles such as questions or assertions. Initial accuracy (50–55%) improved to 76% through role-specific training and contextual embedding. This method reduces manual effort and evaluator bias while enhancing scalability and objectivity in assessing collaborative learning. Tested across languages and cultures, it shows promise for international application. The approach also advances discourse analysis capabilities, contributing to learning analytics and supporting interventions to improve educational outcomes.

Keywords: Learning analytics, Generative AI, Collaborative Learning, Automatic Role Estimation, Discourse Analysis

1. Introduction

Previous studies show that group discussions enhance not only understanding of content but also communication and metacognitive skills (Coffey, 2012; Lee, 2014). In learning analytics, collaborative learning and dialogue modeling using NLP are active areas of research (Chen, 2022; Dowell, 2022). Analyzing utterances helps evaluate student comprehension, but traditional NLP (natural language processing) lacks adequate tools, often requiring manual, subjective analysis.

To address this, various approaches have been explored: role detection via computational linguistics (Venkatesha, 2024), interaction analysis tools (Marcos, 2005), and transformers for utterance classification (Ubani, 2022). Group Communication Analysis (Dowell, 2019) and structured role scripts (Weinberger, 2003) also contribute.

This study differs by using generative AI to automate role estimation, enabling objective, scalable analysis of speaking roles and group dynamics, even with large datasets.

2. Method

2.1 *Transcription of Speech Data*

In this study, voice-based group discussions were held on Zoom, and the learners' voice data was recorded separately for each speaker. The voice data was then converted into text data using Hylable's transcription function (Japanese and English) or OpenAI's Whisper (other languages) and then converted into Japanese using machine translation.

2.2 *Automatic estimation of speech roles using generative AI*

In this experiment, the authors used ChatGPT4o to estimate the role of statements. First, we chronologically arranged all statements made during the discussion and entered one statement per cell in a Google spreadsheet. For the cell contents, we used GPT for Sheets and Docs to input the statement content and an analysis request prompt into ChatGPT, and the output result was displayed in another cell. The prompt instructed the system to return the most appropriate role of the statement among negation, affirmation, question, reason, and assertion. In addition, to improve the accuracy of role estimation, we provided the two statements that preceded the statement in question to determine the context. In addition, we defined the role and a specific example. We paid attention to protecting the personal information of the participants by setting the data entered in the experiment so that it would not be used for learning ChatGPT.

2.3 *Experiment*

The experiment was conducted in Japan, the United States, and Belgium (operated remotely from Japan) from November to December 2024. The Japanese subjects were students who agreed to participate in the experiment and were recruited from the "Educational Information Engineering" course at Sophia University, which I teach. In the United States, we asked Professor Brian Beatty of San Francisco State University to cooperate with the experiment, and the subjects were students taking his course. In Belgium, we asked Professor Mia De Wilde of Thomas More University College to cooperate with the experiment, and the subjects were students in her laboratory. A total of 12 subjects were targeted, four in each of the three countries. The experiment was conducted for each country. Four students were divided into two in a face-to-face environment and two in an online environment (participating individually in a separate room), and connected via Zoom for six high-flex discussions in which all face-to-face and online subjects participated. Each discussion lasted 10 minutes, and three types of formats (brainstorming, consensus building, and debate) were conducted twice each with the camera on and off. Due to the number of consenting subjects, the study was conducted twice in Japan with different participants, once each in the United States and Belgium.

3. Results

3.1 *Frequency Distribution of Utterance Roles*

Using the transcription of speech data and automatic estimation of speech roles by the generation AI described in Chapter 2, we estimated the speech roles collected in the experiment. Figure 1 shows the role estimation results for 105 speeches in one discussion. This process requires manual effort to copy and paste the speech into Google Spreadsheet. However, the queries to ChatGPT and reply input are automated, and the estimation results can be obtained in a few minutes. This research also aims for an "international comparison of speech roles in group discussions in a hyflex environment." However, the results and considerations of the international comparison itself will be discussed in a separate manuscript.

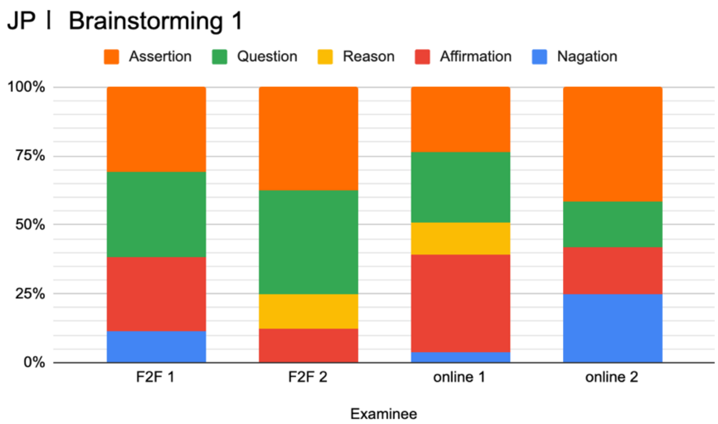

Figure 1. Example frequency distribution of utterance roles

3.2 *Improvement of Role Estimation Accuracy*

In this experiment, we manually verified the estimation by ChatGPT and finally achieved an accuracy of about 76%. In the initial stage, only the content of each comment and the five target roles were input to ChatGPT. As a result, the estimation accuracy was low at 50%-55%. In the next stage, the comment was to be estimated, and the two preceding comments were inputted. As a result, the estimation accuracy improved to 60%-65%. Furthermore, when explanations and examples of the comment roles were provided as additional information, the estimation accuracy improved to about 76%. As seen from this, providing sufficient surrounding and background information is important when instructing the generation AI to process.

4. Discussion & Conclusion

This study demonstrated that generative AI can objectively estimate speaking roles in group discussions, offering a practical tool for evaluating educational dialogues—something previously difficult to assess objectively. The role estimation achieved an average accuracy of 76%, which improved with added context such as prior utterances and role definitions. Although raw text was used this time, future improvements could include filtering out noise (e.g., “ah,” “um”) and better identifying speaker intent.

This method may serve future uses in formative evaluation and external feedback by analyzing biases in roles, topic effects, and temporal changes. Further challenges include accounting for utterance chains (e.g., vocabulary propagation) and interpersonal dynamics. As NLP evolves, accuracy is expected to improve significantly.

Acknowledgments

We thank Professor Brian Beatty of San Francisco State University and Professor Mia De Wilde of Thomas More University College for their interest in this work and contribution to the experiment.

References

Chen, B. & Teasley, S.D. (2022). Learning Analytics for Understanding and Supporting Collaboration, Lang, C. et al. (eds) (2022). Handbook of Learning Analytics (2nd Ed.), 86–95, Society for Learning Analytics Research.

Coffey, G. (2012). Literacy and technology: Integrating technology with small group, peer-led discussions of literature. *International Electronic Journal of Elementary Education*, *4*(2), 395–405.

Dowell, N. M., Nixon, T. M., & Graesser, A. C. (2019). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. Behavior research methods, 51, 1007–1041.

Dowell, N. & Kovanović, V. (2022). Modeling Educational Discourse with Natural Language Processing, Lang, C. et al. (eds) (2022). Handbook of Learning Analytics (2nd Ed.), 105–119, Society for Learning Analytics Research.

Lee, Y. & Ertmer, P. A. (2014). Examining the effect of small group discussions and question prompts on vicarious learning outcomes. *Journal of Research on Technology in Education*, *39*(1), 66-80.

Marcos, J. A., Martinez, A., & Dimitriadis, Y. (2005). Towards adaptable interaction analysis tools in CSCL. In 12th International Conference on Artificial Intelligence in Education (AIED) 2005 (pp. 4-pages).

Ubani, S., & Nielsen, R. (2022). Classifying different types of talk during collaboration. In International Conference on Artificial Intelligence in Education (pp. 227–230). Springer International Publishing.

Venkatesha, V., Nath, A., Khebour, I., Chelle, A., Bradford, M., Tu, J., ... & Krishnaswamy, N. (2024). Propositional extraction from natural speech in small group collaborative tasks. In Proceedings of the 17th International Conference on Educational Data Mining (pp. 169–180).

Weinberger, A. (2003). Scripts for computer-supported collaborative learning. Effects of social and epistemic cooperation scripts on collaborative knowledge construction.