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

DiaRoBERTa: A Multi-Party Dialogue Model for Multi-Skill Recognition in Classroom Collaborative Problem Solving Discussions

Shuqing LIU[a*], Li CHEN[b], SijieXiong[a], Haiqiao LIU[a], Cheng TANG[c] & Atsushi SHIMADA[c]
[a]*The Graduate School of Information Science and Electrical Engineering,Kyushu University ,Japan*
[b] *Division of Math, Sciences, and Information Technology in Education, Osaka Kyoiku University, Japan.*
[c]*The Faculty of Information Science and Electrical Engineering, Kyushu University,Japan*
*liu.shuqing.718@s.kyushu-u.ac.jp

Abstract: Collaborative problem-solving (CPS) integrates cognitive and social dimensions, which makes it a critical skill in educational contexts. However, existing models often treat dialogues as linear sequences, which limits their ability to capture the graph-like structures and intricate interrelations inherent in CPS discussions. These limitations hinder the accurate representation of speaker roles, conversational transitions, and dependencies within multi-party interactions. To address this challenge, we propose DiaRoBERTa, a novel model tailored for CPS dialogue classification. Building on Multi-Party Conversation (MPC) techniques, DiaRoBERTa leverages the ability of MPC to model dialogues as interconnected graphs rather than as mere linear sequences. Specifically, MPC techniques allow us to incorporate specialized markers that encode speaker roles and conversational transitions, thereby explicitly capturing the complex relational dynamics among participants. This enables the model not only to maintain the sequential order of utterances but also to represent the non-linear dependencies and interaction patterns critical for understanding CPS dialogues. Additionally, a targeted loss function addresses class imbalance, improving performance on underrepresented categories.Experiments on a dataset of middle school classroom discussions demonstrate that DiaRoBERTa outperforms baseline models, achieving a 1.5% improvement over existing methods. These results highlight DiaRoBERTa’s effectiveness in adapting sequence-based approaches to handle graph-like structures and its scalability for CPS skill assessment. By addressing key limitations of current methods, DiaRoBERTa provides a robust framework for advancing automated CPS dialogue analysis in collaborative educational settings.

Keywords: Collaborative Problem-Solving, Multi-Party Conversations, Natural Language Processing

1. Introduction

As social challenges grow increasingly complex, Collaborative Problem-Solving (CPS) has emerged as a critical skill for success in the 21st century(OECD, 2017). Defined as the process by which individuals or groups work collectively to achieve common goals (Hesse et al., 2015), CPS fosters essential cognitive and social skills, including critical thinking, perspective-taking, and collaboration.

The significance of CPS is particularly evident in education, where effective CPS fosters student engagement, improves interaction quality, and enhances learning outcomes (Hwang & Chen, 2023; Xu et al., 2023), with social skills playing a critical role in facilitating collaborative processes and ensuring effective communication among participants (Wang et al., 2023; He et al., 2024; Lee et al., 2015).

The classification of CPS skills is important for understanding how students use these skills in CPS learning. Traditionally, this process depends on manual coding, which ensures high reliability but requires significant time, labor, and coder expertise (Castleberry & Nolen, 2018). In contrast, pre-trained language models such as BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019) utilize attention mechanisms and leverage the extensive semantic knowledge acquired during pretraining. These models demonstrate superior performance in dialogue classification tasks (Pugh et al., 2021), which underscores their potential to enhance the analysis of CPS dialogue data.

This study aims to improve the classification of collaborative problem-solving (CPS) dialogues by addressing the limitations of sequence-based models in capturing the complex interactions among multiple speakers. To achieve this, we introduce DiaRoBERTa, a novel model designed specifically for CPS dialogue classification. Building on Multi-Party Conversation (MPC) techniques, DiaRoBERTa integrates special markers to enrich sequence representations, explicitly encoding speaker roles and conversational transitions. Additionally, the model incorporates a targeted loss function to address class imbalance, ensuring more robust classification of CPS-specific social skills.

The contributions can be summarized as follows:

- This study proposes an automatic CPS dialogue classification approach through a tailored DiaRoBERTa model that effectively captures social and cognitive skills.
- This study introduces a novel MPC-based sequence representation that explicitly encodes speaker roles and transition dynamics, addressing the limitations of traditional sequence-based models in CPS dialogue classification.
- This study incorporates a targeted loss function that significantly improves the model's ability to address class imbalance, enhancing performance on underrepresented categories.

2. Categorization of Collaborative Problem - Solving Skills

2.1 Categorization of Collaborative Problem-Solving Skills

Based on the framework proposed by Hesse et al., this study categorizes CPS into five core skill categories: participation, perspective-taking, social regulation, task regulation, and learning and knowledge building (Chen et al., 2024), as shown in Figure 1. These categories are chosen for their practicality in assessment, observation, and instruction.

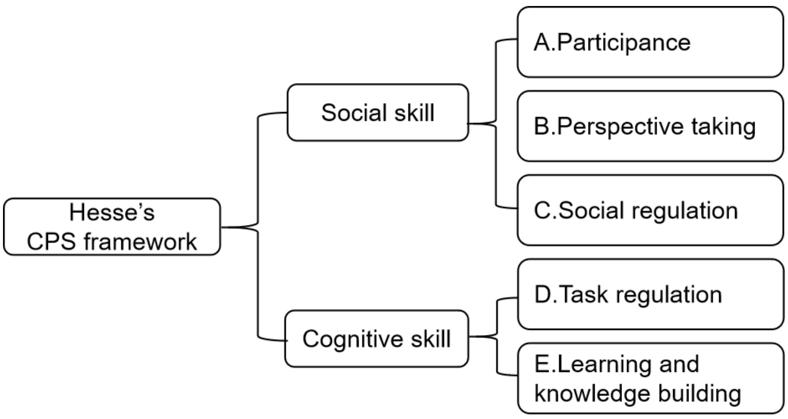

Figure 1. Hesse's Collaborative Problem - solving Framework

The social domains focus on managing interpersonal interactions. Participation refers to active engagement in group discussions, reflecting an individual's willingness to share ideas and consider others' input. Perspective-taking highlights the ability to integrate different viewpoints and reconsider problems from diverse perspectives. Social regulation involves managing group dynamics by resolving conflicts and coordinating actions to achieve shared goals.

The cognitive domains address task-related processes. Task regulation refers to the ability to structure problem-solving activities, manage resources, set goals, and explore solutions to complex challenges. Learning and knowledge-building capture the capacity to connect information from various sources, reflect on progress, and refine problem-solving strategies.

In this study, we annotate dialogue sequences collected from collaborative activities according to these five categories to identify key CPS skills. Dialogue segments that do not match any of the five defined categories are labeled as N (None).

2.2 Input Data and Custom Markers

To enhance the model's ability to capture dialogue context and speaker dynamics, we introduce two custom markers: [SPK] marks the beginning of an utterance and explicitly indicates which participant is speaking. [THN] marks a transition between speakers, explicitly indicating that the preceding and following utterances are delivered by different individuals. Given a dialogue with multiple utterances, the input sequence is constructed as:

$$T_{\text{input}} = \{s_{\text{pre1}}[\text{SPK}]u_{\text{pre1}}[\text{THN}]\dots s_{\text{current}}[\text{SPK}]u_{\text{current}}\}$$

Here, s denotes speaker identifiers, and u represents utterances.

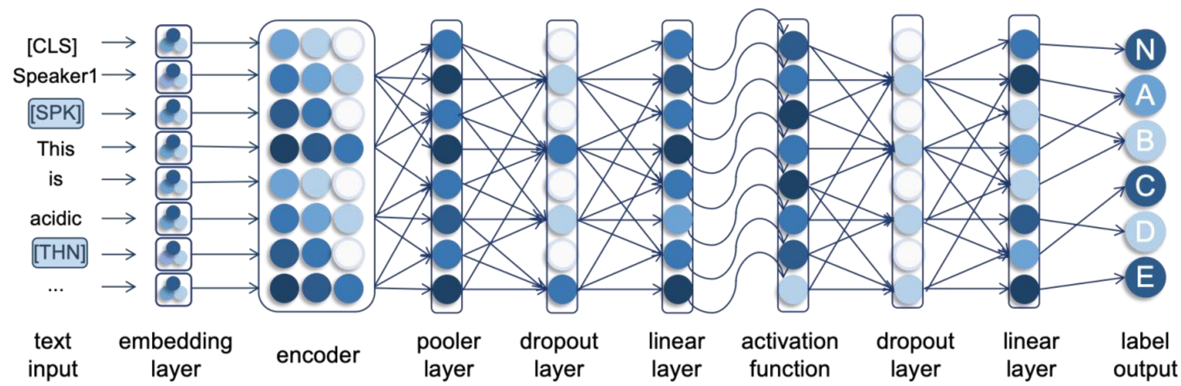

Figure 2. The Architecture of the Network.

2.3 Model Architecture

Our proposed model builds upon the pre-trained RoBERTa architecture and introduces enhancements tailored for dialogue classification tasks. Specifically, we incorporate custom markers to model speaker roles and conversational transitions and utilize focal loss to address data imbalance challenges. The overall network structure is depicted in Figure 2.

3. Experiments and Results

3.1 Dataset

The dataset consisted of transcripts from 40 middle school chemistry classroom discussions, containing 1,310 conversational sentences. Each sentence was manually annotated into one of six CPS categories by a professional teacher, covering both social and cognitive skills. To ensure balanced evaluation, three discussion rounds (138 sentences) were selected as the test set, with the remaining 1,172 sentences used for training. Each discussion included multiple dialogue rounds, with an average of 32.75 sentences per round. The average sentence length across the dataset is 14.24 words.

3.2 Model and Training Details.

The DiaRoBERTa model was designed for six-class classification tasks, incorporating [SPK] and [THN] markers to explicitly model speaker roles and dialogue transitions. Training was conducted on an NVIDIA 3090 GPU with a batch size of 16, learning rate of 4×10^{-5}, and focal loss to address class imbalance. The training process spanned nine epochs, with each epoch requiring approximately one minute.

3.3 Results

As shown in Table 1, DiaRoBERTa achieved the highest accuracy (85.51%) and F1-score (85.09%), surpassing BERT, BART (Lewis et al., 2019), and RoBERTa by over 1.5 percentage points, and it significantly surpasses general-purpose models such as Qwen 2.5 (Qwen et al., 2025), ChatGPT- o3 (OpenAI et al., 2025) and Deepseek-R1 (DeepSeek-AI, 2025) in performance as well.

Table 1. *Performance Comparison of Models*

Model	Accuracy	F1-Score
Qwen 2.5	21.01%	14.56%
ChatGPT o3	35.51%	19.69%
Deepseek R1 671B	47.10%	40.06%
BERT	81.16%	80.53%
BART	80.43%	80.07%
RoBERTa	82.61%	81.66%
DiaRoBERTa	**85.51%**	**85.09%**

An ablation study was conducted to evaluate the contribution of [SPK] and [THN] markers. The results in Table 2 demonstrate that the SPK + THN Form (DiaRoBERTa) achieves the best performance, which emphasizes the significance of combining speaker and transition information for CPS classification.

These results confirm that the combination of [SPK] and [THN] markers plays a crucial role in enhancing the classification of CPS dialogues.

Table 2. *Performance Comparison of Models*

Model	Accuracy	F1-Score
Utterance Form	83.33%	82.69%
Speaker Form	83.33%	82.45%
Context Form	82.61%	81.24%
SPK+THN Form(DiaRoBERTa)	85.51%	85.09%

4. Conclusion

This study presents DiaRoBERTa, a model for classifying CPS dialogues to identify social skills, which outperforms traditional models like BERT and RoBERTa by improving classification accuracy from 82.6% to 85.5%, thus highlighting its ability to capture the complexities of collaborative interactions.

Future research can explore integrating multimodal features for better CPS interaction understanding, applying DiaRoBERTa to diverse educational datasets for generalization evaluation, integrating classroom behavior and learning outcome data for comprehensive analysis of students' social skills, and leveraging advanced methodologies like graph neural networks and dynamic attention mechanisms to capture complex social dynamics.

This work not only contributes to the field of educational data analysis, but it also offers a scalable framework for assessing collaborative interactions in real-world settings, paving the way for further innovations in AI-enhanced education.

Acknowledgements

This work was supported by JST CREST Grant Number JPMJCR22D1 and JSPS KAKENHI Grant Number JP22H00551 and JP24K16759 , Japan.

References

Castleberry, A., & Nolen, A. (2018). Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning*, 10(6), 807–815.

Chen, L., Taniguchi, Y., Shimada, A., & Yamada, M. (2024). Exploring behavioral and strategic factors affecting secondary students' learning performance in collaborative problem solving-based STEM lessons. *SAGE Open*, 14(2), 21582440241251641. https://doi.org/10.1177/21582440241251641

DeepSeek-AI (2025). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. URL: https://arxiv.org/abs/2501.12948

Devlin, J., Chang, M. -W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long and Short Papers) (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Organisation for Economic Cooperation and Development. (2017). Pisa 2015: Collaborative problem-solving framework. *Technical report. OECD Publishing*. https://doi.org/10.1787/19963777

He, J., Ren, S., & Zhang, D. (2024). The relationship between personal-collaborative motivation profiles and students' performance in collaborative problem solving. *Large-scale Assessments in Education*, 12, 34. https://doi.org/10.1186/s40536-024-00219-6

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A Framework for Teachable Collaborative Problem Solving Skills. In P. Griffin & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 37–56). Springer Netherlands.

Hwang, G. -J., & Chen, P. -Y. (2023). Effects of a collective problem-solving *promotion*-based flipped classroom on students' learning performances and interactive patterns. *Interactive Learning Environments*, 31(5), 2513–2528. https://doi.org/10.1080/10494820.2019.1568263

Lee, D., Huh, Y., & Reigeluth, C. M. (2015). Collaboration, intragroup conflict, and social skills in project-based learning. *Instructional Science*, 43, 561–590.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V. and Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR*, abs/1907.11692.

OpenAI, :, El-Kishky, A., Wei, A., Saraiva, A., Minaiev, B., Selsam, D., Dohan, D., Song, F., Lightman, H., Clavera, I., Pachocki, J., Tworek, J., Kuhn, L., Kaiser, L., Chen, M., Schwarzer, M., Rohaninejad, M., McAleese, N., o3 contributors, Mürk, O., Garg, R., Shu, R., Sidor, S., Kosaraju, V. and Zhou, W. (2025). Competitive programming with large reasoning models.

Pugh, S. L., Subburaj, S. K., Rao, A. R., Stewart, A., Andrews-Todd, J., & D'Mello, S. (2021). Say what? Automatic modeling of collaborative problem solving skills from student speech in the wild. In *Proceedings of the 14th International Conference on Educational Data Mining* (EDM 2021).

Qwen et al. (2025) . Qwen2.5 technical report.

Wang, M., Yu, R., & Hu, J. (2023). The relationship between social media-related factors and student collaborative problem-solving achievement: An HLM analysis of 37 countries. *Education and Information Technologies*, 28, 14071–14089. https://doi.org/10.1007/s10639-023-11763-z

Xu, E., Wang, W., & Wang, Q. (2023). The effectiveness of collaborative problem solving in promoting students' critical thinking: A meta-analysis based on empirical literature. *Humanities and Social Sciences Communications,* 10, 16. https://doi.org/10.1057/s41599-023-01508-1