Crafting a Responsible Dialog System for Collaborative Learning Environments

Margaret Perkoff[1], **Emily Doherty**[1], **Jeffrey B. Bush**[1], **Leanne Hirshfield**[1]

[1]University of Colorado Boulder
Boulder, Colorado 80309 USA
Margaret.Perkoff@colorado.edu, Emily.Doherty@colorado.edu, Jeffrey.Bush@colorado.edu,
Leanne.Hirshfield@colorado.edu

Abstract

This work-in-progress explores the use of large language models (LLMs) within a Computer Supported Collaborative Learning (CSCL) environment to enhance collaborative learning outcomes. To design a dialog policy-driven conversational system for small group work settings, we gather stakeholder feedback through a Wizard-of-Oz (WoZ) study and dialog analysis sessions with educational experts. Grounded in pedagogical frameworks of CSCL, our system aims to align with educational goals while addressing the limitations of current conversational AI evaluations that overlook education-specific objectives. Our preliminary laboratory study involves real-time interventions by a simulated agent (JIA), steering the ongoing dialog policy design. We discuss future work that involves transitioning to classroom implementations to ensure real-world validation while adapting the system to foster enhanced collaboration and knowledge building within authentic educational settings.

Introduction

Computer Supported Collaborative Learning (CSCL) has evolved considerably over the last several decades. CSCL systems improve collaboration outcomes and can support regulated learning on collaborative tasks (Salomon and Perkins 1998; Järvelä and Hadwin 2013). In particular, this work focuses on the class of internal CSCL tools, or ones that facilitate collaboration in real-time while being anchored to the task itself (Morris et al. 2010). Morris and colleagues stress that for internal CSCL tools to be effective, they must incorporate elements of instructional practices that a teacher may use to encourage collaborative learning such as roles, scripts, and prompts. Going deeper, DeBarger and colleagues make a theoretical and empirical argument that synergy between teaching routines and classroom technology supports collaboration (DeBarger et al. 2011). With the improvement in the quality of natural language generation (NLG) models, in particular large language models (LLMs), more tutoring and educational systems are incorporating this technology (Kuhail et al. 2023). However, LLMs are not typically trained to be used in the education domain, they are typically pre-trained on large quantities of text scraped from social media datasets and then fine-tuned later on for domain-specific vocabulary. Although fine-tuning can improve their ability to recognize specific terms, it does not ensure that the LLM will produce output that is helpful in the learning context.

In this paper, we propose a design for a dialog policy-driven conversational system to improve collaboration in a small group work setting. From a technical perspective, we incorporate a dialog policy and a controllable response generation (CRG) model to constrain the output of the system so it is suitable for a CSCL environment. Our initial dialog policy design draws from collaboration with experts in teaming science, collaborative learning, and education. We also report on work in progress of a pilot WoZ lab study where we solicit feedback from student participants. Insights from this work will inform the ongoing policy design, which we anticipate will result in higher levels of collaboration in our future lab and classroom studies.

Responsible Dialog System Design

At the core of any conversational software is an underlying dialog system that provides responses to the user based on information from the ongoing dialog. We can imagine a system that takes in the dialog context (in the form of speech, video, or text input), performs some analysis to determine the state of the dialog, and outputs a coherent response. To be successful in CSCL environments, this system must cohere with lesson goals, adhere to the classroom values, and improve the learning experience for both students and teachers. To achieve these goals, our system incorporates two components: (1) a dialog policy and (2) a CRG model (Fig. 1).

The use of dialog policies in neural conversation systems has been shown to improve the quality of responses (Hedayatnia et al. 2020). Our dialog policy consists of theoretical dialog states and actions grounded in measurable discourse features. A dialog state represents the current status of the ongoing collaboration whereas the dialog action indicates the type of intervention that the agent should take. To determine these state-action mappings, we facilitated a coworking session between domain experts. During this session, we analyzed annotated transcripts of university students engaged in a collaborative task, identifying the represented dialog state and potential actions they would take in that scenario. For our system to recognize the dialog states,

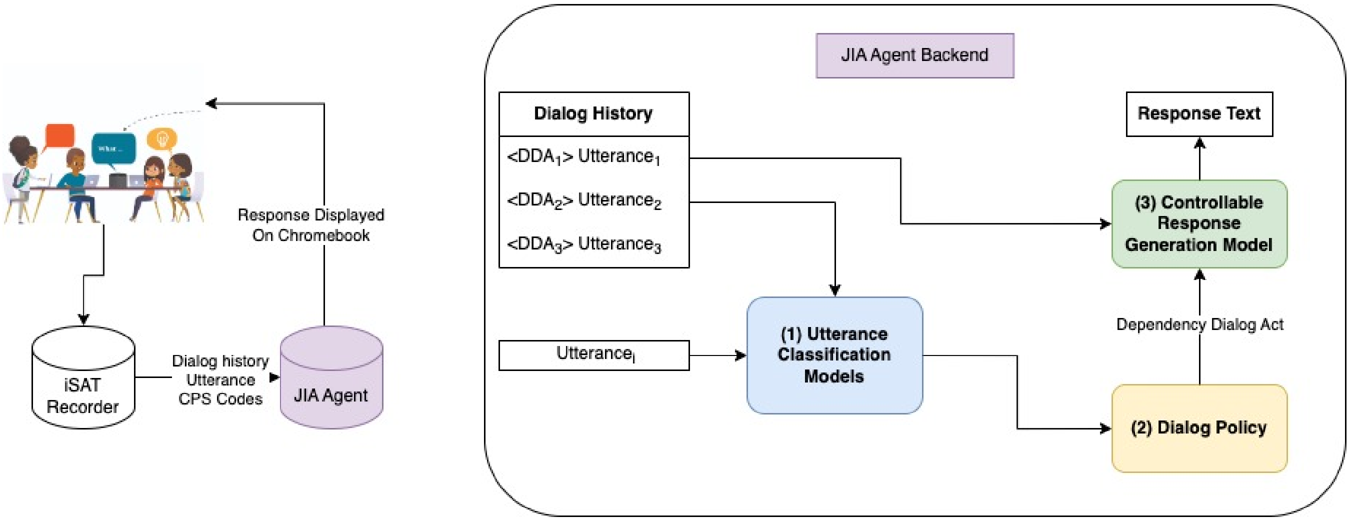

Figure 1: The proposed dialog system. Live speech data is transcribed via our proprietary recorder (Dickler et al. 2022). In (1) the sentences are classified for utterance-level dialog features. These features are used by the dialog policy (2) to determine the dialog state which is passed on to the CRG model (3) to generate the appropriate agent action.

they are grounded in utterance-level discourse features, specifically Dependency Dialog Acts (DDA) (Cai et al. 2023) and Collaborative Problem Solving (CPS) (Andrews-Todd et al. 2022) codes. Each dialog action is also associated with a particular DDA that will be concatenated with the conversation history as input to a CRG model. CRG models can produce responses that are comparable to a template-based model (Walker et al. 2007). In future work, we will fine-tune several pre-trained response generation models on our lab study data to evaluate which one performs the best in this setting.

Evaluating Conversational Systems in the Education Domain

One of the major questions around conversational AI concerns what constitutes a “good” response. Social-bot systems value metrics assessing how engaging a system is, or how many turns the user chooses to continue the conversation for. More recent LLM-based systems measure ”good” responses using a combination of Sensibility, Specificity, and Interestingness (Adiwardana et al. 2020; Zhang et al. 2020; OpenAI 2023). While these measures are valuable for assessing whether or not a response makes sense in a given context, they are not targeted toward education-specific goals. As such, educational conversational systems need evaluation criteria aligned with the needs of teachers and students. Our lab study seeks to develop and validate these criteria by soliciting feedback from students through both objective and subjective data.

Current Study Design

This study, approved by the university’s Institutional Review Board, involves students aged 12 to 17. Upon informed consent, students work in groups of 2-3 (target n = 25 groups) on collaborative Jigsaw tasks centered around brainstorming and knowledge sharing, which were adapted from tasks used in the local curriculum (Cao et al. 2023; Biddy et al. 2021). Employing the WoZ paradigm, participants interact with a simulated computer-based entity known as JIA (Jigsaw Interactive Agent), controlled in real-time by an expert confederate. The confederate receives live video and audio feeds of the group and provides real-time interventions through pop-up messages from JIA. Additionally, real-time speech analysis via our recorder is conducted to label utterances with CPS features and classify speech as on- or off- topic/task (Breideband et al. 2023; Ganesh et al. 2023). Following the task, participants discuss their perceptions of JIA, intervention delivery, and potential applications in a focus group session. Participants also complete surveys on constructs including social loafing (Linnenbrink-Garcia, Rogat, and Koskey 2011), trust (Merritt 2011), and team processes (Marks, Mathieu, and Zaccaro 2001; Luciano, DeChurch, and Mathieu 2018). Real-time data collection and post-task assessments will be triangulated to provide insights into JIA’s impact on collaboration dynamics and inform the ongoing dialog policy design.

Next Steps and Conclusions

Gleaning insights from the pilot study and policy design sessions with domain experts, we will iteratively identify dialog states and determine the corresponding agent responses. This study provides a baseline of students’ perceptions of agent interventions. As we continue with implementing the CRG model for agent output, we can compare its effectiveness to this baseline and adjust accordingly. Our future work will focus on transitioning from lab-based simulations to classroom implementations. This transition will involve adapting the dialog system based on feedback from educators and students, aiming to foster seamless support for enhanced collaboration and knowledge building.

Acknowledgments. This research was supported by NSF National AI Institute for Student-AI Teaming (iSAT) under grant DRL 2019805. The opinions expressed are those of the authors and do not represent views of NSF.

References

Adiwardana, D.; Luong, M.-T.; So, D. R.; Hall, J.; Fiedel, N.; Thoppilan, R.; Yang, Z.; Kulshreshtha, A.; Nemade, G.; Lu, Y.; and Le, Q. V. 2020. Towards a Human-like Open-Domain Chatbot.

Andrews-Todd, J.; Steinberg, J.; Pugh, S. L.; and D'Mello, S. K. 2022. Comparing Collaborative Problem Solving Profiles Derived from Human and Semi-Automated Annotation. *Grantee Submission*.

Biddy, Q.; Chakarov, A. G.; Bush, J.; Elliott, C. H.; Jacobs, J.; Recker, M.; Sumner, T.; and Penuel, W. 2021. A professional development model to integrate computational thinking into middle school science through codesigned storylines. *Contemporary issues in technology and teacher education*, 21(1): 53–96.

Breideband, T.; Bush, J.; Chandler, C.; Chang, M.; Dickler, R.; Foltz, P.; Ganesh, A.; Lieber, R.; Penuel, W. R.; Reitman, J. G.; Weatherley, J.; and D'Mello, S. 2023. The Community Builder (CoBi): Helping Students to Develop Better Small Group Collaborative Learning Skills. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '23 Companion, 376–380. New York, NY, USA: Association for Computing Machinery.

Cai, J. Z.; King, B.; Perkoff, M.; Dudy, S.; Cao, J.; Grace, M.; Wojarnik, N.; Ganesh, A.; Martin, J. H.; Palmer, M.; Walker, M.; and Flanigan, J. 2023. Dependency Dialogue Acts – Annotation Scheme and Case Study.

Cao, J.; Dickler, R.; Grace, M.; Bush, J. B.; Roncone, A.; Hirshfield, L. M.; Walker, M. A.; and Palmer, M. S. 2023. Designing an AI Partner for Jigsaw classrooms. In *Workshop on Language-Based AI Agent Interaction with Children (AIAIC'2023)*.

DeBarger, A. H.; Penuel, W. R.; Harris, C. J.; and Schank, P. 2011. Teaching Routines to Enhance Collaboration Using Classroom Network Technology. In *Techniques for Fostering Collaboration in Online Learning Communities: Theoretical and Practical Perspectives*, 224–244. IGI Global.

Dickler, R.; Foltz, P.; Krishnaswamy, N.; Whitehill, J.; Weatherly, J.; Bodzianowski, M.; Perkoff, R., E. Margaret amd Southwell; Pugh, S.; Bush, J.; Chang, M.; Hirshfield, L.; Showers, D.; Ganesh, A.; Li, Z.; Danilyuk, E.; He, X.; Khebour, I. K.; and Dey, I. 2022. iSAT Speech-Based AI Display for Small Group Collaboration in Classrooms. Interactive event at the Artificial Intelligence in Education Conference (AIED 2022).

Ganesh, A.; Chang, M. A.; Dickler, R.; Regan, M.; Cai, J.; Wright-Bettner, K.; Pustejovsky, J.; Martin, J.; Flanigan, J.; Palmer, M.; and Kann, K. 2023. Navigating Wanderland: Highlighting Off-Task Discussions in Classrooms. In *Artificial Intelligence in Education*, 727–732. Springer Nature Switzerland.

Hedayatnia, B.; Gopalakrishnan, K.; Kim, S.; Liu, Y.; Eric, M.; and Hakkani-Tur, D. 2020. Policy-Driven Neural Response Generation for Knowledge-Grounded Dialog Systems. In Davis, B.; Graham, Y.; Kelleher, J.; and Sripada, Y., eds., *Proceedings of the 13th International Conference on Natural Language Generation*, 412–421. Dublin, Ireland: Association for Computational Linguistics.

Järvelä, S.; and Hadwin, A. F. 2013. New Frontiers: Regulating Learning in CSCL. *Educ. Psychol.*, 48(1): 25–39.

Kuhail, M. A.; Alturki, N.; Alramlawi, S.; and Alhejori, K. 2023. Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1): 973–1018.

Linnenbrink-Garcia, L.; Rogat, T. K.; and Koskey, K. L. K. 2011. Affect and engagement during small group instruction. *Contemp. Educ. Psychol.*, 36(1): 13–24.

Luciano, M. M.; DeChurch, L. A.; and Mathieu, J. E. 2018. Multiteam Systems: A Structural Framework and Meso-Theory of System Functioning. *J. Manage.*, 44(3): 1065–1096.

Marks, M. A.; Mathieu, J. E.; and Zaccaro, S. J. 2001. A Temporally Based Framework and Taxonomy of Team Processes. *AMRO*, 26(3): 356–376.

Merritt, S. M. 2011. Affective Processes in Human–Automation Interactions. *Hum. Factors*, 53(4): 356–370.

Morris, R.; Hadwin, A. F.; Gress, C. L. Z.; Miller, M.; Fior, M.; Church, H.; and Winne, P. H. 2010. Designing roles, scripts, and prompts to support CSCL in gStudy. *Comput. Human Behav.*, 26(5): 815–824.

OpenAI. 2023. GPT-4 Technical Report. *arXiv [cs.CL]*.

Salomon, G.; and Perkins, D. N. 1998. Individual and Social Aspects of Learning. *Review of Research in Education*, 23: 1–24.

Walker, M.; Stent, A.; Mairesse, F.; and Prasad, R. 2007. Individual and domain adaptation in sentence planning for dialogue. *J. Artif. Intell. Res.*, 30(1): 413–456.

Zhang, Y.; Sun, S.; Galley, M.; Chen, Y.-C.; Brockett, C.; Gao, X.; Gao, J.; Liu, J.; and Dolan, B. 2020. DIALOGPT : Large-Scale Generative Pre-training for Conversational Response Generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 270–278. Online: Association for Computational Linguistics.