EasyChair Preprint

№ 10018

Exploring Linguistic Indicators of Social Collaborative Group Engagement

Loris Jeitziner, Lisa Paneth, Oliver Rack, Carmen Zahn and Dirk Wulff

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

May 9, 2023

Exploring Linguistic Indicators of Social Collaborative Group Engagement

Loris T. Jeitziner, Lisa Paneth, Oliver Rack, Carmen Zahn
loris.jeitziner@fhnw.ch, lisa.paneth@fhnw.ch, oliver.rack@fhnw.ch, carmen.zahn@fhnw.ch
University of Applied Sciences and Arts Northwestern Switzerland

Dirk U. Wulff
wulff@mpib-berlin.mpg.de
Max-Planck-Institut für Bildungsforschung

Abstract: This study takes a NLP approach to measuring social engagement in CSCL-learning groups. Specifically, we develop linguistic markers to capture aspects of social engagement, namely sentiment, responsiveness and uniformity of participation and compare them to human ratings of social engagement. We observed small to moderate links between NLP-markers and human ratings that varied in size and direction across the different groups. We discuss measurement and prediction of social collaborative group engagement using natural language processing.

Introduction and Methods

Research in *Computer-supported collaborative learning* (CSCL) is adopting novel technological approaches that enable the analysis of new types of data. One of these approaches is the computational analysis of text, also known as natural language processing (NLP). Research in CSCL has begun to employ NLP to study the verbal communication of groups (Cress et al., 2021, Wise et al., 2021), but this approach has yet to be applied in the context of collaborative group engagement. To fill this gap, we investigate linguistic markers derived using NLP as predictors of collaborative group engagement in CSCL-learning.

We present preliminary findings focusing on the prediction of social engagement within an observational study of CSCL-learning groups. In this study, 6 groups of 3 to 4 members (N = 20) in an online learning setting were tasked with solving a hidden profile assignment. The participants were provided with shared and unshared information about a fictional "murder case." Afterward, they were given 45 minutes to collectively identify the "murderer".

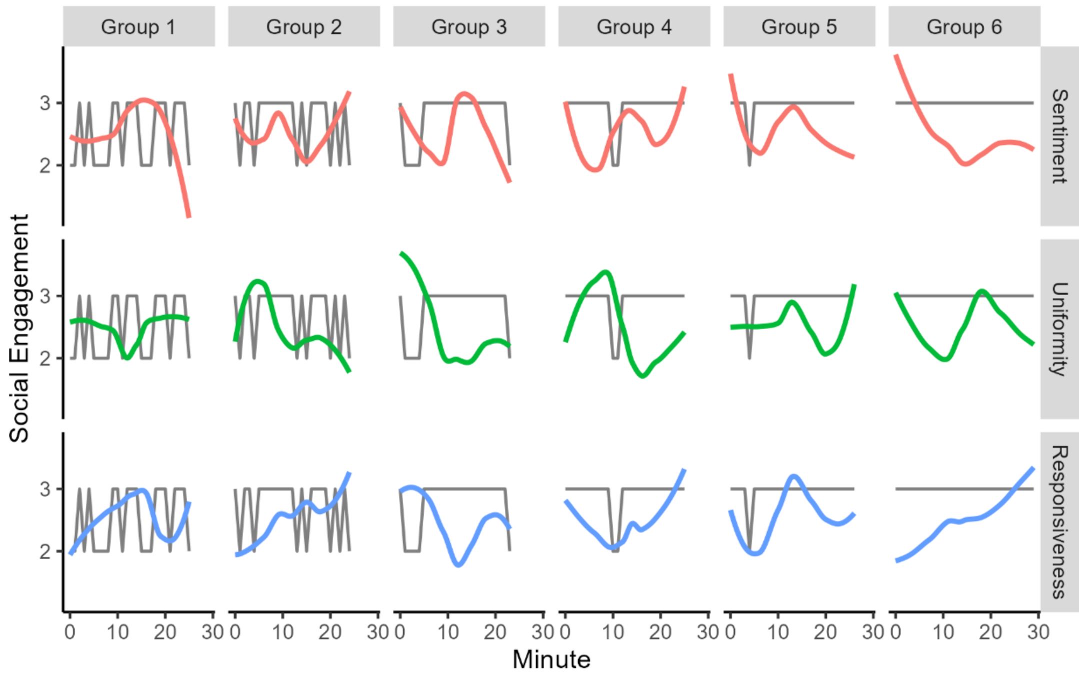
Social engagement was measured using the method by Sinha et al. (2015). The method consists of segmenting observations of videotaped student groups into one-minute intervals and rating these segments regarding the learning group's participation uniformity, responsiveness, and sentiment of communication. For example, a group that converses disrespectfully (sentiment), ignores each other (responsiveness), and is dominated by one group member (participation uniformity) would be rated low on social engagement. Social engagement was measured as either low, moderate, or high.

To predict social engagement, we developed an automatized method to extract linguistic markers that may capture sentiment, responsiveness, and participation uniformity. Sentiment was predicted using a sentiment dictionary approach that computes for each interval an average sentiment valence based on the words available in the sentiment dictionary. Participation uniformity was predicted by calculating the within-interval variance of the number of words produced by each participant. Finally, responsiveness was predicted by calculating the semantic similarity of participants' within-interval utterances using Latent Semantic Analysis (Günther et al., 2015).

Results and Discussion

Figure 1 shows the temporal development of three engagement dimensions and the corresponding linguistic markers separately for each of the six groups. To quantify the predictive potential of the linguistic markers in capturing the engagement dimension, we calculated Spearman correlations within groups. Beginning with sentiment, we observed mixed results with correlations ranging from small positive (r = .10) to small negative correlation (r = -.19). In other words, linguistic sentiment of group communication was not systematically related to human ratings of social engagement.

Figure 1
The temporal development of the social engagement ratings (grey lines), sentiment (red), uniformity (green), responsiveness (blue). The values underlying the linguistic markers were first z-standardized and smoothed using LOESS.



Concerning uniformity, we observed consistently negative associations ranging from a small (r = -.06) to (r = -.43). Thus, the lower the variance of the number of words produced by each group member, the higher the human rating of social engagement. Finally, concerning responsiveness, we observed mixed results with correlations ranging from medium positive (r = .22) to medium positive (r = -.28). Thus, the semantic similarity of group utterances did not consistently predict human ratings of social engagement.

All in all, our preliminary results suggest, at best, small associations between human ratings of social group engagement and linguistic markers of CSCL-learning groups. However, several factors may have limited our ability to detect stronger or more reliable associations. First, our analysis was based on a rather small sample of groups. Second, following Sinha et al. (2015), social engagement was measured on only three levels, providing a rather coarse measurement of social group engagement. Moreover, human raters only used two of the three levels to judge the three groups resulting in minimal variance for three of the six groups. Third, the linguistic markers used in this study represent only one of several possible implementations of sentiment, uniformity, and responsiveness. Consequently, we conclude that there is considerable opportunity for future research to investigate and improve NLP-based approaches to predicting social and other types of group engagements.

References

Cress, U., Rosé, C., Wise, A. F., & Oshima, J. (Eds.). (2021). *International Handbook of Computer-Supported Collaborative Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-030-65291-3

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun-An R package for computations based on Latent Semantic Analysis. *Behavior research methods*, *47*(4), 930-944.

Sinha, S., Rogat, T. K., Adams-Wiggins, K. R., & Hmelo-Silver, C. E. (2015). Collaborative group engagement in a computer-supported inquiry learning environment. *International Journal of Computer-Supported Collaborative Learning*, *10*(3), 273–307. https://doi.org/10.1007/s11412-015-9218-y

Wise, A. F., Knight, S., & Shum, S. B. (2021). Collaborative learning analytics. In *International handbook of computer-supported collaborative learning* (pp. 425–443). Springer.