Advancing Computational Grounded Theory for Audiovisual Data from Mathematics Classrooms

Cynthia D'Angelo, University of Illinois at Urbana-Champaign, cdangelo@illinois.edu
Elizabeth Dyer, Middle Tennessee State University, edyer@mtsu.edu
Christina Krist, University of Illinois at Urbana-Champaign, ckrist@illinois.edu
Josh Rosenberg, University of Tennessee, Knoxville, jmrosenberg@utk.edu
Nigel Bosch, University of Illinois at Urbana-Champaign, pnb@illinois.edu

Abstract: This poster will discuss early findings from a project that is developing theory-based approaches to combine computational methods and qualitative grounded theory in order to analyze classroom video data of middle school mathematics classrooms. These early findings involve the feasibility of using out-of-the-box implementations of video and audio processing algorithms for analysis of video and audio data, focusing on methods to capture instances of collaboration and student–teacher interactions.

Introduction

We view classroom learning environments as complex social systems in which learning—shifts in knowledge, its collective use, and the related patterns of interaction that demonstrate knowledge development in use—is an emergent outcome (Sherin & Star, 2011). Thus, understanding learning in classroom settings requires carefully attending to the dynamics of talk and interaction, especially over time.

Current research methodologies require Herculean efforts to conduct analyses that simultaneously attend to complexity and nuance at a large scale, especially using audio and video as rich sources of data. For example, there are strong qualitative traditions that do actively attend to—and even prioritize—visuospatial and/or acoustic features (e.g., Jordan & Henderson, 1995; Streeck, Goodwin, & LeBaron, 2011). However, these methods are incredibly arduous and time-consuming, making it all but impossible for individuals to carry out more than a few rich case studies. Qualitative studies that look across multiple contexts (e.g., comparing across 100 classrooms) and long time scales (e.g., tracking changes across multiple school years) are incredibly rare.

Given the importance and relevance of visual and auditory features of classroom interactions to those studying learning, there is a need for new methodologies for analyzing the social and visuospatial dimensions of classrooms in video/audio data, especially with the potential to do so at scale. We focus in this project on STEM classrooms in light of recent calls to support rigorous disciplinary talk and interaction in these subject areas.

The underlying premise of the project is that computational techniques can be used to extract relevant features from data—such as those that qualitative researchers would typically annotate when looking at video and audio data—amongst those features that humans can make meaning from and interpret. Accordingly, the project has three main objectives: 1) involve the development of computational techniques to extract features and explore patterns of visuospatial and auditory features from videos of STEM classrooms; 2) apply a methodological framework for utilizing computational grounded theory in novel ways to video data to analyze those patterns; and 3) demonstrate the use of these methods with large video datasets from STEM classrooms.

Methodological Framework

Computational grounded theory (Nelson, 2017) is a methodological framework that combines computational techniques with the commitments and aims of grounded theory. This methodology was developed as a response to the challenges of using either computational or grounded theory approaches on their own: the results of computational analyses are often hard to interpret in context or are not theoretically meaningful, and the results of grounded theory methods are often challenging to replicate or to scale. Computational grounded theory involves three steps: 1) pattern detection using computational exploratory analysis, 2) hypothesis refinement using human-conducted interpretive analysis, and 3) pattern confirmation. These steps generally reflect a movement from more inductive methods to those that are primarily deductive. The result of the three steps of computational grounded theory is a collection of clearly defined themes that emerge from a process of grounded conceptual development and which reflect both unanticipated empirical patterns and the aims, priorities, and subjectivities of the analyst.

A central challenge of this work is to develop methods that model the complexity, nuance, and variability found in STEM classrooms, but which are also interpretable in terms of the computational output, representations of extracted features, and the original data sources. This poster will present from the results from our first step of using this computational grounded theory approach: exploratory computational analyses, including any less-than-desirable solutions.

Data Source

The *Learning Through Teaching* (LTT) dataset will be the primary dataset used for the development of the computational methods (Dyer, 2016). This dataset includes 106 videotaped mathematics lessons from 10 teachers of grades 8-12 across two school years that was designed to capture student and teacher interaction and discourse. The videotaped lessons were spread out across a period of 3-7 months within in each school year, which allows for longitudinal analysis. Video recordings were collected from three camera angles and 8-10 additional audio sources were collected from different student groups. The data captured uses equipment and configurations common to classroom video data collection (Goldman, Zahn, & Derry, 2014; van Es et al., 2015). The extensive nature of the LTT data (i.e., the large number of teachers, observations, and video/audio angles) provides flexibility to explore how the amount of data and its quality matter for computational analyses.

Exploratory Computational Analyses

We are utilizing audio/video processing methods to extract higher-level information from low-level audiovisual data. For audio, we are applying the software program Praat (Boersma, 2001) to extract variables related to voice pitch, speaking turns and distribution, and other characteristics of speech. For video, we are applying a computer vision software package, called OpenPose (Cao, Simon, Wei, & Sheikh, 2017), which provides estimates of where students and teachers are in the classroom, their proximity to each other, and the degree of movement they exhibit. These variables from audio and video data will then be analyzed via computational grounded theory to discover possible learning activities like group formation, group discussion, and teacher–student interaction.

Results and Discussion

This poster will present early findings of the project that involve the feasibility of carrying out exploratory computational analyses using implementations of common algorithms for analysis of video and audio data. We will discuss the suitability of these algorithms and some of the modifications that need to be made in implementing these approaches with this type of classroom video data. For example, our initial video processing efforts revealed the need to downscale high-resolution video, given that processing speeds were slower than 1/300th of real-time playback speed. Also, we found that typical classroom video data presents many challenges with occlusion and subjects moving in and out of frame that these algorithms cannot easily handle. The poster format will allow for visual examples of many of the computer vision algorithms and in what ways they do and do not work for working towards answering theoretically meaningful research questions using classroom video data.

References

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-345.

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7291–7299.

Dyer, E. B. (2016). *Learning through Teaching: An Exploration of Teachers' Use of Everyday Classroom Experiences as Feedback to Develop Responsive Teaching in Mathematics* (Dissertation, Northwestern University).

Goldman, R., Zahn, C., & Derry, S. J. (2014). Frontiers of Digital Video Research in the Learning Sciences: Mapping the Terrain. In K. R. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 213–232). New York, NY: Cambridge University Press.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences, 4*(1), 39-103.

Nelson, L. K. (2017). Computational Grounded Theory: A Methodological Framework. *Sociological Methods & Research.*

Sherin, B. L., & Star, J. R. (2011). Reflections on the study of teacher noticing. In M. G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 66–78). New York, NY: Routledge.

Streeck, J., Goodwin, C., & LeBaron, C. (Eds.). (2011). *Embodied interaction: Language and body in the material world*. Cambridge University Press.

van Es, E. A., Stockero, S. L., Sherin, M. G., Van Zoest, L. R., & Dyer, E. B. (2015). Making the Most of Teacher Self-Captured Video. *Mathematics Teacher Educator, 4*(1), 6–19.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. DRL-1920796.