Apr 20th, 12:00 AM - 12:00 AM

Towards NLP-Based Conceptual Modeling Frameworks

David Shuttleworth
Old Dominion University

Jose Padilla
Old Dominion University

TOWARDS NLP-BASED CONCEPTUAL MODELING FRAMEWORKS

Shuttleworth, David[1], Padilla, Jose. J.[2]

[1]Old Dominion University
[2]Virginia Modeling, Analysis, and Simulation Center
dshut002@odu.edu, jpadilla@odu.edu

ABSTRACT

This paper presents preliminary research using Natural Language Processing (NLP) to support the development of conceptual modeling frameworks. NLP-based frameworks are intended to lower the barrier of entry for non-modelers to develop models and to facilitate communication across disciplines considering simulations in research efforts. NLP drives conceptual modeling in two ways. Firstly, it attempts to automate the generation of conceptual models and simulation specifications, derived from non-modelers' narratives, while standardizing the conceptual modeling process and outcome. Secondly, as the process is automated, it is simpler to replicate and be followed by modelers and non-modelers. This allows for using a common process and generating similar "blueprints" facilitating communication and collaboration efforts. Overall, NLP presents an opportunity for the M&S community to engage with stakeholders and scholars across domains in the simulation development process, lowering entry barriers and increasing participation.

Keywords: conceptual modeling, conceptualization, natural language processing

INTRODUCTION

Conceptual modeling exists within a spectrum of activities that lead to a simulation model. Beginning with the formation of mental models about some problem of interest to analyzing simulation outcomes, conceptual modeling – and resulting conceptual models – is key in the M&S process (Robinson, 2008). However, we are faced with the notion that "conceptual modeling is still more an art than a science" (Thalheim, 2018). Making that process "less art and more science" is an important endeavor to facilitate activities, such as theory-to-simulation traceability, that have implications in efforts of communication/collaboration in M&S teams and verification and validation of simulations. Theoretically, methodologically, and in practice, conceptual modeling has several challenges: lack of agreement on what conceptual modeling is and what it does; lack of agreement on what conceptual models are and what they provide; and how conceptual modeling is performed in practice among a group of multi-disciplinary participants.

This research focuses on transitioning problem descriptions to visual diagrams using NLP. The paper highlights two parts of the research: capturing diagrams from potential users and NLP-generated artifacts from narratives. User-generated diagrams provide an empirical baseline of what individuals extract from narratives that can be used for training NLP models and validating NLP-generated conceptual models.

METHODS AND PRELIMINARY RESEARCH

A qualitative research approach based on Glaw et al. (2017) is used to collect and analyze empirical data. Through interviews, insight is gathered about the artifacts generated by research participants; how decisions were made on what to include, and why a particular schema was used. The artifacts are assessed to identify, classify, and organize the elements contained within. These elements and their connections provide the basis for separating and organizing various concepts such as actions, events, actors, and factors. A collection of user-generated model artifacts is shown in Figure 1 and is from an initial group of four individuals presenting their perspectives of a short narrative about school transportation during the COVID-19 pandemic.

The proposed approach focuses on forming a bridge between informal descriptions, or narratives, that capture the problem of interest and visual representations in support of developing conceptual models. For example, Figure 2 illustrates several fragments of artifacts at varying levels of formality including a concept map, a flow diagram, and a formal sequence diagram. The naïve concept map is produced using an NLP-based entity and relation extraction, while the flow and sequence diagrams are generated using a Large Language Model (LLM). The LLM-produced diagrams are representative of the initial set of artifacts providing a baseline to build upon.

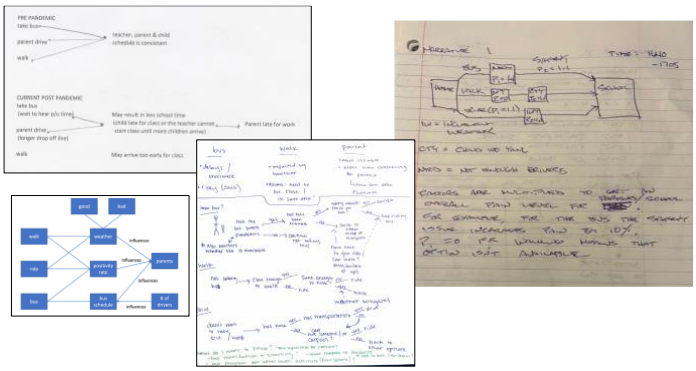

Figure 1. Example artifacts used to derive representative schema and content.

Figure 2. Example artifacts automatically produced via NLP.

CONCLUSION

Conceptual modeling is a key process in the design and development of models and simulations. Using NLP, we explore how to further the science of conceptual modeling moving towards facilitating theory-to-simulation traceability – knowledge of how observations, theories, and data become a simulation model, including assumption generation – and communication and collaboration when developing models in interdisciplinary teams. Interdisciplinary work is becoming increasingly common as we address challenges such as climate change, refugee studies, and community resilience where scholars and practitioners in engineering, social science, climate science, anthropologists, non-governmental organizations, and governments need to share not only information but more importantly how individuals perceive the problem. Common conceptual models aid communication, understanding, and collaboration as perspectives from each participant are explicit and unambiguous without complex engineering artifacts.

REFERENCES

Glaw, X., K. Inder, A. Kable and M. Hazelton. 2017. Visual Methodologies in Qualitative Research. International Journal of Qualitative Methods, 16(1): 1-8.

Robinson, S. 2008. Conceptual Modelling for Simulation Part I: Definition and Requirements. Journal of the Operational Research Society, 59(3): 278-290.

Thalheim, B. 2018. Conceptual Model Notions – A Matter of Controversy. Enterprise Modelling and Information Systems Architectures, 13(1): 9-27.