Understanding Narrative: Computational Approaches to Detecting Narrative Frames

Eva Portelance
eva.portelance@mail.mcgill.ca
McGill University, Canada

Andrew Piper
andrew.piper@mcgill.ca
McGill University, Canada

Understanding narrative structure at a large scale remains a challenging problem within the field of cultural analytics and computational linguistics. Our aim with this project is to develop novel methods to study the pacing of narrative scene changes and the overall distribution of different plotlines within novels. Being able to analyze such narrative features at large scale can give us insights into the way different genres, time periods, or cultures favor different modes of storytelling. In this project we formalize definitions of narrative scenes and implement new methods of detection and clustering using computational methods.

The project involves three steps: creating an operational concept of a narrative frame; the algorithmic segmentation of narratives by frames; and the predictive clustering of frames into larger-scale "plotlines."

We define a "frame" as a significant shift of three variables in a given text window: entities, actions, and objects, which we represent using POS tagging as proper names, verbs, and nouns. We measure significant lexical shift of our three primary variables in a sliding textual window of 1000 words with increments of 100-word shifts. We test window-size and variable selection relative to human annotation to determine the best performing model. We resolve frames into "plotlines" using hierarchical clustering, also demonstrated in our poster. "Frames" serve as inferred textual units and "plotlines" as aggregated clusters of frames.

We have tested the performance of different com-bination of variable selection on nine 12,000-word passages from novels of different genres from a range of time periods (from 1818 to 2011). To date, our algorithm outperforms the current state-of-the-art in Hearst's Texttiling algorithm (1994; 1997) when it comes to placing breaks in the narrative event progression. The performance of our system relative to human performance on the same task (F1 82%, Precision 81%, Recall 86%), shows an F1 score of 69% with 71% precision and 67% recall, where ⅔ annotators agreement is considered to be a true frame boundary. Hearst's method applied to the same problem performs at a significantly lower rate (F1 18%, Precision 18%, Recall 19%). While the overall problem remains challenging we show significant improvement over state of the art systems at detecting narrative segments. Nevertheless, the imperfect accuracy suggests that scene changes have a number of subtle variables that are not exclusively tied to vocabulary or character shifts, which indicate further avenues for future research.

Our poster will present our formalization of narrative events, the results and approach of the segmentation task and the clustering models used. We see this project as a crucial contribution to the larger study of narrative form across different literary genres and time periods.

Bibliography

Hearst, M. (1994). "Multi-Paragraph Segmentation of Expository Text." *Proceedings of the 32nd Meeting of the Association for Computational Linguistics*, Los Cruces, NM.

Hearst, M. (1997). "TextTiling: Segmenting Text into Multi-Paragraph Subtopic Passages." *Computational Linguistics*, 23 (1), pp. 33-64.