University *of Ljubljana*
Faculty *of Computer and Information Science*

Literacy situation models knowledge base creation

Andrej Drofenik, Enej Bačić and Sebastjan Tkavc

Abstract

The abstract goes here. The abstract goes here.

Keywords

Keyword1, Keyword2, Keyword3 ...

Advisors: Slavko Žitnik

Introduction

The main goal of this task is to create a knowledge base based on situation models [1] from selected English short stories of varying lengths and levels of vocabularic complexities. We will tackle the extraction of main characters and their role in the respected work, as well as the relationships between them. Time permitting, we will perform further analysis of patterns of word-usage to describe the themes of the short stories selected. We plan to perform a combination of semantic analysis and spatio-temporal analysis.

Related work

Extensive research has been done on narrative comprehension, the requirements of situation models and implementation of such models.

A recent study [2] has shown that local semantic relations significantly influence recall of paired sentences in L2 readers. Results show that the global casual relations and local semantic relations have a large impact on a reader's memory. This provides insight to assessing text meaning.

As semantic relations are beneficial to meaning assessment in text, we looked at an overview of the field of semantic analysis in natural language processing. The paper by Salloum et al. [3] provides an insight on methods such as latent semantic analysis, explicit semantic analysis and sentiment analysis and the overall importance of semantic analysis.

The paper by Zwaan et al. [4] researched the importance of three dimensions of situational continuity. These consist of temporal, spatial and casual continuity. The authors have also shown that readers simultaneously monitor more than one dimension under normal reading instruction, temporal and casual having the most impact.

In the work of Dasgupta et al. [5] automatic extraction of cause-effect relations using their proposed bi-directional LSTM model with an additional linguistic layer. They achieved better performance than other methods. A product of the research is also an annotated dataset in the sense of cause-effect relations.

The paper Extraction and Analysis of Fictional Character Networks: A Survey [6] provides information on the entire process of character networks. It explains the steps necessary to construct such a network, such as character identification, interaction detection, graph extraction. It outlines the current situation in the field, the methods and performance.

Datasets:

- We have an initial dataset containing 7 short stories, ranging from about 1700 to 8700 words
- A dataset of 12 longer short stories, ranging from about 7600 to 60000 words, freely available on Project Gutenberg [7]
- The Event StoryLine dataset for Casual and Temporal Relation Extraction [8, 9, 10, 11]
- Corpus of Common sense stories and possible semantic parses [12]

- The possibility of extracting short stories from the Reedsy short stories blog [13]. Might pose licensing issues, couldn't find any info on page.

Road-map

We will begin with exploring the available datasets, annotations and the ways in which they can be useful. We will perform the basic text processing and clustering on the acquired short stories, following with managing annotation of the data. This will provide us with some insight on how to proceed. Our plan is to identify the main characters of the short stories and their role in them. We plan to explore semantic analysis of our dataset (literature indicates that local semantic context significantly impacts the meaning assertion):

- latent semantic analysis
- sentiment analysis (RNN, Bert)

The literature shows that humans usually keep track of more dimensions than just one. Our plan would be to perform spatio-temporal analysis and see if the combination of such dimensions results in meaningful performance.

References

[1] Rolf A. Zwaan. Situation models: The mental leap into imagined worlds. *Current Directions in Psychological Science*, 8(1):15–18, 1999.

[2] Shingo Nahatame. Revisiting second language readers' memory for narrative texts: The role of causal and semantic text relations. *Reading Psychology*, 41(8):753–777, 2020.

[3] Said A. Salloum, Rehan Khan, and Khaled Shaalan. A survey of semantic analysis approaches. In Aboul-Ella Hassanien, Ahmad Taher Azar, Tarek Gaber, Diego Oliva, and Fahmy M. Tolba, editors, *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, pages 61–70, Cham, 2020. Springer International Publishing.

[4] Rolf A Zwaan, Joseph P Magliano, and Arthur C Graesser. Dimensions of situation model construction in narrative comprehension. *Journal of experimental psychology: Learning, memory, and cognition*, 21(2):386, 1995.

[5] Tirthankar Dasgupta, Rupsa Saha, Lipika Dey, and Abir Naskar. Automatic extraction of causal relations from text using linguistically informed deep neural networks. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 306–316, 2018.

[6] Vincent Labatut and Xavier Bost. Extraction and analysis of fictional character networks. *ACM Computing Surveys*, 52(5):1–40, Sep 2020.

[7] Project gutenberg.

[8] Github - tommasoc80/eventstoryline: Event storyline corpus - annotated data, baselines and evaluation scripts, evaluation data.

[9] Tommaso Caselli and Piek Vossen. The storyline annotation and representation scheme (StaR): A proposal. In *Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016)*, pages 67–72, Austin, Texas, November 2016. Association for Computational Linguistics.

[10] Tommaso Caselli and Piek Vossen. The event StoryLine corpus: A new benchmark for causal and temporal relation extraction. In *Proceedings of the Events and Stories in the News Workshop*, pages 77–86, Vancouver, Canada, August 2017. Association for Computational Linguistics.

[11] Tommaso Caselli and Oana Inel. Crowdsourcing storylines: Harnessing the crowd for causal relation annotation. In *Proceedings of the Workshop Events and Stories in the News 2018*, pages 44–54, 2018.

[12] Nasrin Mostafazadeh, Nathanael Chambers, Xiaodong He, Devi Parikh, Dhruv Batra, Lucy Vanderwende, Pushmeet Kohli, and James Allen. A corpus and cloze evaluation for deeper understanding of commonsense stories. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 839–849, 2016.

[13] 25000+ best short stories to read online for free with reedsy prompts.