Discourse Relation Parsing for Long Dialogues

Duaa Tariq
Columbia University.

Introduction: Discourse relation parsing is a research area in Natural Language Processing for analyzing discourse structure by identifying relations. Most previous works are limited to written texts which are well-structured, unlike spontaneous dialogues. I will be contributing a dataset consisting of discourse relations annotated for long dialogues, specifically transcripts of TV show episodes from the Forever Dreaming collection (Chen, 2022).

Methods: Previous research on dialogues includes the STAC dataset (Asher, 2016) which is an annotated multi-party chat log for an online game. After a testing trial of the STAC corpus, I defined discourse relations for the transcripts. I have then annotated an episode using Inception, an online web application. Based on my findings, I revised the annotations and redefined relations. I have noted relevant challenges or new concepts. I also recorded the frequencies of the different types of relations.

Results: I found that certain relations such as "comment" and "continuation" were more prevalent due to the structure of longer dialogues. I observed the use of rhetoric and sarcasm in utterances which were more difficult to define, as expected with creative language. I compiled the data onto a table as well as a bar graph showing the frequencies of 21 different discourse relation types.

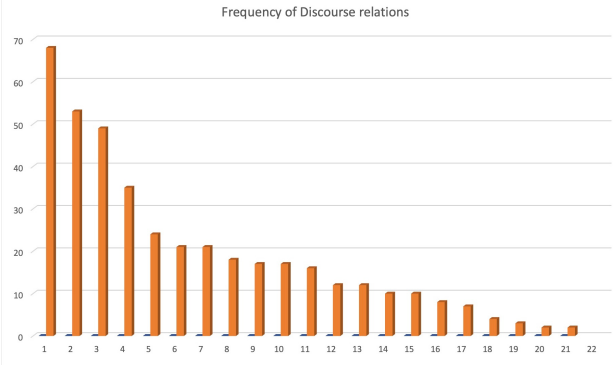

Figure 1 (Sample 1)

Conclusions: There is a major difference between long multi-party dialogues and other written texts. However, to contribute a meaningful dataset, a sample size of at least 1000 episodes is likely required. Relation labels may also still need to be redefined and a variety of TV show genres should be explored.

References:

1) Tools- Inception: Technische Universität Darmstadt -- Computer Science Department -- INCEpTION -- 23.8 (2022-06-29 21:24:41, build 4b271961

2) Forever Dreaming Transcripts: Chen, M., Chu, Z., Wiseman, S., & Gimpel, K. (2022, June 6). *SummScreen: A dataset for abstractive screenplay summarization*. arXiv.org.RetrievedJuly28,2022,from https://arxiv.org/abs/2104.07091

3) STAC data set: Asher, N., Hunter, J., Morey, M., Farah, B., & Afantenos, S. (n.d.). *Discourse structure and dialogue acts in multiparty dialogue: The stac corpus*. ACL Anthology. Retrieved July 28, 2022, from https://aclanthology.org/L16-1432/

4) Li, J., Liu, M., Qin, B., & Liu, T. (2022, January 20). *A survey of discourse parsing - frontiers of computer science*. SpringerLink.RetrievedJuly28,2022,from https://link.springer.com/article/10.1007/s11704-021-0500-z

5) Shi, Z., & Huang, M. (n.d.). *A deep sequential model for discourse parsing on multi-party dialogues*. Proceedings of the AAAI Conference on Artificial Intelligence. Retrieved July28,2022,from https://ojs.aaai.org/index.php/AAAI/article/view/4680

Acknowledgements:
Yilun (Bobby) Hua
Professor Kathleen McKeown