Ph.D. Research Summary

Chuyuan Li

My research focuses on discourse analysis, mainly in dialogue settings. A document is not a random and independent text spans, but instead sequences of ordered and related sentences which together make coherent and meaningful documents: this organization is called discourse structure [6]. I am particularly interested in understanding the connection between sentences: how they interact with each other, what is the inner logic, and how can we represent the coherent structure.

In Natural Language Processing (NLP), discourse analysis is language processing beyond the sentence boundary. It refers to the retrieval of the inherent structure of coherent text, which include different levels of analysis such as topic structure: lexical signals and work distribution indicate topic shifts, referential structure: coreference links between pronouns and entities in order to create local coherence, and coherence-relational structure: two text spans are linked together with specific semantic relation using explicit or implicit connectives [19].

Discourse processing is a high-level form of linguistic analysis that examines the inter-sentence organization of a text, taking into consideration semantic and pragmatic context, as opposed to focusing solely on word or sentence-level coherence as in lexical or syntactic analysis. My research journey started with simple discourse markers that show local textual coherence using, for instance, rhetorical connectives and dialogue acts, I then progressed to complete discourse analysis with discourse structure extraction using distant signals from auxiliary tasks.

I present an overview of my main projects in Fig. 1. With the objective of better understanding discourse structure in dialogues, I performed two directions of research: the former is discourse structure discovery where I made use the discourse markers for classification tasks. Two tasks were conducted in cognitive impairment field: the first one being Schizophrenia detection, which led to two publications [1, 9] and a few communication talks including French national Health and Language Seminar and Semantics and Pragmatics of Dialogue Workshop (SemDial 2021); the second one being depressive detection with one international publication [10]. The second direction is discourse structure prediction where I aim to directly predict a formalized discourse structure (graph-like structure) using few (semi-supervised) or no (unsupervised) annotated data. This project was a collaborative work between myself and my collaborators at the University of British Columbia in Vancouver, during my six-month internship at the UBC NLP group. It resulted in a publication in a highly regarded international NLP conference [11].

Figure 1: Research projects overview.
References


