Empirical Analysis for Disambiguating Organization names

Name

Yuya Sawada

Laboratory's　name

Natural Language Processing Laboratory

Supervisor's　name

Taro Watanabe

Abstract (should be within 1st page)

News literature often reports information related to organizations, with frequent mentions of organization names. To enable quick access to relevant information in a knowledge base through organization names, it is crucial to disambiguate entity mentions by pseudonyms and synonyms. In this dissertation, we focus on companies that are part of organizations and tackle analyzing the challenges of disambiguating company names.

Firstly, we created a dataset that annotated company names from 3025 news articles and discussed that linking companies has several technical challenges: unknown companies, other companies with the same, and temporal changes. In experiments, we implemented comprising a company extraction and a similarity base entity linking model using the pretrained language model, and our system demonstrates better performance than the existing company linking system. However, the results of experiments suggest that our system has still room for improvement in the challenges. We then focus on the disambiguation of temporal changes in company names and aim to structurize the changes in company names. Specifically, we propose a novel event extraction task to extract information related to company name changes as events. This task defines five types of events related to company name changes and mergers in company history documents, and we created a dataset using company history documents in Japanese annual securities reports. In experiments, while LLMs can correctly extract events to a certain extent, even with a small number of examples, there are challenges for extracting events that involve the relationship between multiple companies that have three factors: low frequency, long distance between arguments, and the ambiguity of procedures.