先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	自然言語処理学 (渡辺 太郎 (教授))		
学籍番号	2011039	提出日	令和 4年 1月 19日
学生氏名	岩田 晟		
論文題目	QA-Span base Zero-Pronoun Identification		
要旨			

 In a pro-drop language, such as Japanese, Chinese, and Arabic, phrases containing nouns called arguments, such as SBJ and OBJ, are omitted, and the omission is called Zero-Pronoun(ZP). The existence of this ZP affects the accuracy of downstream tasks in NLP. Therefore, for pro-drop language-based NLP, the problem of ZP needs to be tackled. In recent years, fine-tuning on pre-trained models with large-scale data has become the mainstream in the field of NLP. Some papers reported that solving non-Question-Answering(QA) tasks with a QA fine-tuning method improved performance and task-specific problems in various non-QA tasks.

  To tackle the ZP problem, the model needs to detect the omission of the ZP, for which the predicate corresponding to the arguments is an essential cue. However, existing methods using pre-training models give only the whole sentence and label each word whether ZP or not.

This paper aims to improve the ZP identification task's performance by formulating the ZP task as QA-task.

We experimented on Japanese and Chinese datasets and confirmed that the proposed QA-based method surpassed the sequence labeling baseline.