JAIST Repository
https://dspace.jaist.ac.jp/

Title	A Study on Joint Extraction of Entities and Relations based on Constructing Differentiated Subtask-Specific Features and Enabling Fine-Grained Information Interaction
Author(s)	王, 堯
Citation	
Issue Date	2024-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18904
Rights	
Description	Supervisor: NGUYEN, Minh Le, 先端科学技術研究科, 修士(情報科学)

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Abstract

Extracting entities and relations from raw texts is a crucial and challenging task in the field of *Information Extraction*. Despite the successes achieved by the traditional approaches, fundamental research questions remain open. First, the subject and the object are assumed to have the same impact on their corresponding relation, ignoring the possibility of their differences when either the subject or the object is a complex entity, such as polysemous words, pronouns, and abbreviations. For instance, given a relational triple <Apple, isHeadquarteredIn, Cupertino>. The polysemous entity "Apple" may have a more important impact on the definition of the relation "isHeadquarteredIn" than the object "Cupertino". Because, if the "Apple" refers to the fruit, their relation will be of a "isProducedIn" type. Second, the information interaction mainly occurs between the subtasks of extracting the entity and relation, leaving the fine-grained interaction among the task-specific features of subjects, relations, and objects unexplored.

Motivated by the aforementioned limitations, we propose a novel model to jointly extract entities and relations. The main novelties are as follows: (1) During the encoding phase, we decouple *the whole task of jointly extracting entities and relations* into three subtasks, namely *named subject recognition*, *relation extraction* and *named object recognition*. Thanks to this, we are able to use fine-grained subtask-specific features. (2) We propose novel inter-aggregation and intra-aggregation strategies to enhance the information interaction and construct individual fine-grained subtask-specific features, respectively. (3) In the decoding phase, we combine subtask-specific features of the subject and the object to predict entities and incorporate them to enhance entity representation in the relation extraction subtask.

In order to well evaluate the effectiveness of the proposed method for jointly extracting entities and relations, we conducted a series of experiments based on seven benchmark datasets by comparing with many representative approaches. The experimental results demonstrate that: (1) when either the subject or the object is a complex entity, it has a greater impact on their corresponding relations than a normal entity. (2) Constructing fine-grained subtask-specific features for extracting the subject, the object, and their relation can improve the extraction ability. (3) Our model outperforms several previous state-of-the-art models. In specific, we increase the accuracy score by +2.7%, +0.1%, +0.6%, and +0.6% in the relation extraction task on ACE2004, ACE2005, ADE, and CoNLL04 datasets and +0.3%, +0.6%, +0.5%, +0.1%, and +0.1% in the entity extraction task on ACE2004, ADE, SciERC, NYT, and WebNLG datasets, respectively.