論 文 内 容 の 要 旨

博士論文題目

Relation Extraction: Perspective from Various Supervised Approaches

(関係抽出:様々な教師あり学習手法に基づく観点)

氏　　名　TRAN Van Hien

(論文内容の要旨)

Information extraction transforms unstructured text into structured information on raw data. A vital step in information extraction is relation extraction, which aims to identify semantic relationships between named entities in text. The extracted relations help construct knowledge bases and support various natural language processing applications such as information retrieval and question answering.

Relation extraction has been widely studied in a fully supervised learning approach by training models on large-scale labeled data. Following this approach, existing supervised models have achieved excellent performance. However, these supervised models cannot solve relation extraction in real-world scenarios, such as recognizing new relations or identifying entities and their relations jointly.

This research focuses on two other supervised approaches for relation extraction task, namely zero-shot relation extraction and end-to-end relation extraction. These two supervised approaches help solve relation extraction in real-world scenarios, which are more realistic and challenging.

The first part of the studies addresses zero-shot relation extraction, which aims to recognize (new) unseen relations that cannot be observed during training. We propose two new methods to improve task performance. In the first method, we present a new model that mainly boosts discriminative feature learning on both sentence and relation spaces. This model is also equipped with a self-adaptive comparator network to judge whether the relationship between a sentence and a relation is consistent. Experimental results show that the proposed method significantly outperforms the state-of-the-art methods. In the second method, we argue that

enhancing the semantic correlation between instances and relations is key to solving the zero-shot relation extraction task effectively. A new model entirely devoted to this goal through three main aspects was proposed: learning effective relation representation, designing purposeful mini-batches, and binding two-way semantic consistency. Experimental results on two benchmark datasets demonstrate that our approach significantly improves task performance and achieves state-of-the-art results.

The second part of the studies concentrates end-to-end relation extraction, which aims to detect entity pairs along with their relations to extract relational triplets. We propose an improved decomposition strategy that overcomes two major problems of the previous decomposition strategy by Yu et al. (2020). Our improved decomposition strategy considers each extracted entity in two roles (head and tail) and allows a model to predict multiple relations (if any) of an entity pair. In addition, a corresponding model framework is presented to deploy our new decomposition strategy. Experimental results show that our method significantly outperformed the previous method of Yu et al. (2020) and achieved state-of-the- art performance on two benchmark datasets. Besides, we also present CovRelex (Tran et al., 2021), a scientific paper retrieval system that can automatically detect both entities with various types and their diverse relations through papers, primarily when COVID-19 articles are published rapidly. The system aims to support users efficiently in acquiring such knowledge across many COVID-19 scientific papers.

氏　名	TRAN Van-Hien

（論文審査結果の要旨）

Information extraction transforms unstructured text into structured information on raw data. A vital step in information extraction is relation extraction that identify semantic relationships between named entities in text which are used to construct knowledge base to support various natural language processing applications. Relation extraction has been widely studied in a fully supervised learning approach by training models on large-scale labeled. However, these supervised models cannot solve relation extraction in real-world scenarios, such as recognizing new relations or identifying entities and their relations jointly.

This research focuses on two challenging real-world relation extraction tasks, namely zero-shot relation extraction and end-to-end relation extraction. Zero-shot relation extraction aims to trecognize unseen relations that cannot be observed during training, while end-to-end relation extraction performs entity pair detection along with their relations to extract relation triplets.

Two methods have been investigated for zero-shot relation extraction: The first method discriminatively learns the features on both the sentence and relation spaces with self-adaptive comparator network. The second method learns representations for instances and relations through contrastive learning framework. The two methods have achieved the state-of-the-art results with significant gains under standard benchmarks. In the studies for end-to-end relation extraction, this thesis considered each extracted entity in two roles, i.e., head and tail, and allows a model to predict multiple relations. Empirical results have shown state-of-the-art performance on two benchmark datasets.

The research in this thesis demonstrated that the proposed relation extraction methods achieved significant gains when compared with strong baselines. In addition, the end-to-end relation extraction model is deployed in a COVID-19 scientific paper retrieval system. The studies are published in one high quality peer-reviewed journal papers and two peer-reviewed international conference papers. The research would have an impact to relevant fields in natural language processing, e.g., name entity recognition and end-tasks, such as question answering. As a result, the thesis is sufficiently qualified as a Doctoral thesis of Engineering.