論 文 内 容 の 要 旨

博士論文題目

Structured Representation Learning for Structured Prediction

(構造化予測のための構造化表現学習)

氏　　名　Yiran Wang

(論文内容の要旨)

Structured prediction tasks, e.g., part-of-speech tagging, (nested) named entity recognition, and constituency parsing, are considered to be fundamental and essential techniques of natural language processing. In recent years, emerging deep-learning models, especially pre-trained language models, have provided fantastic ways to obtain informative representation, and have been continuously refreshing the leaderboards of these tasks. However, existing models typically employ universally applicable representation learning techniques, often overlooking the unique characteristics inherent to each specific task. Furthermore, lack of interpretability also keeps them a black box to humans, and the inability to explain their decision-making mechanism hindered researchers from further improving them.

This dissertation mainly focuses on leveraging the task-specified characteristics of these structured prediction tasks to learn structured and interpretable representations for solving these tasks.

First of all, the first method factorizes representation according to the hierarchical structures of the nested named entity recognition task. A carefully designed algorithm is introduced to explicitly exclude the harmful influence on the best path of previous levels. By additionally introducing the chunk selection strategies and switching the encoding to be the innermost first scheme, the proposed method obtained level-wise representation and also pushed the performance to better results.

Moreover, the second method also factorizes the representation for the conventional structured prediction tasks. With the proposed contrastive hashing methods, narrowing the representation bottleneck to be only 24 bits becomes possible without significantly

sacrificing performance. These learned discrete bits demonstrated completely preserving the necessary features of the downstream tasks, therefore, providing researchers with a more interpretable tool to analyze the internal mechanism of the black-box neural networks.

The main contribution of this dissertation is the proposal of two representation learning methods to learn structured representation for structured prediction tasks. Numerous experiments and discussions are further provided to show the effectiveness and efficiency of the methods.

氏　名	Yiran Wang

（論文審査結果の要旨）

Structured prediction tasks, e.g., part-of-speech tagging, (nested) named entity recognition, and constituency parsing, are considered to be fundamental and essential techniques of natural language processing. In recent years, emerging deep-learning models, especially pre-trained language models, have provided fantastic ways to obtain informative representation, and have been continuously refreshing the leaderboards of these tasks. However, existing models typically employ universally applicable representation learning techniques, often overlooking the unique characteristics inherent to each specific task. Furthermore, lack of interpretability also keeps them a black box to humans, and the inability to explain their decision-making mechanism hindered researchers from further improving them.

This thesis mainly focuses on leveraging the task-specified characteristics of these structured prediction tasks to learn structured and interpretable representations for problem solving. The first method factorizes representation according to the hierarchical structures of the nested named entity recognition task. A carefully designed algorithm is introduced to explicitly exclude the harmful influence on the best path of previous levels together with the chunk selection strategies and level-wise information encoding schemes. The second method also factorizes the representation for the conventional structured prediction tasks using contrastive hashing methods. The learned discrete bits completely preserved the necessary features of the downstream tasks without significantly sacrificing performance. Therefore, it provides researchers with a more interpretable tool to analyze the internal mechanism of the black-box neural networks.

The two studies are published in one high quality peer-reviewed journal paper and two peer-reviewed international conference papers. The research would have an impact not only to the natural language processing demanding structured outputs and interpretable representations, but to the relevant fields of machine learning, e.g., vision or speech. As a result, the thesis is sufficiently qualified as a Doctoral thesis of Engineering.