様式８

論　文　内　容　要　旨

報告番号	甲　先　第　391　号	氏　名	薛嗣媛
学位論文題目	Research on Intention Detection in Dialogue System (対話システムにおける意図認識に関する研究)		

内容要旨

With the advent of speech recognition technology, the applications of intelligence interaction systems have become important in daily life. The products like Siri of Apple, Google Home, and Amazon Echo can help the user complete some basic tasks efficiently. The spoken language understanding (SLU) module is an indispensable component in these dialogue systems. A typical SLU module is designed to transform the spoken language into a specific semantic template that human language can be well-understood by the dialogue system. After that, the dialogue management module can facilitate future actions according to detection results in the SLU module. The role of the intention detection task in SLU is to discriminate the implicit intention by recognizing the intents of received utterances. The intent is a semantic label attached with each utterance in dialogue, which represents the user's intention and concise utterance interpretation. Therefore, the intention detection task is crucial to enhance the spoken language understanding performance in the dialogue system. However, the various expressions of user's intents and constantly emerging novel intents make the annotating time-consuming and laborious, building massive obstacles for extending the model to new tasks. Therefore, the zero-shot intention detection task attracts interest in industry and academia.

This thesis mainly studies and analyzes intention detection task and their application of zero-shot intent detection learning. This thesis first introduces the background knowledge of intention detection and zero-shot intention detection. Besides, the motivation and its significance of intention detection have expatiated. In the end, the organization of the whole thesis is illustrated.

Then, we describe the experiments of intention detection in detail. According to the real situation, it is challenging to study the spoken language because of some attributes of natural language. Firstly, the sparsity of semantic information and obscure slang in spoken language makes the model difficult to interpret thoroughly. Secondly, the same underlying utterances have different tags or multiple tags, which gives rise to ambiguity in classifying intention labels. Thirdly, the prior works of multi-class classification of intention detection exploit SoftMax to train an encoder on labeled training data. The learned features are optimized under the supervision of SoftMax, which cannot be sufficiently distinguished because it does not consider the intra-class compactness of features. Therefore, we propose a triplet training framework learns discriminative utterance feature

by using the same weights on different inputs. The triplet loss function infers a non-linear mapping in the resulting latent space, and the inter-class sample distances are maximized based on a certain margin. The results illustrate that the proposed method can effectively improve the recognition performance of these datasets and achieves new state-of-the-art results on single-turn task-oriented datasets (Snips dataset, Facebook dataset), and a multi-turn dataset (Daily Dialogue dataset).

Furthermore, we want to realize the discrimination of unknown intentions. In the chapter 3, we illustrate the zero-shot intent detection in detail. In zero-shot intent detection task, we propose an intention-enhanced attentive Bert capsule network (IE-BertCapsNet) to learn hierarchical utterance features to detect intentions. In particular, we design the intention-enhanced BERT model as the primary capsule to learn a more intention-concentrated hierarchical utterance feature. Then, we aggregate them to discriminate intents by utilizing an attentive capsule network and routing by agreement mechanism. The self-attention mechanism in the capsule improves the model to learn the different contributions of the capsules. Coupling with the large margin cosine loss function, the proposed model can identify discriminative features by forcing the whole network to minimize inter-class distance and minimize intra-class distance. Finally, we leverage the IE-BertCapsNet's feature extraction ability and knowledge transferring capability to conduct unknown intention detection task. Extensive experiments on five benchmark task-oriented datasets in four languages demonstrate that the proposed model can achieve competitive performance that can better discriminate known intents and detect unknown intents.

In the final section, we comprehensively conclude the advantage and weakness of this thesis based on the existing research. In the future, we would like to improve the performance of intent detection with higher stability and scalability and expand our work to meet real-life requirements like generalized zero-shot intention detection and improve its performance. Besides, we would like to explore more variants of capsule network structure to satisfy the pressing needs of other natural language processing tasks.