

# SFC: FEW-SHOT TEXT CLASSIFICATION VIA SIMILARITY FUSED WITH CLASSIFICATION SYSTEM

Jingwen Huang

{hanmei613}@gmail.com

## ABSTRACT

Building conversational chatbot system has become a popular solution to sharing the work of customer service under various business scenarios. One of the most important applications is the single-turn conversational chatbot aiming at user intent detection, which essentially equals to short-text classification problem in the field of Natural Language Processing. However, each time for a new service, the chatbot system often needs to perform well in few-shot setups due to lack of domain-specific data, which is still quite hard even if we use powerful pretrained model like Roberta. Therefore, in this paper, we propose SFC, a system fusion of both similarity model and classification model to overcome this challenge. Our main contributions are: 1) transfer learning and negative sampling based on sentence-pair are utilized as remedy for lack of data; 2) multi-task learning is involved to achieve faster training speed and better performance; 3) model ensembling of classification and similarity model guarantees inference speed while keeping high accuracy. Additionally, we also conduct extensive experiments on four public datasets in few-shot setup (i.e., with only 5 to 20 training data per class). The experimental results show that our system can steadily outperform several competitive baselines by 2 percent in average accuracy.

**Index Terms**— stress, detection, attention, neural network

## 1. INTRODUCTION

Single-turn conversational chatbots are designed to transform existing tasks that rely on human agents, such as classifying customers' questions or queries to find the corresponding answer, into an automatic process based on intent classification model. Since user queries are usually much shorter than paragraphs or documents, the chatbot can actually be turned into a short-text Classification task in Natural Language Processing. Moreover, at the initial stage of building chatbots, it's usually extremely hard to collect sufficient data for each class. In this way, building a single-turn conversational chatbot become a short-text classification problem under few-shot setting (Yu et al., 2018).

Short texts are usually more ambiguous in comparison with long texts since they don't contain enough context-

tual information, which poses a great challenge for classification (Chen et al., 2019). In addition, the few-shot scenario (Yu et al., 2018) adds even more difficulties to the classification task since there is no enough information for the model to learn for each class. Comparing to non-pre-trained neural network such as convolutional neural networks (CNNs) (Kim, 2014) and long short term memory networks (LSTMs) (Mousa and Schuller, 2017), the recently introduced pre-trained language models on large corpus like BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019) has been proven more powerful in solving many NLP tasks including short-text classification for deficient data (Madabushi, H. T., 2020). Especially for few-shot scenarios, transfer learning based on pre-trained model tends to do more help to the negative effects brought by scarcity of training data. Due to the fact that RoBERTa is an improved version of BERT, we build our SFC chatbot system using RoBERTa as pre-trained context-dependent embeddings. To our knowledge, the most common approach is adding a softmax classifier to the top of RoBERTa (i.e., RoBERTa classifier), which turns the problem into a simple text classification task. However, despite the fact that RoBERTa has achieved amazing results in many Natural Language Processing tasks, we still believe it has much more potential under few-shot setting.

Therefore, in this paper we propose to involve similarity model (i.e., sentence-pair classifier) based on RoBERTa, which can score the semantic similarity between two sentences by multiple cross-attention mechanism (Barkan et al., 2019). A natural idea to enhance model performance in few-shot setting is to further pre-train RoBERTa with target domain data for transfer learning (Sun et al., 2019). However, it's always not easy to find domain-specific data for certain service/product if we try to further pre-train a RoBERTa classifier directly. In comparison, it's relatively feasible to obtain dataset for semantically duplicate sentence pair identification task. That is to say, we can obtain a further pre-trained similarity model to identify the class with highest semantic similarity level to each user query. Afterwards, we can fine-tune the similarity model using sentence pairs sampled from target task dataset based on negative sampling strategy (Bamler and Mandt, 2020), which can provide us with more features to learn from limited amount of data.

We also apply multi-task learning (Caruana, 1993; Col-

lobert and Weston, 2008) in training process to enhance the training speed and performance. We set up two different objectives for training. The first one is the regular sentence-pair similarity score which help the model learn the semantic similarity between a query to each existing data point. The second target is to learn the similarity score between a query and all the data points of a certain class as a whole. Specific knowledge contained in these two tasks can be fully explored to obtain a faster training speed and higher accuracy.

Another contribution of this work is that we ensemble the similarity model with classification model at the inference step. In contrast with common model ensembling methods, the classification model are used as an auxiliary model in our system to select promising sentence-pair candidates for similarity model. The experiment results on 4 short-text classification datasets in various few-shot settings show that our system can outperform single model baseline by at least 2 percent on average. Besides, we can also control the inference time for one single query to be within 0.5 seconds, which makes the system applicable in real-life single-turn chatbot scenario.

## **2. ATTENTION BASED NEURAL NETWORK**

## **3. CONCLUSION**

In this work, we successfully applied two widely used attention techniques into lexical stress detection problem. Our proposed model directly takes time series features as input that it could fully explore input information; and the network structure intrinsically support words with different number of syllables, without expanding short words, to reduce input approximation. Experiments also demonstrate the improvements over baselines systems.