icsPLMs: Exploring Pre-trained Language Models in Intelligent Customer Service (Student Abstract)

Shixuan Liu[1], **Chao Wang**[2], **Shuangyong Song**[2]
[1]Australian National University, ACT 2601, Canberra, Australia
[2]China Telecom Corporation Ltd. Data&AI Technology Company, Beijing, China
shixuan.liu@anu.edu.au, {wangc17,songshy}@chinatelecom.cn

Abstract

Pre-trained language models have shown their high performance of text processing in intelligent customer service platforms. However, these models do not leverage domain-specific information. In this paper, we propose icsPLMs optimized for ***i***ntelligent ***c***ustomer ***s***ervice on both word and sentence levels. Our experimental results represent that using targeted strategies can further improve the performance of pre-trained language models in this field.

Introduction

In recent years, there has been a large demand for intelligent assistants on customer service platforms which could automatically provide solutions for users. The intelligent assistants can not only help users rapidly, but also reduce the pressure for customer service staff. In details, the inefficiency and long-waiting problems of traditional customer service can be solved by offering common solutions.

The most important and essential part of an intelligent assistant is identifying users' intents and providing appropriate responses, where query matching (Yu et al. 2018) and intent classification (Song, Chen, and Shi 2017) tasks are its applications. For most natural language processing tasks, the paradigm of pre-training coupled with fine-tuning has become the current mainstream approach such as StructBert (Wang et al. 2019). Although these models provide good generalization ability, the textual information of intelligent customer service domain is not fully exploited. eMLM (Sosea and Caragea 2021) model provides a masking method towards sentiment words which improve the performance on emotion-related tasks. Nevertheless, their approach is limited by using existing sentiment lexicon, which cannot be directly applied to other domains.

In this work, we focus on optimizing pre-training strategies and provide pre-trained language models that sufficiently explore textual information in **I**ntelligent **C**ustomer **S**ervice field called icsPLMs. Our contributions can be summarised as follow:

1. We propose an automatic domain lexicon construction method that can be integrated well with eMLM strategy and applied to intelligent customer service field.
2. We adopt a multi-task learning mode that combines generic and dialogue adaptive sentence-level tasks to obtain better sentence representations.

Approaches

Domain word masking: We propose a intent-oriented domain lexicon construction method that automatically extract domain words in intelligent customer service field. A remarkable phenomenon is that the attention mechanism of pre-trained language models pays more attention to domain words of input text, e.g., order, query, complaint in customer service field. Therefore, when fine-tuning language models for intent classification by a large number of annotated utterances, we extract the words with high attention scores in the last Transformer layer as domain words to construct the domain lexicon as shown in the left part of figure 1. Firstly, an open source Chinese word segmentation tool, Jieba [1] is used to segment a correctly classified sample and obtain all its words. And the words that hit tokens with high attention scores will be treated as domain words. Besides, the masking ratio of domain words is improving during masked language modelling process. In brief, 50% of masking words in a sample will be domain words, if the number of domain words is not enough, this strategy will be adjusted to mask all domain words and the remaining words are still randomly masked.

Multi-task training: We take an approach similar to StructBert(Wang et al. 2019), but adapted to the corpus of the intelligent customer service domain. Three different sentence-level tasks are established to form a paradigm of multi-task training. **Query to query (QQ) matching task** divides query pairs with three different levels of matching based on intent and sub-intent. Assuming we have two queries Q_1 and Q_2, level 1 means they have same intent and sub-intent, level 2 means they have same intent and different sub-intent and level 3 means they have different intent and sub-intent. Models need to recognize these three levels and learn their relationships. Similar to **QQ** task, **query to answer (QA) matching task** has a similar mode. For correct Q_{correct} and A_{correct} pair, level 1 means Q_{correct} and A_{correct} are matching, level 2 means Q_{correct} is paired with

[1]https://github.com/fxsjy/jieba

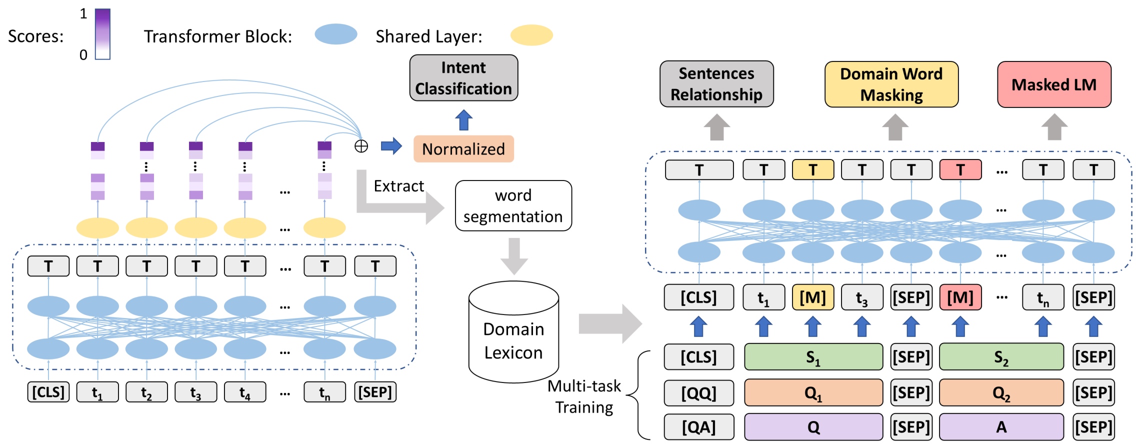

Figure 1: Overall architecture of icsPLMs

Models	QM	IntC	ComIntC
StructBert	87.50	90.82	82.75
icsPLM	88.02	91.15	83.16
icsPLM-MT	**88.50**	**91.86**	**84.78**

Table 1: The performance of our approaches with StructBert and icsPLMs on downstream tasks

a randomly selected A_{random} and level 3 means A_{correct} and Q_{correct} are in reverse order. In addition to these two tasks, we also use a generic corpus to train the sentence-level task in StructBert(Wang et al. 2019) to maintain the generic ability. In the pre-training stage, we use Apache Lucene [2] for constructing QQ, QA datasets and training corpus for these three tasks have an approximate same ration.

Experiments

Datasets: Our datasets are from JD Instant Messaging intelligence platform (*JIMI*) (Song et al. 2022). 600 thousand data collected from real-world user utterances with 1560 fine-grained intents are used to collect 23 thousand domain words. The total pairs are about 250 million for all sentence level tasks. We also construct three datasets from *JIMI* for evaluation, where **QM** is query matching task of intelligent customer service with 130 thousand, 32 thousand, 26 thousand for train, validation and test. **IntC** is an intent classification task composed of 63 user intent labels, with data distribution by 100 thousand for training, 20 thousand for validation, 20 thousand for testing. **ComIntC** is complicated intent classification task which only contains complicated utterances whose lengths are more than 30. 343 thousand, one thousand and 19 thousand are the number of train, validation and test datasets respectively.

Experimental results: Table 1 represents the comparison for our approaches on downstream tasks. We use the official open source StructBert as baseline, where icsPLM only using domain word masking and icsPLM-MT using both domain word masking and multi-task training methods. For evaluation metric, AUC score is used for **QM** while accuracy is used for **IntC** and **ComIntC**. It can be seen that both icsPLMs and icsPLMs-MT outperform the baseline StructBert. The most significant improvement is observed on ComIntC (+2.03%) task, while for **QM** and **IntC**, they also represent 1.00% and 1.04% increases. Both domain word masking and multi-task training could enhance performance of pre-trained language models.

Conclusion and Future Work

In this paper, we propose two approaches to optimize pre-trained language models in the field of intelligent customer service. Our analysis proves the effectiveness of these two methods on three downstream tasks. In the future, we will extend these methods to more domain-specific fields.

References

Song, S.; Chen, H.; and Shi, Z. 2017. Intention classification of user queries in intelligent customer service system. In *IALP '17*, 83–86.

Song, S.; Zou, B.; Lin, J.; Yu, X.; and He, X. 2022. Beyond QA: 'Heuristic QA' Strategies in JIMI. In *DASFAA '22*, 425–429.

Sosea, T.; and Caragea, C. 2021. eMLM: A New Pre-training Objective for Emotion Related Tasks. In *ACL-IJCNLP '21*, 286–293.

Wang, W.; Bi, B.; Yan, M.; Wu, C.; Xia, J.; Bao, Z.; Peng, L.; and Si, L. 2019. StructBERT: Incorporating Language Structures into Pre-training for Deep Language Understanding. In *ICLR '19*.

Yu, J.; Qiu, M.; Jiang, J.; Huang, J.; Song, S.; Chu, W.; and Chen, H. 2018. Modelling Domain Relationships for Transfer Learning on Retrieval-Based Question Answering Systems in E-Commerce. In *WSDM '18*, 682–690.

[2]https://lucene.apache.org/core/