Bingqian Li

Information

Male Age 21

Tel/WeChat: 15158818825 E-mail: libingqian@bupt.edu.cn

Educational Background

Beijing University of Posts and Telecommunications (BUPT) (2020-2024)

School of Computer Science - Computer Science and Technology

Weighted average **93.57,** ranking **1/429,** GPA: **3.88/4**

Skills

English proficiency: **TOFEL: 109 CET-6: 646 CET-4: 671**

Exceptional proficiencies in **C++, Python, Latex** and **Pytorch**

Research

- **Prof.Yiqun Liu Information Retrieval Lab at Tsinghua University 10/2022-now**
- **Research Domain:** Information Retrieval and Natural Language Processing

I am interning in the Legal Search group of the laboratory, where I mainly studied the role of legal elements in the field of legal search.

The impact of legal elements in case retrieval is usually less considered, and the annotated data in the legal field is very limited. In order to assist the Dense retrieval, we skillfully use the unsupervised generative pretraining method to input the case documents into the bert model, and input the output cls part into two decoders to the law article and the reasoning part respectively.

On the basis of this model, I also innovatively introduced a pretraining strategy for constructing positive and negative examples within the text, mainly targeting structural in specific domains. I add transformer decoders on both sides of the decoder in the original model to fuse the encoder output and decoder input and form a cross attention matrix. Using this matrix, combined with common alignment strategies in translation tasks, key elements (such as charges) in the structured text input by the decoder can be mapped back to the input part of the encoder, and the tokens of these parts can be meanpooled as positive examples, while the remaining tokens can be meanpooled as negative examples, and compared with the CLS vector for learning. This method has significant improvement in the evaluation indicators of the pretraining model.

- **Prof.Xiaofeng Tao Beijing University of Posts and Telecommunications 06/2022-10/2022**
- **Research Domain:** Multimodality of vision and text

I mainly worked with my senior in the detection of satirical multimodal information. Satiric multimodal information detection mainly uses a paragraph of text and corresponding images to confirm whether there is satirical information, and identifies the satirical object in the image through anchor boxes.

For this task, under the guidance of my doctoral senior, I conducted research on papers and completed the construction of the dataset. After preparations of the dataset, I trained the image part of the multimodal dataset using Swintransformer and Yolox's head to predict the anchor box.

I am mainly responsible for dataset construction and CV-side model code debugging tasks. Also, I use Exact match, an indicator in text detection to evaluate the performance of each model on out dataset.

the fifth author, DocMSU: A Comprehensive Benchmark for Document-level Multimodal Sarcasm Understanding

Honors and Awards

Honors: First-class Scholarship, Lude Scholarship

Awards: China Undergraduate Mathematical Contest in Modeling [Second Prize]

Mathematics competition of Chinese College Students [Second Prize]

Lanqiao National Programing Contest (Group C++) [Second Prize]

National English Competition for College Students [First Prize]