3 Preliminary Works and Research Plan

Text classification task is an enduring research topic, including spam detection, sentiment analysis and other directions. For spam detection tasks, the traditional machine learning algorithm steps are to first construct feature vectors through feature engineering, and input the feature vectors into the classification model (Bayesian classifier, SVM, etc.) for classification. In recent years, with the development of deep learning, classification tasks based on deep learning model methods have been widely used in spam detection.

I have not explored the spam detection methodology in my previous study, but I have found some inspiring and related resources:

To solve the problem that the classification model performs poorly on short text due to data sparsity. Zeng et al. [7] proposed a topic storage network for short text classification which has a novel topic storage mechanism that can encode the latent topic representation of the class label.

Wang et al.[8] used a novel sequence generation framework to annotate hashtags via viewing the hashtag as a short sequence of words. Moreover, they also proposed to jointly model the target posts and the conversation contexts initiated by them with bidirectional attention to address the data sparsity issue in processing short microblog posts.

Liao et al. [9] researched unsupervised aspect extraction and explored how words appear in global context (on sentence level) and local context (conveyed by neighbouring words) and proposed a novel network to discover aspect words.

Wang et al. [10] proposed a sequence-to-sequence (seq2seq) based neural key phrase generation framework which can helps alleviate the data sparsity that widely exhibited in social media language.

Zeng et al. [11] studied dynamic online conversation recommendation and proposed a model which can captures the temporal aspects of user interests. What's more, this model also can cater for cold start problem where conversations are new and unseen in training.

Since the development of spam detection technology, there has been a relatively mature model framework for use. For pure text spam, SVM and other methods have achieved good recognition results. However, with the development of identification technology, the forms of spam become more diverse. For example, some spam messages add words to images to avoid system detection, in view of this situation, the detection effect of pure text spam detection system is often poor. On the other hand, through the above literature, we can find that the sparsity of data is also one of the important reasons that affect spam detection, and the lack of data will lead to the problem of cold start. For some common spam keywords, the detection system can quickly detect, but for the spam keywords that have not appeared, the detection system often can't detect it. Therefore, I want to propose a multimodal learning framework combining text information and image information to detect spam.

Compared with only using single mode data, the fusion of multi-modal data can train a more accurate and robust classifier. The data set used in our framework mainly includes text information and image information. The text information may include the content of e-mail, email header information, URL link information, keywords and other aspects. For image information, it needs to be artificially classified, and the data image can be classified into spam image and non-spam image. For multimodal data, the labelling process needs a lot of time and effort. Therefore, the learning framework we adopt should be based on unsupervised learning or semi-supervised learning. Here, we propose a semi-supervised learning framework.

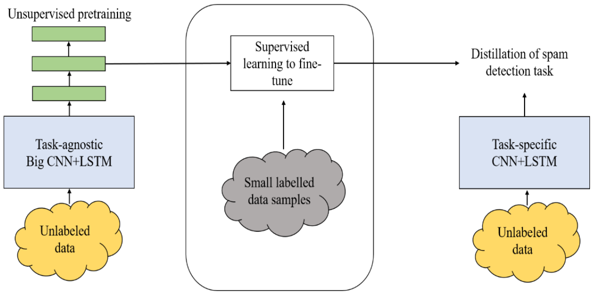
Unsupervised pretraining
Supervised learning to fine-tune
Distillation of spam detection task
Task-agnostic Big CNN+LSTM
Small labelled data samples
Task-specific CNN+LSTM
Unlabeled data
Unlabeled data

Figure 1: Main structure of spam detection based on semi-supervised learning

As Figure 1 reveals, the advantage of semi-supervised learning method is that there is no need to label a large data set, only a small part of the data can be labelled for training. Therefore, how to make good use of unlabelled data is an important problem in the research. For a small number of labelled data, supervised learning is still needed. Because the text information involved in spam detection contains sequence information, therefore, using LSTM structure to process the data can make use of the sequence information in the text, so as to get better detection results. our task also contains image information and the convolution neural network is better for processing image information. In order to integrate the above two advantages, I want to propose a new neural network based on CNN + LSTM for model training, and get the advantages of our network through experimental comparison.