WUST at NTCIR-16 FinNum-3 Task

Yuxuan Liu
School of Computer Science and Technology, Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology
Wuhan, China
L823500180@163.com

Maofu Liu
School of Computer Science and Technology, Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology
Wuhan, China
liumaofu@wust.edu.cn

Mengjie Wu
School of Computer Science and Technology, Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology
Wuhan, China
1749565249@qq.com

ABSTRACT
This article introduces how we deal with the FinNum-3 task of NTCIR-16. In the FinNum-3 task, the relationship between a numeral and a given label is the object of classification. In one text, given a target numeral and its offset in the text, models need to judge whether the given target numeral is in-claim or out-of-claim. In the experiments, we use the BiLSTM architecture to detect the in-claim or out-of-claim of the target numeral in two kinds of financial texts.

KEYWORDS
Financial numeral, Financial Numeral Classification, Numerical Understanding, Natural Language Understanding

ACM Reference Format:

TEAMNAME
WUST

SUBTASK
Investor’s Claim Detection (Chinese), Manager’s Claim Detection (English)

1 INTRODUCTION
WUST team participated in the NTCIR-16 FinNum-3 task. This report introduces the models and methods we used in this task and discusses the experimental results. Numerical-related information in financial text is the focus. Argument mining is a popular study direction in natural language processing. Since there is still none of argument mining study in financial domain, Chen et al. [4] explore the task that detects the claims from the reports written by professional stock analysts. In FinNum-3, the organizer introduced this novel task called NumClaim to detect whether the target numeral is in claim or not. Our team regards this task as a text classification problem. We chose the BiLSTM model to complete classification task and add a kind of numeral encoding. As [claim] shows the relation of claim and category of numerals, we use joint learning to improve the model performance. The rest of this report is organized as follows. Section 2 shows the related work of number classification in financial domain. Section 3 introduces related models and methods. Section 4 shows the official experimental results and our analysis. Finally, some conclusions are drawn in Section 5.

2 RELATED WORK
In the field of financial research, the study of text information has become more and more common. With the technological advancement of natural language processing and artificial intelligence, more and more studies are exploring these connections from the perspective of big data [11][14]. Chen et al. [3] propose number attachment task to identify the relation between the mentioned stock and the numerals in a financial tweet. Chen et al. [4] detect the claims from the reports written by professional stock analysts in financial domain. Numerals play an important role in information expression in the financial texts as numerals shows the fine-grained information in the text[5][6]. There are three mainstream methods, statistical models in statistics, regression models in econometrics, and machine learning-based models in computer science [9]. Machine learning (ML) models will take high-dimensional data as input, usually connect the features of different information sources into one feature vector, which will be applied to machine learning to explore the relationship between the information. Such as neural networks [12], Bayesian classifiers [1] and support vector machines [2]. LSTM has been proven to achieve better classification results in text data. Considering Context semantic relationship in this experiment, we decided to use BiLSTM architecture as the classifier.

3 MODELS
3.1 Robustly Optimized BERT Pretraining Approach
Robustly Optimized BERT Pretraining Approach [10](RoBERTa) is based on the improvement of BERT [7]. BERT is a bi-directional encoding representation model derived from the transformers model, and RoBERTa is improved on the basis of BERT with three training aspects improvements. First, the next sentence prediction (NSP)
task is removed. Second, a dynamic mask is used. RoBERTa uses a
dynamic mask, which generates a new mask pattern each time a
sequence is entered into the model. In this way, in the process of
continuous input of a large amount of data, the model will gradu-
ally adapt to different masking strategies and learn different lan-
guage representations. Third, RoBERTa considers using a larger
byte level BPE vocabulary to train BERT. This vocabulary contains
50K subword units without any additional preprocessing or word
segmentation for input. These improvements make RoBERTa rep-
resentation better extended to downstream tasks than BERT. In the
process tokenizing, we need to add some tokens to the vocabulary
of RoBERTa to ensure that the target numeral can be divided into
single characters to promise the correct offset information of the
target numeral in the experiments. In this report, we call batch size
as bz. In BiLSTM model, we use text vector as the embedding, and
the dimension of a batch is bz × 512 × 768.

3.2 LSTM Model

We conduct experiments employing Bi-directional Long Short-Term
Memory (BiLSTM) as baseline models. The BiLSTM model [13] con-
sists of a BiLSTM layer with hidden size of 500 and a multi-layer
perceptron.

BiLSTM is an improvement of LSTM. The forward LSTM is com-
bined with the backward LSTM to form BiLSTM. The architecture
of LSTM model is shown in Figure 2. The calculation process of
each step is shown in Figure 3, 4, 5 and 6.

\begin{align}
    f_t &= (W_f [h_{t-1}, x_t] + b_f) \\
    i_t &= (W_i [h_{t-1}, x_t] + b_i) \\
    C_t &= \tanh(W_C) [h_{t-1}, x_t] + b_C \\
    C_t &= f_t * C_{t-1} + C_t \\
    o_t &= (W_o [h_{t-1}, x_t] + b_o) \\
    h_t &= o_t * \tanh(C_t)
\end{align}
3.3 Numeral Encoding (NE)

We represent the target numeral with discrete representation and add a distributed representation to present the target numeral position information. In Figure 1, they are represented by NR (Numerical Representation) and PR (Position Representation) modules respectively. That is, we use bag-of-words model and set the bag-of-words size to 11. For each target numeral, we get a $1 \times 11$ tensor to represent the digit (0–9) and the decimal point, and concatenate the $1 \times 11$ tensor with a $1 \times 7$ tensor of the target numeral position information. In the experiments, we use the FNN architecture which output dimension is 4 to encode the numeral information. In Figure 6, it is represented by NE (Numeral Encoding) module, and concatenate the encoded numeral information with the context information, then we get the embedding whose last dimension is 772.

3.4 Joint Learning (JL)

Joint learning referring to learn multiple tasks at the same time allows the two tasks (claim detection and numeral category) to share knowledge in the learning process, and improving the performance and generalization ability of the model by using the correlation of multiple tasks. The joint learning task architecture designed in the experiments is shown in Figure 6. We use FNN model to conduct numeral category classification tasks. BiLSTM effectively captures the input context features. The two tasks share the parameters of the BiLSTM layer, which can improve the ability of the model to understand the tense information of the main tasks, so as to improve the performance of the model.

4 EXPERIMENTS

4.1 Dataset and Evaluation Metrics

In this experiment, we used the NumClaim datasets which are analyst’s report in Chinese and earnings conference call in English proposed by NTCIR-16. In the analyst’s report data set in Chinese, the quantity of training dataset, development dataset and test dataset are 4219, 925 and 4691 respectively. In the earnings conference call data set in English, the quantity of training dataset, development dataset and test dataset are 8837, 1191 and 2383 respectively.

We use Adam [8] as the optimizer, and the learning rate is set to 0.0001, the batch size is set to 64. According to official evaluation criteria, we use the micro-F1 and macro-F1 score to evaluate the experimental results. The calculation process of Macro-F1 is to first calculate F1 of each category separately, and then average F1 of each category, with the same weight of each category. The formula of F1 is as follows:

$$F1 = \frac{2 \ast (\text{precision} \ast \text{recall})}{\text{precision} + \text{recall}}$$

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

TP represents that the sample is positive and predicted to be positive. TN represents that the sample is negative and predicted to be negative. FP represents that the sample is negative but predicted to be positive and FN represents that the sample is positive but predicted to be negative.

4.2 Experimental results

Table 1 and Table 2 show the results of the two data sets respectively. Our results are WUST_1. It shows that for all results, Micro-F1 is higher than Macro-F1 on the whole, because Macro-F1 adds the influence factor of uneven data distribution.

Figures 7, 8, 9 and 10 show the data distribution of the training set and test set of the two tasks respectively. It can be clearly seen that the data distribution of the two tasks is uneven, especially Manager’s Claim Detection task. Since we have not handled the data imbalance problem, it can be clearly reflected in Macro-F1 in Table 1. There is a big gap between Macro-F1 and micro-F1 in our results. However, our experimental results and the data distribution of the training set can show that our model can still learn the characteristics of the proposed task without dealing with the data imbalance.

5 CONCLUSIONS

In this report, we employ the BiLSTM architecture to detect if the target numeral is in or not in claim with joint learning of the target
numeral category classification. We proposed a certain representation method which can be realized by neural networks to represent the target numeral information. Regrettfully, since we do not adopt methods or measures to handle the problem of uneven data distribution, our results are not good and have a big gap between Macro-F1 score and Micro-F1 score. The experimental results show that our model could understand the task through the learning of context semantics.

REFERENCES


Figure 9: Test Data Distribution of Earnings Conference Call.

Figure 10: Test Data Distribution of Earnings Conference Call.