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 NTCIR16. 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, Numeral Understanding, Natural Language Understanding

ACM Reference Format:
Yuxuan Liu, Maofu Liu, and Mengjie Wu. 2022. WUST at NTCIR-16 FinNum-3 Task. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

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. Numeral-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.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

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 gradually adapt to different masking strategies and learn different language 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 representation 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] consists of a BiLSTM layer with hidden size of 500 and a multi-layer perceptron.

BiLSTM is an improvement of LSTM. The forward LSTM is combined 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.

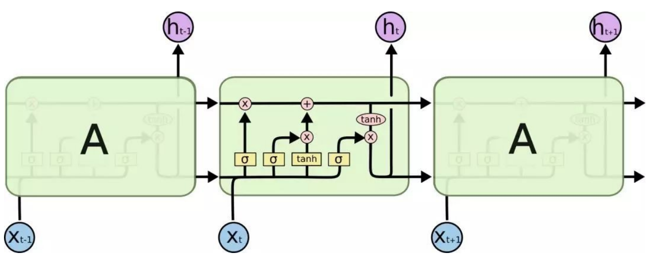

Figure 1: LSTM Model Architecture.

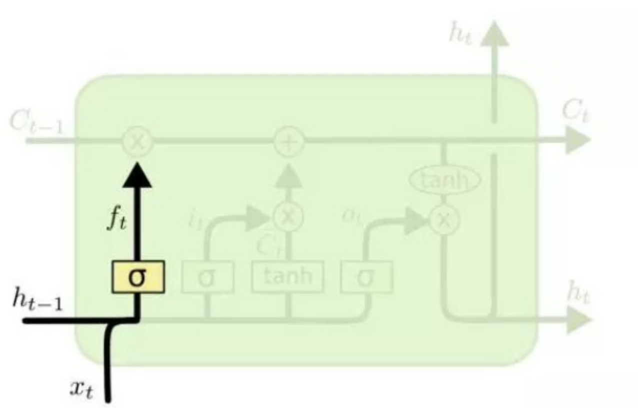

Figure 2: Compute Forgetting Gate.

$$f_t = (W_f[h_t - 1, x_t] + b_f) \quad (1)$$

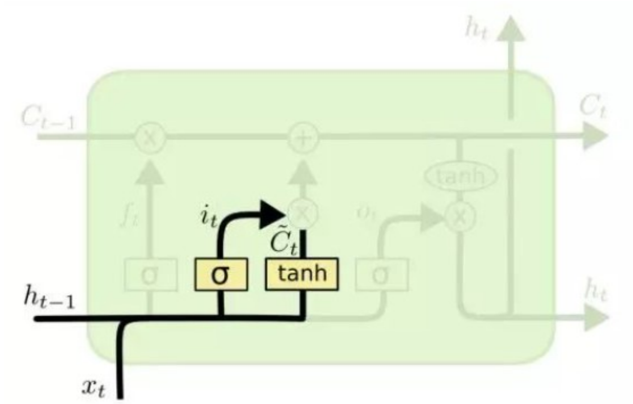

Figure 3: Compute the Memory Gate and Temporary Cell State.

$$i_t = (W_i[h_t - 1, x_t] + b_i) \quad (2)$$

$$C_t = tanh(W_C)[h_t - 1, x_t] + b_c \quad (3)$$

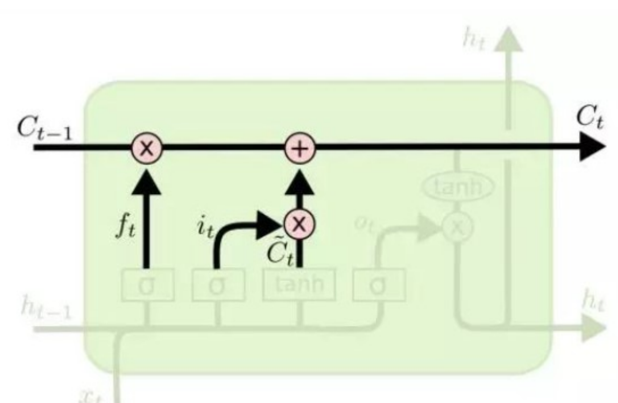

Figure 4: Compute the Current Cell State.

$$C_t = f_t * C_{t-1} + C_t \quad (4)$$

$$o_t = (W_o[h_t - 1, x_t] + b_o) \quad (5)$$

$$h_t = o_t * tanh(C_t) \quad (6)$$

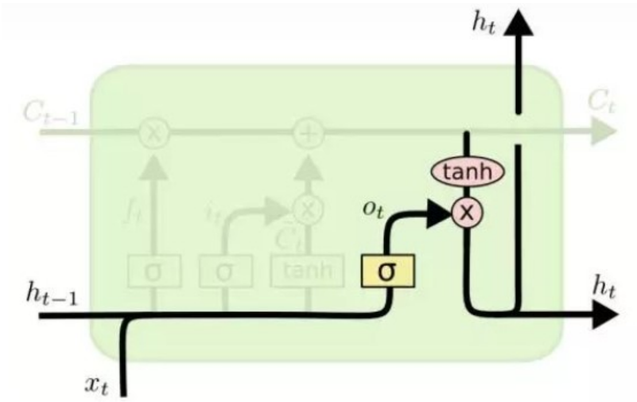

Figure 5: Compute the Output Gate and the Hidden Layer State at the Current State Time.

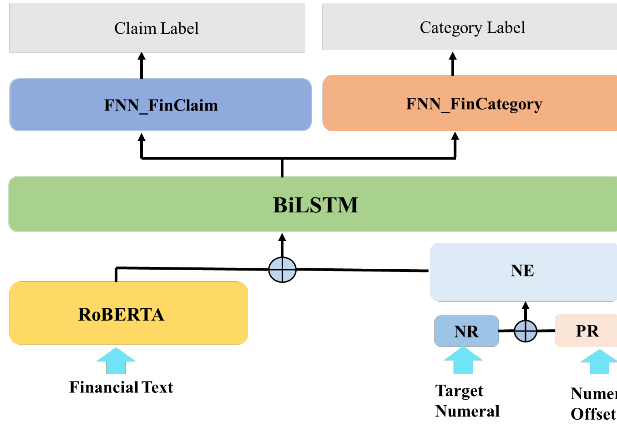

Figure 6: Joint Learning Model Architecture.

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 (Numeral 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 1×11 tensor to represent the digit (0–9) and the decimal point, and concatenate the 1×11 tensor with a 1×7 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:

$$F1 = \frac{2 * (precision * recall)}{precision + recall} \tag{7}$$

$$precision = \frac{TP}{TP + FP} \tag{8}$$

$$recall = \frac{TP}{TP + FN} \tag{9}$$

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

Table 1: Experimental results on analyst's report.

Submission	Claim Detection Micro-F1	Macro-F1	Numeral Category Micro-F1	Macro-F1
CapsNet	80.32%	69.9%	62.59%	20.99%
WUST_1	**84.89%**	**75.70%**	**56.13%**	**17.35%**
CYUT_2	91.73%	86.76%	-	-
TMUNLP_2	91.11%	87.76%	94.03%	72.99%
CYUT_3	92.16%	88.20%	-	-
CYUT_1	92.11%	88.80%	-	-
TMUNLP_1	92.82%	89.56%	94.31%	73.68%
TMUNLP_3	92.75%	89.68%	94.67%	73.89%
IMNTPU_2	94.14%	91.64%	-	-
IMNTPU_3	95.20%	92.91%	-	-
IMNTPU_1	95.31%	93.18%	-	-

Table 2: Experimental results on earnings conference call.

Submission	Claim Detection Micro-F1	Macro-F1	Numeral Category Micro-F1	Macro-F1
CapsNet	89.97%	56.36%	49.64%	26.50%
BERFIN_2	85.10%	68.26%	-	-
WUST_1	**93.37%**	**71.72%**	**48.76%**	**24.02%**
BERFIN_1	94.67%v80.26%		-	-
LIPI_2	95.17%	81.33%	-	-
LIPI_1	95.09%	82.82%	-	-
LIPI_3	95.59%	84.73%	-	-
CYUT_1	94.67%	85.53%	-	-
Passau21_1	96.01%	87.12%	-	-
CYUT_2	95.64%	87.49%	-	-
CYUT_3	96.43%	87.88%	-	-
IMNTPU_1	96.18%	88.39%	-	-
JRIRD_2	96.73%	89.55%	89.76%	72.84%
IMNTPU_2	96.73%	89.86%	-	-
JRIRD_1	97.15%	90.80%	89.68%	72.94%
JRIRD_3	97.27%	91.03%	89.26%	69.11%

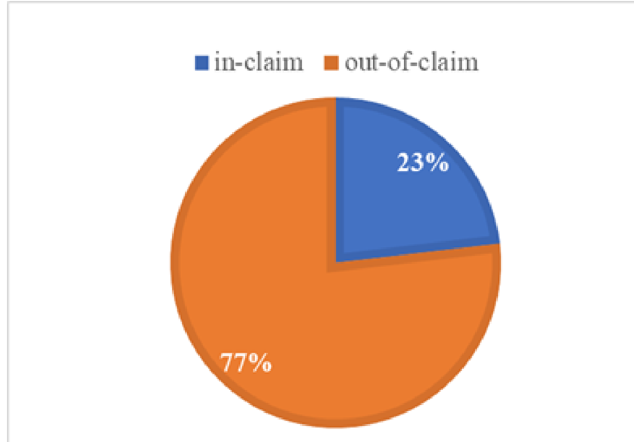

Figure 7: Training Data Distribution of Analyst's Report.

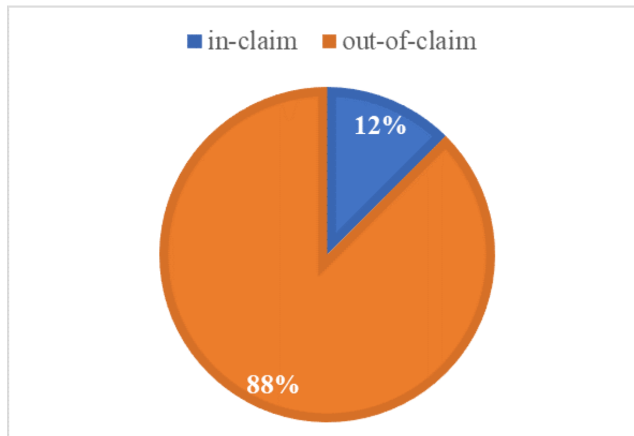

Figure 8: Training Data Distribution of Earnings Conference Call.

numeral category classification. We proposed a certain representation method which can be realized by neural networks to represent the target numeral information. Regretfully, 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

[1] John Binder, Daphne Koller, Stuart Russell, and Keiji Kanazawa. 1997. Adaptive probabilistic networks with hidden variables. *Machine Learning* 29, 2 (1997), 213–244. https://linkspringer.53yu.com/article/10.1023/A:1007421730016

[2] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. 1992. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*. 144–152. https://dl.acm.org/doi/abs/10.1145/130385.130401

[3] Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2019. Numeral Attachment with Auxiliary Tasks. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) *(SIGIR'19)*. Association for Computing Machinery, New York, NY, USA, 1161–1164. https://doi.org/10.1145/3331184.3331361

[4] Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. NumClaim: Investor's Fine-grained Claim Detection. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 1973–1976. https://dl.acm.org/doi/abs/10.1145/3340531.3412100

[5] Chung-Chi Chen, Hen-Hsen Huang, Yow-Ting Shiue, and Hsin-Hsi Chen. 2018. Numeral understanding in financial tweets for fine-grained crowd-based forecasting. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, 136–143. https://ieeexplore.ieee.org/abstract/document/8609586/

[6] Chung-Chi Chen, Hen-Hsen Huang, Hiroya Takamura, and Hsin-Hsi Chen. 2019. Numeracy-600K: Learning Numeracy for Detecting Exaggerated Information in Market Comments. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 6307–6313. https://doi.org/10.18653/v1/P19-1635

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[8] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014). https://arxiv.53yu.com/abs/1412.6980

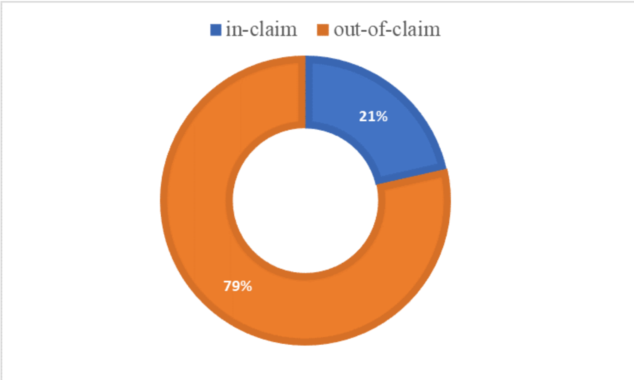

Figure 9: Test Data Distribution of Earnings Conference Call.

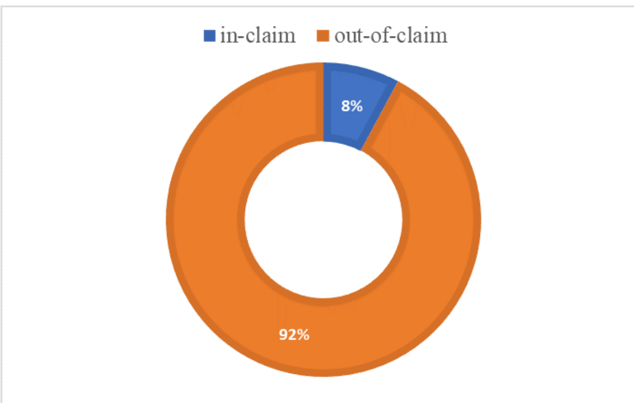

Figure 10: Test Data Distribution of Earnings Conference Call.

[9] Qing Li, Yan Chen, Jun Wang, Yuanzhu Chen, and Hsinchun Chen. 2017. Web media and stock markets: A survey and future directions from a big data perspective. *IEEE Transactions on Knowledge and Data Engineering* 30, 2 (2017), 381–399. https://ieeexplore.ieee.org/abstract/document/8068217/

[10] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR* abs/1907.11692 (2019). arXiv:1907.11692 http://arxiv.org/abs/1907.11692

[11] Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah, and David Chek Ling Ngo. 2014. Text mining for market prediction: A systematic review. *Expert Systems with Applications* 41, 16 (2014), 7653–7670. https://www.sciencedirect.com/science/article/abs/pii/S0957417414003455

[12] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2008. The graph neural network model. *IEEE transactions on neural networks* 20, 1 (2008), 61–80. https://ieeexplore.ieee.org/abstract/document/4700287/

[13] Shu Zhang, Dequan Zheng, Xinchen Hu, and Ming Yang. 2015. Bidirectional long short-term memory networks for relation classification. In *Proceedings of the 29th Pacific Asia conference on language, information and computation*. 73–78. https://aclanthology.org/Y15-1009.pdf

[14] Ilya Zheludev, Robert Smith, and Tomaso Aste. 2014. When can social media lead financial markets? *Scientific reports* 4, 1 (2014), 1–12. https://wwwnature.53yu.com/articles/srep04213