ENHANCING LANGUAGE MODELS FOR FINANCIAL RELATION EXTRACTION WITH NAMED ENTITIES AND PART-OF-SPEECH

Menglin Li & Kwan Hui Lim
Singapore University of Technology and Design
menglin_li@mymail.sutd.edu.sg, kwanhui_lim@sutd.edu.sg

ABSTRACT

The Financial Relation Extraction (FinRE) task involves identifying the entities and their relation, given a piece of financial statement/text. To solve this FinRE problem, we propose a simple but effective strategy that improves the performance of pre-trained language models by augmenting them with Named Entity Recognition (NER) and Part-Of-Speech (POS), as well as different approaches to combine these information. Experiments on a financial relations dataset show promising results and highlights the benefits of incorporating NER and POS in existing models. Our dataset and codes are available at https://github.com/kwanhui/FinRelExtract.

1 INTRODUCTION

Financial Relation Extraction (FinRE) is an important task for the financial domain, which involves the identification of key entities in a piece of text and the relation between these entities Kaur et al. (2023); Hillebrand et al. (2022). Figure 1 shows an illustration of this problem. The more generic task of Relation Extraction also has great significance for the general field of Natural Language Processing Wang et al. (2022); Nayak et al. (2021). The Bidirectional Encoder Representations from Transformers (BERT) Devlin et al. (2019) and its variants have been proposed and shown good results for FinRE and related tasks Hillebrand et al. (2022); Qiu et al. (2023); Vardhan & Chaudhary (2023). For example, KPI-BERT Hillebrand et al. (2022) utilizes BERT with a RNN-based pooling mechanism for FinRE in the context of Key Performance Indicators and their relations. Others like Qiu et al. (2023) build upon R-BERT Wu & He (2019) by leveraging both sentence-level and entity-level information for the FinRE task, while Vardhan & Chaudhary (2023) utilize a similar approach but with XLNET Yang et al. (2019) as the backbone architecture, with additional entity-type markers in the text. Building upon these works, we leverage on pre-trained language models and propose various strategies for augmenting additional information into the standard financial text.

In this paper, we make a few contributions: (i) We propose a simple but effective strategy of using Named Entity Recognition (NER) and Part Of Speech (POS) to improve the performance of pre-trained language models on the FinRE task; (ii) We study the effects of considering NER and POS in these models, as well as different ways of combining this information; and (iii) We perform a series of experiments and ablation study that shows the strong performance of our proposed method.

2 METHODS

Intuition and Research Questions. Our intuition is that pre-trained language models can be enhanced with additional NER and POS information, as the FinRE task involves identifying the relation between two entities. For example in the "Google:acquired:Fitbit" relation, NER is useful for identifying entities (e.g., Fitbit and Google) while POS is useful for identifying verbs that are potentially relations (e.g., acquired). As such, we consider the following two Research Questions (RQs) when developing our proposed model, namely: (RQ1) How can we utilize NER and POS to improve pre-trained language models for FinRE? and (RQ2) What is the best strategy for combining NER and POS with its original text content?

Proposed Model. The foundation of our model is a RoBERTa Liu et al. (2019) backbone, with various strategies for incorporating NER and POS, alongside the original financial text. Given a financial text $T = (t_1, t_2, ...t_n)$ with n tokens, we first extract the corresponding list of NER tokens $E = (e_1, e_2, ...e_n)$ and POS tokens $P = (p_1, p_2, ...p_n)$. For combining text and NER, we use the Text T replaced by NER tokens E that are not "Others", resulting in $TrN = (t_1, e_2, ..., e_17, ...t_n)$ where e_2 and e_17 are NER tokens among the original text. This text with replacement by NER sequence TrN is then concatenated with the POS sequence P, now denoted $TrNP$, before being fed into a RoBERTa backbone. Figure 2 shows our proposed model architecture. In our later ablation study, we also study the effect of different combinations of these text, NER and POS components.

3 Experiments and Results

Dataset and Experiment Setup. We experimented using the REFinD dataset Kaur et al. (2023), which comprises 28,676 annotated instances of 22 different financial relations between eight entity types based on US financial filings. REFinD contains 20,070, 4,306 and 4,300 instances for the training, development and test sets, respectively. We fine-tune our proposed model and the baselines (described later) with the AdamW optimizer using a learning rate of $lr2e-05$, decay of 0.1, dropout of 0,1, batch size of 32 and for 5 epochs. Early stopping is employed and the best model selected from the development set. Micro-F1 and Macro-F1 scores are reported on the test set.

Results: Models and Baselines. In addition to our proposed model, we evaluated it against various other baseline models with the Text (T) input, i.e., the text represents the financial statement. These other models include: SpanBERT Joshi et al. (2020), FinBERT Yang et al. (2020), BERT Devlin et al. (2019), XLM-RoBERTa Conneau et al. (2020), DistilBERT Sanh et al. (2019) and ALBERT Lan et al. (2019), and the results are shown in Table 1. The results show the superior performance of our proposed model against the baselines, with absolute improvements of 0.1144 in Micro-F1 and 0.2134 in Macro-F1, against the best performing baseline models. In addition, we conducted further experiments to compare the same baseline models using the same input, i.e., Text with replacement of NER + POS ($TrNP$), as shown in Table 2. Despite using the same input, the results show that our proposed model still offers the best performance compared to other baseline models, with the financial-related FinBERT being a close second in terms of both Micro-F1 and Macro-F1 scores. An additional ablation study is also presented and discussed in Appendix A.3.

Table 1: Experimental Results of Proposed Model against Baselines using Text (T).

Model	Micro-F1	Macro-F1
SpanBERT	0.6556	0.2505
FinBERT	0.6477	0.3806
BERT	*0.6577*	0.3311
XLM-RoBERTa	0.6407	*0.3373*
ALBERT	*0.6577*	0.2910
DistilBERT	0.6463	0.3281
Proposed Model	**0.7721**	**0.5507**

Table 2: Experimental Results of Proposed Model against Baselines using Text with replacement of NER + POS (TrNP).

Model	Micro-F1	Macro-F1
SpanBERT	0.7572	0.3504
FinBERT	0.7609	*0.5341*
BERT	0.7702	0.4612
XLM-RoBERTa	0.7693	0.4597
ALBERT	0.7447	0.4622
DistilBERT	*0.7712*	0.4800
Proposed Model	**0.7721**	**0.5507**

4 Conclusion and Future Work

We propose a simple but effective strategy to enhance pre-trained language model for the FinRE task, by replacing financial text with NER tokens, then concatenating with its corresponding POS tokens. Experimental results on the REFinD dataset show the promising performance of this approach. Our study also shows that this NER replacement strategy performs better than simply concatenating the NER tokens, and NER has a bigger role in improving performance than POS for such models. Future directions include a more in-depth study on the effects of different types of NER and POS tokens, as well as adaptation to other domains such as for geospatial predictions and recommendations Li et al. (2023); Halder et al. (2024).

5 URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

REFERENCES

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Édouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 8440–8451, 2020.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, 2019.

Sajal Halder, Kwan Hui Lim, Jeffrey Chan, and Xiuzhen Zhang. A survey on personalized itinerary recommendation: From optimisation to deep learning. *Applied Soft Computing*, pp. 111200, 2024.

Lars Hillebrand, Tobias Deußer, Tim Dilmaghani, Bernd Kliem, Rüdiger Loitz, Christian Bauckhage, and Rafet Sifa. Kpi-bert: A joint named entity recognition and relation extraction model for financial reports. In *Proceedings of the 26th International Conference on Pattern Recognition (ICPR)*, pp. 606–612. IEEE, 2022.

Ali Jabbari, Olivier Sauvage, Hamada Zeine, and Hamza Chergui. A french corpus and annotation schema for named entity recognition and relation extraction of financial news. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 2293–2299, 2020.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S Weld, Luke Zettlemoyer, and Omer Levy. Spanbert: Improving pre-training by representing and predicting spans. *Transactions of the association for computational linguistics*, 8:64–77, 2020.

Simerjot Kaur, Charese Smiley, Akshat Gupta, Joy Sain, Dongsheng Wang, Suchetha Siddagangappa, Toyin Aguda, and Sameena Shah. Refind: Relation extraction financial dataset. *arXiv preprint arXiv:2305.18322*, 2023.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. In *Proceedings of the International Conference on Learning Representations*, 2019.

Menglin Li, Kwan Hui Lim, Teng Guo, and Junhua Liu. A transformer-based framework for poi-level social post geolocation. In *European Conference on Information Retrieval*, pp. 588–604. Springer, 2023.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

Tapas Nayak, Navonil Majumder, Pawan Goyal, and Soujanya Poria. Deep neural approaches to relation triplets extraction: A comprehensive survey. *Cognitive Computation*, 13:1215–1232, 2021.

Le Qiu, Bo Peng, Yu Yin Hsu, and Emmanuele Chersoni. A simple approach to financial relation classification with pre-trained language models. In *Proceedings of the 4th Workshop on Knowledge Discovery from Unstructured Data in Financial Services (KDF)*, 2023.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.

Soumya Sharma, Tapas Nayak, Arusarka Bose, Ajay Kumar Meena, Koustuv Dasgupta, Niloy Ganguly, and Pawan Goyal. Finred: A dataset for relation extraction in financial domain. In *Companion Proceedings of the Web Conference 2022*, pp. 595–597, 2022.

N Harsha Vardhan and Manav Chaudhary. imetre: Incorporating markers of entity types for relation extraction. *arXiv preprint arXiv:2307.00132*, 2023.

Hailin Wang, Ke Qin, Rufai Yusuf Zakari, Guoming Lu, and Jin Yin. Deep neural network-based relation extraction: an overview. *Neural Computing and Applications*, pp. 1–21, 2022.

Shanchan Wu and Yifan He. Enriching pre-trained language model with entity information for relation classification. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pp. 2361–2364, 2019.

Yi Yang, Mark Christopher Siy Uy, and Allen Huang. Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*, 2020.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019.

Dongxu Zhang and Dong Wang. Relation classification via recurrent neural network. *arXiv preprint arXiv:1508.01006*, 2015.

Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D Manning. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 35–45, 2017.

A Appendix

A.1 Additional Details about the Financial Relation Extraction (FinRE) Problem

The Financial Relation Extraction (FinRE) problem involves identifying the relation between entities in a financial statement, i.e., a piece of text relating to finance matters. While the eventual ground truth label is in the form of a financial relation tag, this task also involves the additional sub-tasks of detecting the key entity types and identifying the relationship between them. Figure 1 shows an example of this problem, where the financial relation label is "Google:acquired:Fitbit", which is in turn based on the entities "Google" and "Fitbit" connected by a "acquired" relation.

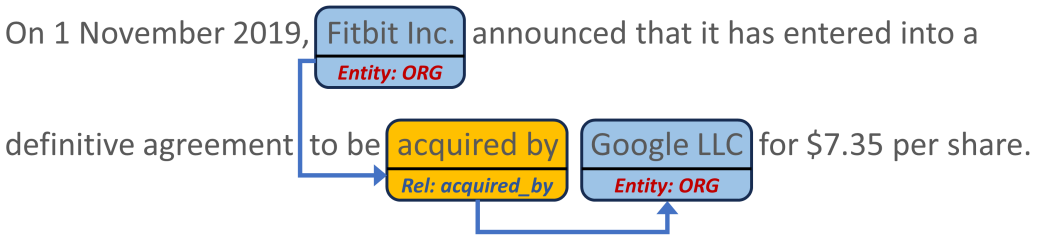

Figure 1: An example of the Financial Relation Extraction (FinRE) problem, with the relation *org:org:acquired_by* extracted from this financial statement.

For our experiments, we use the REFinD dataset Kaur et al. (2023), comprising 28,676 annotated instances with 22 different financial relations. There are also other relevant datasets for the FinRE tasks, such as TACRED Zhang et al. (2017), KBP37 Zhang & Wang (2015), FinRED Sharma et al. (2022) and CorpusFR Jabbari et al. (2020). Among these dataset, TACRED is the largest with 119,474 instances and 42 relations and CorpusFR the smallest with 1,754 instances and 20 relations. REFinD is the second largest among these datasets but poses a distinct advantage over the largest TACRED in that the REFinD only comprises 45.5% instances of the "no relation" type, while TACRED has a larger proportion at 79.5%. Thus, our choice of REFinD for our experiments.

A.2 Additional Details about Proposed Model

Figure 2 provides an illustration of our proposed model architecture. As mentioned in Section 2, the foundation of our model is a RoBERTa Liu et al. (2019) backbone and we build upon the roberta-base implementation on Hugging Face (https://huggingface.co/roberta-base) for this purpose. Thereafter, we propose and explore various strategies for incorporating NER and POS information to enhance the performance in the FinRE task. We selected RoBERTa over BERT and other variants due to the overall stronger performance of the former over a range of text-related tasks, via the implementation of a modified pre-training methodology. While the training time for RoBERTa might be longer, we do not consider this a significant issue as we are using RoBERTa as a pre-trained language model and fine-tuning it using our NER and POS augmented approach.

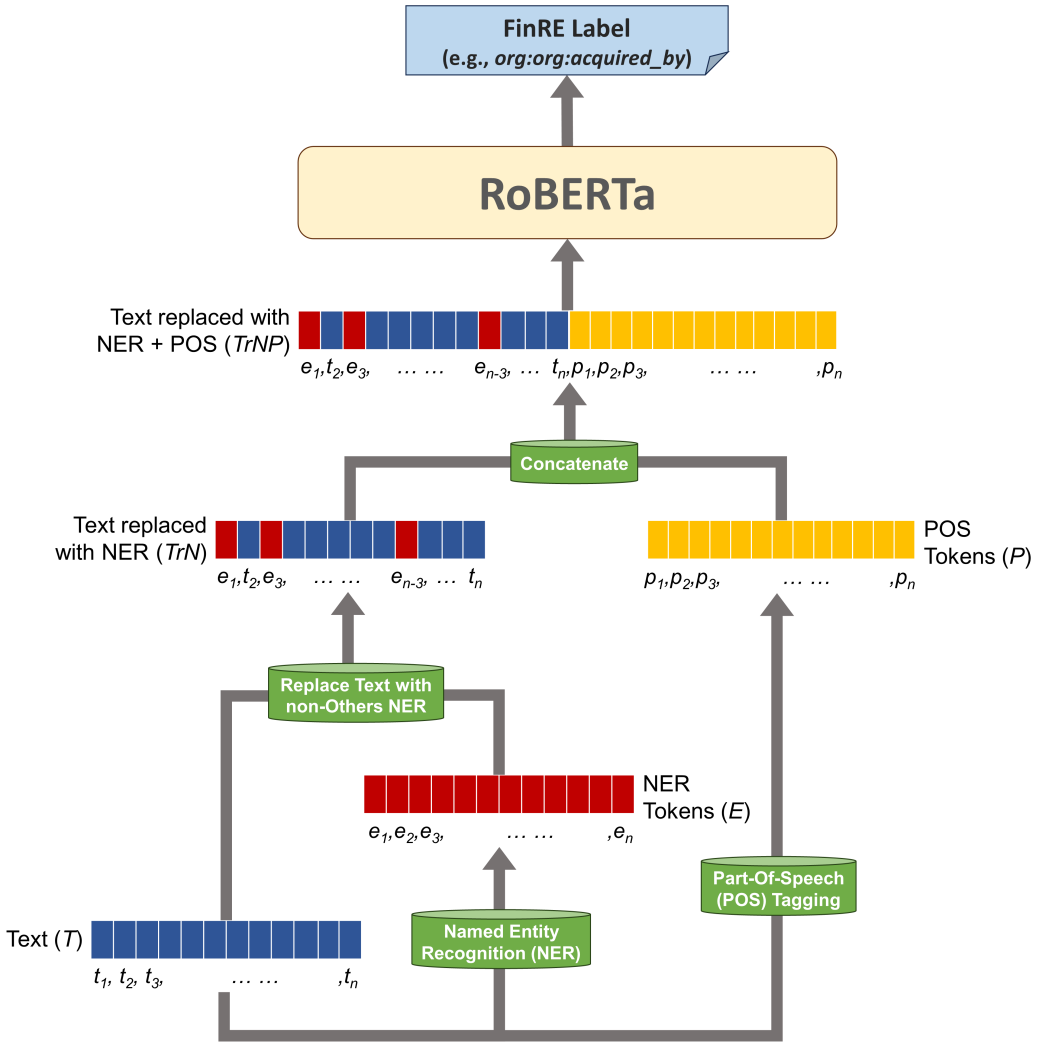

Figure 2: Architecture of Our Proposed Model for the FinRE Task.

Our input is in the form of a financial text $T = (t_1, t_2, ...t_n)$ with n tokens. Using this financial text T input, we utilize the spaCy library (https://spacy.io/) to extract the list of NER tokens $E = (e_1, e_2, ...e_n)$ and POS tokens $P = (p_1, p_2, ...p_n)$, as indicated by the green boxes labelled "Named Entity Recognition (NER)" and "Part-Of-Speech (POS) Tagging" in Figure 2. For both NER and POS, we use the tag/token type and not the actual entity names, i.e., "[ORG]" and not "Microsoft". This sequence of NER tokens E (e.g., "[ORG]") are then used to replace their text counterparts T,

resulting in a sequence TrN with the same length. For example, if e_2 and e_17 are the detected NER tokens, then the resulting sequence will be $TrN = (t_1, e_2, ..., e_{17}, ...t_n)$. Following which, we then concatenate this text with replacement by NER sequence TrN with the POS sequence P, separated by the [SEP] [CLS] tokens typically used in BERT and its variants. Finally, this concatenated sequence $TrNP$ is provided as input into a RoBERTa model and trained using the provided FinRE labels.

A.3 Additional Ablation Study and Discussion

As part of our ablation study, we performed additional experiments to evaluate the different components of Text (T), NER (N), POS (P), Text with replacement by NER (TrN). We evaluated six combinations of these components as shown in Table 3, e.g., TNP refers to RoBERTa using the Text, NER and POS components as a concatenated input. These results show the effectiveness of including both NER and POS, particularly our strategy of replacing text with NER tokens (TrNP), compared to simply concatenating NER to the original text (TNP).

Table 3: Ablation Study (T = Text, N = NER, P = POS, TrN = Text with replacement of NER).

Component	Micro-F1	Macro-F1
T	0.6507	0.3231
TN	0.7265	0.5342
TP	0.6605	0.3546
TNP	0.7379	0.4702
TrN	*0.7640*	**0.5559**
TrNP	**0.7721**	*0.5507*

In the previous paragraph, we have seen that Table 3 highlights the effectiveness of replacing the financial text with NER tokens and concatenating the POS token sequence. We further note that both strategies of replacement by NER tokens and concatenating POS token sequence are also effective individually in improving performance, although NER tokens have a larger effect than POS tokens. This trend is evident in how TN and TrN have a larger improvement over T in terms of both Micro-F1 and Macro-F1, as compared to TP that has a much smaller improvement over T.