A Simple Approach to Financial Relation Classification with Pre-trained Language Models

Le QIU
lani.qiu@connect.polyu.hk
Hong Kong Polytechnic University
Kowloon, Hong Kong SAR, China

Bo PENG
peng-bo.peng@polyu.edu.hk
Hong Kong Polytechnic University
Kowloon, Hong Kong SAR, China

Yu-Yin HSU
yu-yin.hsu@polyu.edu.hk
Hong Kong Polytechnic University
Kowloon, Hong Kong SAR, China

Emmanuele CHERSONI
emmanuele.chersoni@polyu.edu.hk
Hong Kong Polytechnic University
Kowloon, Hong Kong SAR, China

ABSTRACT

The paper serves as an experimental report submitted to the KDF.SIGIR 2023 shared task on relation extraction, focusing on the REFinD dataset. Motivated by recent advancements on Pre-trained Language Models (PLMs), we propose a simple, yet effective approach that leverages popular PLMs such as BERT, and RoBERTa to address this challenge. The approach capitalizes on the inherent capabilities of PLMs to encode sequences and enrich the semantics of the representations at the entity level. We go beyond the lexical and semantic levels by incorporating supplementary information to tackle the challenges in this task of financial relation classification. In the paper, we detail and justify the approach and report the results of our ablation studies.

CCS CONCEPTS

• Information systems → Information retrieval; • Computing methodologies → Information extraction.

KEYWORDS

financial relation extraction, relation classification, shortest dependency path (SDP)

ACM Reference Format:
Le QIU, Bo PENG, Yu-Yin HSU, and Emmanuele CHERSONI. 2023. A Simple Approach to Financial Relation Classification with Pre-trained Language Models. In *Proceedings of 4th Workshop on Knowledge Discovery from Unstructured Data in Financial Services.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Relation extraction (RE) targets one of the fundamental challenges in natural language processing (NLP), which is to comprehend the intricate connections between entities. Given a sequence s, RE extracts relationship triplets like <e_1, r, e_2> that describe a predefined relationship r between two entities e_1 and e_2. For example, for s = *Jobs created Apple*, an RE system outputs a triplet <*Apple*, created_by, *Jobs*>. By automatically detecting and classifying meaningful relationships between entities, RE has the potential to retrieve structured information from unstructured textual data, bridging the gap between natural language and machine-understandable language. RE thus has the potential for multiple downstream applications, such as information retrieval, question-answering, sentiment analysis, and knowledge base construction.

In light of the shared task, our work falls into the category of relation classification (RC). In this case, entities e_1 and e_2 in a relation triplet <e_1, r, e_2> are known, which allows us to skip the steps of named entity recognition (NER) and entity linking. The task of RC to predict the relation r [11] is a subtask of relation extraction or an intermediate step in a pipeline approach to RE.

Motivated by recent work in pre-trained large language models (PLMs), we have sought to devise a simple approach to this challenge with PLMs. In light of this objective, we finally present a RoBERTa-based architecture that incorporates enriched entity-level information, dependency information, and external features to address financial relation classification. While acknowledging that our proposed approach may not represent state-of-the-art (SOTA) methods, we emphasize its simplicity and effectiveness. Our intention is to strike a balance between complexity and performance, delivering a solution that is both comprehensible and capable of achieving commendable results in financial relation classification tasks.

The rest of the paper is organized as follows: Section 3 clarifies the proposed approach theoretically. Section 4 presents the data set of the shared task, the evaluation results of the proposed method, and the ablation study, which provides justification for our method. Finally, the Conclusion section summarizes our work and outlines potential directions for future research.

2 RELATED WORK

Various approaches have been developed to address the challenges of relation classification, ranging from traditional rule-based methods and statistical models to recent deep learning approaches. In deep learning, relation classification can be structure-oriented or semantic-oriented [4, 8]. Structure-oriented methods focus on model architectures. For instance, Zeng et al. [11] employed convolutional deep neural networks (CNNs) to automatically extract features at lexical and sentence levels without the complicated preprocessing used in statistical approaches; Zhang and Wang [12] proposed a recurrent or recursive neural networks (RNN) for relation classification, especially between long distance entities; Zhang et al. [15] introduced dependency trees and built a graph convolutional neural network (GCN) for RE.

Some researchers have integrated multiple approaches to fully exploit the respective advantages: for example, RNN (LSTM, GRU, etc.) can learn temporal and context features, while CNN effectively captures local patterns. Other studies have achieved superior performance by combining RNN and CNN structures in their relation classification experiments [3, 13, 14]. Semantic-oriented approaches explore the capability of text embeddings for relation extraction tasks. The dominant paradigm based on PLMs especially encourages this kind of approach. For example, Baldini Soares et al. [1] found that using text representations from PLMs is a simple and effective strategy for RE tasks. Wu and He [9] proposed the R-BERT model, which leveraged BERT [2] to capture the semantics of the sequence and entity mentions, and it outperformed the previous work approaches in the SemEval2010 task 8 dataset. Likewise, Zhang et al. [16] incorporated knowledge graphs (KGs) into BERT to enrich the representations of NLP tasks, including relation classification.

Inspired by the above-mentioned recent work in relation extraction with PLMs, we adopt a simple approach to this challenge. In the end, we propose a RoBERTa-based architecture incorporating internal and external features to address financial RC tasks.

3 METHODOLOGY

Figure 1 illustrates the basic architecture of our proposed approach. Similar to R-BERT [9], we used a PLM as the backbone, experimented on multiple models, and finally decided on the RoBERTa, and took the text representations at the sentence and entity levels as the main features for classification, along with external features. Notice that we refer to text representations that come from the PLM directly as internal features and others as external features.

Given a sequence *s* with entities e_1 and e_2, we inserted a *[CLS]* tag at the beginning of the sequence, and another two special tokens, <e_i> and </e_i>, at both ends of the entities as location markers, so that it facilitates the language model to capture the location information of the entities, which is believed to be vital for RC tasks [9]. We avoided special characters like # or $ used by Wu and He [9], to prevent confusion about the location makers and the in-text characters (e.g., $ conflicts with the dollar symbol, especially critical in the financial texts). For example, *s = Jobs created Apple* becomes *s' = [CLS]* <e_1> *Jobs* </e_1> *created* <e_2> *Apple* </e_2>.

Taking *s'* as the initial input, we then had the last hidden state output from the PLM as *H* and the last hidden state of the first token, i.e., *[CLS]*, as H_0. Usually, H_0 represents the entire sequence during classification tasks, but here we used an averaged *H* to indicate the sentence representations, hoping that the averaged *H*, which captures more semantics, especially for long sequences.

Based on Wu and He [9], we extracted vectors to represent the target entities, not merely the entity-level information. We also considered the shortest dependency path (henceforth **SDP**) between the words composing the entities, which is essential for relationship identification in most cases [5]. For instance, given *s = Apple, the tech company, was founded by Jobs.*, the SDP between entities *Apple* and *Jobs* is shown as the dashed-line arrows in Figure 2. The nodes *founded* and *by* along this path are SDP words. Instead of averaging separately as in [9], we compressed the semantics of the two entities and SDP words into one vector. This method helps to model the interactions or intricate connections within the fragment. Moreover, we added the entity pair group as an external feature to alleviate the issue of relation distribution imbalance.

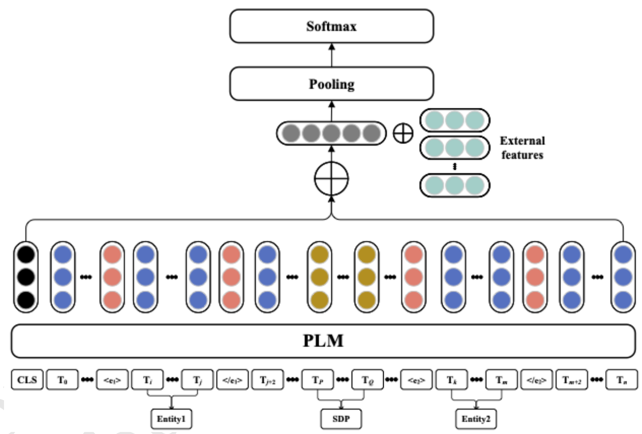

Figure 1: The architecture of our approach

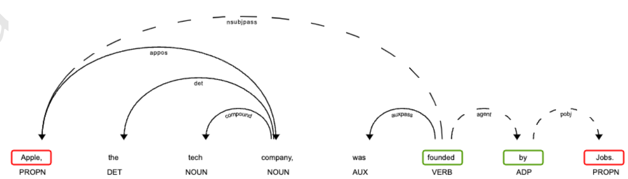

Figure 2: A dependency tree example

4 EXPERIMENTS

4.1 Dataset

REFinD [6], released by the task organizers, is a large dataset for financial relation extraction. The dataset is built on the 10-X reports from trade companies and specifically tailored for finance-related relation extraction tasks. With a large collection of 29,000 instances, the dataset encompasses 22 predefined relations across eight types of entity pairs. Notably, the dataset offers comprehensive annotations, including named entity recognition (NER) tags, part-of-speech (POS) tags, dependency information, etc. The rich annotations greatly simplify the preprocessing work, paving the way for further explorations beyond the goals of this shared task. However, a primary problem with the dataset, as shown in Figures 3 and 4, is that it presents a noticeable imbalance in terms of relation

Table 1: Experimental settings

Optimizer	AdamW
Loss function	Cross entropy
Max sequence length	384
Learning rate	2e-5
Training epoch	5 ± 2
Dropout rate	0.1

and sentence length distribution. Such imbalances pose significant challenges for the relation classification (RC) task, requiring careful consideration and specialized strategies to mitigate their impact. Nonetheless, addressing these challenges can yield valuable insights and advancements in the field of financial relation extraction.

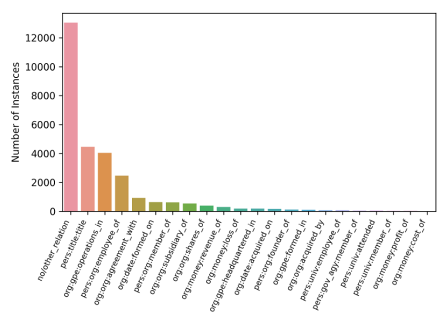

Figure 3: Relation Distribution (from https://refind-re.github.io/)

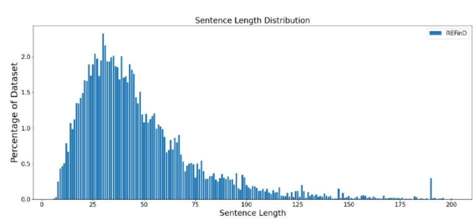

Figure 4: Sentence Length Distribution (from https://refind-re.github.io/)

The REFinD dataset consists of a train set, a validation set, and a public test set. The task organizers have also released a private test set without labels.

4.2 Experimental Settings

Table 1 presents the settings for our experiments, including hyper-parameters, loss functions, and optimizers. Note that the max sequence length has a remarkable impact on the performance (neither 128 nor 512 outperforms 384).

Table 2: Performance of the approach

Test set	Macro-F1	Weighted-F1	Official	Gap
Public	0.6141	0.7734	0.7482	-0.0034
Private	-	-	0.6894	-0.0602

Table 3: Performance of R-BERT with different PLMs

PLM	Macro-F1
R-BERT	0.54
R-BERT with FinBERT	0.56
R-BERT with RoBERTa	0.58

4.3 Evaluation Results

For evaluation, we employed macro-F1 or weighted-F1 as the primary metrics. Besides, we also reported the official evaluation scores (details of the metric have not been released yet). In Table 2, we present the scores achieved on the two test sets, along with the performance gap to the top systems on the leaderboard.[1] It is worth noting that while our approach may not be among the top places, the margin by which it falls behind is relatively small, indicating that the approach is effective.

4.4 Ablation Studies

To offer empirical evidence supporting the approach, we conducted an ablation study to identify the crucial components relevant to the financial RC tasks. This study enabled us to systematically analyze the individual contributions of various components and determine their significance in the overall performance of the approach. By dissecting and evaluating these components, we hoped to enhance the transparency and interpretability of the approach while strengthening its empirical foundation.

To determine the optimal PLM for text representations, our study commenced with an extensive evaluation with R-BERT, a simpler architecture that only focused on entity and sentence representations proposed by Wu and He [9]. Specifically, w applied R-BERT to the REFinD dataset and examined its performance in with various PLMs, including BERT [2], RoBERTa [7] and the domain-adapted FinBERT model [10]. Additionally, we performed experiments involving feature selection and feature fusion techniques (primarily the average strategy on internal features before the final concatenation operation). Furthermore, we thoroughly explored the annotations available within the REFinD dataset and investigated different fusion strategies. Results on the public test set are presented in Table 4 (Denotation reference: EV= entity vectors, GI = entity pair group, S = separate average, U = union average).

Table 3 shows that RoBERTa outperforms the domain-adapted FinBERT [10], a PLM specifically pre-trained on financial texts. The observed improvement, however, is modest. This highlights the inherent limitations of relying solely on sentence and entity representations for RC challenges. The interplay of lexical, semantic,

[1] Please refer to https://codalab.lisn.upsaclay.fr/competitions/11770#results for the details of the ranking.

Table 4: Performance with different features

Features	Average strategy	Best Macro-F1
EV	S	0.58
EV + POS	S	0.59
EV + SDP + POS	U	0.61
EV + SDP + NER	U	0.59
EV + SDP + POS + NER	U	0.61
EV + GI	S	0.61
EV + SDP + GI	U	0.63

and syntactic information across different relation groups requires that we incorporate additional internal or external information to tackle the classification problem. In Table 4, we can see that SDP (an internal feature) and the entity pair group (an external feature) emerge as key factors for the relation classification task. The presence of overlapping lexical, semantic, and syntactic information among different relation groups underscores the significance of an entity pair indicator at the decision boundary. The main idea would be that a pipeline approach is likely to surpass the joint extraction approach in a domain-specific RE task. By looking at the SDP words and integrating them with the target entities during vector extraction, we not only exploit the dependency relation, but are able to model their spatial positing and relative distances.

5 CONCLUSION

This study presents a straightforward approach to the relation extraction (RE) challenge on the REFinD dataset by employing PLMs, especially the RoBERTa model. The method revolves around utilizing sentence-level and entity-level representations for classification, while also incorporating features based on the dependency path and entity pair information. Inspired by the insightful results from ablation studies, we intend to delve deeper into the field of financial relation extraction. In particular, we plan to extend our investigation by leveraging the REFinD dataset for other NLP tasks, such as NER and entity linking. Furthermore, we aim to enhance the approach by incorporating additional dependency-related information.

REFERENCES

[1] Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the Blanks: Distributional Similarity for Relation Learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 2895–2905. https://doi.org/10.18653/v1/P19-1279

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[3] Xiaoyu Guo, Hui Zhang, Haijun Yang, Lianyuan Xu, and Zhiwen Ye. 2019. A single attention-based combination of CNN and RNN for relation classification. *IEEE Access* 7 (2019), 12467–12475.

[4] Xu Han, Tianyu Gao, Yankai Lin, Hao Peng, Yaoliang Yang, Chaojun Xiao, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2020. More data, more relations, more context and more openness: A review and outlook for relation extraction. *arXiv preprint arXiv:2004.03186* (2020).

[5] Lei Hua and Chanqin Quan. 2016. A shortest dependency path based convolutional neural network for protein-protein relation extraction. *BioMed research international* 2016 (2016).

[6] Simerjot Kaur, Charese Smiley, Akshat Gupta, Joy Sain, Dongsheng Wang, Suchetha Siddagangappa, Toyin Aguda, and Sameena Shah. 2023. REFinD: Relation Extraction Financial Dataset. *arXiv preprint arXiv:2305.18322* (2023).

[7] 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. *arXiv preprint arXiv:1907.11692* (2019).

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

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

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

[11] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. 2014. Relation classification via convolutional deep neural network. In *Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers*. 2335–2344.

[12] Dongxu Zhang and Dong Wang. 2015. Relation classification via recurrent neural network. *arXiv preprint arXiv:1508.01006* (2015).

[13] Lei Zhang and Fusheng Xiang. 2018. Relation classification via BiLSTM-CNN. In *Data Mining and Big Data: Third International Conference, DMBD 2018, Shanghai, China, June 17–22, 2018, Proceedings 3*. Springer, 373–382.

[14] Xiaobin Zhang, Fucai Chen, and Ruiyang Huang. 2018. A combination of RNN and CNN for attention-based relation classification. *Procedia computer science* 131 (2018), 911–917.

[15] Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018. Graph Convolution over Pruned Dependency Trees Improves Relation Extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, 2205–2215. https://doi.org/10.18653/v1/D18-1244

[16] Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. ERNIE: Enhanced Language Representation with Informative Entities. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 1441–1451. https://doi.org/10.18653/v1/P19-1139