DCU-ML at the FinNLP-2022 ERAI Task: Investigating the Transferability of Sentiment Analysis Data for Evaluating Rationales of Investors

Chenyang Lyu
School of Computing
Dublin City University
Dublin, Ireland
chenyang.lyu2@mail.dcu.ie

Tianbo Ji
School of Computing
Dublin City University
Dublin, Ireland
tianbo.ji2@mail.dcu.ie

Liting Zhou
School of Computing
Dublin City University
Dublin, Ireland
liting.zhou@dcu.ie

Abstract

In this paper, we describe our system for the FinNLP-2022 shared task: Evaluating the Rationales of Amateur Investors (ERAI). The ERAI shared tasks focuses on mining profitable information from financial texts by predicting the possible Maximal Potential Profit (MPP) and Maximal Loss (ML) based on the posts from amateur investors. There are two subtasks in ERAI: Pairwise Comparison and Unsupervised Rank, both target on the prediction of MPP and ML. To tackle the two tasks, we frame this task as a text-pair classification task where the input consists of two documents and the output is the label of whether the first document will lead to higher MPP or lower ML. Specifically, we propose to take advantage of the transferability of Sentiment Analysis data with an assumption that a more positive text will lead to higher MPP or higher ML to facilitate the prediction of MPP and ML. In experiment on the ERAI blind test set, our systems trained on Sentiment Analysis data and ERAI training data ranked 1st and 8th in ML and MPP pairwise comparison respectively. Code available in this link.

1 Introduction

Financial Opinion Mining (Chen et al., 2021b,a), the focus of the FinNLP-2022 shared task ERAI, has attracted the attention of the Natural Language Processing (NLP) community in recent years (El-Haj et al., 2021; Mariko et al., 2022; Lyu et al., 2022) for its potential use in financial analytic such as stock movement and volatility prediction (Chen, 2021). The FinNLP-2022 shared task ERAI (Chen et al., 2022) targets at extracting profitable information from financial documents particularly the posts from amateur investors. In the shared task, ERAI aims to predict the Maximal Potential Profit (MPP) and Maximal Loss (ML) conveyed by the posts from amateur investors as such mined opinions could be possibly used to analyse the financial market.

To tackle this task, we firstly frame it as a text-pair classification task where the input consists of two documents from different amateur investors. And the output is the label of whether the first document will lead to higher MPP or lower ML. Second, we take advantage of Sentiment Analysis data that have been shown to be useful in financial NLP (Chen, 2021; Wan et al., 2021; Valle-Cruz et al., 2022). Moreover, sentiment data are rich-resource and can be easily obtained (Liu, 2012) and the ERAI data that is in relatively small scale could benefit from it. Specifically, we use sentiment analysis data with an assumption that more positive text would give a higher profitable outcome. Practically, we build the ERAI-like dataset based on sentiment analysis data via iteratively sample two documents from sentiment analysis corpus, if the sentiment polarity of the first document is more positive than the second document then we think the first document would lead to higher MPP as well as higher ML (as a more positive document could mean a more aggressive action which could lead to higher MPP but also with high risk resulting in higher ML). Then we use the ERAI-like sentiment data to pre-train our model, of which the basic architecture is a text-pair classification model, and further fine-tune it with the ERAI training data.

In experiment, we employ BERT-Chinese (Devlin et al., 2019) as our base model since ERAI data is in Chinese. We experimented with three different strategies for training our model: 1) BERT-Senti: only use sentiment data thus fully relying on the transferability of sentiment data; 2) BERT-ERAI: only use ERAI training data; 3) BERT-Senti+ERAI: fine-tune our model after training it using sentiment data. We submit the three systems trained based on the above strategies to ERAI Pairwise Comparison blind test set. We submit the first system for ERAI Unsupervised Ranking evaluation as it is only trained on sentiment data thus it's an unsupervised system. The experimental results on ERAI

Proceedings of the Fourth Workshop on Financial Technology and Natural Language Processing (FinNLP), pages 116 - 121
December 8, 2022

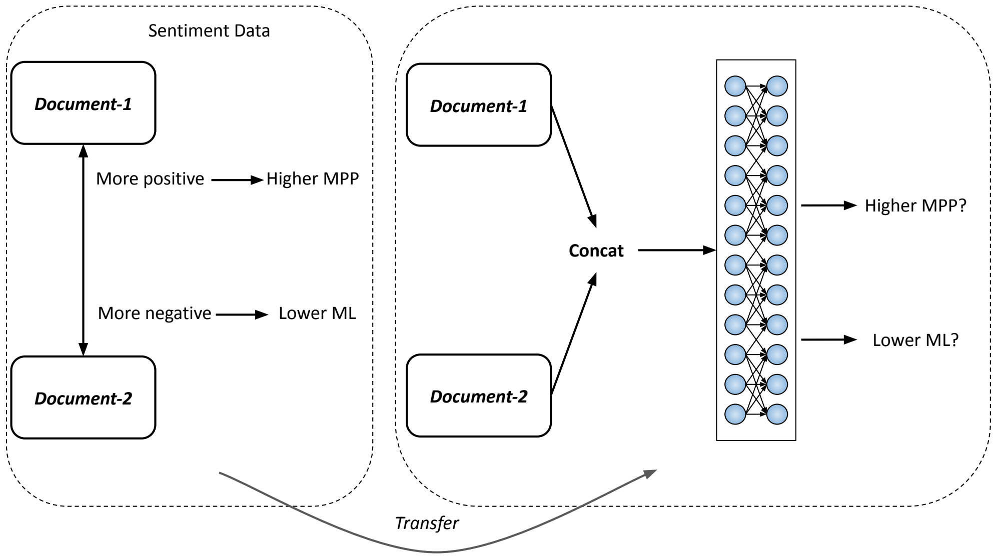

Figure 1: The main architecture of our model. We transfer the sentiment data to the ERAI text-pair classification task.

blind test set show that our BERT-Senti achieves 1st and 8th in ML and MPP pairwise comparison with accuracy of 59.77% and 52.87% respectively. Our BERT-Senti achieves average MPP and ML of 13.97% and -8.25% for Unsupervised Ranking, which ranked at 10th and 13th respectively.

2 Methodology

The main architecture of our proposed approach is shown in Figure 1, where we take advantage of the sentiment data to construct a ERAI-like dataset based on an assumption that a more positive document would lead to higher MPP and a more negative document would lead to higher ML. We pre-train our model based on the ERAI-like sentiment data followed by fine-tuning with the ERAI training data. The resulting systems are submitted to ERAI Pairwise Comparison and Unsupervised Ranking for evaluation.

2.1 Transferring Sentiment Data

We propose to utilize sentiment data as it has been shown that sentiment polarity information can be useful for Financial Opinion Mining (Chen, 2021; Valle-Cruz et al., 2022). Specifically, we assume that a more positive document would lead to higher MPP and a more negative document would lead to lower ML. Based on this assumption, we build our ERAI-like data via iteratively sampling two documents from sentiment corpus, if the first document has more positive sentiment polarity then we assign the document-pair with higher MPP label and higher ML label. The detailed process is shown in Algorithm 1.

Algorithm 1: The process of constructing ERAI-like based on sentiment corpus

S: Sentiment Corpus
$examples = []$
for *i in iteration* **do**
  Sample d_1 from S
  Sample d_2 from $S - d_1$
  if d_1.*sentiment* > d_2.*sentiment* **then**
    $d.text1 = d_1$
    $d.text2 = d_2$
    $d.MPP = 1$
    $d.ML = 0$
  end
  if d_1.*sentiment* < d_2.*sentiment* **then**
    $d.text1 = d_1$
    $d.text2 = d_2$
    $d.MPP = 0$
    $d.ML = 1$
  end
  $examples.append(d)$
end

2.2 Pairwise Comparison

Based on the ERAI-like dataset built in Section 2.1 and ERAI training data, we adopt three strategies to train our BERT model: 1) only use ERAI-like sentiment data built in Section 2.1 and therefore produce an unsupervised system; 2) only use ERAI training data; 3) firstly pre-train using ERAI-like sentiment data followed by fine-tuning with ERAI training data. These three strategies result three corresponding systems: BERT-Senti, BERT-ERAI and BERT-Senti+ERAI. We train the BERT model using our data as a text-pair classification task in which two documents are concatenated and assigned with different segment ID. The prediction head consists of two layers: one for predicting whether the first document would lead to higher MPP (1), the other one for predicting whether the first document would lead to higher ML (0).

2.3 Unsupervised Ranking

For unsupervised ranking, we employ our BERT-Senti system since it is only trained on our ERAI-like sentiment data thus BERT-Senti is an unsupervised system. However, the output of our systems in Section2.2 only indicate whether the first document would lead to higher MPP or ML (boolean value) with a real-valued number. To address such a gap, we reshape the Unsupervised Ranking task as a text-pair classification task where we compare the MPP and ML prediction of each document to all other documents in Unsupervised Ranking dataset. The document with more predictions of higher MPP and lower ML with obtain a higher rank. The process is shown in Algorithm 2.

Algorithm 2: Unsupervised Ranking based on pairwise comparison

```
U: Unsupervised Ranking Corpus
for d in U do
    for d' in U − d do
        if d.MPP > d'.MPP then
            d.MPP+ = 1
        end
        if d.ML < d'.ML then
            d.ML+ = 1
        end
    end
end
sort(U, key = MPP)
sort(U, key = ML)
```

3 Experiment

3.1 Data

The training set and test set of ERAI Pairwise Comparison task contain 200 and 87 examples respectively, the test set of the Unsupervised Ranking task contains 210 examples. We shown some examples from ERAI Pairwise Comparison training set with corresponding English translation in Table 3. The sentiment analysis data we used is from (Zhang and LeCun, 2017), which is a fine-grained sentiment classification dataset based on news in Chinese [1]

[1] lfeng in https://github.com/zhangxiangxiao/glyph#download

Systems	MPP	Systems	ML
Jetsons_1	62.07%	**DCU-ML_1**	59.77%
Yet_1	57.47%	**DCU-ML_3**	59.77%
Yet_2	57.47%	PromptShots_2	54.02%
Yet_3	57.47%	uoa_1	54.02%
LIPI_2	57.47%	aimi_1	52.87%
LIPI_1	54.02%	LIPI_2	50.57%
fiona	54.02%	fiona	48.28%
DCU-ML_1	52.87%	LIPI_3	48.28%
DCU-ML_3	52.87%	**DCU-ML_2**	45.98%
uoa_1	51.72%	PromptShots_1	45.98%
DCU-ML_2	51.72%	LIPI_1	44.83%
Jetsons_3	49.43%	Jetsons_2	41.38%
aimi_1	48.28%	PromptShots_3	41.38%
PromptShots_2	48.28%	Yet_1	40.23%
Jetsons_2	47.13%	Yet_2	40.23%
PromptShots_3	47.13%	Yet_3	40.23%
PromptShots_1	47.13%	Jetsons_1	37.93%
LIPI_3	44.83%	Jetsons_3	36.78%

Table 1: The evaluation results for ERAI Pairwise Comparison task, where our systems are **DCU-ML_1**, **DCU-ML_2**, **DCU-ML_3**, which correspond to BERT-Senti, BERT-ERAI and BERT-Senti+ERAI respectively

that has 5 classes (Very Negative, Negative, Neutral, Positive, Very Positive).

3.2 Training Setup

We employ BERT (Devlin et al., 2019) which has shown superior performance across many NLP tasks (Zhang et al., 2020; Bommasani et al., 2021)as our base model. Our implementation is based on BERT-Chinese (Devlin et al., 2019; Cui et al., 2020) from Huggingface (Wolf et al., 2020). We train our system with a learning rate of 2×10^{-5} for 2 epochs for BERT-Senti and 20 epochs for BERT-ERAI and BERT-Senti+ERAI, the batch size is set to 64 for BERT-Senti and 4 for the other systems. We use a maximum gradient norm of 1. The optimizer we used is AdamW (Loshchilov and Hutter, 2019), for which the ϵ is set to 1×10^{-8}. We perform early stopping when the performance on validation set degrades.

3.3 Results

The evaluation results on the blind test sets for ERAI Pairwise Comparison and Unsupervised Ranking are shown in Table 1 and Table 2. The results in Table 1 show that our BERT-Senti and BERT-Senti+ERAI outperform BERT-ERAI, which show the effectiveness of the transferability of sentiment data. Moreover, our BERT-Senti and BERT-Senti+ERAI outperform all other systems in

Systems	Average MPP of Top 10% Posts	Systems	Average ML of Top 10% Posts
PromptShots_2	24.39%	Baseline	-2.46%
PromptShots_3	23.76%	Yet_3	-3.24%
PromptShots_1	22.53%	LIPI_1	-4.11%
LIPI_2	18.27%	aimi_1	-4.17%
Baseline	17.61%	Yet_1	-4.35%
LIPI_1	17.46%	LIPI_3	-5.56%
UCCNLP_3	14.81%	Yet_2	-5.77%
Yet_3	14.61%	UCCNLP_3	-5.85%
aimi_1	14.02%	UCCNLP_1	-6.22%
DCU-ML_1	13.97%	UCCNLP_2	-6.77%
UoA_1	12.35%	PromptShots_1	-7.80%
Yet_2	12.10%	LIPI_2	-7.81%
LIPI_3	11.83%	**DCU-ML_1**	-8.25%
UCCNLP_2	11.34%	UoA_1	-9.39%
UCCNLP_1	11.10%	PromptShots_3	-12.33%
Yet_1	8.52%	PromptShots_2	-13.04%

Table 2: The evaluation results for ERAI Unsupervised Ranking task, where our submitted is **DCU-ML_1**, which corresponds to BERT-Senti.

Document-1	Document-2	MPP Label	ML Label
中壽可以準備賣給開發金了，除權息前應該可以完成 *(Zhongshou can prepare to sell it to the development gold.)*	中 壽 今 天 發 動 攻 勢 ， 往*34*靠攏 *(Zhongshou launched the offensive today and moved closer to 34.)*	0	0
有在往上動的感覺了 各位覺的呢 *(I feel like moving up What do you think about it?)*	永豐金融卷減少了*1000*多張,會不會停損在最高點啊 *(The Yongfeng Financial Volume has been reduced by more than 1,000 pieces. Will it stop at the highest point?)*	0	0
低接買盤開始浮現,不過近期也應該是盤整(除非有新的進度消息) *(Low buying the market has begun to emerge, but it should also be consolidated recently (unless there is new progress news))*	宏和一開盤,一路往上衝,漲的有點太高,希望能穩穩漲就好 *(As soon as Honghe opened, rushing up all the way, the rise was a bit too high, I hope to rise steadily)*	1	1

Table 3: Examples from ERAI Pairwise Comparison training set with English translation, where 0 represents *lower* MPP and *lower* ML for Document-1.

ML prediction with an accuracy of 59.77%. The results of BERT-Senti and BERT-Senti+ERAI are the same, we think the possible reason could be that the relatively small scale of test set (87 examples) introduces little variance on performance. In Unsupervised Ranking task, our submitted system BERT-Senti achieves an average MPP and ML of 13.97% and -8.25 respectively, which indicates the need for further improvement. We think the possible reason for that BERT-Senti fails to select documents with higher MPP and lower ML could be that sentiment data only provides a binary estimation for which document leads to higher MPP or lower ML, which is not precise. Besides, the noises in the prediction of Pairwise Comparison also makes it more difficult for accurately identifying the MPP and ML for documents.

4 Conclusion

In this paper, we proposed to use sentiment analysis data to enhance the ERAI shared task, results show that our proposed approach achieves superior performance in Pairwise Comparison, showing the effectiveness of our method. The results on Unsupervised Ranking task indicate there is still room for further improvement.

Limitations

Our method relies on a strong assumption that a more positive document would lead to higher MPP and a more negative document would lead to lower ML. However, this is an empirical assumption which needs more careful investigation before further using.

Acknowledgements

This work was funded by Science Foundation Ireland through the SFI Centre for Research Training in Machine Learning (18/CRT/6183).

References

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. 2021. On the opportunities and risks of foundation models.

Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2021a. Financial opinion mining. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts*, pages 7–10.

Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2021b. *From Opinion Mining to Financial Argument Mining*. Springer Briefs in Computer Science. Springer.

Chung-Chi Chen, Hen-Hsen Huang, Hiroya Takamura, and Hsin-Hsi Chen. 2022. Overview of the finnlp-2022 erai task: Evaluating the rationales of amateur investors. In *Proceedings of the Fourth Workshop on Financial Technology and Natural Language Processing*, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Qinkai Chen. 2021. Stock movement prediction with financial news using contextualized embedding from bert. *arXiv preprint arXiv:2107.08721*.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. 2020. Revisiting pre-trained models for Chinese natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 657–668, Online. Association for Computational Linguistics.

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 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Mahmoud El-Haj, Paul Rayson, and Nadhem Zmandar, editors. 2021. *Proceedings of the 3rd Financial Narrative Processing Workshop*. Association for Computational Linguistics, Lancaster, United Kingdom.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Chenyang Lyu, Tianbo Ji, Quanwei Sun, and Liting Zhou. 2022. Dcu-lorcan at fincausal 2022: Span-based causality extraction from financial documents using pre-trained language models. In *Proceedings of the The 4th Financial Narrative Processing Workshop in the Thirteenth Language Resources and Evaluation Conference*, pages 116–120, Marseille, France. European Language Resources Association.

Dominique Mariko, Hanna Abi-Akl, Kim Trottier, and Mahmoud El-Haj. 2022. The financial causality extraction shared task (fincausal 2022). In *Proceedings of the The 4th Financial Narrative Processing Workshop @LREC2022*, pages 105–107, Marseille, France. European Language Resources Association.

David Valle-Cruz, Vanessa Fernandez-Cortez, Asdrúbal López-Chau, and Rodrigo Sandoval-Almazán. 2022. Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods. *Cognitive computation*, 14(1):372–387.

Xingchen Wan, Jie Yang, Slavi Marinov, Jan-Peter Calliess, Stefan Zohren, and Xiaowen Dong. 2021. Sentiment correlation in financial news networks and associated market movements. *Scientific reports*, 11(1):1–12.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Xiang Zhang and Yann LeCun. 2017. Which encoding is the best for text classification in chinese, english, japanese and korean? *arXiv preprint arXiv:1708.02657*.

Zhuosheng Zhang, Hai Zhao, and Rui Wang. 2020. Machine reading comprehension: The role of contextualized language models and beyond.