Financial Opinion Mining

Chung-Chi Chen, [1] Hen-Hsen Huang,[2,3] Hsin-Hsi Chen[1,3]

[1] Department of Computer Science and Information Engineering
National Taiwan University, Taiwan
[2] Institute of Information Science, Academia Sinica, Taiwan
[3] MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
cjchen@nlg.csie.ntu.edu.tw, hhhuang@iis.sinica.edu.tw
hhchen@ntu.edu.tw

1 Type and Length

We will provide a **three-hour introductory** tutorial, named Financial Opinion Mining.

2 Goal of the Tutorial

When it comes to financial opinion mining, bullish and bearish come into people's minds. However, more fine-grained information will be missed if we only focus on the market sentiment analysis of financial documents. Thanks to the recent "CS + X" trend, more interdisciplinary cooperation exists between computer science and other domains. In the "NLP + Finance" community, lots of recent works pay their attention to in-depth analysis of different kinds of financial documents rather than market sentiment prediction. For example, our previous works (Chen et al., 2018, 2019a) find that the numeral information extracted from financial social media data is comparable to the price targets extracted from professional analysts' reports. Keith and Stent (2019) analyze the pragmatic and semantic features in the earnings conference calls and discuss how these features influence the investor's decision-making process. Zong et al. (2020) point out the difference between the textual factors and cognitive factors by comparing the accurate and inaccurate professional analysts' reports. The above-mentioned works conclude the necessity of capturing fine-grained opinions in the financial narratives. As the increasing interest of our community on this topic, recently, more and more related workshops spring up in the leading conferences, including FinWeb-2021 in the Web Conference, FinNLP-2021 in IJCAI, FinIR-2020 in SIGIR, and FNP-2020 in COLING.

In this tutorial, we will show where we are and where we will be to those researchers interested in this topic. We divide this tutorial into three parts, including coarse-grained financial opinion mining, fine-grained financial opinion mining, and possible research directions. This tutorial starts by introducing the components in a financial opinion proposed in our research agenda (Chen et al., 2021b) and summarizes their related studies. We also highlight the task of mining customers' opinions toward financial services in the FinTech industry, and compare them with usual opinions. Several potential research questions will be addressed. The audiences of this tutorial will gain an overview of financial opinion mining and figure out their research directions.

3 Tutorial Outline

We will cover the following topics based on recent works published in representative conferences and workshops. Both technical details and the application scenarios will be introduced. The contrast of financial opinion mining with general opinion mining will also be discussed. The characteristics of different kinds of financial documents will be listed.

3.1 Coarse-grained Financial Opinion Mining

The topic of the first session gives the overview of financial opinion mining, including the investor's opinion and the customer's opinion. We start with sentiment analysis in the financial domain. The comparison between the general sentiment and the market sentiment will also be discussed (Loughran and McDonald, 2011; Chen et al., 2020b). The lexicons for the sentiment analysis (Bodnaruk et al., 2015; Li and Shah, 2017; Sedinkina et al., 2019) in financial documents and the applications of adopting sentiment analysis results (Bollen et al., 2011; Du et al., 2019; Lin et al., 2020) will be included. This session also covers the sentiment analysis of financial narratives from different resources, including formal documents such as financial statements

and professional analyst's reports and informal documents such as blogs and social media platforms. The overview of applications on stock movement prediction and volatility forecasting will also be presented.

3.2 Fine-grained Financial Opinion Mining

The second session will focus on the fine-grained financial opinion mining, which is the recent trend in this field and also the research interest of the presenters. This session will start by the discussion of the aspect analysis of financial narratives (Maia et al., 2018; Chen et al., 2019a). The numeral in the textual data (Naik et al., 2019; Wallace et al., 2019; Chen et al., 2018, 2019a, 2020c) and the numeracy of the neural network models (Spithourakis and Riedel, 2018; Chen et al., 2019b) attract lots of attentions recently. In the financial narrative, the proportion of numerals are higher than that of other domains' documents. Without numerals, more important information will be missed. Thus, we summarize the related works for understanding the numerals in financial documents and provide a systematic analysis on these studies. The linguistic features of different kinds of financial documents will also be discussed (Keith and Stent, 2019; Zong et al., 2020), which can provide insights for the future works on feature engineering. The results of cross-document inference and comparison are also included (Chen et al., 2018; Keith and Stent, 2019).

3.3 Possible Research Directions

In the last session, we will discuss four possible research directions for future works (Chen et al., 2020a), including (1) argument mining in finance, (2) opinion quality evaluation, (3) implicit influence inference, and (4) opinion tracking in time series. We will link the proposed directions with the latest progress of NLP. For example, when introducing the ideas of argument mining in finance, we will provide a brief overview of current development on argument mining (Cabrio and Villata, 2018; Lawrence and Reed, 2019), and further present some instances for discussing the relation between current works and the proposed directions in financial opinion mining (Chen et al., 2020c). When discussing opinion quality evaluation, we will cover the studies of online review quality evaluation (Eirinaki et al., 2012; Wei et al., 2016; Ocampo Diaz and Ng, 2018), and show the difference between dealing with online reviews and

dealing with financial opinions.

The audience will be inspired by this tutorial and find an interesting research direction for their work. With the discussion on the possible research directions, many novel ideas will be figured out during this tutorial.

4 Recommended Small Reading List

We recommend the audiences to read the following papers, which will be discussed in the tutorial.

- For understanding the difference between general sentiment analysis and financial sentiment analysis: Loughran and McDonald (2011)

- For having the picture of the basic application scenario: Bollen et al. (2011)

- The importance of numerals in the financial documents: Chen et al. (2018)

- For capturing the idea and the intent of fine-grained opinion mining: Keith and Stent (2019)

- For conceiving the proposed research directions: Chen et al. (2021a)

5 Presenters

Chung-Chi Chen[1] is currently a postdoctoral researcher at the MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan. He got the Ph.D. degree in the Department of Computer Science and Information Engineering at National Taiwan University. He received the M.S. degree in Quantitative Finance from National Tsing Hua University, Taiwan. His research focuses on opinion mining and sentiment analysis in finance. He is the organizer of FinNum shared task series in NTCIR (2018-2022) and the FinNLP workshop series in IJCAI (2019-2021). He is the presenter of the AACL-2020 "Natural Language Processing in Financial Technology Applications" tutorial and the presenter of the EMNLP-2021 "Financial Opinion Mining" tutorial. His work has been published in ACL, WWW, SIGIR, IJCAI, and CIKM, and served as PC members in ACL, AAAI, EMNLP, CIKM, and WSDM. He won the 1st prize in both the Jih Sun FinTech Hackathon (2019) and the Standard Chartered FinTech competition (2018), and the 2nd prize in both the Jih Sun FinTech

---

[1]http://cjchen.nlpfin.com/

Hackathon (2018) and the E.SUN FHC FinTech Hackathon (2017).

Hen-Hsen Huang[2] is an assistant research fellow at the Institue of Information Science, Academia Sinica, Taiwan. His research interests include natural language processing and information retrieval. His work has been published in ACL, SIGIR, WWW, IJCAI, CIKM, COLING, and so on. Dr. Huang received the Honorable Mention of Doctoral Dissertation Award of ACLCLP in 2014 and the Honorable Mention of Master Thesis Award of ACLCLP in 2008. He served as the registration chair of TAAI 2017, the publication chair of RO-CLING 2020, and as PC members of representative conferences in computational linguistics including ACL, COLING, EMNLP, and NAACL. He was one of organizers of FinNum Task at NTCIR and FinNLP Workshop at IJCAI.

Hsin-Hsi Chen[3] received the Ph.D. degree in electrical engineering in 1988 from National Taiwan University, Taipei, Taiwan. Since August 2018, Hsin-Hsi Chen has been a distinguished professor in the Department of Computer Science and Information Engineering, National Taiwan University. He was conference chair of IJCNLP 2013, program co-chair of ACM SIGIR 2010, senior PC members of ACM SIGIR 2006, 2007, 2008 and 2009, area/track chairs of AAAI 2020, EMNLP 2018, ACL 2012, ACL-IJCNLP 2009 and ACM CIKM 2008, and PC members of many conferences (IJCAI, SIGIR, WSDM, ACL, COLING, EMNLP, NAACL, EACL, IJCNLP, WWW, and so on). He will be conference chair of ACM SIGIR 2023. He received Google research awards in 2007 and 2012, MOST Outstanding Research Award in 2017, and the AmTRAN Chair Professorship in 2018.

References

Andriy Bodnaruk, Tim Loughran, and Bill McDonald. 2015. Using 10-k text to gauge financial constraints. *Journal of Financial and Quantitative Analysis*, pages 623–646.

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8.

Elena Cabrio and Serena Villata. 2018. Five years of argument mining: a data-driven analysis. In *IJCAI*, pages 5427–5433.

Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2019a. Numeral attachment with auxiliary tasks. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1161–1164.

Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020a. Fine-grained opinion mining in financial data: A survey and research agenda. *arXiv preprint arXiv:2005.01897*.

Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020b. Issues and perspectives from 10,000 annotated financial social media data. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 6106–6110.

Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020c. Numclaim: Investor's fine-grained claim detection. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*.

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

Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2021b. A research agenda for financial opinion mining. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 1059–1063.

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)*, pages 136–143. IEEE.

Chung-Chi Chen, Hen-Hsen Huang, Hiroya Takamura, and Hsin-Hsi Chen. 2019b. Numeracy-600K: Learning numeracy for detecting exaggerated information in market comments. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6307–6313, Florence, Italy. Association for Computational Linguistics.

Chi-Han Du, Ming-Feng Tsai, and Chuan-Ju Wang. 2019. Beyond word-level to sentence-level sentiment analysis for financial reports. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1562–1566. IEEE.

Magdalini Eirinaki, Shamita Pisal, and Japinder Singh. 2012. Feature-based opinion mining and ranking. *Journal of Computer and System Sciences*, 78(4):1175–1184.

Katherine Keith and Amanda Stent. 2019. Modeling financial analysts' decision making via the pragmatics and semantics of earnings calls. In *Proceedings of the 57th Annual Meeting of the Association*

---

[2]http://www.cs.nccu.edu.tw/~hhhuang/
[3]http://nlg.csie.ntu.edu.tw/advisor.php

for Computational Linguistics, pages 493–503, Florence, Italy. Association for Computational Linguistics.

John Lawrence and Chris Reed. 2019. Argument mining: A survey. *Computational Linguistics*, pages 1–54.

Quanzhi Li and Sameena Shah. 2017. Learning stock market sentiment lexicon and sentiment-oriented word vector from StockTwits. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 301–310, Vancouver, Canada. Association for Computational Linguistics.

Sheng-Chieh Lin, Wen-Yuh Su, Po-Chuan Chien, Ming-Feng Tsai, and Chuan-Ju Wang. 2020. Self-attentive sentimental sentence embedding for sentiment analysis. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1678–1682. IEEE.

Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Macedo Maia, Siegfried Handschuh, André Freitas, Brian Davis, Ross McDermott, Manel Zarrouk, and Alexandra Balahur. 2018. Www'18 open challenge: financial opinion mining and question answering. In *Companion Proceedings of the The Web Conference 2018*, pages 1941–1942.

Aakanksha Naik, Abhilasha Ravichander, Carolyn Rose, and Eduard Hovy. 2019. Exploring numeracy in word embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3374–3380, Florence, Italy. Association for Computational Linguistics.

Gerardo Ocampo Diaz and Vincent Ng. 2018. Modeling and prediction of online product review helpfulness: A survey. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 698–708, Melbourne, Australia. Association for Computational Linguistics.

Marina Sedinkina, Nikolas Breitkopf, and Hinrich Schütze. 2019. Automatic domain adaptation outperforms manual domain adaptation for predicting financial outcomes. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 346–359, Florence, Italy. Association for Computational Linguistics.

Georgios Spithourakis and Sebastian Riedel. 2018. Numeracy for language models: Evaluating and improving their ability to predict numbers. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2104–2115, Melbourne, Australia. Association for Computational Linguistics.

Eric Wallace, Yizhong Wang, Sujian Li, Sameer Singh, and Matt Gardner. 2019. Do NLP models know numbers? probing numeracy in embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5307–5315, Hong Kong, China. Association for Computational Linguistics.

Zhongyu Wei, Yang Liu, and Yi Li. 2016. Is this post persuasive? ranking argumentative comments in online forum. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 195–200, Berlin, Germany. Association for Computational Linguistics.

Shi Zong, Alan Ritter, and Eduard Hovy. 2020. Measuring forecasting skill from text. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5317–5331, Online. Association for Computational Linguistics.