

Chapter 7

Perspectives and Conclusion



In this book, we started by describing the components of financial opinions as well as opinion sources in the financial domain. Then, we surveyed options for modeling financial opinions, and discussed in detail one fundamental characteristic of financial narratives: numerals. We also listed numerous applications and research directions. We have thus described financial opinion mining and have provided essential examples to illustrate various concepts. In this final chapter, we organize future directions and summarize the ideas in the book.

7.1 Future Directions

Table 7.1 highlights research topics on which few studies have been conducted. In Chap. 2, components such as the validity period of a financial opinion currently lack a good definition. Also lacking are in-depth experiments and analyses using argument mining in the financial domain. Because the argumentative units and the structure in Fig. 2.7 are crucial for fine-grained financial opinion mining, we suggest future studies start from (R1), (R2), and (R3) in Table 7.1. These research topics are related to organizing the information needed for financial opinion mining.

In Chap. 3, we discuss the various sources of financial opinions by provider. Ideally, all kinds of financial opinions could be organized using a single method. However, since the characteristics of each opinion depend on the provider of that opinion, we must use taxonomies or methods that reflect the characteristics of each provider. Chap. 4 emphasizes the importance of quality evaluation and influence estimation. These two components link a financial opinion with the target financial instrument. The quality and influence of a financial opinion help us judge whether we should consider the given opinion in the decision-making process. In addition

Table 7.1 Summary of research topics in financial argument mining

Index	Section	Research topic
R1	2.1	Extracting/estimating the validity period of a financial opinion
R2	2.2	Relation linking for elementary argumentative units in a financial opinion
R3	2.3	Analyzing relations between financial opinions
R4	4.2	Evaluating the quality of a financial opinion
R5	4.3	Estimating the influence of a financial opinion
R6		Implicit information inference
R7	5.2	General numeral attachment in financial narratives
R8	5.3	Exploring model numeracy
R9	6.1	Detection of false financial information
R10		Generation of financial analysis reports
R11	6.2	Financial opinion-based personalized recommendation
R12	6.3	Improving services for both employees and customers
R13	7.1	Organizing multimodal financial data
R14		Borrowing the proposed structures to other domains

to these features, it is also important to be able to produce inferences based on the given facts. These topics correspond to (R4), (R5), and (R6).

Chapter 5 demonstrates the central role that numerals play in financial narratives. We have discussed many of the challenges when working with financial social media data, but these are only some of the topics in this research direction. For example, general numeral attachment is another topic that merits future study. Also, the modeling of numeracy has attracted the attention of researchers; in the financial domain in particular, this is essential. Further development of numeracy would improve the performance of downstream financial tasks. This corresponds to (R7) and (R8).

Many application scenarios are proposed in Chap. 6. One of the jobs of a professional analyst is to verify information that has been collected. Fake information is currently a highly active topic in the research community. However, it is important to differentiate someone's subjective opinion from fake or false information; the task in this case becomes judging between trustworthy opinion and mere hyperbole or exaggeration. This can be accomplished by analyzing the components of a financial opinion. For an analyst, his/her final task is to produce a report; likewise, one goal of the proposed research would be to produce a report that passes the Turing test. In Table 7.1, the corresponding indexes are (R9) and (R10). The extracted financial opinions would then facilitate the development of financial services such as (R11) and (R12). Thus all of these scenarios depend on the results of fine-grained financial opinion mining.

Below, we mention research topics that were not mentioned in previous chapters. The first concerns multimodal data in financial opinions. In previous chapters, we mainly focused on textual data as well as some audio data. However, images are also

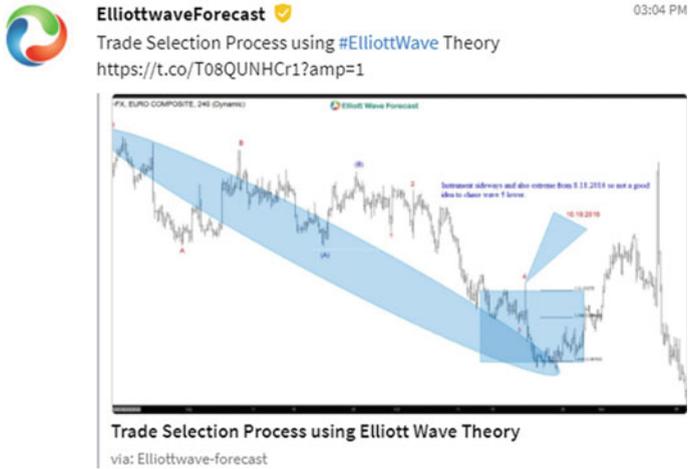


Fig. 7.1 Financial opinion expressed as an image

important ways to express financial opinions, especially on social media platforms. Figure 7.1¹ shows an image that expresses an opinion based on technical analysis. If we were to analyze only the textual data in this tweet, we would not find any opinion from the writer. However, an examination of the image reveals the method and price level that the writer is seeking to communicate. Indeed, in some cases, investors present their analysis of price movement via price charts, which often include expectations about future price movements. Thus image analysis in financial opinion mining is another topic that merits research.

Figure 7.1 shows another important issue: external reference of opinions. This occurs when users share abstracts of their blog posts on Twitter-like platforms; some include links to news articles for reference. Such external references are a common challenge in the analysis of social media data. In this instance, analyzing free-form websites is also an interesting topic.

Figure 7.2² shows another image-related instance, containing a slide released by a company for an earnings conference call. Slides like this may include statistical diagrams to visualize data. Understanding this kind of data is important and also helps when working on analysts' reports. Although most reports include diagram descriptions, it remains an open question as to whether capturing information from images will improve the performance of downstream tasks.

The left-hand side of Fig. 7.2 is further evidence of the importance of numerals in the financial domain. Managers and investors regularly discuss numbers, especially accounting ratios. Thus, as mentioned in Chap. 5, even for text mining, we should carefully analyze numeral information when working with financial narra-

¹<https://stocktwits.com/ElliottwaveForecast>.

²<https://www.deltaww.com/zh-TW/Investors/Analyst-Meeting>.

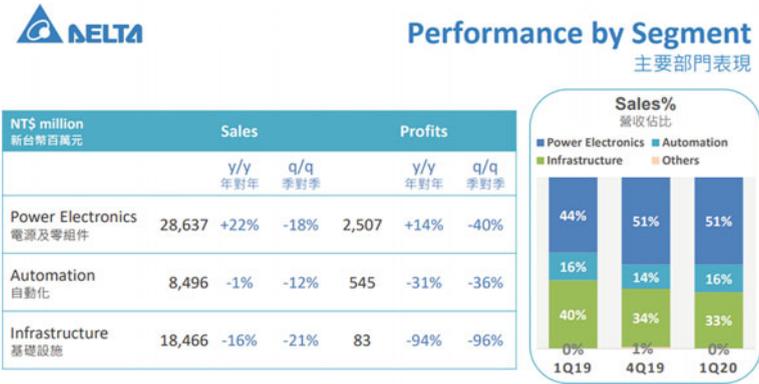


Fig. 7.2 A slide from an earnings conference call

tives. Figure 7.2 also shows tables in financial documents, another important issue. Tables are a straightforward way by which to represent structured data. Tables are common in financial documents, especially formal documents. Lamm et al. [3] propose a dataset and method for parsing numeral information in Penn Treebank Wall Street Journal articles [4]. Data mining methods can be used on such data after it has been translated into structured form. Recent studies have focused on encoding tabular data [1, 5]. Capturing both textual and tabular data may bring machines closer to human-level financial document understanding.

When numerals are mentioned, one topic that comes to mind is math word problems (MWP) [2]. In financial opinion mining, this is not as important, because managers and investors provide already-calculated results in their talks and posts; they do not ask readers to calculate the information needed. However, methods for MWP can be adopted to address (R8) in Table 7.1. This would further advance the performance of numeral understanding in financial narratives.

Finally, we seek to emphasize that the notions proposed in this book can be used in other domains. Although we use financial opinions here as an example, future work can draw from studies on fine-grained financial opinion mining for other target domains. Below, we use scientific article writing and clinical document analysis as examples.

We can use the structure in Fig. 2.6 for all kinds of persuasive narratives because it is based on the concept of argumentation mining. It can also be used to review and analyze scientific articles. In these articles, experimental results are the premises based upon which the authors produce claims. During the paper review process, one task is determining whether the given experimental results support the authors' claims. Given all of the claims, the authors further conclude their work's contribution, which is similar to the main claim in financial narratives. The only difference is that the authors of scientific articles draw conclusions, and the authors of financial analysis reports make predictions. The basic concept, however, remains the same.

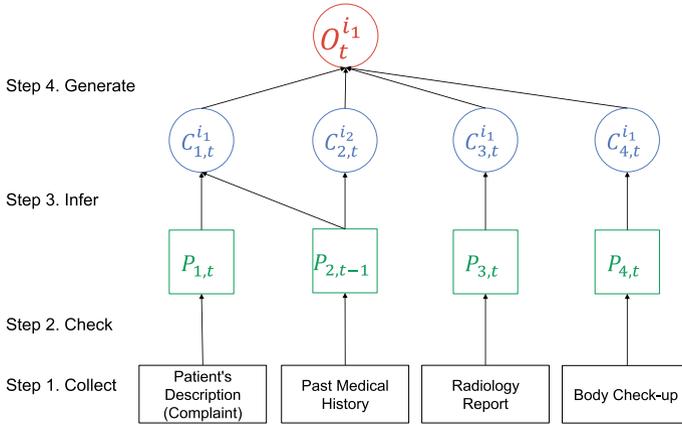


Fig. 7.3 Applying the workflow of argument mining in the clinical scenario

The other case is the decision-making process for different domains. In Fig. 6.4, we show the workflow of a professional analyst. Figure 7.3 uses the same flow for a clinical case. Doctors collect the necessary data as clues for diagnosis. Some data are unstructured, such as complaints and past medical history, whereas the body check-up results may be represented in a structured form. The radiology report may contain image data. After collecting data, doctors check whether the data makes sense or is incorrect, after which it becomes the premises for diagnosis. Different data may lead to various illnesses (i_1 and i_2), and doctors may produce different claims based on different combinations of premises. Doctors enter their final decisions in the medical record. Thus the ideas in this book can be used in other domains.

7.2 Conclusion

Although opinion mining has been discussed for a long time, it continues to attract attention. Continued advances in NLP techniques and infrastructure have facilitated better performance in general opinion mining tasks than ever before; now is the time to address domain-specific cases. To this end we have provided an overview of financial opinion mining in this book. Beyond sentiment analysis, we have laid out a blueprint from financial opinion mining to financial argument mining. Notions from argumentation mining are adopted to form the framework of financial opinion mining. We have discussed sources of financial opinions and characteristics of financial opinions from different sources, and we have also surveyed the literature to identify unexplored issues. We have also introduced a prominent domain-specific characteristic in financial narratives: numerals. Last, we have proposed application scenarios of financial opinion mining given current FinTech trends. Thus far, we have separated financial opinion tasks into several sub-tasks. We believe that addressing these

sub-tasks one by one will enhance the machine's ability to understand financial documents. Additionally, the proposed notions will help to make the decision-making process of machines more explainable. Addressing the issues proposed will bring us closer to our ultimate aim: the AI analyst.

Here, we emphasize that our goal is not to predict the price movements of financial instruments; rather, the goal in financial opinion mining is to empower machines to understand financial narratives and further provide professional-level rational analysis. Since price movement is random, it is not necessary to use backtesting results to evaluate all of the work on financial opinion mining. That is, although end-to-end prediction of the outcomes (sales or price movements) of a company can be considered a sub-task of financial opinion mining, it is not the final goal of this research.

Finally, global change hinges on opinion; opinion mining is thus essential to understanding these changes. This also applies to the financial domain. Financial opinion mining is necessary to understand the changes in financial markets. It is our hope that the ideas in this book inspire readers. We intend to provide the foundations for bringing our community closer to professional-level language understanding and generation in the financial domain.

References

1. Herzig, J., Nowak, P. K., Müller, T., Piccinno, F., Eisenschlos, J.: TaPas: weakly supervised table parsing via pre-training. In: Proceedings of the Fifty-Eighth Annual Meeting of the Association for Computational Linguistics (Online, July 2020), pp. 4320–4333. Association for Computational Linguistics, Stroudsburg
2. Kushman, N., Artzi, Y., Zettlemoyer, L., Barzilay, R.: Learning to automatically solve algebra word problems. In: Proceedings of the Fifty-Second Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Baltimore, Maryland, June 2014), pp. 271–281. Association for Computational Linguistics, Stroudsburg
3. Lamm, M., Chaganty, A., Manning, C. D., Jurafsky, D., Liang, P.: Textual analogy parsing: What's shared and what's compared among analogous facts. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (Brussels, Belgium, Oct.–Nov. 2018), pp. 82–92. Association for Computational Linguistics, Stroudsburg
4. Taylor, A., Marcus, M., Santorini, B.: An overview. In *Treebanks*. The Penn Treebank, pp. 5–22. Springer, Berlin (2003)
5. Yin, P., Neubig, G., Yih, W.-T., Riedel, S.: TaBERT: Pretraining for joint understanding of textual and tabular data. In: Proceedings of the Fifty-Eighth Annual Meeting of the Association for Computational Linguistics (Online, July 2020), pp. 8413–8426. Association for Computational Linguistics, Stroudsburg

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

