SPRINGER BRIEFS IN COMPUTER SCIENCE

Chung-Chi Chen
Hen-Hsen Huang
Hsin-Hsi Chen

From Opinion Mining to Financial Argument Mining

OPEN ACCESS

Springer

SpringerBriefs in Computer Science

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

**Indexing: This series is indexed in Scopus, Ei-Compendex, and zbMATH **

More information about this series at http://www.springer.com/series/10028

Chung-Chi Chen · Hen-Hsen Huang ·
Hsin-Hsi Chen

From Opinion Mining to Financial Argument Mining

Springer

Chung-Chi Chen
Department of Computer Science
and Information Engineering
National Taiwan University
Taipei, Taiwan

Hen-Hsen Huang
Department of Computer Science
National Chengchi University
Taipei, Taiwan

Hsin-Hsi Chen
Department of Computer Science
and Information Engineering
National Taiwan University
Taipei, Taiwan

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-981-16-2880-1 ISBN 978-981-16-2881-8 (eBook)
https://doi.org/10.1007/978-981-16-2881-8

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

This book is dedicated to all contributors in this field.

Preface

Human being's behaviors are led by personal opinions and others' views. To explain and predict their behaviors, capturing the opinions is one of the possible approaches. As one of the important topics in the natural language processing (NLP) community, opinion mining, aka sentiment analysis, has attracted much attention in the past decade. Argument mining, an extension of opinion mining, has rapidly emerged as a hot research topic in recent years. Not only to capture someone's opinion, but also argument mining aims to investigate the reason behind the opinion. In the financial domain, argument mining can be applied to understand the public's expectations for the market, providing valuable information for investment and other close applications. However, no single silver bullet for opinion and argument mining can deal with all domain-specific challenges because each domain has its own characteristics, especially the highly specialized financial domain. To facilitate the development of the technologies and applications in the financial domain, this book gives an overview from coarse-grained sentiment analysis to fine-grained financial argument mining.

This book provides a foundation for newcomers to understand the challenges and methods in financial opinion mining and to indicate the road map for researchers to achieve professional-level financial opinion understanding. Because the financial market changes with the participants' behaviors (e.g., buying or selling), the opinions of market participants become a crucial clue when analyzing the movement of financial instruments' prices. In this book, we adopt the notions of argument mining for an in-depth analysis of the opinions of financial market participants. We first define financial opinion in terms of basic components, and then determine the structures within an opinion and among opinions. A survey shows where we are now with the introductions of both classical approaches in general opinion mining and the latest works in financial opinion mining. In particular, the recent advances in the deep learning approach have led to substantial progress in many areas of artificial intelligence such as NLP and FinTech. This book will cover the related cutting-edge technologies including numeracy understanding, argument mining and financial document processing. Several unexplored research questions and potential application scenarios are also presented in the research agenda, pointing out where we are going. We hope the insights of this book can inspire researchers in both academics and industry, and further prompt them to join the field of financial argument mining.

Although this book is absorbed in financial opinions, the proposed concepts, which merge opinion mining and argument mining, can also be applied to other domains. We look forward to seeing new findings and more novel extensions based on the proposed ideas.

Taipei, Taiwan Chung-Chi Chen
April 2021 Hen-Hsen Huang
Hsin-Hsi Chen

Acknowledgments This book was supported partially by the Ministry of Science and Technology, Taiwan, under grants Most 109-2218-E-009-014, MOST 109-2634-F-002-034 and MOST 110-2634-F-002-028.

Contents

Chapter 1
Introduction

Financial opinion mining is a branch of traditional opinion mining and sentiment analysis which shares the basic notions of traditional approaches and adds its own domain-specific characteristics. In Sect. 1.1, we start with a common definition of general opinion mining after which we briefly overview traditional research directions. In Sect. 1.2, we compare financial opinion mining and general opinion mining, and in Sect. 1.3, we explain the motivation behind capturing financial opinions. We conclude the chapter with an overview of the structure of this book in Sect. 1.4.

1.1 Opinion Mining and Sentiment Analysis

Life is a series of choices, each of which is informed by personal opinions. A person's opinion may influence the opinions of others, and in turn influence the decisions they make. Thus a better understanding of people's opinions would make it possible for us to predict behaviors and guess a person's next steps. For example, every four years, we attempt to predict the outcome of the US presidential election. If we were able to capture every voter's opinion, we would be able to accurately predict the election results. However, thus ascertaining all opinions before an election is a difficult problem. We hence must use approximate approaches such as surveys to identify trends. After 2000, with the development of the Web and the increase in information shared by users, researchers began to investigate *opinion mining* methods to collect information that was once unattainable. In a common definition, an opinion is represented as a quintuple [6]:

$$(e, a, s, h, t) \, ,$$

C.-C. Chen et al., *From Opinion Mining to Financial Argument Mining*,
SpringerBriefs in Computer Science,
https://doi.org/10.1007/978-981-16-2881-8_1

Fig. 1.1 A product review from Amazon, where the five-star label indicates the opinion holder possesses a positive opinion toward the PlayStation 5 Console

in which an opinion holder h holds an opinion about entity e at time t with sentiment s under aspect a. Based on this definition, opinion mining is also termed *sentiment analysis*.

Although these five components, in particular aspect and sentiment, have been discussed for nearly two decades now [5, 8], they remain the focus of much active research [12, 13] due to the wide variety of potential applications. Figure 1.1 shows an example of an opinion, in this case a product review from Amazon. To simply judge the overall sentiment of the review writer, we can treat the five-star rating as a label indicating that the opinion holder possesses a positive sentiment toward the PlayStation 5 Console. Upon further investigation of the review's contents, we find that the opinion holder possesses a positive sentiment toward the new controller but a negative sentiment toward the bold design and plastic stand. Components e, h, and t, in turn, are relatively easy to extract from the platform metadata, which explains why the focus of most research remains on aspect-based sentiment analysis. The example in Fig. 1.1 shows that the sentiment s can vary depending on which aspect of the product (i.e., entity e) is in question. Potential task settings include the following:

1. Two-class classification (positive/negative)
2. Three-class classification (positive/neutral/negative)
3. Classification with discrete degrees (one-star to five-star)
4. Regression with continuous sentiment scores (0 to 1 or -1 to 1)

After extracting the opinion components, the problem becomes how to evaluate the usefulness and helpfulness of the opinion to readers. Figure 1.2 shows a review with little information. As with humans when making decisions, this kind of opinion may not be useful. The figure also shows a common approach for evaluating the opinion for a product: the "Helpful" button allows readers to annotate the review from a helpfulness aspect. These labels are then used for training supervised models [10]. Note however that false information or opinion spam also exists on online platforms.

Fig. 1.2 The "Helpful" button allows readers to praise the review from a helpfulness aspect

Detecting this kind of opinion is an area of active research in opinion mining [3]. Both content analysis [11] and spam detection [4, 9] are important research topics. However, opinions with little information are not necessarily opinion spam. Although the review in Fig. 1.2 is not useful for readers, the customer did purchase the product (Verified Purchase).

After sorting out the opinions and constructing quintuples from the various sources, we can (1) summarize opinions for a certain entity, (2) submit queries to search for opinions, and (3) compare opinions. The tasks mentioned in this section illustrate the work done over the past two decades on opinion mining and sentiment analysis.

1.2 Financial Opinion Mining

In this book, we define a financial opinion as an opinion related to a financial instrument. A financial opinion also has the five components mentioned in Sect. 1.1. One major difference is that sentiment in a financial opinion is termed *market sentiment* (bullish/bearish), which is different from sentiment (positive/negative) in general opinion mining research. For example, an investor holding a bullish position may possess negative sentiment because the price is falling. Studies have been done which contrast general sentiment and market sentiment, yielding the following findings:

- Three-quarters of the negative words in the Harvard Dictionary are not negative words in financial narrative [7].
- Bullish words in the financial domain are sometimes labeled as neutral words in general sentiment dictionaries [1].
- Positive sentiment does not always lead to bullish market sentiment [2].

Financial opinion is different from general opinion in that many financial opinions focus on forecasting the future instead of describing an experience. Many general opinions such as product reviews are based on the experience of using certain products. In contrast, financial opinion predicts future phenomena based on whatever facts are available. We define financial opinions in such a way as to yield an overall view from opinion analysis to the interaction between opinions and financial

Table 1.1 Notations used in this book and associated information extracted from Fig. 1.3

Notation	Denotation	Example in Fig. 1.3
e	Target entity, i.e., mentioned financial instrument	$AAPL
s	Market sentiment	Bullish
h	Opinion holder	William
t^p	Publishing time	1/3/20 11:44 PM
t^v	Validity period of an opinion	1/6/20–1/10/20 (this week)
$M^e_{t^p}$	Market information set of e before t^p	Close price: 297.32
a	Analysis aspect	Technical analysis
d	Degree of market sentiment	[1.91%, 3.26%]
C	A set of claims	Price target: [303, 307]
P	A set of premises	Chart is setup to RUN
q	Opinion quality	Low
ip	Influence power	Low

Fig. 1.3 Investor opinion shared on Stocktwits, a social media platform for finance

instruments. Table 1.1 shows the components of a financial opinion. In this book, we discuss financial opinion mining using these components.

Here, we go through the components using the instance shown in Fig. 1.3, which is a post from Stocktwits, a social media platform for finance. First, e denotes the target financial instrument ($AAPL) that the opinion holder (h, i.e., William) is discussing, and s denotes the market sentiment (bullish) of h on e. Temporal information is crucial for financial documents. A financial opinion can include two kinds of temporal information: the publishing time of the document (t^p, i.e., 1/3/20 11:44 PM) and the validity period of the opinion (t^v). In this example, the validity period of the price, which ranges from 303 to 307, is “this week”, which means that we should not take this tweet into account after one week. In most opinion mining tasks, opinions have no such validity period. However, due to the dynamic nature of the financial market, financial opinions do have validity periods, even the opinions of professional stock analysts are the same. Most reports from professional analysts have validity periods under one year.

Market information before t^p ($M^e_{t^p}$) may also be mentioned by the investor. Even it is not mentioned, recording market information can help us better understand the financial opinion. For example, if the writer in Fig. 1.3 does not provide the “bullish” tag, we can compare 303–307 with the close price (297.32) to infer that this investor has a bullish market sentiment about e.

In this book, we adopt the notions of argument mining to represent the full picture of financial opinion mining. In Chap. 2, we discuss this in detail. We can consider the market sentiment to be the main claim, which may consist of several claims (C). In each claim (c), there may exist several premises (P) that support the claim from different aspects (a); with each claim has its degree (d) of market sentiment. The quality of the opinion (q) and the influence power of the opinion (ip) should be evaluated. For example, a professional analyst's report may be of greater quality than a social media post and thus exert greater influence on the market.

1.3 Why Study Financial Opinion Mining?

Having described the components of a financial opinion, we now lay out the motivation for capturing financial opinion and thus why we seek to extract these components. We begin with the financial market operation model. Figure 1.4 shows an example of an order book, which lists the interests of buyers and sellers at a given time toward a given financial instrument. The figure lists the prices at which investors are willing to buy or sell, along with the quantity at each price level. Note that the deal price moves from 496.5 to 497.0 in only ten seconds; the quantities at different price levels change as well. Is it that during these ten seconds, the fundamental information of the company has suddenly changed, for instance the earnings per share? If not, what has caused the deal price to move from 496.5 to 497.0 so quickly? Below are some possible scenarios.

- Because there exists an arbitrage opportunity, the trading algorithm or trader sends the order at \$497.
- A new investor sends a new order at \$497.
- Some investors change their willingness to buy at prices lower than \$497 to higher than \$497.

Regardless of the rationale, we find that the change in the financial market is caused by changes in investor opinions. In connection to this, note that automatic trading algorithms are constructed based on human beings, and the rationales behind these algorithms can be viewed as opinions. In the example in Fig. 1.4, these ten

Fig. 1.4 Comparison of an order book at two time points. The change in the financial market is caused by changes in investor opinions

Deal price = 496.5 at the time t

Buy		Sell	
Quantity	Price	Price	Quantity
24	496.5	497.0	156
123	496.0	497.5	245
236	495.5	498.0	299
1,244	495.0	498.5	347
275	494.5	499.0	697

Deal price = 497.0 at t + 10 seconds

Buy		Sell	
Quantity	Price	Price	Quantity
20	496.5	497.0	100
200	496.0	497.5	232
120	495.5	498.0	399
983	495.0	498.5	347
200	494.5	499.0	400

seconds have resulted in changes not only to the deal price but also to the quantity at each price level. This shows that investor opinions are always changing. Indeed, ideally, given the ability to accurately capture all investor opinions, we would be able to perfectly predict market movements.

Financial opinion mining is one way to analyze the financial market and provide a rationale for market movements. For example, stock prices in energy and travel sectors surged in 2020 because many investors believed that the Pfizer vaccine could resolve the COVID-19 crisis.

Thus, we see that financial opinion mining is more complex than general opinion mining tasks: we seek to understand the decision process of all kinds of investors, regardless of whether they are (1) professional or amateur, (2) rational or irrational, or (3) well-informed or ill-informed. Even if two investors are provided with the same information, they could make different decisions under different rationales. Also, two bullish opinions may have different amounts of confidence or cause different degrees of impact on the market. These phenomena continue to complicate financial opinion mining.

Although we focus on financial opinion mining in this book, similar concepts can be adopted in other domains. We propose application scenarios in other domains in Chap. 7. In sum, solving the issues in financial opinion mining would provide solutions for other opinion-oriented tasks as well.

1.4 Overview of the Book

In Chap. 2, we describe in detail the components of financial opinions and raise several examples for reference. We further use the notions of argument mining to understand the structure of a single financial opinion. We also propose structures between opinions and those between opinions and financial instruments. In Chap. 3, we discuss opinions from various sources, including managers, professionals, social media users, and journalists, and then mention possible research directions for each kind of source. In Chap. 4 we explain how current techniques are used to extract opinion components and link relations between opinions. We also discuss opinion quality and the evaluation of influence. Because numerals contain much useful information in financial narratives, we discuss several numeral-related tasks in Chap. 5. Following this, in Chap. 6 we lay out application scenarios for financial opinion mining in the financial technology (FinTech) industry. We then conclude in Chap. 7, highlighting future directions and unexplored issues and suggesting approaches to adopting the notions proposed in this book to other domains.

References

1. Chen, C.-C., Huang, H.-H., Chen, H.-H.: NTUSD-Fin: a market sentiment dictionary for financial social media data applications. In: Proceedings of the First Financial Narrative Processing Workshop (FNP 2018) (2018)
2. Chen, C.-C., Huang, H.-H., Chen, H.-H.: Issues and perspectives from 10,000 annotated financial social media data. In: Proceedings of the Twelfth Language Resources and Evaluation Conference, pp. 6106–6110 (2020)
3. Chen, Y.-R., Chen, H.-H.: Opinion spam detection in Web forum: a real case study. In: Proceedings of the Twenty-Fourth International Conference on World Wide Web, pp. 173–183 (2015)
4. Chen, Y.-R., Chen, H.-H.: Opinion spammer detection in Web forum. In: Proceedings of the Thirty-Eighth International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 759–762 (2015)
5. Dave, K., Lawrence, S., Pennock, D.M.: Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceedings of the Twelfth International Conference on World Wide Web, pp. 519–528 (2003)
6. Liu, B.: Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, vol. 5, issue 1, pp. 1–167. Morgan & Claypool Publishers, San Rafael (2012)
7. Loughran, T., McDonald, B.: When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. J. Financ. **66**(1), 35–65 (2011)
8. Nasukawa, T., Yi, J.: Sentiment analysis: capturing favorability using natural language processing. In: Proceedings of the Second International Conference on Knowledge Capture, pp. 70–77 (2003)
9. Noekhah, S., binti Salim, N., Zakaria, N. H.: Opinion spam detection: using multi-iterative graph-based model. Inf. Process. Manag. **57**(1), 102140 (2020)
10. Ocampo Diaz, G., Ng, V.: Modeling and prediction of online product review helpfulness: a survey. In: Proceedings of the Fifty-Sixth Annual Meeting of the Association for Computational Linguistics (Melbourne, Australia, July 2018), pp. 698–708. Association for Computational Linguistics, Stroudsburg
11. Ren, Y., Zhang, Y.: Deceptive opinion spam detection using neural network. In: Proceedings of COLING 2016, the Twenty-Sixth International Conference on Computational Linguistics: Technical Papers (Osaka, Japan, Dec. 2016), The COLING 2016 Organizing Committee, pp. 140–150
12. Xu, L., Bing, L., Lu, W., Huang, F.: Aspect sentiment classification with aspect-specific opinion spans. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (Online, Nov. 2020), pp. 3561–3567. Association for Computational Linguistics, Stroudsburg
13. Xu, L., Li, H., Lu, W., Bing, L.: Position-aware tagging for aspect sentiment triplet extraction. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (Online, Nov. 2020), pp. 2339–2349. Association for Computational Linguistics, Stroudsburg

Chapter 2
Modeling Financial Opinions

In this chapter we lay out the primary background of a financial opinion and the relation between opinions and financial instruments. Together, these constitute an overall picture of opinion-based market interaction. Following this discussion we propose several research issues. First, in Sect. 2.1, we discuss the components in a financial opinion one by one, as well as potential research directions; we also explain why we need to extract components and estimate their quality (or influence). After recognizing the components in an opinion, in Sect. 2.2 we identify the relationship between components based on the notions of argument mining. Then, in Sect. 2.3, we present how argumentation structures between financial opinions are formed by linking each opinion structure. We close the chapter in Sect. 2.4 with the interaction between the financial market and opinions.

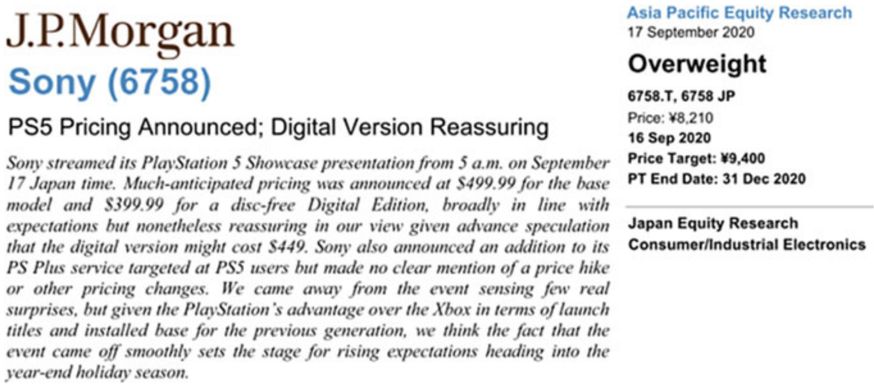
2.1 Opinion Components

2.1.1 Target Entity

As mentioned in Chap. 1, there are 12 components in a financial opinion, that is, an opinion related to a financial instrument. The first important component is the subject of discussion: the target entity. By definition, any monetary contract, including debt, equity, foreign exchange, and derivatives, can be the financial instrument. Because stock is the most common case, we mainly use stocks' examples in this book. The same concepts can be employed for other financial instruments.

In financial narratives, investors tend to tag the target entity with a unique ticker symbol. For example, investors use 6758 to represent the stock of Sony Corporation in Japan. The equity of a given company may be listed on multiple stock

C.-C. Chen et al., *From Opinion Mining to Financial Argument Mining*,
SpringerBriefs in Computer Science,
https://doi.org/10.1007/978-981-16-2881-8_2



J.P.Morgan

Sony (6758)

PS5 Pricing Announced; Digital Version Reassuring

Sony streamed its PlayStation 5 Showcase presentation from 5 a.m. on September 17 Japan time. Much-anticipated pricing was announced at $499.99 for the base model and $399.99 for a disc-free Digital Edition, broadly in line with expectations but nonetheless reassuring in our view given advance speculation that the digital version might cost $449. Sony also announced an addition to its PS Plus service targeted at PS5 users but made no clear mention of a price hike or other pricing changes. We came away from the event sensing few real surprises, but given the PlayStation's advantage over the Xbox in terms of launch titles and installed base for the previous generation, we think the fact that the event came off smoothly sets the stage for rising expectations heading into the year-end holiday season.

Asia Pacific Equity Research
17 September 2020

Overweight

6758.T, 6758 JP
Price: ¥8,210
16 Sep 2020
Price Target: ¥9,400
PT End Date: 31 Dec 2020

Japan Equity Research
Consumer/Industrial Electronics

Fig. 2.1 A professional analyst report about the target entity Sony with the ticker symbol 6758 JP

Fig. 2.2 A post by a financial social media user showing their opinion on the target entity Sony with the ticker symbol SNE

exchanges, for instance the Tokyo Stock Exchange, New York Stock Exchange, and the London Stock Exchange. The ticker symbols of the equity of Sony Corporation in these exchanges are 6758, SNE, and SON, respectively. In this case, in some financial documents identifying the target entity is straightforward. Figures 2.1 and 2.2 show documents written by a professional analyst and a financial social media user, respectively. Use of ticker symbols (6758 JP in Fig. 2.1 and SNE in Fig. 2.2) for the mentioned target entity reflects investor consensus.

2.1.2 *Market Sentiment*

In the general domain, sentiment can be positive, neutral, or negative, whereas in the financial domain, market sentiments are bullish, neutral, or bearish. On most financial social media platforms, writers can provide a market sentiment label—either bullish or bearish—before posting their opinions. Figure 2.2 depicts an example with a bullish label. Note that bullish (bearish) market sentiment means the writer thinks the price of the target entity will rise (fall).

In some cases, including analyst reports, the definition of market sentiment is slightly different. It can differ, for instance, across various institutions, as shown in Fig. 2.1, where market sentiment is overweight, neutral, or underweight. Such a rating is given based on a comparison with other stocks. For instance, according to the definition of J.P. Morgan, the meanings of these market sentiment labels are as follows:

- Overweight: The target entity will outperform the average return of the stocks that have been analyzed by this analyst or this team in the next six to twelve months.
- Neutral: The target entity will perform according to the average return of the stocks that have been analyzed by this analyst or this team in the next six to twelve months.
- Underweight: The target entity will underperform the average return of the stocks that have been analyzed by this analyst or this team in the next six to twelve months.

In this case, an overweight rating does not mean the price will rise. It simply means that the target stock may outperform other stocks, either by rising more or by falling less.

Simple market sentiment is used in the reports of other analysts, who use buy, hold, and sell to represent their market sentiments. This kind of definition is based on the expected return of the target entity. If the return is expected to go up (down), they recommend that their customers buy (sell) the target stock. A more complex setting is also common, in which analysts set a threshold for going up and down. For example, they assign a "buy" ("reduce") label to the stock if and only if the expected return is higher (lower) than 10% (−10%). For expected returns between 10% and −10%, they assign a "hold" rating.

In summary, market sentiment can be represented in various ways; its definition is typically provided within the reports or platforms themselves.

2.1.3 Opinion Holder

The same opinion held by different people may have different influences on the market. For example, the opinions in Figs. 2.1 and 2.2 are bullish opinions about the equity of Sony, but the opinion holders are different. The opinion in Fig. 2.1 is likely to be read by far more people than that in Fig. 2.2, which indicates the importance of recognizing and analyzing the opinion holder. Indeed, one important research topic is determining whether a given opinion is coming from a trustworthy opinion holder. The opinion holder's wider network may also influence the trustworthiness of an opinion.

We classify the opinion holders into the following groups:

- By opinion source: managers, professionals, social media users, and journalists.
- By expertise: professional investors and amateur investors.
- By historical performance: accurate investors and inaccurate investors.

In Chap. 3, we further discuss opinions from different sources.

2.1.4 Publishing Time and Validity Period

With financial opinions, temporal information is much more important than in opinions from other domains. For instance, while opinions about the PlayStation 5 Console from 2020 may still be useful for those who want to buy the PlayStation 5 Console two years later, in 2022, bullish opinions on the equity of Sony in 2020 will most likely be worthless in 2022. This explains the need to note the publishing time and estimate the validity period of financial opinions.

In most cases, the publishing time is easily obtained from the title of the document (for analyst reports) or from the platform metadata (social media posts). The publishing time helps us arrange opinions in order and can be used to link opinions with market data. For example, the price target of an investor and the close price of the target entity are paired to evaluate the degree of investor sentiment. Note that the price target is the price level that investors think the price of a financial instrument will be at.

The validity period is also an important concept in financial opinions. In the report depicted in Fig. 2.1, the publishing time is 17 Sep 2020, and the analyst has set the "PT End Date" to 31 Dec 2020. However, most financial opinions do not provide an exact validity period, which complicates the estimation of the validity period of financial opinions; this remains an open problem. When all opinions are viewed on a timeline, temporal information plays a crucial role. More details are provided in Sect. 2.4.

2.1.5 Market Information

Investors analyze financial instruments based on the information available prior to the publishing time of their opinion. In many cases, especially with social media posts, market information is known to investors and is not included in their posts. For example, they only state "$SNE Target 150 March 2021" and do not provide a sentiment label. Understanding that "150" is the price target of $SNE, we must ascertain the close price of $SNE in order to infer the investor's sentiment. That is, if the close price of $SNE is higher (lower) than 150, this investor possesses a bearish (bullish) sentiment about $SNE. This not only shows the importance of recording the market information before the publishing time, but also indicates that the numerals in financial narratives are crucial for understanding financial opinions. Indeed, Chap. 5 is devoted entirely to research on this topic.

2.1.6 Aspect

Basically, investors analyze the financial instruments from two aspects: fundamental and technical. Based on financial or economic factors such as financial statements, fundamental analysis is used to evaluate the value of the target financial instrument.

Table 2.1 Taxonomies of aspects proposed in FiQA-2018 [3] and NumAttach [1]

FiQA-2018 [3]			NumAttach [1]
Level 1	Level 2		
Corporate	Price action	Strategy	Asset
Stock	Technical analysis	Strategy	Liability
Economy	Coverage	Legal	Equity
Market	Risks	Fundamentals	Income
	Financial	Market	Economics
	Sales	Volatility	Indicator
	Signal	Insider activity	Pattern
	Dividend policy	Reputation	
	Options	Conditions	
	M&A	Regulatory	
	Rumors		

Technical analysis, in turn, uses historical data such as price or trading volume to predict price movement.

These two aspects can be further extended into various subcategories. For example, investors can base their analysis on many different parts of the financial statement, including assets, liability, or equity terms [1]. Events such as mergers and acquisitions (M&A) and lawsuits can also be considered as different aspects [3]. Different technical analysis methods can be adopted for different aspects. Aspects proposed in the literature are listed in Table 2.1.

2.1.7 *Elementary Argumentative Units*

In this section, we introduce the elementary argumentative units of financial opinions based on Toulmin's argumentative model [4], shown in Fig. 2.3. *Claim* and *premise* are two basic units of an argument: claim is the subjective view of the investor, and premise is the objective fact used to support the claim.

Warrant is the background knowledge that causes an investor infer the claim based on the premise, and *backing* is used to support the warrant. Assume that the analyst states a claim of EPS growth based on a premise of improved margins via

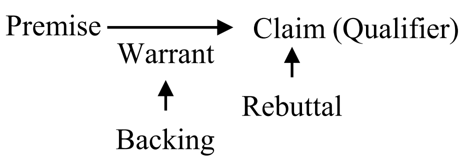


Fig. 2.3 The relation between a premise and a claim viewed by Toulmin's argumentative model

labor efficiency. The warrant is this: more efficient labor will help us produce more in the same amount of time, which will lead to increased income. In this case, the backing is simply accounting common sense. Here, warrant and backing are implicit information in the argumentation. That is, generally, warrant and backing are not written down in the argumentative documents.

In argumentative models, the qualifier represents the strength of the claim, and can be the investor's confidence. In Fig. 2.1, the price target can be taken as a proxy for the confidence of the analyst. The qualifier can also be considered as the degree of market sentiment. Finally, the *rebuttal* is composed of counterarguments meant to defeat the claim. We explain the rebuttal in detail in Sect. 2.3 when we construct the argumentation structure between opinions.

2.1.8 Opinion Quality

As mentioned above, the qualifier represents the confidence of investors in their opinions. Another evaluation metric is the quality of the opinion. Figures 2.1 and 2.2 show that opinions may have different weights with different investors due to their quality. Note that the evaluation of the quality of a financial opinion is still an open problem: the interpretation of the opinion is affected by the rationality of the inference, the writing style of the opinion holder, and so on.

In this book, the quality of a financial opinion is determined based on the rationality between the claims and the premises. That is, we evaluate whether the specific premises are trustworthy, and further determine whether the inference from these premises is reasonable given the claims. This is an objective evaluation of the premises supporting the investor's analysis, as opposed to the subjective confidence of the investor. An investor may be very confident about a certain trading strategy, but sometimes the setting of the strategy may not make sense to others. This raises another research question: whether rational analysis always lead to profitable results? Since there is little discussion in this direction, this topic is worthy of investigation. Because quality is related to the argumentative units, we illustrate the relation between all opinion components in Sect. 2.2.

2.1.9 Influence

In financial opinion mining, we seek to predict market movement based on investor opinions. We must thus judge whether the given opinion will influence the market, and how much of an impact it will have. Although an investor may provide sound analysis, it is possible that this investor has not entered the market, or that no other investors view the analysis. In such a case, is it prudent to consider this opinion when analyzing the financial instrument that is the subject of this analysis? On the other hand, a sensational article headline or a social media post with false —or even fake—information can have a big impact on the market. Therefore, to understand how the opinion will influence the market, the influence power of the opinion must be considered.

2.2 Argumentation Structure in Opinions

After defining all the components of a financial opinion, we now construct a graph that shows the relationships of these components. Figure 2.6 shows the argumentation structure of the analysis in Fig. 2.4. In this report, the main claim (*MC*) on Michaels stock is overweight, which is the final market sentiment (s) of this opinion. The

J.P.Morgan

Michaels

Overweight

MIK, MIK US

Price (16 Sep 20): $10.18

Price Target (Dec-20): $16.00

We expect the following: (1) Long-term targets of LSD SSS, MSD EBIT growth and HSD EPS growth driven by MIK's ongoing retail 101, omni-channel, and makers/Pro initiatives to drive topline/share (see bullet below) with the opportunity to improve margins through labor efficiency, merchandising rigor, inventory flow disciplines, cost leverage, and sourcing/private label expansion. **(2)** Capital/investment spending to remain relatively consistent with history given modest new store growth and a highly manageable omni-channel investment cycle (i.e., no need for a big supply chain or tech stack buildout); MIK targeted 2.5-3.0% of sales for capex on its last analyst day. **(3)** In terms of capital allocation, at the last analyst day MIK also targeted excess free cash flow solely to share repurchases (and we highlight its current FCFE yield of 23% and FCFF yield of 13%). However, new management has rightfully acknowledged that the company's financial leverage (5.5x gross debt to EBITDAR on our '21 estimates) is holding back its valuation given algorithmic trading and some value investors' aversion to leverage. This suggests some of the FCF could be dedicated to debt paydown, as well as repo, and hence our view of a HSD EPS growth rate vs. the old team's target of 10-15%. Notably, with MIK currently refinancing its term loan and the peak holiday inventory build happening now, we see the potential for MIK to start share repurchases as early as November, although we have not modeled any until 1Q (with 2021 embedding a total repo of 16MM shares for ~$250MM). **(4)** Recall, we believe that MIK was comping in the 20's QTD when it reported on September 3rd. Given the shift of back-to-school spending into September and lateral checks, we believe MIK's trend has sustained to slightly accelerated this month, although it remains unclear if management will speak to QTD.

Fig. 2.4 Arguments of a professional analyst

analyst makes six claims (c) to support the main claim, most of which are supported by one or more premises (p).

Below we introduce the basic concepts behind different argumentation structures. Firstly, the structure from p_1 to the MC is termed a *sequential structure*, where w denotes the persuasiveness of the premise for the claim. Secondly, the structure of (p_2, p_3, c_2, c_3) is a named linked argument, where p_2 supports c_2 and c_2 is also supported by c_3 with p_3. Thirdly, claims such as c_4 may not be supported by any premises. Fourthly, the structure of (p_4, c_5, c_6) is a divergent argument, where two claims are supported by the same premise. Lastly, the full argumentation structure is a hybrid structure.

In Fig. 2.6, parameter w denotes the weight of the premise supporting the indicated claim. Many proxies could be used as w. For example, the warrant for inferring the premise to the claim is one possible proxy. The rationale behind using this premise to support the claim is also a possible proxy. Parameter w thus influences q, the quality or qualifier of the claim, and q has a further impact on the main claim. That is, w influences the trustworthiness or quality of the financial opinion.

Based on the above rationales, the relationships between the opinion components can be listed as follows. The investors can make claims from different aspects to support the main claim. Thus, the aspect is related to individual claims instead of linking to the main claim directly. Additionally, the market sentiments of the claims can differ from the main claim. For example, investors may consider both bullish and bearish perspectives to come to their final decisions. The validity period of the main claim and the claims may be different, because investors may take both short- and long-term influence into account. Because other investor's opinions may become the premise of the other opinion holders, the opinion holder of the main claim may be different from that of the premises. Finally, the opinion quality of the main claim will be influenced by q of the claims and w of the warrants or premises that directly support the main claim.

Previous work shows that modeling the argumentation structure in this way is useful for evaluating the quality of persuasive essays [5] and the persuasiveness of online debates [2]. However, few studies adopt this idea to analyze investor opinions. In this section, we not only provide an example of representing investor opinion as an argumentation structure, but also show that we can evaluate the rationality of each node pair in the structure and assign weights to the edges. Given rationality or quality scores, the argumentation structure becomes a directed weighted graph. This kind of structure also better reflects an investor's behavior when reading a report.

2.3 Argumentation Structure Among Opinions

As mentioned in Sect. 2.1.7, investors regularly debate price movements. Figure 2.5 shows the argumentation structure of the opinions expressed during a discussion conducted on an online forum. The original poster makes a claim about TSM's price and backs this up with several premises from different aspects. The first reply, R1,

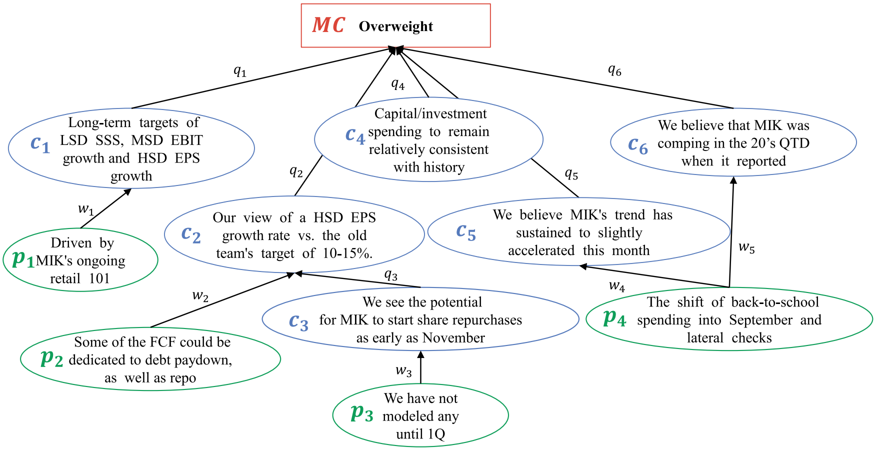


Fig. 2.5 Argumentation structure of the report in Fig. 2.4

which agrees with the original post, can be considered as supporting the main claim of the original post. The second reply, R2, supports one of claims of the original post. The third and the fourth replies, R3 and R4, attack the main claim of the original post from different aspects. In this case, R3 and R4 are rebuttals of the claim in the original post.

Because the components mentioned in Sect. 2.2 are inherent in a financial opinion, support or attacks from other opinions may not influence those components. Interaction between opinions at time t can be considered as the premises of other opinions at time $t + 1$. In contrast to analyzing a single financial opinion, the readers of the thread in Fig. 2.5 treat the discussion as an opinion, and consider it based on the concepts outlined in Sect. 2.2. We discuss this in detail in Sect. 2.4.

As in an online debate platform on which debaters discuss a given topic over several rounds, posters in online financial forums discuss the possible price movement directions over several rounds from different aspects. This makes it possible for us to adopt the concept of supports and attacks from argument mining to evaluate the persuasiveness of the original post. We can further construct a larger argumentation graph, where all arguments of the investors are connected using edges denoting bullish/bearish stances toward certain financial instruments. Comparing the rationales from the investors from both stances allows us not only to link opinions from different investors and different documents to a graph, but also to formulate an explanation of the decision process.

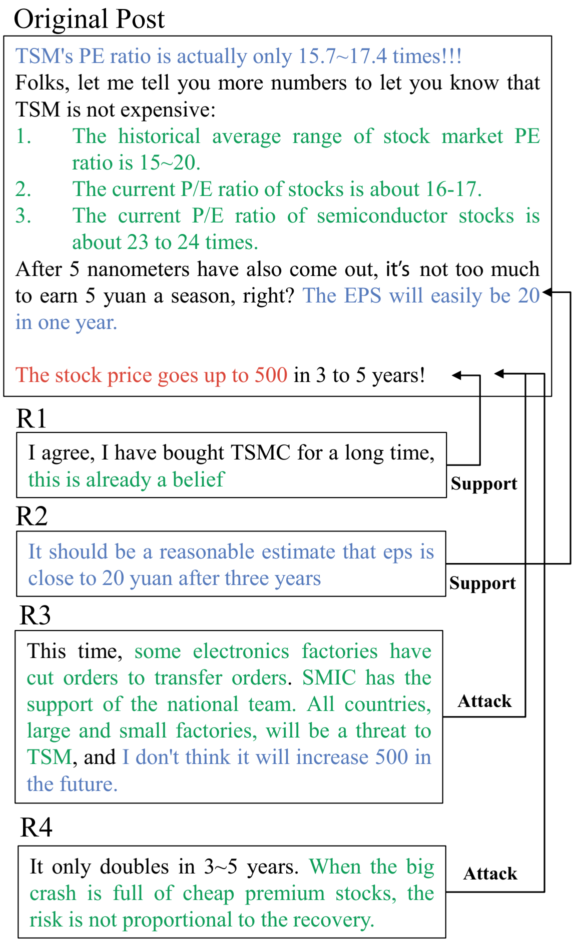


Fig. 2.6 Argumentation structure among opinions. The span in red represents the main claim of the original post, spans in blue denote claims, and spans in green denote premises

2.4 Relations Among Opinions and Target Entities

Investor opinions are linked to the target financial instrument (e) and may influence outcomes—such as the stock price—in the next time step. Figure 2.7 shows an example discussing the relations of opinions (O) and financial instruments (e), where U and D denote bullish and bearish, respectively. UI denotes an investor with bullish opinion and long e, DI denotes an investor with bearish opinion and short e, and UN denotes an investor with bullish opinion who takes no actions in the market. At time t in Fig. 2.7, the facts related to e_1 ($P^{e_1}_{1,t}$, $P^{e_1}_{2,t}$, $P^{e_1}_{3,t}$, and $M^{e_1}_t$) are considered the premises, where $M^{e_1}_t$ denotes market information such as the close price of e_1. For

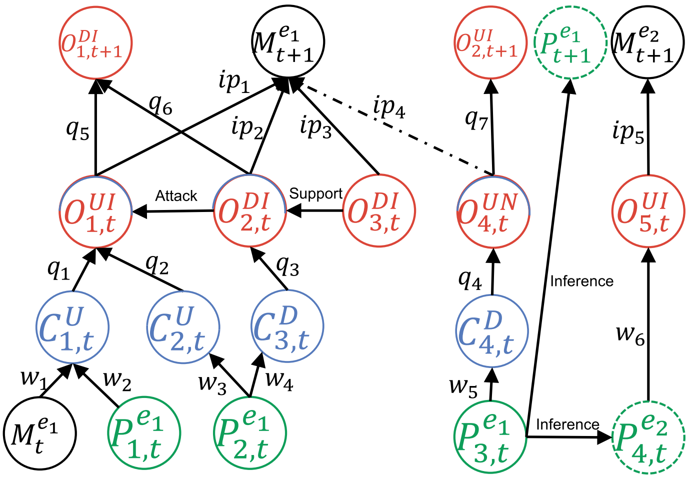


Fig. 2.7 Relations among opinions and target entities

example, claim $C^U_{1,t}$ is based on $M^{e_1}_t$ and $P^{e_1}_{1,t}$. Because good news does not always lead to increased stock prices, the same premise may lead to different claims. The structure of $P^{e_1}_{2,t}$, $C^U_{2,t}$, and $C^D_{1,t}$ is an example of this case.

After investors form their opinions based on the given facts, they may take actions $O^{UI}_{1,t}$, $O^{DI}_{2,t}$, and $O^{DI}_{3,t}$ in the market—or they may do nothing ($O^{UN}_{4,t}$). This leads to a problem. This example includes two bullish opinions and two bearish opinions. Should we therefore conclude that the investors currently have neutral attitudes about e_1? If we remove $O^{UN}_{4,t}$ from the market, will the stock price fall due to the two bearish opinions? Consider an example. If the opinion holder of $O^{UI}_{1,t}$ buys 1,000 shares and the opinion holders of both $O^{DI}_{2,t}$ and $O^{DI}_{3,t}$ only short 5 shares, the influence power (*ip*) of $O^{UI}_{1,t}$ may be greater than that of others. This shows the importance of evaluating the *ip* of an opinion. Since this is little discussed in the literature, it is still an open problem.

The opinions at time t not only influence the market at time $t + 1$, but also become the premises for opinions at time $t + 1$. The opinion holder of $O^{UI}_{1,t}$ may change his/her view from bullish to bearish ($O^{DI}_{1,t+1}$) after considering the attack of $O^{DI}_{2,t}$ on the original opinion (i.e., $O^{UI}_{1,t}$). Although $O^{UN}_{4,t}$ may not influence the stock price, the rationale of this opinion may become the premise of someone's opinion in the next time step (i.e., $O^{UI}_{2,t+1}$). Thus, another interesting topic for future work is how to construct a graph that represents the interaction between opinions over time.

Last, although $P^{e_1}_{3,t}$ is a fact related to e_1, it may also be implicitly related to other entities (e_2). That is, an investor may make a claim about e_2 based on information about e_1. Additionally, investors can also infer possible events for e_1 at time $t + 1$ ($P^{e_1}_{t+1}$) based on the given facts at time t ($P^{e_1}_{3,t}$). In Chap. 4, we will discuss how to infer implicit relations between entities given the results of previous work.

2.5 Summary

In this chapter, we provide an overview of the opinion-based financial market, introducing the inherent components of a financial opinion and adopting the concept of argument mining to link financial opinions. We propose an overall picture of financial opinions and the financial instruments. In the rest of this book, we further discuss the sources of opinions and the methods explored before based on the notions proposed in this chapter.

Since there is much discussion about the operations of the financial market, the ideas in this chapter are just one of the possible pictures of the market. We seek to provide an opinion-based point of view so that readers can understand the goal of this book.

References

1. Chen, C.-C., Huang, H.-H., Chen, H.-H.: Numeral attachment with auxiliary tasks. In: Proceedings of the Forty-Second International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1161–1164 (2019)
2. Li, J., Durmus, E., Cardie, C.: Exploring the role of argument structure in online debate persuasion. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (Online, Nov. 2020). Association for Computational Linguistics, pp. 8905–8912 (2020)
3. Maia, M., Handschuh, S., Freitas, A., Davis, B., McDermott, R., Zarrouk, M., Balahur, A.: WWW'18 Open Challenge: financial opinion mining and question answering. In: Companion Proceedings of the The Web Conference 2018, pp. 1941–1942 (2018)
4. Toulmin, S.E.: The Uses of Argument. Cambridge University Press, Cambridge (2003)
5. Wachsmuth, H., Al Khatib, K., Stein, B.: Using argument mining to assess the argumentation quality of essays. In: Proceedings of COLING 2016, the Twenty-Sixth International Conference on Computational Linguistics: Technical Papers, pp. 1680–1691 (2016)

Chapter 3
Sources and Corpora

In this chapter, we focus on the sources of financial opinions; we group these sources by the opinion holders: insiders (Sect. 3.1), professionals (Sect. 3.2), social media users (Sect. 3.3), and journalists (Sect. 3.4). Each opinion holder may have his/her own goals when expressing opinions, resulting in different opinions from unique viewpoints. In this chapter we discuss related research topics and findings, including opinion mining related work in both the finance and computer science domains.

3.1 Insiders

Before introducing the opinions of different opinion holders, it is necessary to understand the process when information is released. Figure 3.1 shows the timeline from the establishment of a fact to that fact becoming well-known. From time t^h to time t^p, the information is known only by a few insiders in the institution. During this period this is called *inside information*. At time t^p, the insider—for instance the manager—publishes the information to the market. Once published, this becomes *public information*. For example, managers naturally know the number of orders for the next three months; this fact is established at time t^h, at which point only the insiders know this information. Note that in most cases, insiders are bound by law to keep this kind of information secret. They must abstain from disclosing insider information and must not use it for trading. This information is not released until it is publicly communicated by managers at time t^p, for instance during earnings conference calls, which may be three months after t^h. Initially, this information may be available only to analysts and other participants in the calls. Then, as they begin to spread the news that they heard in the call, the information gradually becomes more widely known.

C.-C. Chen et al., *From Opinion Mining to Financial Argument Mining*,
SpringerBriefs in Computer Science,
https://doi.org/10.1007/978-981-16-2881-8_3

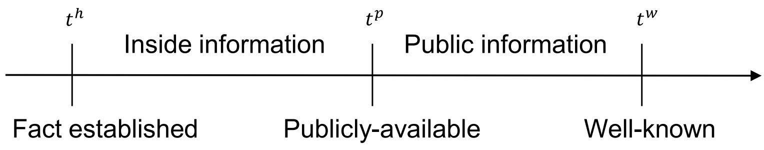


Fig. 3.1 The timeline from the establishment of a fact at t^h to that fact becoming public at t^p and then well-known at t^w

Table 3.1 Information sources from stock market insiders

Source	Meaning
Form 10-K	An annual, detailed report on company operations. This report is required by the supervising agency
Form 10-Q	A quarterly report on company operations. Unlike the 10-K report, some information in the 10-Q report is unaudited
Form 8-K	Used to publish unscheduled events or changes in the company's operations
Annual general meeting	A mandatory meeting held to relay the previous year's operations and present the future directions of the company. Shareholders express their opinions on operations by voting in this meeting
Earnings conference call	Generally held quarterly, this call provides a forum for managers to relay company operations to investors
Speeches or interviews	Managers may be invited to share their view on the industry or be interviewed about company operations. These public speeches may also contain their personal opinions

Given this process, the opinions of managers and other insiders are clearly most crucial when analyzing financial instruments. In this section, we use the stock market as an example, and then extend the concept to other financial instruments. In the stock market, insiders are managers in a company. Since divulging insider information is prohibited by the company and trading based on insider information is forbidden by governments, in most cases we are limited to mining public information. Table 3.1 shows the possible sources of opinions from managers. Note that sources such as Form 10-K provide only historical financial information about the company, such as the previous year's earnings. Below, we discuss the findings of previous work, which uses the sources in Table 3.1. Source names such as Form 10-K, Form 10-Q, and Form 8-K follow the U. S. Securities and Exchange Commission. In other countries, although the names of these reports may differ, their meanings remain the same. Relevant forms not listed here can be found in the EDGAR database,[1] which additionally contains all regulatory reports for the listed companies.

Loughran and McDonald [24] find that in the Harvard Dictionary, about three-quarters of the words considered to be negative words in the general domain are not negative in the financial domain. They propose six word lists for financial narratives

[1] https://www.sec.gov/edgar/search/.

from the following aspects: negative, positive, uncertainty, litigious, strong modal, and weak modal. Based on these word lists, their experimental results show that the more negative words there are in the 10-K, the lower the excess returns near the report release date are. All word lists are significantly related to stock return volatility. In addition, negative, uncertainty, and litigious word lists are significantly related to fraud lawsuits. Thus, the negative and positive word lists seem to simply reflect events that have already occurred; likewise, the litigious list does not concern opinions. It is the uncertainty and strong/weak modal word lists that concern implicit information, and thus reveal manager opinions.

The Management Discussion and Analysis (MD&A) section in the 10-K report is considered an important part for analyzing the manager's opinions on both past operations and future directions of the company. Wang et al. [37] adopt the word lists of Loughran and McDonald [24] to extract textual features from the MD&A. Their work shows that sentiment words in MD&A are highly correlated with volatility, i.e., company risk. Rekabsaz et al. [28] propose a fusion method with textual data in both the 10-K report and the market data. Their model outperforms GARCH [14] and the SVM model presented by Wang et al. [37].

10-Q reports, in turn, contain information that is similar to that in the 10-K reports. These reports cover operations in the previous quarter, and also contain an MD&A section. Here is the statement from Apple Inc.'s 10-Q report in Q3 2020.[2]

> This section and other parts of this Quarterly Report on Form 10-Q ("Form 10-Q") contain forward-looking statements, within the meaning of the Private Securities Litigation Reform Act of 1995, that involve risks and uncertainties. Forward-looking statements provide current expectations to future events based on certain assumptions and include any statement that does not directly relate to any historical or current fact. For example, this Form 10-Q describes forward-looking statements which regard the potential future impact of the COVID-19 pandemic on the Company's business and results of operations. Forward-looking statements can also be identified by words such as "future," "anticipates," "believes," "estimates," "expects," "intends," "plans," "predicts," "will," "would," "could," "can," "may," and similar terms. Forward-looking statements are not guarantees of future performance and the Company's actual results may differ significantly from the results discussed in the forward-looking statements.

This statement shows the importance of the MD&A section, and also indicates that the section contains manager opinions based on the given facts. From this statement, we see that financial opinions focus mainly on forward-looking views as opposed to explaining what has already happened.

To retrieve the latest information about a company, we look for 8-K reports about unscheduled events. Although the report itself contains no opinion, as illustrated in Fig. 2.7, it is fundamental to an informed financial opinion. Thus, automatic extraction of events in the 8-K report is related to financial opinion mining.

Zheng et al. [44] propose Doc2EDAG, a document-level event extraction method for extracting financial events from 8-K reports (event-related announcements) in Chinese. They focus on five event types: equity freezes, equity repurchases, under-

[2] https://s2.q4cdn.com/470004039/files/doc_financials/2020/q3/_10-Q-Q3-2020-(As-Filed).pdf.

weight equities, overweight equities, and equity pledges. They achieve F1-scores of 70.2%, 87.3%, 71.8%, 75.0%, and 77.3% for these event types, respectively.

After understanding the company events at time t, investors often seek to infer what will happen next, i.e., the events at time $t + 1$. Based on 8-K reports, Zhai and Zhang [43] propose future event forecasting, which they formulate as a sequence-to-sequence task. For model input they use known (past) event sequences, and train the models to generate future event sequences. Their experimental results show that forecasting a company's future events remains a difficult problem.

In addition to regulatory documents, managers' public speeches and other communication also provide meaningful cues for investors by which to analyze a company's operations. Annual general meetings and earnings conference calls are the most common meetings between managers and investors. Both meetings can reveal managers' opinions. Although the agendas of annual general meetings are always recorded, the discussions are not always transcribed. In this part, we use the earnings conference calls to discuss what can be known from such communication. Transcriptions of earnings conference calls are also publicly available on sites such as Seeking Alpha.[3]

Professional analysts often update their reports after attending earnings conference calls. Based on what they learn from the call, they either maintain or change their market sentiment toward the stock of the company. Keith and Stent [18] model analysts' decisions via features extracted from earnings conference calls, and show that semantic features (Doc2Vec [22] and bags of words) are more predictive than both market features and pragmatic features (named entities, predicates, sentiments, etc.). They also suggest using the whole document instead of a selection of parts such as the Q&A section. Price et al. [26] show that sentiment in earnings conference calls is significantly related to abnormal returns and trading volume, and the Q&A section in the earnings conference calls has more explanatory power than the document as a whole. Ye et al. [42] use multi-round Q&A features in their model, which outperforms the model of Theil et al. [35] in 3-day, 7-day, and 15-day volatility prediction.

Many studies use the audio and transcriptions of earnings conference calls to predict stock volatility. Qin and Yang [27] feed both verbal and vocal features to a contextual bidirectional LSTM model, and further merge these features to predict volatility. They show that using both audio and textual data is significantly better than only using either audio or textual data for 3-day, 7-day, and 15-day volatility prediction. Yang et al. [41] follow Qin and Yang's work [27] and propose a hierarchical transformer-based model under a multi-task setting. They show that jointly learning the average n-day and single-day volatility improves model performance. Their results also indicate that with their architecture, audio information may not be needed for 15-day and 30-day forecasting. Sawhney et al. [30] use graph convolution networks to further improve 3-day and 7-day results.

The above studies show the importance of insider opinions. In the foreign exchange market, insiders can be members of central banks such as the Federal Reserve Board of Governors in the U. S. For example, speeches given by the Chair of

[3] https://seekingalpha.com/earnings/earnings-call-transcripts.

the Federal Reserve always attract investor attention, because they reveal the attitude toward the U. S. Fed Funds Target Rate. Some studies [1, 2] use the Beige Book—the Summary of Commentary on Current Economic Conditions—as a source, and show that the content of the Beige Book is significantly predictive of GDP growth and aggregate employment. Sadique et al. [29] indicate that the tone in the beige book influences stock market volatility and trading volume.

The Minutes of the Federal Open Market Committee is another important source from which to mine opinions from important decision-makers. Stekler and Symington [33] use keywords to construct an index to reflect the sentiment (optimistic/neutral/pessimistic) of the Federal Reserve System (the Fed). They also consider the degree of sentiment and separate the keywords into several classes. They show that the proposed index facilitates the capturing of cues for forecasting the future economic environment. Ericsson [15] show that Stekler and Symington's index can be used to forecast the real US GDP growth rate in the Green Book, another Fed publication. All of the aforementioned sources and other related sources can be downloaded from the official website of the Federal Reserve System.[4]

In summary, researchers analyze the information at time t^p in Fig. 3.1 to capture past facts. Additionally, investors also attempt to mine (predict) inside information based on publicly-available information, because the tone or expressions of insiders sometimes discloses (implies) information that they have not yet published. Generally, because insiders have more information than other market participants, their opinions are considered the most important. That is why professional analysts frequently contact the CEO or CFO of the companies directly: Brown et al. [5] show that over half of 365 surveyed analysts visit or contact the CEO or CFO more than four times a year.

However, does the market always follow insider opinions? That is, are their opinions always correct? Han and Wild [16] show that when managers report good news about the company, they tend to release forecasts that are more optimistic than those of analysts. Jelic et al. [17] indicate that when earnings decline, management earnings forecasts become more inaccurate, based on their statistics of Malaysian initial public offerings (IPOs) from 1984 to 1995. Findings of previous works thus indicate that even given insider opinions, we must still evaluate the quality of these opinions based on the premises and facts given.

3.2 Professionals

In the financial domain, many knowledgeable people are considered professionals, including professors in finance departments, analysts in financial institutions, economists, and so on. A financial analyst is one such professional who collects as much information as possible and further analyzes the value of the financial instrument based on this information. Vukovic et al. [36] show that the Russian stock market

[4] https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm.

significantly reflects analysts' recommendations, which shows the importance of the professional opinions. In this section, we focus on the opinions of financial analysts.

As mentioned in Sect. 3.1, analysts visit or contact CEOs or CFOs directly to get the latest information. This is in contrast to common investors, who cannot expect to get such first-hand information from managers. Such privileged access for professionals explains their influence on market investors. Professionals generally share their opinions via analysis reports; sometimes they also give speeches or interviews. This is unlike the regulatory reports of companies, which must be purchased. For example, investors and researchers can download analysts' reports using systems like Bloomberg Terminal or Thomson Reuters Eikon, but using these systems is often costly.

Other studies focus on the interaction between companies and analysts. Cohen et al. [11] show that a company calling on many bullish analysts during earnings conference calls may actually be a cue for poor future earnings. The findings in Keith and Stent [18] may explain this. They analyze the behavior of analysts in earnings conference calls and present the following findings:

- In the question-answering section, bullish analysts are called on earlier to ask questions than other analysts with neutral or bearish sentiment toward the company.
- Bullish analysts ask more positive questions in the earnings conference call, and ask more questions about organizations.
- Bearish and neutral analysts ask more about past events.

These studies not only show that companies do care about the opinions of professional analysts but also indicate that these analysts' opinions (questions) can influence the company's future asset price.

Also, similar backgrounds and knowledge for professionals is no guarantee that their opinions will also be similar: differing analysis methods or information can result in different opinions and in reports with different levels of accuracy. Zong et al. [45] order analyst reports by their accuracy in earnings forecasting, and compare the semantic features of the 4,000 most accurate reports with those of the 4,000 most inaccurate reports. They find that the number of uncertain statements, the amount of future temporal orientation, and the number of negative words are significantly associated with inaccurate reports. Accurate reports, in turn, use more cardinal numbers, nouns, and positive words. Accurate reports focus more on past events as opposed to describing present and future events. They also use the BERT architecture [12] to identify whether a given report is accurate or inaccurate, yielding accuracies from 64% to 70%. Their work provides insight on how to evaluate the quality of analysts' opinions.

Professional opinions influence the market and other investors. Additionally, companies respect the opinions of professionals. Thus, their reports make it possible not only to understand their opinions but also to glean useful information from the interaction between analysts and insiders.

3.3 Social Media Users

Anyone can be a social media user. Insiders and professionals may have public or private accounts on social media platforms. Information posted using their public accounts can be considered as coming from insiders or analysts. However, information posted using private accounts, which could be anonymous, would be considered at the same level as posts from non-professionals. In this section, we focus on the information coming from users whose background we cannot easily discern: most social media users fit this criterion. Although the opinion of an individual social media user may not be as influential as that of an insider or a financial analyst, the opinions of a group of social media users could represent the view of amateur investors, i.e., non-professional investors. Because the price of a financial instrument moves based on all market participants, the view of such amateur investors clearly should also be considered when making investment decisions.

Some studies use the general sentiment of social media data as a feature when predicting price movements. Bollen et al. [4] show that the mood or general sentiment of Twitter users is correlated to the Dow Jones Industrial Average Index, in particular the mood from calm and happy aspects. Si et al. [32] adopt a Dirichlet process mixture (DPM) model [34] to analyze the aspect of the tweets, and use this to conduct aspect-based sentiment analysis, showing that adding their features to models improves the accuracy of movement predictions for the S&P 100 index.

In previous work [6], we show the difference between general and market sentiment via financial social media data collected from StockTwits, and propose NTUSD-Fin, a market sentiment dictionary for financial social media data.[5] Li and Shah [23] also use the StockTwits data to construct a market sentiment dictionary. They show that using their proposed dictionary for market sentiment analysis yields better results than other dictionaries. Xu and Cohen [40] directly use the tweets collected from StockTwits and enhance the proposed model with historical market data. Their results show that considering temporal information and adding historical market data both facilitate stock movement prediction. Although their approach does not analyze the market sentiment of each tweet, they still use the opinion of social media users to predict stock movements. Additionally, they released the SockNet dataset[6] for future research.

In addition to analyzing social media users, some studies compare the relations or performance between the opinions of social media users and those of professional analysts. Eickhoff and Muntermann [13] show that when considering opinions from social media platforms, the more platforms are used, the more accurate the results are. They also show that diversity in user ages can decrease accuracy. Based on logit models, they indicate that the opinions of social media users can be used to predict the opinions of professional analysts, and vice-versa. In previous work [7], we compare price targets of professional analysts with those of social media users, yielding the following findings:

[5] http://ntusdfin.nlpfin.com/.

[6] https://github.com/yumoxu/stocknet-code.

- Social media users tend to set more progressive price targets.
- Given the same trading strategy—follow the price targets of investors to buy/sell stocks and use the same stop-loss setting—backtesting results are similar between professional analysts and social media users.
- We also evaluate the informativeness of other kinds of opinions from social media users, including the predicted support or resistance price level, buy-side cost, and sell-side cost. We find that these opinions provide incremental information for trading, especially 3-day and 5-day trading [8].

From this perspective, Fig. 3.1 raises the question: what kind of information do social media users have? The general understanding is that most social media users get information later than insiders and professionals, that is, they get the information at time t^w. However, because anyone could be a social media user, information published at time t^p may eventually be made available on social media platforms as well. Sometimes, insider information or information that has not yet been officially published can be found on these platforms. Chiarella v. United States, 445 U.S. 222 (1980)[7] is an interesting real-world case. Although at the time there were no social media platforms, it may be that information from social media platforms could also be considered hearsay. Below is the syllabus of the case provided by the U.S. Supreme Court:

> Petitioner, who was employed by a financial printer that had been engaged by certain corporations to print corporate takeover bids, deduced the names of the target companies from information contained in documents delivered to the printer by the acquiring companies and, without disclosing his knowledge, purchased stock in the target companies and sold the shares immediately after the takeover attempts were made public.

In the current era, if Petitioner were to share this information on a social media platform, could this be detected and then considered as useful information for trading? This would be an interesting research direction for future work. This case suggests that inside information may find its way to social media platforms too.

We seek to highlight one characteristic of the opinions of social media users. In general, insiders and professionals do not base their decisions on faulty premises or misinformation. However, social media users may use false or fake information to form their opinions. Thus, when analyzing the opinions of social media users, it is essential to determine whether their premises are in fact correct. Given 10,000 annotated financial social media data,[8] we find that over 93% of users on StockTwits, a Twitter-like financial social media platform, failed to provide reasons (premises) for their claims [9], which naturally makes it difficult to check their premises. Presumably, the primary reason for this omission is the word limit (280 words per tweet) of this kind of platform. This suggests that one solution would be to instead use a blog or some other online forum as a source.

Several studies show evidence supporting the usefulness of the wisdom of the crowd in the financial domain. This is thus important information that should be considered in this era.

---

[7] https://supreme.justia.com/cases/federal/us/445/222/.

[8] http://finsome.nlpfin.com/.

3.4 Journalists

Journalists are different from other professionals in the financial domain: in contrast to other professionals, who often share their opinions, journalists focus on collecting and summarizing information. Their main focus is to provide the latest news and publish this information far and wide. Thus, journalists seldom share their own opinions. Below is a list of the kinds of information that can be gleaned from journalistic publications such as news articles or magazines:

- Latest published facts: This could be a summary of an earnings conference call or news of an certain unscheduled event.
- Opinions and editorials[9]: In newspapers or magazines, these contain the opinion of the writer. In these cases, the opinion holder is the writer, and we can consider this opinion to be a professional opinion.
- Professional opinions: In addition to editorials, opinions can also be found within news articles. For example, after an earnings conference call, the journalist may interview professional analysts and list their opinions at the end of the article, in an effort to share the facts released in the earnings conference call.
- Hot topics trending on social media: For example, the article entitled "He turned $5,000 into nearly half a million with the help of Tesla options—now he's all in on just two stocks"[10] discusses a hot topic on Reddit and also shares the opinions of the social media users.

Thus, in contrast to other sources, in most news articles, we focus on extracting opinions from other investors instead of the journalist's own opinions, in which case identifying the opinion holder gains additional importance.

In NTCIR-7, Seki et al. [31] propose a dataset for multilingual opinion mining, one of the subtasks of which is opinion holder extraction. Many studies on general sentiment analysis propose methods for this [3, 10, 19–21, 25, 38, 39]. These methods and their findings also apply in financial opinion mining. We survey these in Chap. 4.

3.5 Summary

In this chapter, we overview the sources of financial opinions based on who is providing the information. We use the stock market as the primary example, and also extend these concepts to the foreign exchange market. Naturally, opinions from insiders are the most important information, because they possess both inside information and public information, which are both crucial for inferring future events such as stock

[9] For example, the Opinion section in The New York Times: https://www.nytimes.com/section/opinion.

[10] https://www.marketwatch.com/story/he-turned-5-000-into-nearly-half-a-million-with-the-help-of-tesla-options-now-hes-all-in-on-just-two-stocks-11606842686.

movements. However, since their opinions may not always be accurate, when considering insider opinions, the most important task is evaluating the quality of the opinion.

The opinions of professionals influence not only the market but also the opinions of other investors. Relevant studies have been conducted on (1) analyzing the interaction between professionals and insiders and (2) observing which features best characterize accurate and inaccurate reports.

After the development of the Web, the wisdom of the crowd became a widely discussed topic. Social media platforms play an important role of opinion sharing for everyone. Many studies have demonstrated the usefulness of opinions from social media users. In the financial domain, however, few studies have discussed how to evaluate individual opinions; they instead focus on using the average of all available opinions. This is a thus a topic that merits further investigation.

It is important to keep in mind that good news does not always lead to rises in a financial instrument's price. Price movement is based on investor opinion. People may have both bullish and bearish opinions on any given fact from various aspects. For example, at first glance, the news "the GPD growth rate is 5.2%" looks like good news. However, if the expected growth rate was 6%, this news is in fact bad news. Thus, more fine-grained analysis is needed to better understand the influence among facts, opinions, and financial instruments.

References

1. Armesto, M.T., Hernández-Murillo, R., Owyang, M.T., Piger, J.: Measuring the information content of the Beige Book: a mixed data sampling approach. J. Money Credit Bank. **41**(1), 35–55 (2009)
2. Balke, N.S., Petersen, D.: How well does the Beige Book reflect economic activity? Evaluating qualitative information quantitatively. J. Money Credit Bank. 114–136 (2002)
3. Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V., Jurafsky, D.: Automatic extraction of opinion propositions and their holders. In: 2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text, vol. 2224 (2004)
4. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. J. Comput. Sci. **2**(1), 1–8 (2011)
5. Brown, L.D., Call, A.C., Clement, M.B., Sharp, N.Y.: Inside the "black box" of sell-side financial analysts. J. Account. Res. **53**(1), 1–47 (2015)
6. Chen, C.-C., Huang, H.-H., Chen, H.-H.: NTUSD-Fin: a market sentiment dictionary for financial social media data applications. In: Proceedings of the First Financial Narrative Processing Workshop (FNP 2018) (2018)
7. Chen, C.-C., Huang, H.-H., Shiue, Y.-T., Chen, H.-H.: Numeral understanding in financial tweets for fine-grained crowd-based forecasting. In: 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI). IEEE, pp. 136–143 (2018)
8. Chen, C.-C., Huang, H.-H., Chen, H.-H.: Crowd view: converting investors' opinions into indicators. In: IJCAI, pp. 6500–6502 (2019)

9. Chen, C.-C., Huang, H.-H., Chen, H.-H.: Issues and perspectives from 10,000 annotated financial social media data. In: Proceedings of the Twelfth Language Resources and Evaluation Conference, pp. 6106–6110 (2020)
10. Choi, Y., Cardie, C., Riloff, E., Patwardhan, S.: Identifying sources of opinions with conditional random fields and extraction patterns. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pp. 355–362 (2005)
11. Cohen, L., Lou, D., Malloy, C.J.: Casting conference calls. Manag. Sci. (2020)
12. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: 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) (Minneapolis, Minnesota, June 2019). Association for Computational Linguistics, pp. 4171–4186
13. Eickhoff, M., Muntermann, J.: Stock analysts vs. the crowd: mutual prediction and the drivers of crowd wisdom. Inf. Manag. **53**(7), 835–845 (2016)
14. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econom.: J. Econom. Soc. 987–1007 (1982)
15. Ericsson, N.R.: Eliciting GDP forecasts from the FOMC's minutes around the financial crisis. Int. J. Forecast. **32**(2), 571–583 (2016)
16. Han, J.C., Wild, J.J.: Stock price behavior associated with managers' earnings and revenue forecasts. J. Account. Res. 79–95 (1991)
17. Jelic, R., Saadouni, B., Briston, R.: The accuracy of earnings forecasts in IPO prospectuses on the Kuala Lumpur stock exchange. Account. Bus. Res. **29**(1), 57–72 (1998)
18. Keith, K., Stent, A.: Modeling financial analysts' decision making via the pragmatics and semantics of earnings calls. In: Proceedings of the Fifty-Seventh Annual Meeting of the Association for Computational Linguistics, pp. 493–503 (2019)
19. Kim, S.-M., Hovy, E.: Identifying opinion holders for question answering in opinion texts. In: Proceedings of AAAI-05 Workshop on Question Answering in Restricted Domains, pp. 1367–1373 (2005)
20. Kim, S.-M., Hovy, E.: Extracting opinions, opinion holders, and topics expressed in online news media text. In: Proceedings of the Workshop on Sentiment and Subjectivity in Text, pp. 1–8 (2006)
21. Ku, L.-W., Lee, C.-Y., Chen, H.-H.: Identification of opinion holders (2009)
22. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. In: International Conference on Machine Learning, pp. 1188–1196 (2014)
23. Li, Q., Shah, S.: Learning stock market sentiment lexicon and sentiment-oriented word vector from StockTwits. In: Proceedings of the Twenty-First Conference on Computational Natural Language Learning (CoNLL 2017) (Vancouver, Canada, Aug. 2017). Association for Computational Linguistics, pp. 301–310
24. Loughran, T., McDonald, B.: When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. J. Financ. **66**(1), 35–65 (2011)
25. Lu, B.: Identifying opinion holders and targets with dependency parser in Chinese news texts. In: Proceedings of the NAACL HLT 2010 Student Research Workshop, pp. 46–51 (2010)
26. Price, S.M., Doran, J.S., Peterson, D.R., Bliss, B.A.: Earnings conference calls and stock returns: the incremental informativeness of textual tone. J. Bank. Financ. **36**(4), 992–1011 (2012)
27. Qin, Y., Yang, Y.: What you say and how you say it matters: predicting stock volatility using verbal and vocal cues. In: Proceedings of the Fifty-Seventh Annual Meeting of the Association for Computational Linguistics (Florence, Italy, July 2019). Association for Computational Linguistics, pp. 390–401

28. Rekabsaz, N., Lupu, M., Baklanov, A., Dür, A., Andersson, L., Hanbury, A.: Volatility prediction using financial disclosures sentiments with word embedding-based IR models. In: Proceedings of the Fifty-Fifth Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1712–1721 (2017)
29. Sadique, S., In, F., Veeraraghavan, M., Wachtel, P.: Soft information and economic activity: evidence from the Beige Book. J. Macroecon. **37**, 81–92 (2013)
30. Sawhney, R., Khanna, P., Aggarwal, A., Jain, T., Mathur, P., Shah, R.: VolTAGE: volatility forecasting via text-audio fusion with graph convolution networks for earnings calls. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 8001–8013 (2020)
31. Seki, Y., Evans, D.K., Ku, L.-W., Sun, L., Chen, H.-H., Kando, N., Lin, C.-Y.: Overview of multilingual opinion analysis task at NTCIR-7. In: NTCIR (2008)
32. Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., Deng, X.: Exploiting topic based Twitter sentiment for stock prediction. In: Proceedings of the Fifty-First Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 24–29 (2013)
33. Stekler, H., Symington, H.: Evaluating qualitative forecasts: the FOMC minutes, 2006–2010. Int. J. Forecast. **32**(2), 559–570 (2016)
34. Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M.: Hierarchical Dirichlet processes. J. Am. Stat. Assoc. **101**(476), 1566–1581 (2006)
35. Theil, C.K., Broscheit, S., Stuckenschmidt, H.: PRoFET: predicting the risk of firms from event transcripts. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19 (7 2019), International Joint Conferences on Artificial Intelligence Organization, pp. 5211–5217
36. Vukovic, D., Ugolnikov, V., Maiti, M.: Analyst says a lot, but should you listen? Evidence from Russia. J. Econ. Stud. (2020)
37. Wang, C.-J., Tsai, M.-F., Liu, T., Chang, C.-T.: Financial sentiment analysis for risk prediction. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing, pp. 802–808 (2013)
38. Wiegand, M., Klakow, D.: Convolution kernels for opinion holder extraction. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 795–803 (2010)
39. Wiegand, M., Schulder, M., Ruppenhofer, J.: Opinion holder and target extraction for verb-based opinion predicates—The problem is not solved. In: Proceedings of the Sixth Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 148–155 (2015)
40. Xu, Y., Cohen, S.B.: Stock movement prediction from tweets and historical prices. In: Proceedings of the Fifty-Sixth Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Melbourne, Australia, July 2018). Association for Computational Linguistics, pp. 1970–1979
41. Yang, L., Ng, T.L.J., Smyth, B., Dong, R.: HTML: hierarchical transformer-based multi-task learning for volatility prediction. In: Proceedings of the Web Conference 2020, pp. 441–451 (2020)
42. Ye, Z., Qin, Y., Xu, W.: Financial risk prediction with multi-round Q&A attention network. In: Bessiere, C. (ed.) Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20 (7 2020), International Joint Conferences on Artificial Intelligence Organization, pp. 4576–4582. Special track on AI in FinTech
43. Zhai, S.S., Zhang, Z.D.: Forecasting firm material events from 8-K reports. In: Proceedings of the Second Workshop on Economics and Natural Language Processing (Hong Kong, Nov. 2019). Association for Computational Linguistics, pp. 22–30

44. Zheng, S., Cao, W., Xu, W., Bian, J.: Doc2EDAG: an end-to-end document-level framework for Chinese financial event extraction. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the Ninth International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (Hong Kong, China, Nov. 2019). Association for Computational Linguistics, pp. 337–346
45. Zong, S., Ritter, A., Hovy, E.: Measuring forecasting skill from text. In: Proceedings of the Fifty-Eighth Annual Meeting of the Association for Computational Linguistics (Online, July 2020). Association for Computational Linguistics, pp. 5317–5331

Chapter 4
Organizing Financial Opinions

In Chap. 2, we discuss what we need to extract and understand when analyzing financial opinions. In Chap. 3, we discuss where to find financial opinions. This chapter concerns how to extract and understand the financial opinions in these sources. Although BERT-like models currently perform well on many NLP tasks, the perspectives and findings from older works are still worth considering for future work. We provide an overall picture of where we are now and also discuss research topics worth exploring.

4.1 Component Extraction

4.1.1 Target Entity and Opinion Holder

As we mentioned in Sect. 2.1.1, investors use the ticker symbol to represent the financial instrument in question. Because of this, in many documents it is not difficult to determine which financial instrument is being talked about. However, not all documents use ticker symbols, especially on social media platforms. Consider "Should I put this next to my MSFT certificate or my AAPL?" in which the writer does not use cashtags "$MSFT" and "$AAPL" to represent the stocks, instead simply using the bare ticker symbols "MSFT" and "AAPL". We address this case by adding the ticker symbols into the tokenizer. Lists of ticker symbols can be downloaded from the stock exchange. However, ambiguities can cause problems with the keyword matching approach. For example, the ticker symbol of ETFMG Travel Tech ETF is "AWAY". As NLP preprocessing usually involves converting all letters to lowercase, this can lead to ambiguities between the general word "away" and the lowercase

C.-C. Chen et al., *From Opinion Mining to Financial Argument Mining*,
SpringerBriefs in Computer Science,
https://doi.org/10.1007/978-981-16-2881-8_4

symbol ticker “away”. The following preprocessing procedure [15] is one good way to work with financial data.

1. Extract special terms such as URLs or ticker symbols.
2. Handle numeral information.
3. Convert to lowercase.

Ambiguities also exist when using the company name instead of its ticker symbol. For example, “Alphabet” is a general word as well as the name of the parent company of Google. To remove ambiguities, in formal documents such as news articles, writers sometimes include both the company name and the ticker symbol, for instance, “Alphabet (GOOGL)”.

Even the same word on different dates can denote different entities. For example, up until 2016, “AWAY” stood for HomeAway.com, but beginning in 2020 it became the ticker symbol of ETFMG Travel Tech ETF. Hence ticker symbols mentioned at different periods may have different interpretations, which means that we must periodically update the ticker symbol list and make sure we are using the right list for the right time. Otherwise, when analyzing older data, if we use the latest ticker symbol list, we could end up assigning opinions to the wrong target entity.

Since not all organization names or financial instruments mentioned in financial opinions follow the above conventions, entity identification is a fundamental problem. Organization names and financial instruments are both named entities. Studies on named entity recognition (NER) in the NLP literature provide many solutions. Below we list some of these for reference.

- Schön et al. [65] propose a guideline for annotating B2B products and suppliers in various documents, and publish the DFKI Product Corpus[1].
- Farmakiotou et al. [28] propose a rule-based method for Greek financial documents. They demonstrate higher F-scores when identifying organization names than when identifying person or location names.
- Alvarado et al. [2] publish a dataset with annotations on loan agreements[2], and show that using a small annotated in-domain dataset yields large improvements in domain-specific NER. However, their results indicate that identifying organization names is more difficult than identifying location or person entities.
- Jabbari et al. [33] publish a French corpus[3] and experiment with the spaCy toolkit.[4] They also show that identifying organization entities is more difficult than identifying person and location names.
- Mai et al. [52] focus on fine-grained NER covering 200 named entity categories.[5] They show that the best-performing model (LSTM + CNN CRF + Dictionary) on the English dataset does not perform the best on the Japanese dataset, which uses many characters in narratives.

[1] https://github.com/DFKI-NLP/product-corpus.

[2] http://people.eng.unimelb.edu.au/tbaldwin/resources/finance-sec/.

[3] http://bit.ly/CorpusFR.

[4] https://spacy.io/.

[5] Tag set: https://nlp.cs.nyu.edu/ene/version7_1_0Beng.html.

- For Chinese, Shih et al. [68] publish the CNEC corpus. Chen and Lee [18] show the difficulty of using a keyword-based strategy to identify organization names in Chinese. Chen and Chen [19] separate named entities into proper name and organization types, and use this pattern to identify organization names.

Within a narrative, the opinion holder is also a kind of named entity. Although some of the NER studies listed above show that identifying person names is easier than identifying organization names, identifying opinion holders involves more than just identifying person names. To identify opinion holders, we must not only recognize the person's name but also link the name with an expressed opinion. In formal documents and on social media platforms, we can identify the opinion holder from the metadata or directly extract it from a certain position in the document. However, as mentioned in Sect. 3.4, some opinions are part of the content in a document, and the opinion holder may or may not be the writer of the document. Below we list studies on opinion holder extraction. Although some do not evaluate their approach on financial documents, the experience they record is still useful.

- Bethard et al. [7] use classification to evaluate whether an SVM model with parse tree features classifies input sentences correctly (propositional opinion, opinion holder, and null). That is, instead of extracting the opinion holder, they seek to determine whether the opinion holder is explicitly mentioned in the input sentence. They achieve results of 56.75 and 47.54% in precision and recall.
- Kim and Hovy [35] propose a maximum entropy model with several syntactic features for opinion holder identification. Their system yields 64% accuracy in experiments conducted on the MPQA dataset,[6] which provides annotated news articles. Choi et al. [20] propose a hybrid model with AutoSlog [61] and a conditional random field (CRF). Their model yields an F1 score of 69.4% on the MPQA dataset.
- Kim and Hovy [36] select opinion-bearing frames from FrameNet[7] [4] and propose a stepwise approach to extract the opinion holder and topic of the given sentence.
- Wiegand and Klakow [77] use different kernels in an SVM model. In experiments conducted on the MPQA 2.0 dataset, their best-performing model yields an accuracy of 94.53% and an F1 score of 62.61%.
- Ku et al. [39] use CRF on a Chinese news dataset (NTCIR-7) [66] and achieve a 73.4% F1 score. They show that over 66% of the opinions in news articles are not the opinions of the author; only 19% are consistently labeled as the author's opinion.
- Lu [48] uses a dependency parser to identify opinion holders and target entities from the NTCIR-7 dataset. The proposed method yields a 75.7% accuracy and a 78.4% F1 score on an opinion holder identification task.

In summary, both target entity extraction and opinion holder extraction can be considered NER tasks. For target entity extraction, a dictionary or knowledge base

[6] https://mpqa.cs.pitt.edu/corpora/mpqa_corpus/.

[7] https://framenet.icsi.berkeley.edu/fndrupal/.

for the financial domain is sometimes necessary to extract domain-specific products or financial instruments. Opinion holder extraction, in contrast, is almost the same as in the traditional task setting. As we mentioned, holders of opinions in news articles are usually not the writer of the article; this is also true in financial opinion mining. Although cases in the news are similar to previous work, there is a paucity of work on opinion holder extraction in financial documents such as earnings conference calls or analysts' reports. An interesting task for future work would be to compare the same task across various types of documents.

4.1.2 Market Sentiment and Aspect

Many studies treat market sentiment analysis and aspect extraction as classification tasks. Liu [46] provides an overview of general sentiment analysis. In this section, we focus on studies in the financial domain.

Many works in this domain [47, 74] use text-based economic indexes with sentiment keywords. They construct indexes using keyword counts, and further analyze the predictability with respect to market data such as price movement or price volatility. Such works are not the focus of this section because we have already discussed the usefulness of these economic indexes in Chap. 3. In this section we instead focus on methods for predicting the market sentiment of a given sentence or document. Below we mention related work.

- Cortis et al. [21] annotate market sentiment scores from -1 to 1 on both social media data and news articles, and publish an annotated dataset for SemEval-2017 Task 5. Jiang et al. [34] augment word2vec embeddings [56] with n-gram, part-of-speech, word cluster, sentiment lexicon, numeral, metadata, and punctuation features. Their ensemble model performed the best in the SemEval-2017 Task 5 social media data track. Mansar et al. [54] achieved the first place in the news article track with a convolutional neural network with features extracted using VADER [32], a rule-based sentiment analysis toolkit.
- Gaillat et al. [30] concatenate (1) the output of a long short-term memory architecture (LSTM) for encoding tweets, (2) LSTM output for the word embedding with general sentiment features, (3) VADER output, and (4) the sentiment degree from the AFINN word list [57] as features. Their model outperforms that of Jiang et al. [34] on the financial social media sentiment analysis task.
- Xing et al. [79] compare the performance of dictionary-based methods and machine learning models on the Yelp dataset [83] and their StockSen dataset. They find that all models make incorrect predictions, and point out several error types, including irrealis mood, rhetoric, dependent opinion, unspecified aspects, unrecognized words, and external references.
- Yuan et al. [82] publish a Chinese news dataset for target-based sentiment analysis, and compare the performance of several baselines. On their dataset, BERT achieves an F1 score of 79.84%.

It is also important to understand why opinion holders are bullish/bearish toward the target entity. Opinion holders may analyze the target entity from different aspects, which can be separated into several categories. The most coarse-grained taxonomy is to classify aspects into fundamental analysis and technical analysis. It remains an open question as to which taxonomy is the most helpful for capturing investor opinion. Below we list related work.

- Maia et al. [53] present a taxonomy for the analysis aspect of financial opinions; this is used in FiQA-2018. Table 2.1 shows the two-level taxonomy used in this dataset. The LSTM model proposed by Shijia et al. [69] yields the best results on this dataset.
- We use a statistics-based method to analyze the words in different aspects of the FiQA-2018 dataset [10]. We find that words that are frequently used in the narrative of certain aspects are useful as keywords for aspect classification.
- In another study [11], we propose a taxonomy for aspects of financial data. We show that using aspect information as an auxiliary task improves performance on numeral attachment, that is, linking the given numeral with the related target entity. Chapter 5 includes a detailed discussion on numeral-related tasks.

In sum, market sentiment analysis can be approached either as classification or regression. As long as we have an annotated dataset for supervised learning, any current state-of-the-art model can be used. However, as shown in Xing et al. [79], domain-specific methods are still necessary, because performance of a given end-to-end model can drop considerably after changing to a domain-specific dataset. Aspect extraction is highly related to nouns in the narrative. For example, a tweet that mentions the word *dividend* is likely to be an opinion that is based on the analysis of the dividend policy aspect. Since financial opinion mining is still at an early stage, few studies discuss aspect-based sentiment analysis. However, the common practice of investors is to analyze financial instruments from different aspects to produce their main claim. Also, even two sets of analysis results produced for a given financial instrument at the same time can be different. Thus one direction for future work is aspect-based financial opinion mining. Although both sentiment and aspect labels are provided in the FiQA-2018 dataset [53] for financial social media data, in the Fin-SoMe dataset [12], we find that over 90% of social media users do not provide the reason, i.e., the aspect, for their claims. In-depth analysis of longer documents or formal reports may yield different findings from those of social media data.

4.1.3 Temporal Information

One common NLP task is extracting temporal information; this can be considered an NER task. In most cases, we achieve very good performance on this task, because people generally express temporal information using patterns. After extracting temporal expressions, researchers attempt to organize the events into a timeline; this is called *temporal relation analysis*. This task is more challenging than just extracting temporal expressions. As we mention in Sect. 2.1.4, the publishing time and validity period are important temporal information in financial opinion mining. Obtaining the publishing time is not difficult, since regardless of source, almost all documents include metadata that reveals the publishing time. In contrast, the validity period of a financial opinion is an unexplored issue. We can borrow techniques developed for temporal relations to find the validity period. Below we list some work on temporal information tasks.

- Pustejovsky et al. [60] propose a guideline for annotating temporal information and relations between time and events. They also published the TIMEBANK corpus, the annotation scheme which later became an ISO standard.
- Verhagen et al. [76] propose the TempEval shared task for understanding temporal information in English documents. In TempEval-3 [75], a rule-based method [73] for extracting temporal expressions in English and Spanish yielded F1 scores of 81.34% and 85.3%, respectively.
- Bethard et al. [6] propose a domain-specific temporal information task with clinical documents in SemEval-2017. MacAvaney et al. [51] achieve an F1 score of 59% for time span extraction in SemEval-2017 with a CRF model.
- We proposed a numerical taxonomy for financial social media data [15] and held a FinNum shared task in NTCIR-14 [17]. Temporal information is one of the categories in this taxonomy. Azzi and Bouamor [3] and Wu et al. [78] enrich the word vector with several tailor-made features for numeral information, and achieve an accuracy of over 98% in the terminal category.

These studies show that extracting temporal information from financial documents is not difficult. However, it is indeed challenging to detect the validity period or maturity date of a financial opinion. Once we have extracted a temporal span from a document, understanding the meaning of the span is a complex task which involves first understanding its context. In the FinNum dataset, from the temporal category we separate out the maturity date of options, which are a kind of financial instrument. Participants' models demonstrated accuracies of 96–98% for fine-grained temporal data, but achieved only 62–75% accuracy when classifying maturity dates [16]. This performance drop shows the difficulty of understanding temporal information.

Finally, we compare the temporal information in financial opinion mining with that in traditional opinion mining. In financial narratives, most investors' opinions are predictions of the future based on the past and present. However, in traditional opinion mining such as product reviews, writers' opinions are related to past experiences only. In clinical documents, most information also relates to the present and the past. Hence, temporal information in financial opinion mining may be more complicated than that in other domains.

Table 4.1 Performances of claim detection and premise detection in analyst reports

Model	Claim detection	Premise detection
CNN	76.15	55.25
BiGRU	77.97	48.62
CapsNet	77.93	52.47
BERT	**79.86**	**57.69**

4.1.4 Elementary Argumentative Units

As mentioned in Sect. 2.1.7, we explain fine-grained financial opinion mining using argument mining. Although segmentation of paragraphs into their elementary argumentative units has been widely discussed in the NLP literature [40], there is little discussion about this for documents in the financial domain. In this section we list work in the argument mining track and list some of our experimental results on financial documents.

- Aharoni et al. [1] publish a dataset for claim and evidence detection. Levy et al. [41] use this dataset to explore context-dependent claim detection, that is, selecting the claim that is related to the given topic. Their CDCD approach selects the most relevant sentences and further locates boundaries using two filters. Their results demonstrate the difficulty of the proposed task. Many extensions of this work come from IBM Project Debater.[8]
- Rinott et al. [62] propose a pipeline approach to detect the evidence—or premise—of a given claim. They classify evidence into three types: study, expert, and anecdotal. Their results show that detecting expert testimony is easier than discerning anecdotal or empirical evidence.
- Daxenberger et al. [23] compare claims from web discourses, persuasive essays, and online comments. They present results for different datasets with several features, and find that keywords such as "should" are crucial cues for neural network models to identify cross-domain claims.
- Chakrabarty et al. [9] use IMO/IMHO (in my (humble) opinion) acronyms as a self-label for Reddit posts, and publish a corpus with 5.5 million claims.[9] They show that using this corpus to fine-tune the language model significantly improves claim detection performance in other datasets.
- Schaefer and Stede [64] publish a corpus[10] with claim and evidence labels on German tweets that contain the keyword "climate". Based on our observations [12], it is not easy to label evidence for claims on financial social media because few social media users provide premises for their claims.

[8] https://www.research.ibm.com/haifa/dept/vst/debating_data.shtml.

[9] https://bitbucket.org/tuhinch/imho-naacl2019/src/master/.

[10] https://github.com/RobinSchaefer/climate-tweet-corpus.

- In previous work [13], we annotate claims in professional stock analysis reports written in Chinese, and publish the NumClaim dataset.[11] We use pointwise mutual information to identify keywords near the investor's claims, and find that words like "estimate", "price target", and "downgrade/upgrade" are frequently used in claim sentences. We extend previous work and annotate the premise(s) for the given claim. Table 4.1 shows the results for different models. We find that detecting claims is easier than detecting premises. This may be because analysts use certain words to express their claims; this echoes the findings of Daxenberger et al. [23].

In sum, the argumentative narrative of an investor may be different from claims or premises in other domains. This is primarily because investors follow convention when writing analysis reports. For example, they use "estimate" or "price target" instead of "should," which is used in other domains. We look at financial opinion mining as a form of argument mining. More fine-grained analysis is needed to better understand domain-specific cases, which leads to the second reason: we find that investors always make claims using estimations, which are represented using numerals. Thus numerals play a crucial role in investor claim detection. In Chap. 5, we discuss this topic in depth.

4.2 Relation Linking and Quality Evaluation

Extracting the components of a financial opinion yields a basic understanding of the opinion. Once extracted, the components—especially the argumentative units—must be linked. In this section, we discuss how to construct an argumentation structure like Fig. 2.6, and further estimate the rationality of using the extracted premises to support claims. The quality of a financial opinion may also influence the accuracy of downstream tasks. However, evaluation of this quality is rarely discussed in the literature. We discuss studies using documents in other domains as an example and suggest directions for evaluating the quality of a financial opinion.

- Stab and Gurevych [70] annotate given argumentative unit pairs with *support* or *non-support* in persuasive essays. Using an SVM model, they achieve an F1 score of 72.2% for relation identification.
- Sakai et al. [63] label given statement pairs with *support* or *non-support* in a dialogue. They experiment on English and Japanese data, and explore several models. An extremely randomized tree with unigram, bi-gram, and tri-gram features performs best on both datasets.
- Stab and Gurevych [71] publish a dataset[12] for parsing argumentation structures in persuasive essays. Their experimental results show that simultaneously learning all subtasks—component classification, relation identification, and argumentation structure—improves the performance of each. Their results also show that relation

[11] http://numclaim.nlpfin.com/.

[12] https://www.informatik.tu-darmstadt.de/ukp/research_6/data/index.en.jsp.

linking is more difficult than component classification. Eger et al. [26] propose the LSTM-ER model, which outperforms the ILP model [71].

- Kirschner et al. [37] propose an annotation guideline for argumentation structures in scientific publications in which sentences are the basic unit. They label relationships between two sentences as *support*, *attack*, *detail*, or *undirected sequence*. In this work, they focus on analyzing the statistics of annotation results.
- Klebanov et al. [38] discuss the relationship between argument structure and essay quality. They conduct experiments using argumentative essays written for the TOEFL test [8], and show that adding argumentation structure features to the model improves the performance of essay quality evaluation.
- Li et al. [42] enhance BERT by encoding argument structure features with the Bi-LSTM model for online debate persuasion prediction. In this case, persuasion can be viewed as a proxy for the quality of the debate text. They use both textual information and argumentation structure to evaluate the quality of online debates.

These studies not only concern methods for argumentative unit relation linking, but also show the usefulness of adding argumentation structure into models for quality evaluation. However, there is little discussion on the quality of more informal data such as those from social media platforms. The most relevant task is online review helpfulness evaluation. Below we list some related work and review experimental results with financial data.

- Ghose and Ipeirotis [31] use ratings left by product review readers who press the "Helpful" button depicted in Fig. 1.2 as the helpfulness label of a given review. They represent a product review using the characteristics of the review writer as well as the readability and subjectivity features of the review. They perform an ablation study which shows that readability better predicts the helpfulness of reviews of products in audio, video, and digital camera categories. For DVD reviews, reviewers' characteristics and subjective features lead to higher AUCs than readability features. This echoes the findings of Danescu-Niculescu-Mizil et al. [22]: the content of a book review is not the only feature that influences votes of review readers.
- Yang et al. [81] approach helpfulness prediction as a regression task. They extract emotion [59] and reasoning [72] features from reviews in book, home, outdoor, and electronic categories, and show that these features improve the performance of review helpfulness evaluation.
- Diaz and Ng [24] survey studies on product review helpfulness modeling and prediction, and provide suggestions for future work.
- Fan et al. [27] use product metadata to enhance neural network models for helpfulness prediction. They select key phrases from the review with product metadata, and further pass the results to the helpfulness predictor. Experiments on Amazon and Yelp datasets support the proposed process.
- Xiong and Litman [80] show that adding helpfulness features to the sentence scoring function improves the performance of extractive summarization of online reviews.

Table 4.2 Results of discriminating premises of analysts from those of amateur investors. (* denotes results that are significantly different from the Sem. model under McNemar's test with $p < 0.05$.)

Feature	Micro-F1	Macro-F1
Dep	62.39	61.54
POS	73.43	73.34
Sem.	88.59	88.59
POS + Dep + Sem.	**90.81***	**90.81***

- Shaar et al. [67] use 2016 US Presidential debate and Twitter corpora to construct a dataset[13] for detecting whether a given claim has already been fact-checked on trustworthy platforms. In this task, given an unverified claim, models rank a set of verified claims from PolitiFact[14] or Snopes[15] to evaluate whether the verified claim supports the unverified input claim. The learning-to-rank model achieves MRRs of 60.8 and 78.8% on the debate and Twitter datasets, respectively.

Although these works do not use financial documents, we believe that these methods could be adapted to the financial domain with minor modifications. For example, online product categories correspond to different financial instruments in the financial market such as stocks and foreign exchanges. Note that a company's stock can be considered a product in the financial market. Additionally, product metadata in financial opinion mining may consist of contracts, market data, or company introductions.

Drawing from previous work, we propose a simple approach for evaluating the opinion quality of financial social media users [14]. We use part-of-speech, dependency, and semantic features to encode the analysis of social media users and professional analysts, and further employ the BiGRU model to determine whether the input sentence was written by a professional analyst. With this experiment we attempt to identify professional-level social media posts. Our rationale is that the more professional-level sentences there are in a social media post, the higher its quality. Table 4.2 shows the results of discriminating analyst and amateur investors' premises. To evaluate the effectiveness of our rationale, we use the following metrics as proxies for financial opinion quality.

For bullish and bearish opinions posted on day t, we calculate the maximum possible profit (MPP) and the maximum loss (ML) as

$$MPP_{bullish} = \frac{\max_{i=t+1}^{T} H_i - O_{t+1}}{O_{t+1}} \tag{4.1}$$

[13] https://github.com/sshaar/That-is-a-Known-Lie.

[14] https://www.politifact.com/.

[15] https://www.snopes.com/.

Table 4.3 Performances of the methods for opinion ranking

Method	Avg. MPP (%)	Avg. ML (%)	RPR
Random	11.94	−17.28	0.69
Popularity	8.88	−8.69	1.02
Proposed approach	17.61	**−3.72**	**4.73**
Analyst	**22.30**	−6.52	3.42

$$ML_{bullish} = \frac{\min_{i=t+1}^{T} L_i - O_{t+1}}{O_{t+1}} \tag{4.2}$$

$$MPP_{bearish} = \frac{O_{t+1} - \min_{i=t+1}^{T} L_i}{O_{t+1}} \tag{4.3}$$

$$ML_{bearish} = \frac{O_{t+1} - \max_{i=t+1}^{T} H_i}{O_{t+1}}, \tag{4.4}$$

where O_t denotes the opening price of day t, H_t denotes a list of the highest prices on day t, L_t denotes a list of the lowest prices on day t, and T is the last day of the backtesting period.

MPP sheds light on the potential profit, and also indicates the potential of the selected opinions. ML, on the other hand, provides information about the downside risk. We use ML to determine whether the opinion was posted at the right time, i.e., whether bullish (bearish) opinions were posted at relatively lower (higher) price levels of the target financial instrument. Finally, the average $MPP/|ML|$, termed RPR, evaluates the expected **R**eturn when investors take an additional one **P**ercent of **R**isk.

Table 4.3 shows the performance of the top 10% of opinions sorted using different methods. Compared with randomly-selected amateur opinions, the top-ranked opinions mined by our approaches outperform for all metrics, in particular the averaged ML. The outcomes of our approaches are also superior to the results of opinions ranked by the number of likes given by social media users (Popularity).

We further compare our results with the statistics of randomly-selected professional analysts. Although analysts identify targets with higher potential profit, the downside risk of trading based on analyst opinions is 1.75 times that of the downside risk of following top-ranked opinions of amateur investors. The RPR of top-ranked opinions using the proposed approach is also better than that of professional analysts. That shows that top-ranked opinions are comparable to the opinions of professional analysts.

Thus our experimental results show that writing style is also a useful feature for opinion quality evaluation. Future work on financial opinion mining can explore the use of features such as opinion readability and subjectiveness as well as the opinion holder's background to evaluate review helpfulness. Our experiments not only provide directions for financial opinion quality evaluation, but also show that evaluating opinion quality is useful for downstream tasks in the financial domain.

In this section, we explore both argumentation structure and opinion quality in other domains, and present evidence for the usefulness of fine-grained argumentative information in downstream tasks; this remains an underdeveloped topic in financial opinion mining. As we show in this section, narratives in the financial domain often differ from those in other fields. Future work can annotate datasets by slightly modifying the guidelines in previous works to fit financial domain narratives. Despite the importance of quality evaluation, most studies on financial opinion mining continue to use the law of large numbers to average sentiment collected from different sources, and do not account for document quality. We can draw from studies on helpfulness evaluation to develop baselines for financial opinion quality evaluation. One step in this research direction is to use tailor-made methods and features for financial documents. Although many studies use prediction accuracy as a proxy for the quality of a financial opinion, annotated benchmark datasets are still necessary because even high-quality reports are not always accurate [84]. In Chap. 5 we discuss characteristics of financial narratives that facilitate future work on domain-specific methods for financial opinion mining tasks.

4.3 Influence Power Estimation and Implicit Information Inference

In this section, we discuss issues from Fig. 2.7, including influence power estimation and implicit information inference. Chapter 3 lists studies that indicate that opinions from different sources predict the future price movement of financial instruments. However, estimating the influence of an opinion on future financial outcomes is still an open issue. Note that just because an opinion is accurate does not mean it possesses great influence; likewise, just because an opinion is highly influential does not mean it is accurate. Most studies take the average of opinions from the same source as the overall opinion for that source. Many studies relate to electronic word-of-mouth (eWOM). Below we list some of such studies, after which we list some studies that estimate the influence of opinions one by one.

- Anindya et al. [31] use ordinary least squares (OLS) regression to estimate the effect of product reviews on future product sales. They show that retail price bears the most significant influence on the sales of the next time step. The standard deviation of the reviews' subjective scores in audio, video, and DVD categories also reveals a significant influence. The number of reviews is also an important fact in digital camera and DVD categories.
- Lin et al. [43] use sentiment on social media platforms to predict the sales of different brands' smartphones. They demonstrate that adding sentiment features improves the performance of downstream tasks. Additionally, they apply a meta-learning framework [29] to further improve prediction accuracy.

- Mariani and Borghi [55] analyze how a hotel's online review features influence its future financial performance. They find that the valence and volume of online reviews positively influence future performance, and that the degree of helpfulness is also an important factor.
- Luca [49] conducts a case study on Yelp.com reviews, and finds that each additional star earned by the restaurant on Yelp yields a 5–9% increase in revenue. However, this applies to individual companies and not restaurant chains. This work also shows that certified reviewers have twice the impact of common reviewers.
- Banerjee et al. [5] use reviewer features as proxies of reviewer trustworthiness, and find that the trustworthiness of the reviewer positively influences his/her online reputation. They thus suggest that companies encourage the most trustworthy reviewers to write reviews of the company's products.

As discussed in Sect. 4.2, many studies have been conducted on e-commerce platforms, but few use financial data to evaluate the quality of financial opinions. This is similar to the case of influence power estimation. The above studies demonstrate the potential of analyzing the influence power of opinions for product sales as well as hotel and restaurant operations. Intuitively, insider opinions outweigh those from social media users. One issue that remains unexplored in financial opinion analysis is evaluating which analyst's opinion has a greater impact on the market, or which social media user's opinion a company should be more concerned about.

Features of opinion holders can proxy the holder's influence power. For example, Warren Buffett's opinion on specific financial instruments is likely to influence more investors and have a greater impact on the market than this author's opinion. Future work can draw from the findings of the studies listed here in the financial domain to sort out the most important opinions from the hundreds and thousands that are posted every day. In Sect. 6.1, we list application scenarios related to information provisioning.

Another topic in Fig. 2.7 is implicit information influence, where, for instance, facts about one company impact the stock price of another company. For example, bad news about Taiwan Semiconductor Manufacturing Co., Ltd. may reflect poor prospects for the semiconductor industry as a whole. Thus, such news could also influence the stock prices of Intel Corporation and Samsung Electronics. An important problem for investors is making this kind of inference to gain a fuller picture of the financial market. Many studies on this problem focus on extracting the relationship between companies from textual data. Below we list some work on this topic.

- Oral et al. [58] extract relations between companies from banking orders. Sender, receiver, and process details in the transactions are extracted to construct a relational graph. They use a BiLSTM model to predict the relation type of the given entity pair.
- Ma et al. [50] link news articles by a bag of proposed features, and encode each news article into a vector. They show that this representation successfully groups related news articles, and they conduct further experiments on the downstream

tasks of stock movement prediction and news recommendation. Their results attest the usefulness of the proposed embedding.

- In previous work [44], we experiment with annotations from professional journalists,[16] in which labels are provided for stocks that are related to the given news article but not mentioned explicitly in the article. We propose a dynamic graph Transformer model to recommend possible stocks given the article. Experimental results show the usefulness of the proposed method. We also conduct experiments on stock movement prediction [45], and produce results that show that additionally taking into account implicitly-related news improves the accuracy of the attention-based model.

These studies show the importance of information inference in financial textual data. That is, even financial instruments that are not mentioned explicitly in an article can be influenced by facts reported in the article. How best to capture this in a neural network is still an open issue. This helps to bring model decisions more in line with those of professional investors, and also yields more accurate predictions, as shown in previous work [45].

In this section, we discuss estimating the influence of an opinion on the target entity and show the importance of inferring implicit information based on the given facts. Another type of information inference is logically infering the next possible event. Ding et al. [25] present a financial event logic graph, a knowledge graph used to infer relations between events. This direction is also important in financial opinion mining. Compared to previous sections, effectively addressing the issues raised in this section—especially information inference—requires more domain knowledge.

4.4 Summary

This chapter proposes directions for organizing financial opinions. We follow the notions proposed in Chap. 2 when discussing related methods. Although many of the studies listed here do not use financial documents as sources, we believe that their models and findings can be adopted in future work on similar tasks with financial documents. The most fundamental step of the proposed framework is the extraction of elementary argumentative units. We suggest that future work extend sentiment analysis to fine-grained opinion mining based on the proposed research directions. We also seek to highlight three crucial tasks that could help models to better approximate human performance: quality evaluation, influence power estimation, and information inference. The argumentation structure in Fig. 2.7 can bring models closer to human-level understanding. Once we are able to build this structure automatically, we will

[16] https://www.moneydj.com/.

be much closer to being able to explain the reasons for market movement. Report generation would then be the next step. In Chap. 6, we discuss possible application scenarios.

References

1. Aharoni, E., Polnarov, A., Lavee, T., Hershcovich, D., Levy, R., Rinott, R., Gutfreund, D., Slonim, N.: A benchmark dataset for automatic detection of claims and evidence in the context of controversial topics. In: Proceedings of the First Workshop on Argumentation Mining (Baltimore, Maryland, June 2014), Association for Computational Linguistics, pp. 64–68
2. Alvarado, J.C.S., Verspoor, K., Baldwin, T.: Domain adaption of named entity recognition to support credit risk assessment. In: Proceedings of the Australasian Language Technology Association Workshop 2015, pp. 84–90 (2015)
3. Azzi, A.A., Bouamor, H.: Fortia1@the NTCIR-14 FinNum Task: Enriched sequence labeling for numeral classification
4. Baker, C.F., Sato, H.: The FrameNet data and software. In: The Companion Volume to the Proceedings of Forty-First Annual Meeting of the Association for Computational Linguistics, pp. 161–164 (2003)
5. Banerjee, S., Bhattacharyya, S., Bose, I.: Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business. Decis. Support Syst. **96**, 17–26 (2017)
6. Bethard, S., Savova, G., Palmer, M., Pustejovsky, J.: SemEval-2017 task 12: Clinical TempEval. In: Proceedings of the Eleventh International Workshop on Semantic Evaluation (SemEval-2017) (Vancouver, Canada, Aug. 2017), Association for Computational Linguistics, pp. 565–572
7. Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V., Jurafsky, D.: Automatic extraction of opinion propositions and their holders. In: 2004 AAAI Spring Symposium on Exploring Attitude and Affect in Text, vol. 2224 (2004)
8. Blanchard, D., Tetreault, J., Higgins, D., Cahill, A., Chodorow, M.: TOEFL11: a corpus of non-native English. ETS Res. Rep. Ser. **2013**(2), i–15 (2013)
9. Chakrabarty, T., Hidey, C., McKeown, K.: IMHO fine-tuning improves claim detection. 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. 558–563 (2019)
10. Chen, C.-C., Huang, H.-H., Chen, H.-H.: Fine-grained analysis of financial tweets. In: Companion Proceedings of the Web Conference 2018, pp. 1943–1949 (2018)
11. Chen, C.-C., Huang, H.-H., Chen, H.-H.: Numeral attachment with auxiliary tasks. In: Proceedings of the Forty-Second International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1161–1164 (2019)
12. Chen, C.-C., Huang, H.-H., Chen, H.-H.: Issues and perspectives from 10,000 annotated financial social media data. In: Proceedings of the Twelfth Language Resources and Evaluation Conference, pp. 6106–6110 (2020)
13. Chen, C.-C., Huang, H.-H., Chen, H.-H.: NumClaim: investor's fine-grained claim detection. In: Proceedings of the Twenty-Ninth ACM International Conference on Information & Knowledge Management, pp. 1973–1976 (2020)
14. Chen, C.-C., Huang, H.-H., Chen, H.-H.: Evaluating the rationales of amateur investors. In: The World Wide Web Conference (2021)
15. Chen, C.-C., Huang, H.-H., Shiue, Y.-T., Chen, H.-H.: Numeral understanding in financial tweets for fine-grained crowd-based forecasting. In: 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 136–143. IEEE (2018)
16. Chen, C.-C., Huang, H.-H., Takamura, H., Chen, H.-H.: Final Report of the NTCIR-14 FinNum Task: Challenges and current status of fine-grained numeral understanding in financial social

media data. In: NII Conference on Testbeds and Community for Information Access Research, pp. 183–192. Springer (2019)

17. Chen, C.-C., Huang, H.-H., Takamura, H., Chen, H.-H.: Overview of the NTCIR-15 FinNum-2 Task: Numeral attachment in financial tweets. Development **850**(194), 1–044 (2020)
18. Chen, H.-H., Lee, J.-C.: Identification and classification of proper nouns in Chinese texts. In: COLING 1996 Volume 1: The Sixteenth International Conference on Computational Linguistics (1996)
19. Chen, K.-J.: Knowledge extraction for identification of Chinese organization names. In: Second Chinese Language Processing Workshop, pp. 15–21 (2000)
20. Choi, Y., Cardie, C., Riloff, E., Patwardhan, S.: Identifying sources of opinions with conditional random fields and extraction patterns. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pp. 355–362 (2005)
21. Cortis, K., Freitas, A., Daudert, T., Huerlimann, M., Zarrouk, M., Handschuh, S., Davis, B.: SemEval-2017 Task 5: Fine-grained sentiment analysis on financial microblogs and news. Association for Computational Linguistics (ACL)
22. Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., Lee, L.: How opinions are received by online communities: a case study on Amazon.com helpfulness votes. In: Proceedings of the Eighteenth International Conference on World Wide Web, pp. 141–150 (2009)
23. Daxenberger, J., Eger, S., Habernal, I., Stab, C., Gurevych, I.: What is the essence of a claim? Cross-domain claim identification. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2055–2066 (2017)
24. Diaz, G.O., Ng, V.: Modeling and prediction of online product review helpfulness: a survey. In: Proceedings of the Fifty-Sixth Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 698–708 (2018)
25. Ding, X., Li, Z., Liu, T., Liao, K.: ELG: An event logic graph. arXiv preprint arXiv:1907.08015 (2019)
26. Eger, S., Daxenberger, J., Gurevych, I.: Neural end-to-end learning for computational argumentation mining. In: Proceedings of the Fifty-Fifth Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 11–22 (2017)
27. Fan, M., Feng, C., Sun, M., Li, P.: Reinforced product metadata selection for helpfulness assessment of customer reviews. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the Ninth International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 1675–1683 (2019)
28. Farmakiotou, D., Karkaletsis, V., Koutsias, J., Sigletos, G., Spyropoulos, C.D., Stamatopoulos, P.: Rule-based named entity recognition for Greek financial texts. In: Proceedings of the Workshop on Computational Lexicography and Multimedia Dictionaries (COMLEX 2000), pp. 75–78 (2000)
29. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. arXiv preprint arXiv:1703.03400 (2017)
30. Gaillat, T., Sousa, A., Zarrouk, M., Davis, B.: FinSentiA: sentiment analysis in English financial microblogs. In: Actes de la Conférence Traitement Automatique de la Langue Naturelle, TALN, p. 271 (2018)
31. Ghose, A., Ipeirotis, P.G.: Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics. IEEE Trans. Knowl. Data Eng. **23**(10), 1498–1512 (2010)
32. Gilbert, C., Hutto, E.: VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International Conference on Weblogs and Social Media (ICWSM-14), vol. 81, p. 82 (2014)
33. Jabbari, A., Sauvage, O., Zeine, H., Chergui, H.: 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)

34. Jiang, M., Lan, M., Wu, Y.: ECNU at SemEval-2017 Task 5: an ensemble of regression algorithms with effective features for fine-grained sentiment analysis in financial domain. In: Proceedings of the Eleventh International Workshop on Semantic Evaluation (SemEval-2017), pp. 888–893 (2017)
35. Kim, S.-M., Hovy, E.: Identifying opinion holders for question answering in opinion texts. In: Proceedings of AAAI-05 Workshop on Question Answering in Restricted Domains, pp. 1367–1373 (2005)
36. Kim, S.-M., Hovy, E.: Extracting opinions, opinion holders, and topics expressed in online news media text. In: Proceedings of the Workshop on Sentiment and Subjectivity in Text, pp. 1–8 (2006)
37. Kirschner, C., Eckle-Kohler, J., Gurevych, I.: Linking the thoughts: analysis of argumentation structures in scientific publications. In: Proceedings of the Second Workshop on Argumentation Mining, pp. 1–11 (2015)
38. Klebanov, B. B., Stab, C., Burstein, J., Song, Y., Gyawali, B., Gurevych, I.: Argumentation: content, structure, and relationship with essay quality. In: Proceedings of the Third Workshop on Argument Mining (ArgMining2016), pp. 70–75 (2016)
39. Ku, L.-W., Lee, C.-Y., Chen, H.-H.: Identification of opinion holders (2009)
40. Lawrence, J., Reed, C.: Argument mining: a survey. Comput. Linguist. **45**(4), 765–818 (2019). Dec
41. Levy, R., Bilu, Y., Hershcovich, D., Aharoni, E., Slonim, N.: Context dependent claim detection. In: Proceedings of COLING 2014, the Twenty-Fifth International Conference on Computational Linguistics: Technical Papers, pp. 1489–1500 (2014)
42. Li, J., Durmus, E., Cardie, C.: Exploring the role of argument structure in online debate persuasion. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 8905–8912 (2020)
43. Lin, Z., Madotto, A., Winata, G.I., Liu, Z., Xu, Y., Gao, C., Fung, P.: Learning to learn sales prediction with social media sentiment. In: Proceedings of the First Workshop on Financial Technology and Natural Language Processing, pp. 47–53 (2019)
44. Liou, Y.-T., Chen, C.-C., Huang, H.-H., Chen, H.-H.: Dynamic Graph Transformer for implicit tag recognition. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (Apr. 2021), Association for Computational Linguistics
45. Liou, Y.-T., Chen, C.-C., Tang, T.-H., Huang, H.-H., Chen, H.-H.: FinSense: an assistant system for financial journalists and investors. In: Proceedings of the 14th International Conference on Web Search and Data Mining (2021)
46. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. **5**(1), 1–167 (2012)
47. Loughran, T., McDonald, B.: When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. J. Finance **66**(1), 35–65 (2011)
48. Lu, B.: Identifying opinion holders and targets with dependency parser in Chinese news texts. In: Proceedings of the NAACL HLT 2010 Student Research Workshop, pp. 46–51 (2010)
49. Luca, M.: Reviews, reputation, and revenue: the case of Yelp.com. Harvard Business School NOM Unit Working Paper, 12-016 (2016)
50. Ma, Y., Zong, L., Yang, Y., Su, J.: News2vec: news network embedding with subnode information. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the Ninth International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (Hong Kong, China, Nov. 2019), Association for Computational Linguistics, pp. 4843–4852
51. MacAvaney, S., Cohan, A., Goharian, N.: GUIR at SemEval-2017 Task 12: a framework for cross-domain clinical temporal information extraction. In: Proceedings of the Eleventh International Workshop on Semantic Evaluation (SemEval-2017), pp. 1024–1029 (2017)
52. Mai, K., Pham, T.-H., Nguyen, M.T., Nguyen, T.D., Bollegala, D., Sasano, R., Sekine, S.: An empirical study on fine-grained named entity recognition. In: Proceedings of the Twenty-Seventh International Conference on Computational Linguistics, pp. 711–722 (2018)

53. Maia, M., Handschuh, S., Freitas, A., Davis, B., McDermott, R., Zarrouk, M., Balahur, A.: WWW'18 Open Challenge: Financial opinion mining and question answering. In: Companion Proceedings of the The Web Conference 2018, pp. 1941–1942 (2018)
54. Mansar, Y., Gatti, L., Ferradans, S., Guerini, M., Staiano, J.: Fortia-FBK at SemEval-2017 Task 5: Bullish or Bearish? Inferring sentiment towards brands from financial news headlines. In: Proceedings of the Eleventh International Workshop on Semantic Evaluation (SemEval-2017), pp. 817–822 (2017)
55. Mariani, M.M., Borghi, M.: Online review helpfulness and firms' financial performance: an empirical study in a service industry. Int. J. Electron. Commer. **24**(4), 421–449 (2020)
56. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. Adv. Neural Inf. Process. Syst. **26**, 3111–3119 (2013)
57. Nielsen, F.Å.: A new ANEW: evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903 (2011)
58. Oral, B., Emekligil, E., Arslan, S., Eryiğit, G.: Extracting complex relations from banking documents. In: Proceedings of the Second Workshop on Economics and Natural Language Processing (Hong Kong, Nov. 2019), Association for Computational Linguistics, pp. 1–9
59. Pennebaker, J.W., Francis, M.E., Booth, R.J.: Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates **71**(2001), 2001 (2001)
60. Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., Ferro, L., et al.: The TIMEBANK Corpus. In: Corpus Linguistics, vol. 2003, p. 40. Lancaster, UK (2003)
61. Riloff, E.: An empirical study of automated dictionary construction for information extraction in three domains. Artif. Intell. **85**(1–2), 101–134 (1996)
62. Rinott, R., Dankin, L., Alzate, C., Khapra, M.M., Aharoni, E., Slonim, N.: Show me your evidence—An automatic method for context dependent evidence detection. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 440–450 (2015)
63. Sakai, K., Inago, A., Higashinaka, R., Yoshikawa, Y., Ishiguro, H., Tomita, J.: Creating large-scale argumentation structures for dialogue systems. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (2018)
64. Schaefer, R., Stede, M.: Annotation and detection of arguments in tweets. In: Proceedings of the Seventh Workshop on Argument Mining, pp. 53–58 (2020)
65. Schön, S., Mironova, V., Gabryszak, A., Hennig, L.: A corpus study and annotation schema for named entity recognition and relation extraction of business products. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (2018)
66. Seki, Y., Evans, D. K., Ku, L.-W., Sun, L., Chen, H.-H., Kando, N., Lin, C.-Y.: Overview of multilingual opinion analysis task at NTCIR-7. In: NTCIR (2008)
67. Shaar, S., Babulkov, N., Da San Martino, G., Nakov, P.: That is a known lie: detecting previously fact-checked claims. In: Proceedings of the Fifty-Eighth Annual Meeting of the Association for Computational Linguistics (Online, July 2020), Association for Computational Linguistics, pp. 3607–3618
68. Shih, C.-W., Tsai, R. T.-H., Wu, S.-H., Hsieh, C.-C., Hsu, W.-L.: The construction of a Chinese named entity tagged corpus: CNEC1.0. In: Proceedings of the Sixteenth Conference on Computational Linguistics and Speech Processing, pp. 305–313 (2004)
69. Shijia, E., Yang, L., Zhang, M., Xiang, Y.: Aspect-based financial sentiment analysis with deep neural networks. In: WWW (Companion Volume) (2018)
70. Stab, C., Gurevych, I.: Identifying argumentative discourse structures in persuasive essays. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 46–56 (2014)
71. Stab, C., Gurevych, I.: Parsing argumentation structures in persuasive essays. Comput. Linguist. **43**(3), 619–659 (2017)

72. Stone, P.J., Bales, R.F., Namenwirth, J.Z., Ogilvie, D.M.: The general inquirer: a computer system for content analysis and retrieval based on the sentence as a unit of information. Behav. Sci. **7**(4), 484 (1962)
73. Strötgen, J., Zell, J., Gertz, M.: HeidelTime: Tuning English and developing Spanish resources for TempEval-3. In: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pp. 15–19 (2013)
74. Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S.: More than words: quantifying language to measure firms' fundamentals. J. Finance **63**(3), 1437–1467 (2008)
75. UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M., Pustejovsky, J.: SemEval-2013 Task 1: TempEval-3: Evaluating time expressions, events, and temporal relations. In: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pp. 1–9 (2013)
76. Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., Pustejovsky, J.: SemEval-2007 Task 15: TempEval temporal relation identification. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), pp. 75–80 (2007)
77. Wiegand, M., Klakow, D.: Convolution kernels for opinion holder extraction. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 795–803 (2010)
78. Wu, Q., Wang, G., Zhu, Y., Liu, H., Karlsson, B.: DeepMRT at the NTCIR-14 FinNum Task: A hybrid neural model for numeral type classification in financial tweets. In: Proceedings of the Fourteenth NTCIR Conference on Evaluation of Information Access Technologies (2019)
79. Xing, F., Malandri, L., Zhang, Y., Cambria, E.: Financial sentiment analysis: an investigation into common mistakes and silver bullets. In: Proceedings of the Twenty-Eighth International Conference on Computational Linguistics, pp. 978–987 (2020)
80. Xiong, W., Litman, D.: Empirical analysis of exploiting review helpfulness for extractive summarization of online reviews. In: Proceedings of COLING 2014, the Twenty-Fifth International Conference on Computational Linguistics: Technical Papers, pp. 1985–1995 (2014)
81. Yang, Y., Yan, Y., Qiu, M., Bao, F.: Semantic analysis and helpfulness prediction of text for online product reviews. In: Proceedings of the Fifty-Third Annual Meeting of the Association for Computational Linguistics and the Seventh International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pp. 38–44 (2015)
82. Yuan, C., Liu, Y., Yin, R., Zhang, J., Zhu, Q., Mao, R., Xu, R.: Target-based sentiment annotation in Chinese financial news. In: Proceedings of the Twelfth Language Resources and Evaluation Conference, pp. 5040–5045 (2020)
83. Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. Adv. Neural Inf. Process. Syst. **28**, 649–657 (2015)
84. Zong, S., Ritter, A., Hovy, E.: Measuring forecasting skill from text. In: Proceedings of the Fifty-Eighth Annual Meeting of the Association for Computational Linguistics (Online, July 2020), Association for Computational Linguistics, pp. 5317–5331

Chapter 5
Numerals in Financial Narratives

Numerals are more common in financial narratives than in documents from other domains, which makes understanding numerals very important when analyzing financial documents. In this chapter, we summarize our work on numerals in financial narratives and share findings from the FinNum shared task series in the 14th and 15th NTCIR Conferences. In Sect. 5.1, we discuss how to understand the meaning of a given numeral, and in Sect. 5.2, we discuss numeral attachment, where we link numerals and named entities. In Sect. 5.3, we show experimental results from downstream tasks that demonstrate the importance of numeral understanding in financial narratives. We conclude by proposing future research directions in Sect. 5.4.

5.1 Numeral Understanding

In Chap. 3, we identified the sources of financial opinion as insiders, professionals, social media users, and journalists. Table 5.1 lists the statistics of numerals in documents from these sources[1]: numerals are common in all kinds of financial documents. Indeed, almost every news article contains at least one numeral. This indicates the importance of numeral information in financial narratives, and explains why we devote an entire chapter to this topic.

In our work [5], we compare numerals in analysis reports with those in documents of other domains (hotel reviews [12] and persuasive essays [11]). Table 5.2 shows the statistics of these datasets. These results demonstrate the importance of numerals in financial documents. Below, we explain why managers, investors, and journalists

[1] We collect earnings calls from Refinitiv, analyst reports from Bloomberg, financial tweets from StockTwits, and financial news from MoneyDJ.

C.-C. Chen et al., *From Opinion Mining to Financial Argument Mining*,
SpringerBriefs in Computer Science,
https://doi.org/10.1007/978-981-16-2881-8_5

Table 5.1 Statistics of numerals appearing in four types of financial documents

	Earnings call	Analysis report	Financial tweet	Financial news	
Unit of measurement	Sentence	Sentence	Tweet	Headline	Article
Instances	13,574	4,952	2,028	75,448	75,448
Instances with numerals	7,499	2,938	1,395	45,073	75,297
Proportion with numerals	55.3%	59.3%	68.8%	59.7%	99.8%

Table 5.2 Statistics of numerals in three datasets from different domains [5]

Source	Analysis report	Hotel review [12]	Persuasive essay [11]
Language	Chinese	Chinese	English
Words	42,594	21,848	97,420
Numerals	5,144	67	111
Proportion with numerals (words)	12.1%	0.3%	0.1%

use so many numerals. First, managers must provide statistics about past operations and provide evidence about the results of future operations. These are generally represented using numerals. For example, instead of vaguely stating, "The company earned a lot last year," managers say, "In 2020 the earnings per share was 4.3, which was 40% higher than that in 2019." When making a claim, they do not say, "The company's future operations are promising;" instead they say, "We expect growth sales next year to be between 20% and 30%." Second, investors analyze financial instruments based on fundamental analysis and technical analysis, both of which predominantly use numerals to represent the results. For example, investors using fundamental analysis pay attention to financial statements; indeed, almost every term in a financial statement is a numeral. Likewise for those conducting technical analysis, which is based on historical price data statistics. Third, since all market participants (managers and investors) pays close attention to numerals, journalists make sure to provide numeric information in their articles.

A numeral is a kind of named entity. The temporal information mentioned in Sect. 4.1.3 is also represented by numerals. Although regular expressions make it easy to extract numerals from textual data, it can be difficult to understand what each numeral means. See for example the following tweet, which contains nine numerals:

(E5.1) $TSLA 256 Break-out thru 50 & 200- DMA (197-230) upper head res (274-279) Short squeeze in progress Nr term obj: 310 Stop loss:239.

These can be separated into monetary numerals (256, 197, 230, 274, 279, 310, and 239) and technical analysis parameters (50 and 200). Of the monetary numerals, 256

Table 5.3 A comprehensive taxonomy of financial numerals

	Earnings calls		Analysis reports [5]		Tweets [8]	
Category	Instances	Ratio	Instances	Ratio	Instances	Ratio
MONETARY: *money*	2,656	19.72%	874	16.99%	736	8.30%
MONETARY: *quote*	–	–	75	1.46%	1,033	11.65%
MONETARY: *change*	753	5.59%	18	0.35%	176	1.98%
MONETARY: *buy price*	–	–	–	–	415	4.68%
MONETARY: *sell price*	–	–	–	–	135	1.52%
MONETARY: *forecast*	–	–	–	–	355	4.00%
MONETARY: *stop loss*	–	–	–	–	35	0.39%
MONETARY: *support or resistance*	–	–	–	–	302	3.41%
PERCENTAGE: *relative*	3,040	22.57%	708	13.76%	767	8.65%
PERCENTAGE: *absolute*	969	7.19%	810	15.75%	346	3.90%
TEMPORAL: *date*	2,647	19.65%	2,134	41.49%	2,653	29.92%
TEMPORAL: *time*	8	0.06%	3	0.06%	365	4.12%
OPTION: *exercise price*	–	–	–	–	132	1.49%
OPTION: *maturity date*	–	–	–	–	70	0.79%
INDICATOR	–	–	–	–	216	2.44%
QUANTITY	2,199	16.33%	278	5.40%	982	11.07%
PRODUCT/VERSION	349	2.59%	136	2.64%	150	1.69%
RANKING	50	0.37%	3	0.06%	–	–
OTHER	798	5.92%	105	2.04%	–	–
	13,469	100.00%	5,144	100%	8,868	100%

is the close price of $TSLA, 197 and 230 are the moving averages of the 50-day and 200-day historical prices, 274 and 279 are the expected resistance price levels based on the this investor's analysis, 310 is the price target, and 239 is the stop-loss price of this investor. In this instance, the taxonomy for numerals in traditional NER tasks is insufficient for us to understand the numerals in financial narratives. Thus, we propose a taxonomy for financial numerals. This is shown in Table 5.3 with various statistics. Below, we explain each category using examples from social media [8].

Monetary numerals belong to the MONETARY category. One example is 110.20 in (E5.2) quoting the price of Facebook's security. These are further divided into the following eight subcategories: *money*, *quote*, *change*, *buy price*, *sell price*, *forecast*, *stop loss*, and *support or resistance*.

> **(E5.2)** $FB (110.20) is starting to show some relative strength and signs of potential B/O on the daily.

To distinguish these subcategories, recall that *money*, *quote*, and *change* are about status, not opinions; other subcategories are about opinions, specifically those of the tweet writer. Numerals such as 'a loss of $3 billion' are put in the *money* subcategory.

Numeral 110.20 in (E5.2) is a *quote*. Numerals describing changes in prices or money are seen as *change*. For example, '$AAPL -$3 today' describes a change in the price of $AAPL.

An individual investor's buying and selling prices help us understand the investor's performance, based on which we assign weights to the opinions of each investor. Thus 137.89 in (E5.3) is a *buying* instance and 36.50 in (E5.4) is an example of *selling*.

(E5.3) $SPY Long 1/2 position 137.89

(E5.4) $KOG Took a small position- hopefully a better outcome than getting kneecapped by $BEXP selling itself dirt cheap at 36.50

Investors sometimes forecast the price of the instruments based on their analysis results. Such monetary prediction numerals are put in the *forecast* subcategory: one such example is 14.35 in (E5.5). This opinion can be considered a summarization of the analysis results which yields information not only about the market sentiment and its degree but also the exact price level. A *stop-loss* price is the price level at which investors close their positions: an example is 17.99 in (E5.1).

(E5.5) $CIEN, CIEN seems to have broken out of a major horizontal resistance. Targets $14.35.

Support or resistance prices predict price movements. Some investors believe that when the price reaches the resistance price, it will then fall, and when the price reaches the support price, it will then rebound. This subcategory helps us identify price movement boundaries: an example of *support or resistance* is 46 in (E5.6).

(E5.6) $CTRP, $46 Breakout Should be Confirmed with Wm%R Stochastic Up

Section 6.1 will include application scenarios with numerals that convey investor opinions.

Financial documents contain many ratio-related numerals, for example, accounting ratios such as P/E ratios and current ratios. All such numerals are classified as PERCENTAGE, and are further divided into the *absolute* subcategory, which indicates the proportion of a certain amount, and the *relative* subcategory, which indicates change relative to the original amount. An example of *absolute* is 167.1 in (E5.7); 1.64, −2.7, −2.5, and −1.6 are examples of *relative*.

(E5.7) no trades today...currently 167.1% net long...ended the day down 1.64% due to $CASY (-2.7%), $NKE (-2.5%), $SRCL (-1.6%) and $JJSF (-1.6%)

As discussed in Sect. 4.1.3, temporal information is crucial in the financial domain. The date that many investors focus may have higher volatility. Thus we seek to capture temporal information that reveals such critical dates and times. Numerals in the TEMPORAL category are further divided into *date* and *time*. An example of *date* is (E5.8); (E5.9) shows *time*.

(E5.8) @DrCooper: $GDX $NUGT $DUST Buying on Weakness (06/30/2015)

(E5.9) $AMRN So what was that @ 11 a.m.?

OPTIONS, which are widely discussed in financial social media, are further divided into *maturity date* and *exercise price*. Such information helps us evaluate investor performance, similar to the MONETARY category's *target* price. *Maturity date* is shown in (E5.10), and *exercise price* is shown in (E5.11) (as $111).

(E5.10) looks like a big feb 18-22 $put spread on $cree.

(E5.11) Bought $FB $111 calls for $0.62.

When investors use technical indicators to analyze price movements, we match analysis result with price using the INDICATOR numerals that they mention. One example is (E5.12), which shows the need to identify the INDICATOR parameter.

(E5.12) $AAPL hit my short term target of the 100 SMA.

QUANTITY information also reveals an investor's position: we assign larger weights to opinions held by those with large positions. Sales quantities are also vital information in accounting. An example of QUANTITY is (E5.13).

(E5.13) $RSOL bought 3500 shares today!

Considering the impact that opinions toward iPhone 6 and iPhone 12 could have on Apple's security shows that PRODUCT/VERSION numbers should also be captured to understand the topic of discussion. An example is (E5.14).

(E5.14) iPhone 6 may not be as secure as Apple thought.. $AAPL

RANKINGS are sometimes mentioned by managers and analysts, such as #1 and #2 in (E5.15), an earnings call. These reflect a company's market position, and are important information for understanding the target company.

(E5.15) The chart on the left here we've shown back in March and it shows the market position of over 75% of our Chemical product sales where we're either #1 or #2 in the market.

Given this taxonomy, we return to Table 5.3 to compare the narratives of different market participants. First, we find that managers rarely discuss the company's stock price, and few analysts use technical analysis in their reports. However, social media users regularly tweet about technical analysis results. From this we can differentiate managers from investors and professionals from amateur investors. Second, numerals reveal the different habits of market participants. Thirty-nine percent of numerals in earnings calls are PERCENTAGES, which constitute only 29% and 12% of analysis reports and social media data, respectively: when describing company operations, managers pay more attention to comparisons rather than only provide the information shown in financial statements. In contrast, investors, especially social media users, use many MONETARY numerals. Third, analysts use more TEMPORAL information than other market participants. Fourth, we find that although managers sometimes mention Quantities, analysts do not seem to focus on this. Additionally, we also find that the unit of Quantities are different between managers' and amateur investors' narratives. Most managers describe the Quantities related to product sales, and many amateur investors talk about the Quantities of financial instruments they buy/sell.

The above numeral categories and statistics suggest many cues that help us better understand numeral information. Below, we discuss findings from the literature for this task. Numeral understanding is formulated as a classification task [6]. Because extracting numerals from textual data is trivial, we focus on classifying the extracted numerals into the proposed categories. In many NLP tasks, Transformer-based language models and BERT-like architectures are currently the state of the art. In numeral understanding of financial social media data, BERT achieves the best performance [24] with 89.72% and 87.98% micro-F1 and macro-F1 scores in a 17-class classification setting. Below we list features that have been proposed:

- **Part-of-speech (POS) tags**: Ait Azzi and Bouamor [1] and Liang and Su [15] extract POS features with CMU ARK Twitter POS Tagger [20] and CoreNLP [18], respectively.
- **Keywords**: Ait Azzi and Bouamor [1] adopt keywords from Chen et al. [6]. Liang and Su [15] propose patterns for (sub)categories.
- **Topic**: Spark [23] uses latent Dirichlet allocation (LDA) [2] to extract features for tweet topics.
- **Position**: Spark [23] uses the position of the target numeral in the tweet.
- **Named entities**: Liang and Su [15] extract named entities using CoreNLP [18].
- **Format**: Integer (float) format information is used as a feature [23, 25]. Co-occurrence format information is extracted via patterns [25].
- **Numeral information**: Spark [23] uses the raw numeral value as well as the log of the raw value and the normalized raw value.
- **Bag-of-characters**: Spark [23] considers the *n* characters nearest the target numeral.
- **Prefixes/suffixes**: Wu et al. [25] use prefixes and suffixes.
- **Brown clusters**: Wu et al. [25] use the *j*-character prefix of the Brown clusters [3] as features.
- **Recognizers.Text type**: Wu et al. [25] adopt the text types extracted by Microsoft. Recognizers.Text.

Given the results of these studies and the analysis of our own work [7], we find that features proposed by Wu et al. [25] (format, prefixes/suffixes, Brown clusters, and Recognizers.Text) perform well in general categories (MONETARY and TEMPORAL). However, handcrafted features used in Ait Azzi and Bouamor [1] could improve performance in finer-grained subcategories such as *relative*, *absolute*, *exercise price*, and even QUANTITY and PRODUCT/VERSION. For future work, we suggest enhancing models with the above features; it is also worth discussing what BERT-like models can and cannot capture when using end-to-end models directly.

In summary, numerals are crucial in financial narratives, and different documents predominantly use different types of numeral information. The literature yields important insights for future work. We will discuss the applications of numeral understanding in Chap. 6.

5.2 Numeral Attachment

After understanding the meaning of each numeral, the task becomes determining which target entity is related to the given numeral. For example, in (E5.16), both \$65 and \$8 are *quotes*. Should we average these and conclude that the close price of \$NE is 36.5 because there is only one target entity? Clearly the answer to this question is no, because \$65 is the price of oil; only \$8 is related to \$NE.

(E5.16) \$NE OK NE, last time oil was over \$65 you were close to \$8. Giddy-up. . .

To address this problem, we define a new task termed *numeral attachment*. In this task, we identify whether the given numeral and the given target entity are related. Taking (E5.16) as an example, when given \$65 and \$NE, the model should output "not attached". When given \$8 and \$NE, the model should output "attached". Table 5.4 describes the NumAttach 2.0 dataset proposed in previous work [9]. Fifty-five percent of financial tweets contain more than one cashtag, and 73% of financial tweets have more than one numeral. Table 5.5 shows the label distribution. "Attached" cases account for the larger proportion (77%); the "not attached" instances account for 23%.

Below we list studies that use the NumAttach dataset and summarize their findings.

- Xia et al. [26] use TF-IDF as features for a SVM model. Their model is 10% better than the majority-vote model under the macro-F1 metric.
- Liang et al. [16] show the results when using BERT only to encode textual data instead of fine-tuning the BERT model. They use BERT word vectors as the input to CNN and BiLSTM models. The experimental results show that dependency features are not useful with the proposed model.
- Chen and Liu [10] discuss the results of the BERT-BiLSTM model with different class weights. Weights (0.8, 0.2), which approximate the dataset distribution, outperform other settings, including (0.99, 0.01) and (0.9, 0.1). They also show the usefulness of paraphrasing tweets by removing meaningless terms that were selected manually.

Table 5.4 Distribution of single-numeral and multi-numeral cashtags

	Single-cashtag	Multi-cashtag
Single-numeral	1,282 (12.40%)	1,427 (13.80%)
Multi-numeral	3,347 (32.37%)	4,284 (41.43%)

Table 5.5 Distribution of attached and not attached labels in both single-numeral and multi-numeral cashtags

	Single-cashtag	Multi-cashtag
Single-numeral	1,204/78	1,017/410
Multi-numeral	3,071/276	3,106/1,178

- Jiang et al. [14] look at fine-tuning techniques. They tune each layer with different learning rates, after which they change the learning rate per iteration using slanted triangular learning rates [13] and cyclical momentum [22] methods. They show that together with the BERT model, these fine-tuning techniques significantly improve performance.
- Moreno et al. [19] propose an ensemble model which uses the *min* between BERT and RoBERTa [17] as the prediction. They discuss the results on performance using different thresholds, and suggest using a threshold of 0.7 rather than 0.5.

Although we are discussing numeral information, the studies mentioned in Sect. 5.1 and those in this section do not take numerals themselves into consideration. That is, the works mentioned above focus on contextual features; few examine the given numerals. For example, a four-figure number is more likely to stand for the year than to denote a percentage; likewise, a four-figure number is more likely to be related to the S&P 500 index than the Dow Jones Industrial Average index. In previous work [4], we propose a text representation for numeral-related tasks which concatenates embeddings for tokens, characters, positions, and magnitudes, as illustrated in Fig. 5.1. We further use Fig. 5.2 to illustrate magnitude embeddings. Given a target number of 1.35, we separate it into individual digits and represent each digit with a one-hot vector containing 11 dimensions to cover 0 to 9 as well as the decimal point. The results of the ablation experiments shown in Table 5.6 demonstrate the usefulness of this representation for numeral attachment.

We also find that co-training with other fine-grained context understanding tasks is helpful for numeral-related tasks. We jointly learn numeral attachment with two auxiliary tasks: (1) whether the tweet contains the reason (Reason-binary), and (2) the aspect of the reason (Aspect). The results in Table 5.7 show that these settings improve the performance of numeral attachment. These findings suggest that finding better representations for numerals would be better than representing them using context alone. In Sect. 5.3, we show other cases for the usefulness of (1) tailor-made numeral representation, and (2) co-training with fine-grained auxiliary tasks.

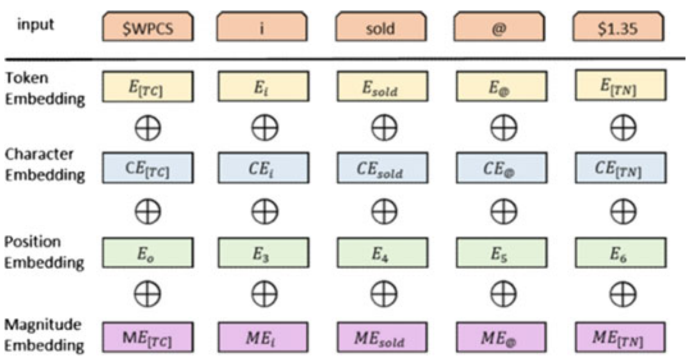


Fig. 5.1 Text representation for numeral-related tasks [4]

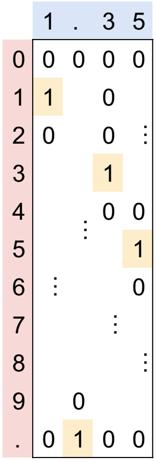


Fig. 5.2 Magnitude embedding [4]

Table 5.6 Ablation analysis of input representation [4]

Token	v	v	v	v
Character		v	v	v
Position			v	v
Magnitude				v
Macro-F1	60.08%	69.59%	69.73%	73.46%

Table 5.7 Ablation analysis for auxiliary tasks [4]

Numeral attachment	v	v	v	v
Reason-binary		v		v
Aspect			v	v
Macro-F1	67.14%	69.97%	66.95%	73.46%

We can further formulate numeral attachment in a more general way. That is, given a numeral, the model should identify the entities that are related to the numeral. Given example (E5.17) from an earnings conference call, it may not be enough to know only that "$53.3" billion" is a MONETARY numeral, and that it is related to this company. The "$53.3" billion" here in fact describes this company's revenue. Thus, the next challenge is extracting the entity described by the given numeral.

(E5.17) We generated $53.3 billion in revenue, a new Q3 record.

Table 5.8 lists instances of general numeral attachment. In (E5.17), since "revenue" is mentioned explicitly, we can link the extracted "52.3 billion" and "the company" with "revenue". Likewise for the "stop loss" case in (E5.1). However, in

Table 5.8 Instances of general numeral attachment

Example	Numeral	Target entity	Relation
(E5.17)	53.3 billion	The company	Revenue (explicit)
(E5.1)	239	$TSLA	Stop loss (explicit)
	256		Quote (implicit)

Table 5.9 Top-ranking numeral-related entities in both earnings calls and analysts' reports

	Earnings call (manager)		Analysis report (investor)	
Rank	Entity	Frequency	Entity	Frequency
1	Revenue	767	Revenue	855
2	Q	371	EPS	481
3	Sales	255	Gross margin	326
4	EPS	221	Profit	275
5	Earnings	165	Operating margin	115
6	Years	154	Price target	108
7	Free cash flow	110	Operating profit	59

cases such as "256" in (E5.1), we cannot extract the named entity to link it with the target numeral. In this instance, the annotations and pre-defined taxonomy introduced in Sect. 5.1 help us determine the implicit information in the narrative.

Manual annotation of the numeral-related entities in the earnings call and analysis report allows us to better understand the use of such named entities. In the earnings calls (English), there are 2,502 unique entities out of 13,469 annotations, and in the analysts' reports (Chinese), there are 1,206 unique entities out of 10,000 annotations. Table 5.9 lists the top-ranking entities, yielding the following findings.

1. Managers report data about operations, including revenue, sales, EPS, earnings, and free cash flow.
2. Investors not only focus on quantitative operation results (revenue and EPS), but also pay attention to accounting ratios (gross margins and operating margins).
3. Managers seldom mention the stock price, but investors often discuss it.

In summary, accurate numeral understanding and numeral attachment facilitates in-depth understanding of numeral information. Information gleaned via these tasks is useful for fine-grained financial opinion mining, because numerals constitute much of the content of financial narratives. For example, instead of merely identifying claim sentences, we can investigate the claims in detail. We can also confirm whether a company that provides more numerals as evidence in its reports indeed has a better outlook than a company that provides little numeral information.

APPLE (NASDAQ:AAPL) ANALYST RATINGS HISTORY

Show: Only the Most Recent Rating From Each Brokerage

Date	Brokerage	Action	Rating	Price Target
1/5/2021	Canaccord Genuity	Boost Price Target	Buy	\$145.00 → \$150.00
1/5/2021	JPMorgan Chase & Co.	Set Price Target	Buy	\$150.00
1/4/2021	Credit Suisse Group	Boost Price Target	Neutral	\$106.00 → \$120.00
1/4/2021	UBS Group	Set Price Target	Neutral	\$115.00
1/4/2021	Royal Bank of Canada	Set Price Target	Buy	\$132.00
12/18/2020	Citigroup Inc. 3% Minimum Coupon Principal Protected Based Upon Russell	Boost Price Target		\$125.00 → \$150.00
12/18/2020	Smith Barney Citigroup	Boost Price Target		\$125.00 → \$150.00
12/16/2020	Morgan Stanley	Boost Price Target	Overweight	\$136.00 → \$144.00

Fig. 5.3 Price targets of professionals collected by an information vendor (MarketBeat)

5.3 Improving Financial Opinion Mining via Numeral-Related Tasks

In the previous sections, we show how to understand the meaning of numerals and how to link the related entities to a given numeral. In this section, we discuss how to use the extracted information and how to improve financial opinion mining by enhancing the numeracy of models. The discussed topics are listed as follows.

- The informativeness of opinions expressed with numerals.
- Claim detection with auxiliary numeral understanding tasks.
- Volatility forecasting using numeral information.
- Enhancing numeracy with magnitude embeddings.

Investors' price targets go beyond bullish and bearish. A price target not only reveals the investor's market sentiment but also shows what price level the investor expects to see in the future. Information vendors like Bloomberg and MarketBeat[2] collect price targets of professional analysts, and show this information in tabular form, as shown in Fig. 5.3, which attests the importance of this information. However, few platforms provide price targets of investors using social media platforms, even though these investors regularly discuss price targets. Models for numeral understanding and numeral attachment could be used to extract such information automatically from investors' tweets to produce an overview similar to Fig. 5.3.

Table 5.10 shows statistics compiled in previous work [6], in which we compare crowd investors and professional analysts' price targets, finding that crowd investors are more progressive, because the difference between their close prices and price

[2] https://www.marketbeat.com/.

Table 5.10 Comparison of crowd investors and professional analysts' price targets [6]

	Crowd	Analyst
Average difference between close price and price target	13.17%	6.75%
Achievement rate	67.03%	74.73%
Duration	3.38 months	2.46 months

Table 5.11 Results of three backtesting strategies [6]

	Crowd	Analyst
Win ratio	68.13%	71.43%
Average profit	11.08%	6.42%
Average loss	−8.43%	−8.40%

targets is larger than that of professional analysts. Table 5.11 shows the experimental results based on the following simple trading rules:

- If the price target is higher (lower) than the close price, long (short) the stock.
- If the close price reaches the price target when the position is held, close this position for profit.
- If the unrealized loss reaches 7%, close the position.

Thus, using fine-grained financial opinion from the crowd yields promising backtesting results. This also demonstrates the informativeness of price targets from both professional analysts and financial social media users.

We also discuss how numeral information affects the performance when extracting financial opinion components. As discussed above, investors do not claim that prices will rise, especially in reports from professionals. They may instead make price target claims. Based on our observation, many such claims are made via estimations. Thus, in previous work [5], we sought to encode the estimation in the given sentence and to determine whether such information would improve the performance of claim detection. Table 5.12 shows the experiment results.

The baselines are the results of directly using entire sentences as the model input. We use the representation from Fig. 5.2 to encode numerals in the sentence, and find that adding numeral information improves claim detection performance in professional analysts' reports. We further use category classification from Sect. 5.1 as the auxiliary task, and find that adding this task further improves performance. This experiment attests the usefulness of numeral understanding for fine-grained semantic analysis in financial narratives, and shows that independently encoding numerals restores information that was not present in the original language model.

Following, we discuss whether the extracted numeral information improves the performance of downstream tasks. Unlike the price target experiment, in the following experiment, we extract accounting metrics from the transcription of the earnings conference call, and use this extracted information for volatility forecasting.

Table 5.12 Performances of claim detection [5]

	CNN	BiGRU	CapsNet
Baseline	77.26%	78.29%	78.68%
+ Numeral information	78.19%	79.06%	80.91%
+ Numeral information & category task	81.35%	81.65%	82.62%

Table 5.13 Statistics of annotations for DNU-GAAP and DNU-Influence

DNU-GAAP			DNU-Influence		
Class	Labels	%	Class	Labels	%
GAAP	3,675	40.68%	Positive	5,467	60.52%
Non-GAAP	432	4.78%	Negative	669	7.41%
Other	4,927	54.54%	Neutral	2,898	32.08%

In addition to category information, we use two other labels for numerals. The first concerns Generally Accepted Accounting Principles (GAAP), which we term domain-specific numeral understanding (DNU-GAAP). Such numerals are assigned one of the following labels.

- GAAP: A GAAP-related numeral
- Non-GAAP: A numeral used for adjusting the metric related to GAAP
- Other

We also use a label concerning the influence of the given numeral toward the related named entity: this task is called DNU-Influence. Table 5.13 shows the statistics of these annotations. We distill sentences from earnings conference calls into these labels. For example, (E5.18) becomes *absolute*/Non-GAAP/Positive.

(E5.18) Our adjusted tax rate is expected to be 20.5.

After converting all of the sentences to the above form, we use a two-layer Transformer to forecast the volatility. Table 5.14 shows the results under the public-available dataset [21]: the proposed method outperforms other baselines in 3-day and 7-day volatility prediction. In this experiment, we use only the context to understanding the meanings of given numerals, and further use the meanings of these numerals for the downstream task. The results again attest the importance of numerals in financial narratives, and also demonstrate that numeral understanding in financial narratives can improve the performance of downstream tasks.

Finally, we highlight the usefulness of magnitude embeddings. We have already discussed the three kinds of financial opinion sources: insiders (earnings conference calls), professionals (analysis reports), and social media users (tweets). Now, we focus on the numerals in news articles. As shown in Table 5.1, almost all financial

Table 5.14 Experimental volatility forecasting results. The evaluation metric is MSE (the lower is the better)

	3-day	7-day	15-day	30-day
MDRM (text only)	1.431	0.439	0.309	0.219
MDRM (text + audio)	1.371	0.420	0.300	0.217
HTML (text only)	1.175	0.372	**0.153**	**0.133**
HTML (text + audio)	0.845	0.349	0.251	0.158
Proposed method	**0.745**	**0.300**	0.232	0.187

news articles contain at least one numeral, and over 59% of news headlines have at least one numeral. Based on this finding, we use a new cloze task: we use the numeral in the headline as the answer, and then remove the numeral from the headline, making the headline without an answer the question stem. As the plausible answers to the question, we select four distinct numerals whose values are closest to the value of the answer. The goal of this task is to test whether the model selects the nearest numeral when given the question stem. The following example demonstrates the idea:

News Article:

Major banks take the lead in self-discipline. The five major banks' newly-imposed mortgage interest rates climbed to **1.986%** in May. ... Also approaching **2%** integer alert ... Up to **2.5%** ... Also increased by **0.04** percentage points from the previous month ... Prevent the housing market bubble from fully starting.

Question Stem:

Driven by self-discipline, the five major banks' new mortgage interest rates are approaching nearly ______%.

Answer Options:

(A) Also increased by **0.04** percentage points from the previous month

(B) The five major banks' newly-imposed mortgage interest rates climbed to **1.986%** in May.

(C) Also approaching **2%** integer alert

(D) Up to **2.5%**

Answer: (C)

We conduct experiments with four models.

- BERT embedding similarity: Uses cosine similarity of token embeddings of question stem and that of answer options. Most similar option is chosen.
- Vanilla BERT: Encodes question stem and answer options using BERT-Large, and generates prediction using multilayer perceptron.
- BERT-BiGRU: Vanilla BERT + BiGRU architecture.

Table 5.15 Numeral cloze results. The symbol * denotes results that are significantly different from the second-best model (BERT-BiGRU) under McNemar's test with $p < 0.05$

Model	Accuracy
BERT embedding similarity	57.30%
Vanilla BERT	66.41%
BERT-BiGRU	67.15%
BERT-BiGRU + numeral encoder	**69.95%***

- BERT-BiGRU + Numeral Encoder: Uses CNN as numeral encoder to extract features for numerals in answer options.

Table 5.15 shows the experimental results. The results attest the usefulness of the numeral encoder, which extracts numeral features independently. The results also show that the proposed techniques and the directions of numeral understanding are essential for the numeracy of neural network models.

The pilot experiments in this section show that regardless of the source (earnings conference call, analysis report, social media data, or news article), numeral information provides information that yields a better understanding of financial documents. Our results also indicate the importance of fine-grained analysis for such numerals. For future work, we suggest adding numeral understanding tasks to models if dealing with financial textual data. We also demonstrate the usefulness of magnitude embeddings; note that their usefulness likely extends to domains other than the financial domain.

5.4 Summary

In this chapter, we present a special characteristic of financial narratives—numerals. First, we show that in all kinds of financial documents, numerals account for over 50% of the sentences (or tweets/articles: see Fig. 5.1). Second, we propose a numeral understanding task, with which we seek to understand the meaning of numerals via context. To this end we propose a taxonomy and annotations, and also survey features used in the literature. Third, we extend the numeral attachment task from our previous work [4] to a more general task. Fourth, we conduct experiments on four tasks and four kinds of documents to show the usefulness of numeral-related tasks and the helpfulness of numeral representation. The experimental results attest the importance of numeral information and as well as the robustness of the proposed methods. In Chap. 6, we discuss applications that involve the extraction of numeral-related opinions.

References

1. Azzi, A.A., Bouamor, H.: Fortia1@the NTCIR-14 FinNum task: enriched sequence labeling for numeral classification
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
3. Brown, P.F., Desouza, P.V., Mercer, R.L., Pietra, V.J.D., Lai, J.C.: Class-based n-gram models of natural language. Comput. Linguist. **18**(4), 467–479 (1992)
4. Chen, C.-C., Huang, H.-H., Chen, H.-H.: Numeral attachment with auxiliary tasks. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1161–1164 (2019)
5. Chen, C.-C., Huang, H.-H., Chen, H.-H.: NumClaim: investor's fine-grained claim detection. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 1973–1976 (2020)
6. Chen, C.-C., Huang, H.-H., Shiue, Y.-T., Chen, H.-H.: Numeral understanding in financial tweets for fine-grained crowd-based forecasting. In: 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 136–143. IEEE (2018)
7. Chen, C.-C., Huang, H.-H., Takamura, H., Chen, H.-H.: Final report of the NTCIR-14 FinNum task: challenges and current status of fine-grained numeral understanding in financial social media data. In: NII Conference on Testbeds and Community for Information Access Research, pp. 183–192. Springer (2019)
8. Chen, C.-C., Huang, H.-H., Takamura, H., Chen, H.-H.: Overview of the NTCIR-14 FinNum task: fine-grained numeral understanding in financial social media data. In: Proceedings of the 14th NTCIR Conference on Evaluation of Information Access Technologies, pp. 19–27 (2019)
9. Chen, C.-C., Huang, H.-H., Takamura, H., Chen, H.-H.: Overview of the NTCIR-15 FinNum-2 task: numeral attachment in financial tweets. Development **850**(194), 1–044 (2020)
10. Chen, Y.-Y., Liu, C.-L.: MIG at the NTCIR-15 FinNum-2 task: use the transfer learning and feature engineering for numeral attachment task. In: Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies (2020)
11. Eger, S., Daxenberger, J., Gurevych, I.: Neural end-to-end learning for computational argumentation mining. In: ACL, Vancouver, Canada, July, pp. 11–22. Association for Computational Linguistics (2017)
12. Eger, S., Daxenberger, J., Stab, C., Gurevych, I.: Cross-lingual argumentation mining: machine translation (and a bit of projection) is all you need! In: COLING, Santa Fe, New Mexico, USA, August, pp. 831–844. Association for Computational Linguistics (2018)
13. Howard, J., Ruder, S.: Universal language model fine-tuning for text classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 328–339 (2018)
14. Jiang, M. T.-J., Chen, Y.-K., Wu, S.-H.: CYUT at the NTCIR-15 FinNum-2 task: tokenization and fine-tuning techniques for numeral attachment in financial tweets. In: Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies (2020)
15. Liang, C.-C., Su, K.-Y.: ASNLU at the NTCIR-14 FinNum task: incorporating knowledge into DNN for financial numeral classification. In: Proceedings of the 14th NTCIR Conference on Evaluation of Information Access Technologies, vol. 192 (2019)
16. Liang, Y.-C., Cheng, Y.-Y., Huang, Y.-H., Chang, Y.-C.: TMUNLP at the NTCIR-15 FinNum-2. In: Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies (2020)
17. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: RoBERTa: a robustly optimized BERT pretraining approach (2019). arXiv preprint arXiv:1907.11692
18. Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, D.: The stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55–60 (2014)

19. Moreno, J.G., Boros, E., Doucet, A.: TLR at the NTCIR-15 FinNum-2 task: improving text classifiers for numeral attachment in financial social data
20. Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., Smith, N.A.: Improved part-of-speech tagging for online conversational text with word clusters. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 380–390 (2013)
21. Qin, Y., Yang, Y.: What you say and how you say it matters: predicting stock volatility using verbal and vocal cues. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, July, pp. 390–401. Association for Computational Linguistics (2019)
22. Smith, L.N.: Cyclical learning rates for training neural networks. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 464–472. IEEE (2017)
23. Spark, A.: BRNIR at the NTCIR-14 FinNum task: scalable feature extraction technique for number classification
24. Wang, W., Liu, M., Zhang, Y., Xiang, J., Mao, R.: Financial numeral classification model based on BERT. In: NII Conference on Testbeds and Community for Information Access Research, pp. 193–204. Springer (2019)
25. Wu, Q., Wang, G., Zhu, Y., Liu, H., Karlsson, B.: DeepMRT at the NTCIR-14 FinNum task: a hybrid neural model for numeral type classification in financial tweets. In: Proceedings of the 14th NTCIR Conference on Evaluation of Information Access Technologies (2019)
26. Xia, X., Wang, W., Liu, M. WUST at NTCIR-15 FinNum-2 task. In: Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies (2020)

Chapter 6
FinTech Applications

The Financial Stability Board defines financial technology (FinTech) as "technology-enabled innovation in financial services."[1] At the 2015 World Economic Forum, experts proposed a taxonomy for financial services[2] that can be classified into six major categories: payments, deposits and lending, market provisioning, capital raising, insurance, and investment management. Because financial opinion mining can be applied to many listed services, we survey various cases in this chapter and show that financial opinion mining is useful and crucial in many financial application scenarios. In Sect. 6.1, we discuss information provision services in the financial domain. In Sect. 6.2, we discuss work on personalized recommendations, which is the goal of emotional banking. In Sect. 6.3, we discuss applications for improving employee efficiency. In this chapter, we demonstrate the importance of financial opinion mining in the financial industry.

6.1 Information Provision

An analyst is a professional information provider who summarizes current events and produces claims based on these events. That is, analysts not only provide the latest market information, but also offer their view based on all available information. We begin this section with the workflow of a professional analyst.

1. **Information collection**: They collect information from sources listed in Chap. 3 such as insiders and news articles.

[1] https://www.fsb.org/wp-content/uploads/P140219.pdf.

[2] http://www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf.

C.-C. Chen et al., *From Opinion Mining to Financial Argument Mining*,
SpringerBriefs in Computer Science,
https://doi.org/10.1007/978-981-16-2881-8_6

Fig. 6.1 Screenshot of Bloomberg Terminal's sentiment analysis function

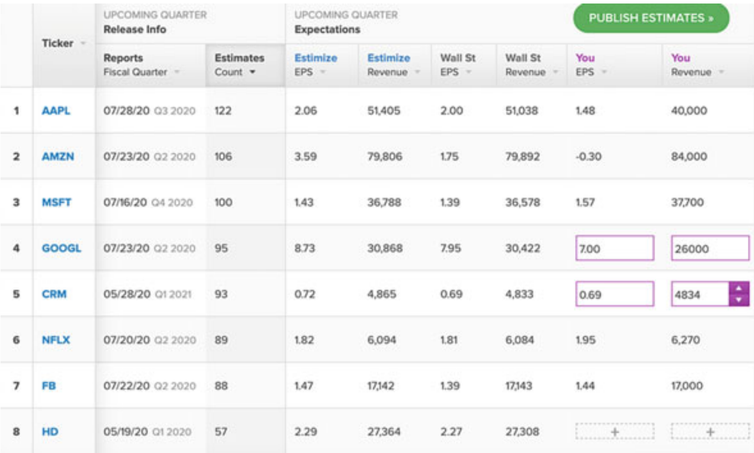
2. **Information verification**: They verify the collected information by visiting companies or via discussion with other analysts.
3. **Influence inference**: They infer the potential influence of each piece of information.
4. **Opinion formulation**: They sort out the important parts to produce claims and generate a report.

A professional analyst thus "connects all the dots" to get the full picture. When developing an information provision service, we seek to provide analysts with automated assistance by doing the trivial, tedious work for them.

Information vendors such as Bloomberg and Refinitiv play an important role in the first step of the analyst's workflow. They provide the latest news, quotes, and analysis reports from other organizations, combining all essential data on one platform. They provide not only raw data but also sort out this raw data to produce structured data. Of the sources listed in Chap. 3, information vendors most often neglect the opinions of social media users, despite the many studies [19, 40] that demonstrate the informativeness of such opinions. Hence one challenge is collecting opinions and presenting them in a structured form similar to what information vendors do for the views and opinions of insiders and professionals.

In financial opinion mining, sentiment analysis is the most common topic. As shown in Fig. 6.1,[3] Bloomberg Terminal demonstrates how to visualize the extracted sentiments of social media users with market data. They show counts of positive and negative tweets alongside historical price data. As mentioned in previous chapters, such sentiment comes from coarse-grained investor opinion. However, there are many details in a financial opinion: we here discuss how to collect fine-grained information.

[3] https://www.bloomberg.com/company/press/bloomberg-and-twitter-sign-data-licensing-agreement/.



	Ticker	UPCOMING QUARTER Release Info		UPCOMING QUARTER Expectations					
		Reports Fiscal Quarter	Estimates Count	Estimize EPS	Estimize Revenue	Wall St EPS	Wall St Revenue	You EPS	You Revenue
1	AAPL	07/28/20 Q3 2020	122	2.06	51,405	2.00	51,038	1.48	40,000
2	AMZN	07/23/20 Q2 2020	106	3.59	79,806	1.75	79,892	-0.30	84,000
3	MSFT	07/16/20 Q4 2020	100	1.43	36,788	1.39	36,578	1.57	37,700
4	GOOGL	07/23/20 Q2 2020	95	8.73	30,868	7.95	30,422	7.00	26000
5	CRM	05/28/20 Q1 2021	93	0.72	4,865	0.69	4,833	0.69	4834
6	NFLX	07/20/20 Q2 2020	89	1.82	6,094	1.81	6,084	1.95	6,270
7	FB	07/22/20 Q2 2020	88	1.47	17,142	1.39	17,143	1.44	17,000
8	HD	05/19/20 Q1 2020	57	2.29	27,364	2.27	27,308	+	+

Fig. 6.2 Screenshot of Estimize, a service that compiles earnings estimations of its users

Estimize[4] is a FinTech company which compiles earnings estimations of its users. Figure 6.2 shows a screenshot. Users fill out forms, which Estimize uses to calculate the average of all users' estimations. With this information, they compare EPS and revenue estimations from both professional investors and social media users. Jame et al. [21] find that the forecasts provided by Estimize's users improve price discovery. Da and Xing [12] analyze Estimize forecasts from a herding perspective to show that the more public information the user accesses, the less the user shares his/her own private opinion. These works also confirm the accuracy of forecasts from crowdsourcing platforms.

In addition to claims about EPS and earnings, investors also produce forecasts such as price targets. Since many financial opinions are expressed in natural language instead of in a tabular form like that in Fig. 6.2, understanding opinions in unstructured form is another research focus. In Sect. 5.3, we show that price targets from social media users are good predictors of stock movement. In previous work [8], we demonstrate how to visualize this information for investors: Fig. 6.3 shows a screenshot of CrowdPT, the resultant system, which makes it easy for investors compare stock prices with price targets. In addition to price targets, some of the categories in Table 5.3 contain informative opinions. Almost all financial opinions can be converted into an index and shown in charts such as those in Figs. 6.1 and 6.3. In previous work [5], we also show that the distribution of returns based on *buy/sell price* and *support or resistance price* signals from social media users is significantly different from that of randomly selected trading days. These systems and studies all inform methods for automatic information collection, and are also examples of ways to visualize such financial information. These studies support the importance of collecting more fine-grained information as opposed to capturing sentiment only.

[4] https://www.estimize.com/.

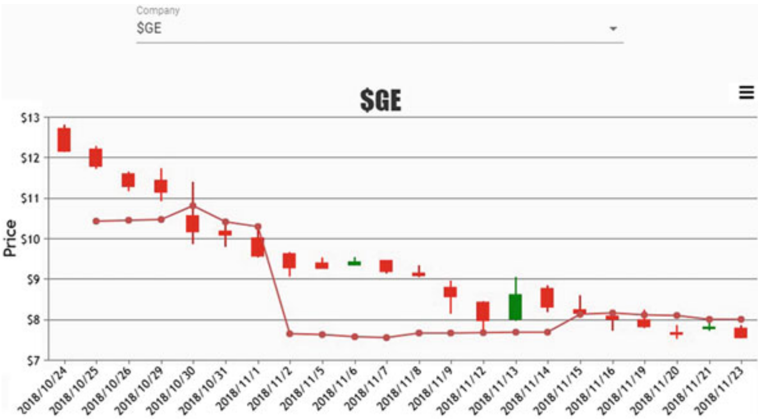


Fig. 6.3 Screenshot of CrowdPT, a system that enables investors to compare stock prices with price targets [8]

Once the information is collected, verification is necessary. Automatic fact-checking is a related research topic. Most objective descriptions of facts in talks or documents released by insiders, professionals, and journalists are correct, reliable information. However, their subjective opinions must be verified. For example, it is important to be able to judge whether a manager's claims in an earnings conference call are rational. It is difficult to design and collect the data needed to train the corresponding models for rationality-checking. In previous work [7], we use market comments to simulate this scenario. According to Chap. 5, numerals are important in financial narratives; managers and investors all focus on numeral information and make claims that include estimations expressed as numerals. In one corresponding verification task, we judge whether a given numeral in a market comment is exaggerated. Take for example (E6.1) and (E6.2): the words in these sentences are the same but the price targets are different. Given a stock which closes at 850, (E6.2)'s price target of 300 is likely an exaggeration. Our experimental results show that models perform well in very irrational cases, but perform worse in instances in which the correct numeral is replaced with a similar value.

(E6.1) We reiterate our buy recommendation and maintain the price target of 900.

(E6.2) We reiterate our buy recommendation and maintain the price target of 300.

Unlike information collected from formal, trustworthy sources, almost all Web information—especially that from social media platforms—must be verified before use. Relevant studies include those on fake news verification [33], fact-checking [16], and even spam detection [22, 34]. The quality evaluation task discussed in Sect. 4.2 is also a related issue.

For the third step—influence inference, in which we estimate the influence of each piece of information—we list some studies that use various kinds of information to predict the impact on future price movement.

- **Financial statements**: Holthausen and Larcker [18] predict excess returns using a logit model with data from financial statements. Their results show that the proposed model earns significant, abnormal returns over the period from 1978 to 1988.
- **Market data**: Liu et al. [28] encode market data at multiple time scales using both an RCNN architecture and a discrete wavelet transform [24] for stock trend prediction. They experiment on two datasets with intra-day price data (the FI-2010 dataset [31] and their CSI-2016 dataset). Their results support the usefulness of considering multi-scale market data for stock trend prediction. Ding et al. [13] experiment on inter-day market data, and also demonstrate the helpfulness of multi-scale representations.
- **Information from insiders**: As mentioned in Sect. 3.1, formal reports and insider talks are all informative and predictive for future returns and risks. For example, Loughran and McDonald [30] show the usefulness of sentiment information for prediction tasks in 10-K reports, and Qin and Yang [32] use earnings conference calls to predict volatility.
- **Investor opinions**: In Sects. 3.2 and 3.3, we discuss the informativeness of professionals' opinions as well as those from users of social media platforms. In Chap. 4, we show how to analyze these opinions. Based on the proposed argumentation structures and the concept of influence power estimation, we can infer the impact of a given opinion on a certain target entity.
- **News**: Event-based market movement prediction has been widely discussed in the NLP community. Many studies use news articles as data sources. For example, Hu et al. [20] propose a hybrid attention network and show that trading based on their model's predictions yields better profits than other baselines. Cheng et al. [10] extract events into tuples, which they then use to construct an event knowledge graph. They also show that their framework is profitable in the stock market.

The above-mentioned studies estimate the probability of future events given financial information. These probabilities can be used in the final step of analysis summarization.

Figure 6.4 illustrates the workflow with the concepts proposed in Chap. 2. We have completed step 3 in the figure. That is, at step 1 we collect information ($I_{i,t}^{e_1}$), and at step 2 we verify this information. That which is identified as fake or exaggerated is removed, and information which is correct becomes the premises ($P_{j,t}^{e_1}$); market data is also a premise. In step 3, we produce inferences based on this verified information. Different models may yield different claims (C^U or C^D). A given model's claims may also vary with the input data. The final step is to summarize the premises and the claims to author a report, which is considered an opinion (O).

The NLP community has proposed datasets and models for use in exploring summarization. For example, Li et al. [26] propose a system that extracts events, then links them, and finally generates a summary based on feature weights. Fabbri et al. [14] publish the Multi-News dataset, which contains more than 50,000 instances, and propose an end-to-end model that merges the pointer-generator network [35] and maximal marginal relevance [3]. For short, text-like tweets, Shapira et al. [36]

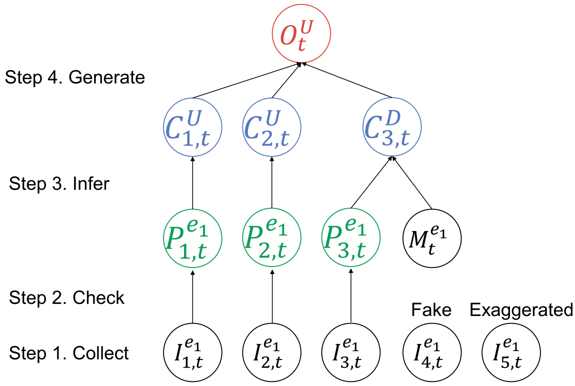


Fig. 6.4 Workflow with the concepts that are introduced in Chap. 2

propose a system based on open knowledge representation [39]. As these works are similar to summarizing premises, future works could borrow their approaches.

Claim generation, however, may be different from premise summarization, because it takes stance into account. Although some studies on argument mining [1, 15, 17] explore claim generation, few generate claims for financial opinions. Many studies in financial opinion mining stop at step 3 in Fig. 6.4. This may be because templates can be used to generate claims. For example, if the model predicts that the price of $AAPL will rise to 200 in the next three months, we can use template (E6.3) to generate the claim (The price target of "$AAPL" is "200".).

(E6.3) The price target of "target entity" is "model's prediction".

Although there are indeed templates that could be used to generate the claims based on the results of step 3, this is still a worthwhile research direction. These claims should be context-aware sentences, and would need to be generated based on the premises. For example, although the price targets are the same in (E6.4) and (E6.5), the meanings of these instances are different: this shows the necessity of exploring context-aware claim generation.

(E6.4) Revenue is expected to decline due to COVID-19, so we lowered our target price to 200.

(E6.5) We adjust our target price to 200, because we believe that the economy rebounds in the second half of the year.

Additionally, the structure and strategy of the resulting report may yield different influences on different readers. For example, Yang et al. [41] analyze the persuasion strategies of crowdfunding posts. This research direction may also be worth exploring when generating professional reports.

In this section, we use the workflow of professional analysts as an example. Every function (information collection, information verification, influence inference, and opinion formulation) could be a service that we provide to customers. For example, we could provide verified information to customers, or we could provide them with model predictions. These functions can be explored based on the concepts discussed here about financial opinion mining. We also illustrate the workflow in Fig. 6.4 based on the ideas proposed in Chap. 2. We suggest that future work follow the proposed steps and rationales to produce innovations in the information provision field.

6.2 Personalized Recommendation

Personalized recommendations are an important function in the next generation of banking, i.e., Bank 4.0 [23]. Neural network models and other advanced architectures yield significant improvements in recommendation. In particular, on platforms like e-commerce platforms that have access to a considerable amount of user data, performance has improved significantly. However, as e-commerce products are different from financial products, we face particular challenges when designing personalized recommendation systems for the financial domain. For example, whereas product prices on e-commerce platforms generally do not change constantly, those for financial instruments in financial markets typically do. Indeed, in the financial market, the prices of stocks, bonds, and options change every day; they can even change repeatedly in the space of a second. Also, on e-commerce platforms, product specifications generally stay the same; for instance, the iPhone 12 Pro uses the Apple A14 Bionic, and will not change to use the Apple A13 Bionic (in most cases). However, in the financial market, a company's operations may change every quarter. As companies are the underlying asset for financial instruments such as stocks and bonds, opinions about the iPhone 12 Pro may still be valuable after a year, whereas opinions about $AAPL are worthless after that same year.

Although some methods can be used for both e-commerce platforms and financial markets we must still account for the characteristics of the financial domain to improve performance. For example, just because someone mentions $AAPL does not mean that should we recommend $AAPL-related tweets to them. Instead, we should first seek to understand why they have mentioned $AAPL. For example, maybe they are interested in stocks that have attained a new 52-week high. In this case, instead of recommending $AAPL-related tweets to them, we should recommend other stocks that have made a new high. In previous work [6], we propose a task called next cashtag prediction, in which we attempt to predict cashtag(s)—that is, stock(s)—that the user will mention in the next five days. We present a tailor-made personalized recommendation method for financial social media platforms. As illustrated in Fig. 6.5, the proposed model uses three kinds of latent vectors:

- **User interest vectors**: The interests of the given user, captured from the tweets posted by the user.
- **Analysis vectors**: Background information on candidate cashtags derived from discussions (tweets) from other users.
- **Chart vectors**: Price data in the form of historical prices and volumes.

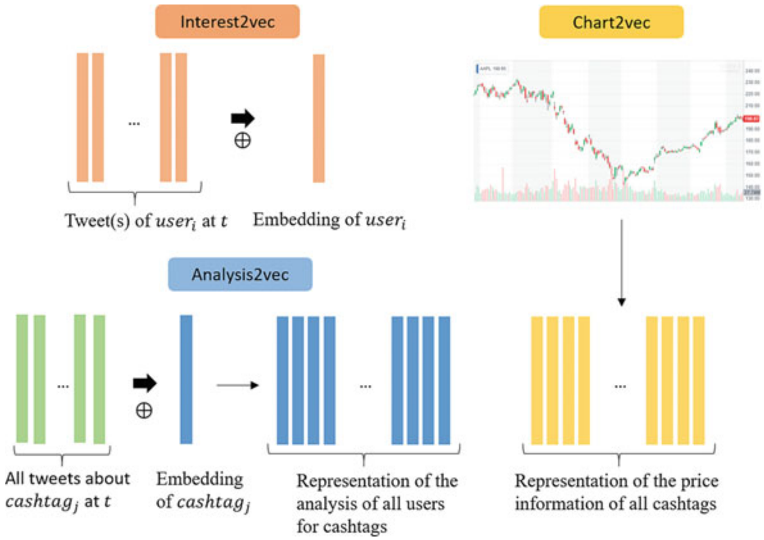


Fig. 6.5 Three kinds of latent vectors, including user interest vectors, analysis vectors, and char vectors

The proposed method achieves a 69.03% $hit@2$ when there are 30 candidate cashtags. This work shows a a direction for personalized investment suggestion.

The following work also provides insights.

- Insurance is also a financial product. Bi et al. [2] present a system for recommending insurance products to cold-start users. They employ user latent features from other domains for the insurance domain, showing the possibility of cross-domain features for financial applications.
- An ideal recommendation system proposes the best product to users based on the user's interests or budget, and also explains its decision. Chen et al. [9] discuss a similar scenario with data from an e-commerce platform. In their system they consider both personalized recommendation and explanation, which are also important in the financial domain. For example, the salesperson not only recommends a fund to the customer, but also explains why the recommended fund is suitable for the customer. In this case, the reason may simply be the salesperson's opinion.

Thus, the consensus of e-commerce-based studies is that customer opinions are essential elements to consider when producing personalized recommendations. This also applies to financial applications. In this section, we have laid out a rough outline of an application for financial opinion in recommendation systems. Previous work also shows that latent features of a given domain can be transferred to the financial domain. Additionally, we show why explaining decisions and recommendations are key functions for future work.

Table 6.1 Customer opinion and implicit relation to stock of credit-card-issuing bank in (E6.6)

Meaning	Example in (E6.6)	Implicit relation to market
Target entity	FlyGo	2887.TW
Market sentiment	–	Bearish
Sentiment	Negative	–
Opinion holder	Lisa	Lisa
Publishing time	2020/1/11	2020/1/11
Validity period of an opinion	–	–
Market information set	Cashback: 1%	Close price: 13.3
Analysis aspect	Cashback	Credit card services
Degree of sentiment	−0.8	−0.3
Set of claims	–	–
Set of premises	Cashback canceled	–
Opinion quality	Low	–
Influence power	Low	Low

6.3 Improving Employee Efficiency

In this section, we discuss how to apply techniques for financial opinion mining to improve employee efficiency in related industries. In previous chapters, we discussed financial opinions about investment and trading; these can be considered investor opinion. In the financial industry, services are important immaterial products, and the opinions on financial services are similar to those in the general domain. We take (E6.6) as an example, where FlyGo is a credit card.

(E6.6) Because the cashback of FlyGo was canceled, I cut it directly.

As shown in Table 6.1, the components defined in Chap. 2 can be used to analyze this opinion. Here, note that the customer's opinion may not provide claims for trading and investment. Thus, we can use positive/negative as the sentiment analysis in the general domain. This kind of opinion may also lack a validity period, because the cashback can change every year. In this case, we must also note the market information, i.e., the FlyGo contract. If the cashback changes in the following year, this negative opinion should not be considered for other users interested in FlyGo. To evaluate the quality of this opinion, we analyze the aspect and degree of sentiment and extract the argumentative units. Influence power in this case may be defined differently from that for an investor's opinion. If (E6.6) were posted by an opinion leader on social media platfroms, the market share of FlyGo could drop. This phenomenon also exists on e-commerce platforms, as we discuss in Sect. 4.3. Related work reviewed in Sect. 4.3 shows that customer opinions influence product sales. Table 6.1 lists information that may be related to the stock of the credit-card-issuing bank implied in (E6.6). This shows that customer opinions are also important in the financial domain.

Because customer service staff face customer opinions daily, customer service in the financial industry is another relevant topic. After extracting customer opinion, we attempt to detect their intent. For example, once we have determined that the target entity is related to credit card services, we could put the call through to the credit card coordinator. That is, we leverage the information we have extracted to detect the customer's intent. Moreover, if we discern that the customer is complaining about the low cashback, we could reduce customer churn by suggesting a better plan for the customer. We can also infer the reasoning behind for customer questions. For example, perhaps the customer asking about the cashback rate ratio of foreign spending is planning to travel overseas. In this case, we could encourage him/her to purchase travel insurance. These scenarios are common cases for financial institutions. Although few studies use data in the financial domain, experience from other domains could be adopted in the future. Below we mention some related work.

- Intent detection is domain-specific. The dataset and the taxonomy of intents should be tailor-made for different scenarios. Casanueva et al. [4] present a dataset containing 13,083 instances over 77 intents in the banking domain. They pre-train the sentence encoder on a conversation response selection task, and show that the proposed model is useful for intent detection. They also experiment with cross-domain intent detection datasets such as CLINCI50 [25] and HWU64 [29] to show the robustness of the proposed method.
- Identity fraud can be viewed as a kind of implicit intent. Wang et al. [37] propose an identity fraud detection framework. Their system asks questions drawn from the original personal knowledge graph, and further detects whether the responder is the correct user based on dialogue interactions. They conduct experiments with a simulated dataset and demonstrate promising pilot results.
- Selecting a proper response to the customer is also important. Wang et al. [38] experiment with debt collection. They select policies based on the dialogue state, and further choose the current state script. Their proposed two-state method outperforms a flow-based method for both single- and multi-round dialogues.

The above studies show that some methods can be used in several domains; intents specific to a certain domain, though, may still necessitate customization. One goal of this research direction is providing automatic customer services. One open issue in the financial domain is how to reply to customers based on their opinions. Given the development of current NLP methods, human-machine cooperation probably remains the most likely method for real-world applications.

Information extracted from various sources can be used to improve the working efficiency of employees. For example, in previous work [27], we proposed FinSense, a system that suggests stocks that are implicitly related to a given news article; a screenshot is shown in Fig. 6.6. When a news article is pasted into the box on the left, FinSense extracts the stocks mentioned in the article and lists them in the middle box. This function is provided for journalists to streamline their job, because they no longer need to provide labels after they complete the article. Nevertheless, they may

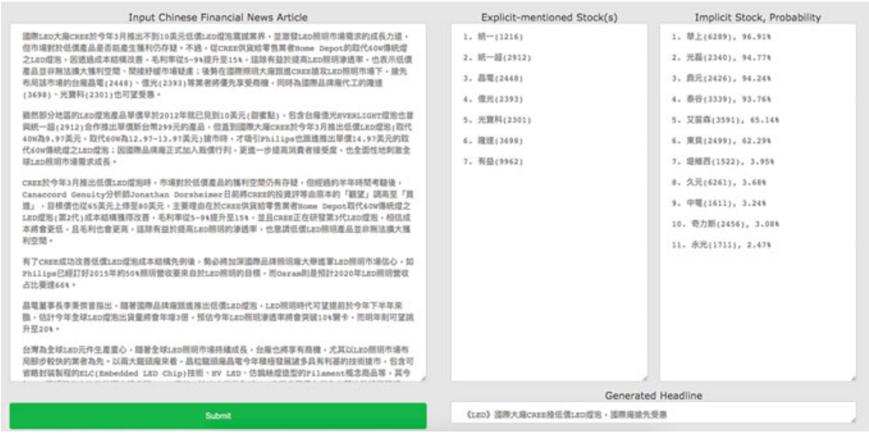


Fig. 6.6 Screenshot of FinSense, a system that suggests stocks that are implicitly related to a news article [27]

need to provide additional labels for implicit stocks, that is, stocks that are related but not explicitly mentioned. FinSense also recommends such stocks, based on the implicit information inference techniques mentioned in Sect. 4.3. Journalists must also compose a headline for the news article. FinSense suggests a headline based on the Transformer model [11]. This figure is thus one example of an application that uses financial opinion mining to improve employee efficiency. Although the recommended tags and headline may need some tweaking, the system does narrow down the journalist's choices.

In the financial domain, we also discuss another type of opinion: customer opinion. We go over scenarios that involve extracting the components of a customer's opinion, and discuss intent detection and dialog generation in customer services as potential applications. As an example, we show how implicit information inference can be used to streamline a journalist's job.

6.4 Summary

In this chapter we describe applications of financial opinion mining. Providing information to the customer is the primary purpose of many financial institutions. We provide a detailed discussion of the workflow of professional analysts, and present selected investment scenarios. In Sect. 6.3, we discuss personalized recommendations and domain-specific features. Various studies show the feasibility of transferring other domains' latent features to the financial field. We show how financial opinion components can be extracted and used to improve employee efficiency. We show relations between customer and investor opinions in the financial domain. Because customer opinions in the financial domain are similar to those in other fields, we believe that they can be leveraged using methods proposed for opinions in other domains. Hence in this book we focus on investor opinions. In the next chapter, we summarize the proposed research directions and show how to apply the results of financial opinion mining research to other domains.

References

1. Atanasova, P., Wright, D., Augenstein, I.: Generating label cohesive and well-formed adversarial claims. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (Online, Nov. 2020). Association for Computational Linguistics, pp. 3168–3177 (2020)
2. Bi, Y., Song, L., Yao, M., Wu, Z., Wang, J., Xiao, J.: DCDIR: a deep cross-domain recommendation system for cold start users in insurance domain. In: Proceedings of the Forty-Third International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1661–1664 (2020)
3. Carbonell, J., Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of the Twenty-First Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 335–336 (1998)
4. Casanueva, I., Temčinas, T., Gerz, D., Henderson, M., Vulić, I.: Efficient intent detection with dual sentence encoders. In: Proceedings of the Second Workshop on Natural Language Processing for Conversational AI (Online, July 2020). Association for Computational Linguistics, pp. 38–45 (2020)
5. Chen, C.-C., Huang, H.-H., Chen, H.-H.: Crowd view: Converting investors' opinions into indicators. IJCAI, pp. 6500–6502 (2019)
6. Chen, C.-C., Huang, H.-H., Chen, H.-H.: Next cashtag prediction on social trading platforms with auxiliary tasks. In: 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 525–527. IEEE (2019)
7. Chen, C.-C., Huang, H.-H., Takamura, H., Chen, H.-H.: Numeracy-600K: learning numeracy for detecting exaggerated information in market comments. In: Proceedings of the Fifty-Seventh Annual Meeting of the Association for Computational Linguistics (Florence, Italy, July 2019). Association for Computational Linguistics, pp. 6307–6313 (2019)
8. Chen, C.-C., Huang, H.-H., Tsai, C.-W., Chen, H.-H.: CrowdPT: summarizing crowd opinions as professional analyst. In: The World Wide Web Conference, pp. 3498–3502 (2019)
9. Chen, T., Yin, H., Ye, G., Huang, Z., Wang, Y., Wang, M.: Try this instead: personalized and interpretable substitute recommendation. In: Proceedings of the Forty-Third International ACM SIGIR Conference on Research and Development in Information Retrieval (New York, NY, USA, 2020), SIGIR '20. Association for Computing Machinery, pp. 891-900 (2020)
10. Cheng, D., Yang, F., Wang, X., Zhang, Y., Zhang, L.: Knowledge graph-based event embedding framework for financial quantitative investments. In: Proceedings of the Forty-Third International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2221–2230 (2020)

11. Chu, J., Chen, C.-C., Huang, H.-H., Chen, H.-H.: Learning to generate correct numeric values in news headlines. In: Companion Proceedings of the Web Conference 2020, pp. 17–18 (2020)
12. Da, Z., Huang, X.: Harnessing the wisdom of crowds. Manag. Sci. **66**(5), 1847–1867 (2020)
13. Ding, Q., Wu, S., Sun, H., Guo, J., Guo, J. Hierarchical multi-scale Gaussian transformer for stock movement prediction. In: Bessiere, C. (ed.) Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20 (7 2020). International Joint Conferences on Artificial Intelligence Organization, pp. 4640–4646. Special Track on AI in FinTech (2020)
14. Fabbri, A., Li, I., She, T., Li, S., Radev, D.: Multi-news: a large-scale multi-document summarization dataset and abstractive hierarchical model. In: Proceedings of the Fifty-Seventh Annual Meeting of the Association for Computational Linguistics (Florence, Italy, July 2019). Association for Computational Linguistics, pp. 1074–1084 (2019)
15. Gretz, S., Bilu, Y., Cohen-Karlik, E., Slonim, N.: The workweek is the best time to start a family – a study of GPT-2 based claim generation. In: Findings of the Association for Computational Linguistics: EMNLP 2020 (Online, Nov. 2020). Association for Computational Linguistics, pp. 528–544 (2020)
16. Hassan, N., Arslan, F., Li, C., Tremayne, M.: Toward automated fact-checking: detecting check-worthy factual claims by ClaimBuster. In: Proceedings of the Twenty-Third ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1803–1812 (2017)
17. Hidey, C., McKeown, K.: Fixed that for you: generating contrastive claims with semantic edits. 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) (Minneapolis, Minnesota, June 2019). Association for Computational Linguistics, pp. 1756–1767 (2019)
18. Holthausen, R.W., Larcker, D.F.: The prediction of stock returns using financial statement information. J. Account. Econ. **15**(2–3), 373–411 (1992)
19. Hsu, T.-W., Chen, C.-C., Huang, H.-H., Chang Chen, M., Chen, H.-H.: Hedging via opinion-based pair trading strategy. In: Companion Proceedings of the Web Conference 2020, pp. 69–70 (2020)
20. Hu, Z., Liu, W., Bian, J., Liu, X., Liu, T.-Y.: Listening to chaotic whispers: a deep learning framework for news-oriented stock trend prediction. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pp. 261–269 (2018)
21. Jame, R., Johnston, R., Markov, S., Wolfe, M.C.: The value of crowdsourced earnings forecasts. J. Account. Res. **54**(4), 1077–1110 (2016)
22. Jindal, N., Liu, B.: Review spam detection. In: Proceedings of the Sixteenth International Conference on World Wide Web, pp. 1189–1190 (2007)
23. King, B.: Bank 4.0: Banking Everywhere, Never at a Bank. Wiley, New York (2018)
24. Lahmiri, S.: Wavelet low-and high-frequency components as features for predicting stock prices with backpropagation neural networks. J. King Saud Univ.-Comput. Inf. Sci. **26**(2), 218–227 (2014)
25. Larson, S., Mahendran, A., Peper, J. J., Clarke, C., Lee, A., Hill, P., Kummerfeld, J. K., Leach, K., Laurenzano, M. A., Tang, L., Mars, J.: An evaluation dataset for intent classification and out-of-scope prediction. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the Ninth International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (Hong Kong, China, Nov. 2019). Association for Computational Linguistics, pp. 1311–1316 (2019)
26. Li, W., He, L., Zhuge, H.: Abstractive news summarization based on event semantic link network. In: Proceedings of COLING 2016, the Twenty-Sixth International Conference on Computational Linguistics: Technical Papers (Osaka, Japan, Dec. 2016). The COLING 2016 Organizing Committee, pp. 236–246 (2016)

27. Liou, Y.-T., Chen, C.-C., Tang, T.-H., Huang, H.-H., Chen, H.-H.: FinSense: an assistant system for financial journalists and investors. In: Proceedings of the 14th International Conference on Web Search and Data Mining (2021)
28. Liu, G., Mao, Y., Sun, Q., Huang, H., Gao, W., Li, X., Shen, J., Li, R., Wang, X.: Multi-scale two-way deep neural network for stock trend prediction. In: Bessiere, C. (ed.) Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20 (7 2020). International Joint Conferences on Artificial Intelligence Organization, pp. 4555–4561. Special Track on AI in FinTech (2020)
29. Liu, X., Eshghi, A., Swietojanski, P., Rieser, V.: Benchmarking natural language understanding services for building conversational agents. In: Tenth International Workshop on Spoken Dialogue Systems Technology (2019)
30. Loughran, T., McDonald, B.: When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. J. Financ. **66**(1), 35–65 (2011)
31. Ntakaris, A., Magris, M., Kanniainen, J., Gabbouj, M., Iosifidis, A.: Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods. J. Forecast. **37**(8), 852–866 (2018)
32. Qin, Y., Yang, Y.: What you say and how you say it matters: predicting stock volatility using verbal and vocal cues. In: Proceedings of the Fifty-Seventh Annual Meeting of the Association for Computational Linguistics (Florence, Italy, July 2019). Association for Computational Linguistics, pp. 390–401 (2019)
33. Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., Choi, Y.: Truth of varying shades: analyzing language in fake news and political fact-checking. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2931–2937 (2017)
34. Rayana, S., Akoglu, L.: Collective opinion spam detection: bridging review networks and metadata. In: Proceedings of the Twenty-First ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 985–994 (2015)
35. See, A., Liu, P. J., Manning, C. D.: Get to the point: summarization with pointer-generator networks. In: Proceedings of the Fifty-Fifth Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1073–1083 (2017)
36. Shapira, O., Ronen, H., Adler, M., Amsterdamer, Y., Bar-Ilan, J., Dagan, I.: Interactive abstractive summarization for event news tweets. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (Copenhagen, Denmark, Sept. 2017). Association for Computational Linguistics, pp. 109–114 (2017)
37. Wang, W., Zhang, J., Li, Q., Zong, C., Li, Z.: Are you for real? Detecting identity fraud via dialogue interactions. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the Ninth International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 1762–1771 (2019)
38. Wang, Z., Liu, J., Cui, H., Jin, C., Yang, M., Wang, Y., Li, X., Mao, R.: Two-stage behavior cloning for spoken dialogue system in debt collection. In: Bessiere, C. (ed.) Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20 (7 2020). International Joint Conferences on Artificial Intelligence Organization, pp. 4633–4639. Special Track on AI in FinTech (2020)
39. Wities, R., Shwartz, V., Stanovsky, G., Adler, M., Shapira, O., Upadhyay, S., Roth, D., Martínez-Cámara, E., Gurevych, I., Dagan, I.: A consolidated open knowledge representation for multiple texts. In: Proceedings of the Second Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics, pp. 12–24 (2017)
40. Xu, Y., Cohen, S. B.: Stock movement prediction from tweets and historical prices. In: Proceedings of the Fifty-Sixth Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Melbourne, Australia, July 2018). Association for Computational Linguistics, pp. 1970–1979 (2018)

41. Yang, D., Chen, J., Yang, Z., Jurafsky, D., Hovy, E.: Let's make your request more persuasive: modeling persuasive strategies via semi-supervised neural nets on crowdfunding platforms. 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) (Minneapolis, Minnesota, June 2019). Association for Computational Linguistics, pp. 3620–3630 (2019)

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

C.-C. Chen et al., *From Opinion Mining to Financial Argument Mining*,
SpringerBriefs in Computer Science,
https://doi.org/10.1007/978-981-16-2881-8_7

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 exaggregation. 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

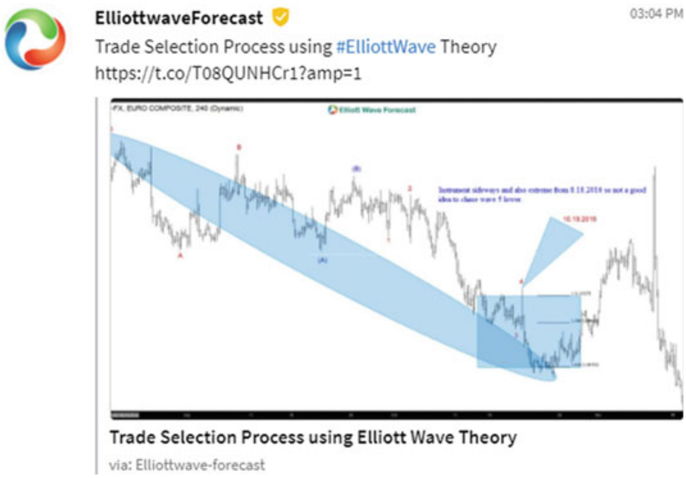


Fig. 7.1 Financial opinion expressed as an image

important ways to express financial opinions, especially on social media platforms. Figure 7.1[1] 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 7.2[2] 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.

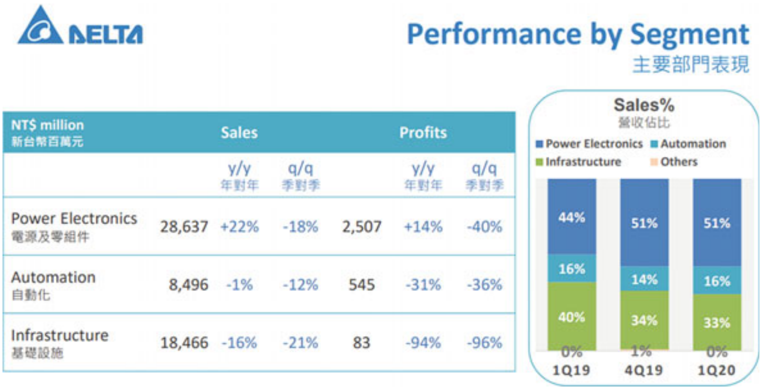


Fig. 7.2 A slide from an earnings conference call

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 (MWPs) [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.

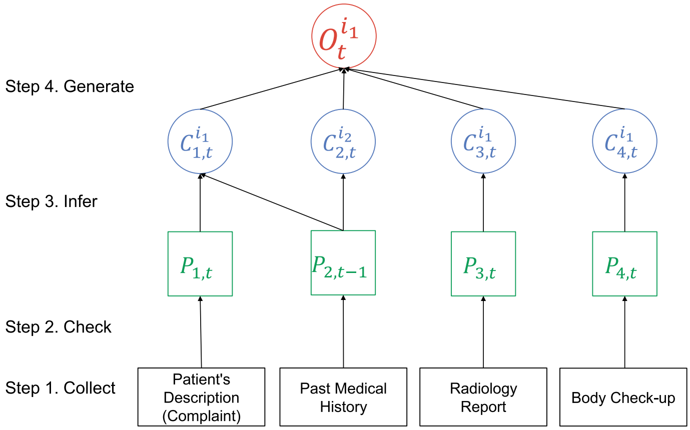


Fig. 7.3 Applying the workflow of argument mining in the clinical scenario

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