

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

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 Corpus¹.
- 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 agreements², 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 corpus³ 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/>.

⁵Tag 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 FrameNet⁷ [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

⁶https://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 unrealistic 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 corpus¹⁰ 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.

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

⁹<https://bitbucket.org/tuhinch/imho-naacI2019/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 dataset¹² 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/>.

¹²https://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 dataset¹³ 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 PolitiFact¹⁴ or Snopes¹⁵ 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

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

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

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

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

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

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 inferring 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.: Fortial@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](https://arxiv.org/abs/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](https://arxiv.org/abs/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](https://arxiv.org/abs/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

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

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

