

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.

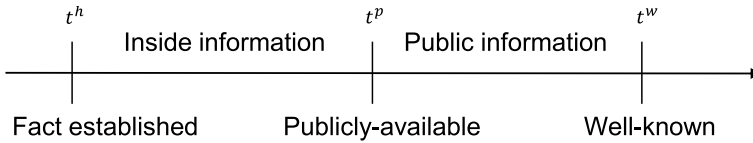


Fig. 3.1 The timeline from the establishment of a fact at t^h to that fact becoming public at t^p and then well-known at t^w

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](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 dataset⁶ 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 editorials⁹: 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

⁹For 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 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 GDP 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.: VoTAGE: 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., Scholder, 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

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.

