

Chapter 5

Numerals in Financial Narratives



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

5.1 Numeral Understanding

In Chap. 3, we identified the sources of financial opinion as insiders, professionals, social media users, and journalists. Table 5.1 lists the statistics of numerals in documents from these sources¹: numerals are common in all kinds of financial documents. Indeed, almost every news article contains at least one numeral. This indicates the importance of numeral information in financial narratives, and explains why we devote an entire chapter to this topic.

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

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

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

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

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

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

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

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

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

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

Table 5.3 A comprehensive taxonomy of financial numerals

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

narratives. Most managers describe the Quantities related to product sales, and many amateur investors talk about the Quantities of financial instruments they buy/sell.

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

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

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

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

5.2 Numeral Attachment

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

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

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

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

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

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

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

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

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

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

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

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

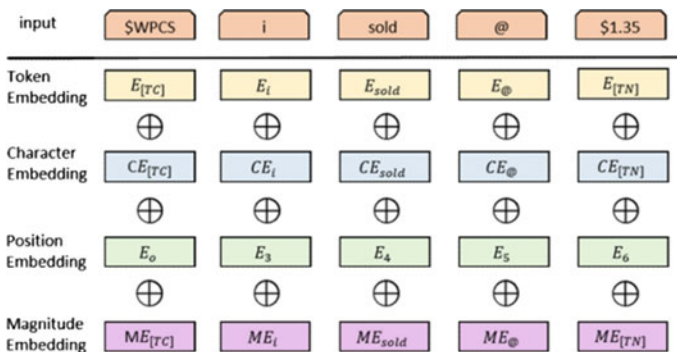


Fig. 5.1 Text representation for numeral-related tasks [4]

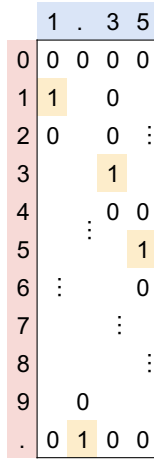


Fig. 5.2 Magnitude embedding [4]

Table 5.6 Ablation analysis of input representation [4]

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

Table 5.7 Ablation analysis for auxiliary tasks [4]

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

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

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

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

Table 5.8 Instances of general numeral attachment

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

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

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

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

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

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

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

APPLE (NASDAQ:AAPL) ANALYST RATINGS HISTORY

Show: ▾

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

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

5.3 Improving Financial Opinion Mining via Numeral-Related Tasks

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

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

Investors’ price targets go beyond bullish and bearish. A price target not only reveals the investor’s market sentiment but also shows what price level the investor expects to see in the future. Information vendors like Bloomberg and MarketBeat² collect price targets of professional analysts, and show this information in tabular form, as shown in Fig. 5.3, which attests the importance of this information. However, few platforms provide price targets of investors using social media platforms, even though these investors regularly discuss price targets. Models for numeral understanding and numeral attachment could be used to extract such information automatically from investors’ tweets to produce an overview similar to Fig. 5.3.

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

²<https://www.marketbeat.com/>.

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

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

Table 5.11 Results of three backtesting strategies [6]

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

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

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

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

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

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

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

Table 5.12 Performances of claim detection [5]

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

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

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

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

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

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

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

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

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

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

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

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

News Article:

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

Question Stem:

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

Answer Options:

- (A) Also increased by **0.04** percentage points from the previous month
- (B) The five major banks' newly-imposed mortgage interest rates climbed to **1.986%** in May.
- (C) Also approaching **2%** integer alert
- (D) Up to **2.5%**

Answer: (C)

We conduct experiments with four models.

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

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

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

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

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

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

5.4 Summary

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

References

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

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

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.

