Knowledge-Based Systems 69 (2014) 14–23

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

News impact on stock price return via sentiment analysis

Xiaodong Li [a], Haoran Xie [a,*], Li Chen [b], Jianping Wang [a], Xiaotie Deng [c]

[a] Department of Computer Science, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong Special Adminstrative Region
[b] Department of Computer Science, Hong Kong Baptist University, Kowloon, Hong Kong Special Adminstrative Region
[c] Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 31 October 2013
Received in revised form 15 April 2014
Accepted 15 April 2014
Available online 26 April 2014

Keywords:
News impact
Stock price return
Sentiment
Prediction
Experiment

ABSTRACT

Financial news articles are believed to have impacts on stock price return. Previous works model news pieces in *bag-of-words* space, which analyzes the latent relationship between word statistical patterns and stock price movements. However, news sentiment, which is an important ring on the chain of mapping from the word patterns to the price movements, is rarely touched. In this paper, we first implement a generic stock price prediction framework, and plug in six different models with different analyzing approaches. To take one step further, we use Harvard psychological dictionary and Loughran–McDonald financial sentiment dictionary to construct a sentiment space. Textual news articles are then quantitatively measured and projected onto the sentiment space. Instance labeling method is rigorously discussed and tested. We evaluate the models' prediction accuracy and empirically compare their performance at different market classification levels. Experiments are conducted on five years historical Hong Kong Stock Exchange prices and news articles. Results show that (1) at individual stock, sector and index levels, the models with sentiment analysis outperform the *bag-of-words* model in both validation set and independent testing set; (2) the models which use sentiment polarity cannot provide useful predictions; (3) there is a minor difference between the models using two different sentiment dictionaries.

1. Introduction

Stock market is an important and active part of nowadays financial market. Both investors and speculators in the market would like to make better profits by analyzing market information. Financial news articles, known as one major source of market information, are widely used and analyzed by investors. In *Big Data* era, the amount of news articles has been increasing tremendously. In front of such a big volume of news pieces, more and more institutions rely on the high processing power of modern computers for analysis. Predictions given by support systems could assist investors to filter noises and make wiser decisions. Therefore, how to model and analyze news articles so as to make more accurate predictions becomes an interesting problem.

Bag-of-words based approaches model news articles by vector space model which translates each news piece into a vector of word statistical measurements, such as the number of occurrences, etc. Machine learning models are then employed to capture the relationship between the word statistical patterns and the stock price movements. Although many *bag-of-words* based approaches have been reported to have prediction power in some previous works [14,41,42], they miss one important ring on the chain of the mapping from textual news articles to the final directional predictions, which is the sentiment of news. As shown in Fig. 1, one important step in the flow is that the news articles are first interpreted by investors and translated into market sentiments; the investors then make their decisions based on the sentiment interpretations; and market prices aggregate the actions of each investor and reflect them in the final price movements. Therefore, integrating the sentiment analysis into the prediction framework would become more critical.

Sentiment analysis models documents from sentiment dimensions. Instead of just measuring word frequency, each word (especially the *colorful* ones that have sentiment polarity) is decomposed and represented by a vector of sentiment features. For example, word "accelerate" can be represented by *strong* and *active* by using Harvard IV-4 psychological dictionary; using the same dictionary, word "accept" can be represented by *positive*, *submit* and *social relation* features. The number of the sentiment dimension is fixed in the dictionary. Each document can be represented by a vector of sentiment values by summing up the sentiment vectors of each word in the document. Making predictions based on the sentiment representation should have many advantages over *bag-of-words*:

* Corresponding author. Tel.: +852 68502340.
E-mail address: hrxie2@gmail.com (H. Xie).

http://dx.doi.org/10.1016/j.knosys.2014.04.022
0950-7051/

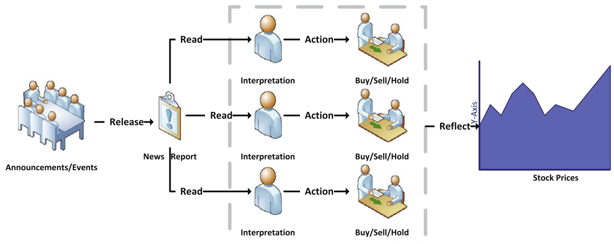

Fig. 1. The general scenario that news impact takes effect on the market prices. (1) Events happen; (2) events are reported; (3) reports are read by investors; (4) investors interpret the information according to their own knowledge; (5) investors take actions based on their interpretations, positions and budgets; and (6) various actions are translated into orders and reflected in stock price movements.

(1) **Reductive dimension**. Comparing with tens of thousands of words that are commonly observed while using words as the features, the sentiment representation largely reduces the dimensions to the order of hundreds, e.g., Harvard IV-4 psychological dictionary has 182 dimensions and Loughran–McDonald financial sentiment dictionary has 6 dimensions; (2) **Affective interpretation**. It is usually hard to interpret the mapping generated in *bag-of-words* space. In contrast, sentiment scores in different dimensions can sometimes give people more straightforward *feelings* about the document.

Although document sentiment analysis has been proposed and employed in many applications, e.g., recommender system [9], few work has been reported in the cross domain of computer science and algorithmic trading. In this paper, we first setup a generic framework that can take market information sets and plug in different customized prediction models. We then implement the prediction models using either sentiment analysis approach, *bag-of-words* approach or sentiment polarity approach, and compare their daily stock price return prediction accuracy on five years of Hong Kong Stock Exchange market data. In order to avoid any biases introduced by the framework, such as sentiment dictionary, the labeling method and the comparison at different market classification levels, etc., we (1) employ two sentiment dictionaries for comparison; (2) discuss and rigorously test different instance labeling methods and (3) compare models' performance at individual stock, sector and index levels in the experiment. The empirical results show that (1) at individual stock, sector and index levels, the models with sentiment analysis outperform the *bag-of-words* model in both validation and independent testing sets; (2) the models which use sentiment polarity cannot provide useful predictions; (3) there is a minor difference between the models using two different sentiment dictionaries.

The rest of the paper is organized as follows. In Section 2, we review the related work in both stock price prediction and sentiment analysis. In Section 3, we illustrate the work flow of the framework and do exploratory investigation on the data set and sentiment dictionary we use. In Section 4, we explain the setup of the experiment and show the parameter tuning related to the framework. And also, we give the experimental results and discussions. In Section 5, we provide our conclusion and future work directions.

2. Related work

The concept of *sentiment* has been known by people for a long time. It refers to *a specific view or notion* [12]. Sentiment analysis refers to *the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials* [16]. In this section, we review the related work about news impact and sentiment analysis in both finance and computer science domain.

2.1. News sentiment analysis in finance domain

The sentiment of news articles and their impacts on stock price returns have been studied in finance domain. Niederhoffer [27] analyzes *New York Times* and classifies 20 years of headlines into 19 predefined semantic categories from extreme-bad to extreme-good. He also analyzes how the markets react to the news of different categories and finds that markets have a tendency to overreact to bad news. Davis et al. [10] analyzes the effects of optimistic or pessimistic language used in news on firms' future performance. Their conclusion has two folds: (1) there is a bias between the readers' expectation and the writers' intension and (2) readers react strongly to both the content and the affective side of the reports which violate their expectations. Tetlock [36] extracts and quantifies the optimism and pessimism of *Wall Street Journal* reports, and observes that trading volume tends to increase after pessimism reports and high pessimism scored reports tend to be followed by a down trend and a reversion of market prices. Tetlock et al. also use Harvard IV-4 psychological dictionary in their work [37], where only positive and negative dimensions are exploited. They analyze the fraction of negative words in *Dow Jones News Service* and *Wall Street Journal* stories about S&P 500 firms from 1980 through 2004.

2.2. Bag-of-words approach in computer science domain

The *bag-of-words* approach has been applied to news impact analysis in financial market for many years. Seo et al. [35] build a TextMiner system (a multi-agent system for intelligent portfolio management), which could assess the risk associated with companies by analyzing news articles. Schumaker and Chen [34] build AZFinText system which is able to give directional forecast of prices based on financial news. *Bag-of-words* approach represents the textual news articles by term vectors and evaluates the "importance" of each term as their weights. After learning the mapping from the word statistical patterns to the outcome labels, the approach makes predictions for future unseen data.

2.3. Sentiment analysis in computer science domain

2.3.1. Sentiment dictionary construction

Applications evaluate word's sentiment mainly by constructing a sentiment dictionary. The construction approach could be briefly categorized as **semi-automatic** and **manual** which are described below:

Semi-automatic. The dictionary is first constructed by some *seed* words that are manually selected. The dictionary is then expanded from the seeds by the rules defined by application.
Manual. The dictionary is purely constructed by linguistic experts. This kind of dictionary is usually smaller in size than the one constructed semi-automatically, but more accurate.

Hatzivassiloglou and McKeown [17] make two hypotheses: (1) adjectives that are separated by "and" have the same polarity and (2) adjectives that are separated by "but" have opposite polarity. They use seed words and classify adjectives into positive and negative groups. Wiebe [40] also evaluates adjectives for polarity classification. He groups adjectives by word's tone and orientation clusters. Kim and Hovy [21] generate the polarity dictionary by using the WordNet to expand the selected seed words. They make two assumptions: (1) synonyms have the same polarity and (2) antonyms have the opposite polarity. The strength of a word polarity is measured by the number of the word's synonyms that are in the dictionary. Polarity of a word is considered as neutral when the word's strength is below a predefined threshold. Nasukawa and Yi [26] give more weights to local sentiment than global sentiment, since they think that global sentiment of a document is hard to evaluate and agree on by human evaluators. Similarly, Godbole et al. [15] also analyze local sentiments. General Inquirer[1] is a computer-assisted approach for content analysis of textual data. The Harvard IV-4 dictionary within the General Inquirer Augmented Spreadsheet contains more than 10,000 words and 182 sentiment dimensions. Loughran and McDonald [23] provide a manually made financial sentiment dictionary[2] which contains 6 sentiment dimensions.

Concept-level approach on sentiment mining has been a trend recently [7,18,19]. Tsai et al. [38] propagate sentiment values on the framework of ConceptNet.[3] They propose a two-step method to integrate random walk and iterative regression to build a concept-level dictionary. Xia et al. [43] analyze the domain adaptation problem and claim that SS-FE which uses both *labeling adaptation* and *instance adaptation* will help improve the overall performance. Cambria et al. [8,5] propose SenticNet which exploits *sentic computing* for concept-level sentiment analysis.[4] Poria et al. [31] enhance the SenticNet 1.0 by assigning an emotion label. We summarize the papers reviewed in Table 1.

2.3.2. Sentiment analysis application

Computer science researchers make their efforts in applying the sentiment analysis in real applications. Turney [39] builds a recommend/not-recommend binary classifier that is based on the sentiment orientation of product reviews. Pang et al. [30] employ naive bayes, maximum entropy classification, and support vector machines to classify movie reviews and achieve an accuracy of 83%. Pang and Lee [29] also identify which sentences in a movie review are of subjective character and improve their sentiment analysis. Mullen et al. [24] build a model combining unigram features and favorability measures. They claim that the model gives high prediction accuracy on the data set of Epinions.com reviews. Fukuhara et al. [13] extract topics from a text archive, and depict temporal trends of sentiments according to the associated time stamps. Devitt and Ahmad [11] propose a cohesion measure based on graph-based text representation to compute the polarity of financial news texts and their impact on financial markets. Abbasi et al. [1] use sentiment analysis to construct sentiment features (both stylistic and syntactic). The features are further selected by entropy weighted genetic algorithm. The final system is evaluated on the data of web forum postings in both English and Arabic languages. Bautin et al. [3] use IBM WebSphere Translation Server (WTS) to translate news written in 8 different languages into English and analyze the sentiment based on Lydia system. Willson et al. [25] argue that word sentiment should be determined in the context and they propose a method that distinguishes words' prior sentiment and contextual polarity based on annotated Multi-perspective Question Answering opinion corpus. With the development of micro-blog, such as Twitter and Weibo, people begin to consider Tweets as a source of customers' opinions. Pak et al. [28] build a classifier on Twitter corpus to classify Tweets into positive, negative and neutral categories. Zhang and Skiena [44] implement a simple market neutral trading strategy that exploits the trading signals generated by sentiment analysis on news and blogs. However, only two sentiment dimensions, i.e. positive and negative, are considered in stock recommendation system. Using multi-dimensional scaling and artificial neural networks, Cambria et al. [6] propose a cognitive model that could make multi-word expressions be organized better in a way of brain-like universe and natural language concepts. Huang et al. [20] evaluate the objective sentiment words in the SentiWordNet sentiment lexicon to improve the performance of word-of-mouth sentiment classifications.

Table 1
Sentiment dictionaries.

Paper	Construction approach	Dimensions
Hatzivassiloglou and McKeown [17]	Semi-automatic	2
Wiebe [40]	Semi-automatic	2
Kim and Hovy [21]	Semi-automatic	2
Nasukawa and Yi [26]	Semi-automatic	2
Godbole Srinivasaiah and Skiena [15]	Semi-automatic	2
Cambria et al. [8,5]	Semi-automatic	2
Harvard IV-4	Manual	182
Loughran and McDonald [23]	Manual	6

3. Sentiment analysis on news impact

3.1. Stock universe

The stocks we investigate are listed in Hong Kong Stock Exchange. As indicated by Zhang and Skiena [44], the correlation between the daily trading volume and the number of referenced news articles becomes larger when the stock's market capitalization is big, which further implies that news coverage on the actively-traded stocks is higher than less active ones. Therefore, in our work, we focus on the Hang Seng Index (HSI)[5] constituents which are big capitalization stocks and traded actively.

There are 4 sectors, i.e., Commerce, Finance, Properties and Utilities, with 50 stocks in HSI until year 2013. Due to the *tyranny of the index* effect[6] [33], we remove the stocks that were added after the starting date (2003-01-01) of our data set from the stock universe according to the change history of the constituents. After preprocessing, 22 stock symbols[7] are left in the universe, which are listed in Table 2 by their sectors.[8]

3.2. Data set

We use a news archive from FINET,[9] a major financial news vendor in Hong Kong. The news archive contains both company-specific and market related news from January 2003 to March 2008. All the news articles were written in English. Each piece of news is tagged with a time stamp showing the time the news is released, which helps classify news by dates. Companies that are related to the news are listed at the end of the news using their stock symbols, which helps establish the mapping from the news articles to stocks and vice versa.

For the stocks in the selected universe, we calculate the average number of their daily news articles, which is shown in Fig. 2. We also find the market weight assigned by HSI and the market capital of each stock, which are also drawn in Fig. 2. It can be calculated that the correlation between "Weight" and "No. of news" ($corr = 0.7990$) and the correlation between "Market Cap." and "No. of news" ($corr = 0.8484$) are high, which means as the weight

1 http://www.wjh.harvard.edu/~inquirer/Home.html.
2 http://www3.nd.edu/mcdonald/WordLists.html.
3 http://conceptnet5.media.mit.edu.
4 SenticNet 1.0 has "polarity" dimension, and SenticNet 2.0 and 3.0-beta have five sentiment dimensions, where "pleasantness", "attention", "sensitivity" and "aptitude" are added.

5 http://www.hsi.com.hk/HSI-Net/HSI-Net.
6 It is observed that during the first few months the price behaviors of the newly added constituents do not appear rational and they are usually mispriced.
7 Symbols are in the format "0000.HK", where "0000" stands for 4 digits.
8 Symbols with strikes are the stocks added during 2003-now.
9 http://www.finet.hk/mainsite/index.htm.

Table 2
Stock universe.

Sector	Commerce	Finance	Properties	Utilities
	0013.HK	0005.HK	0001.HK	0002.HK
	0019.HK	0011.HK	~~0004.HK~~	0003.HK
	~~0027.HK~~	0023.HK	0012.HK	0006.HK
	0066.HK	0388.HK	0016.HK	~~0836.HK~~
	~~0135.HK~~	~~0939.HK~~	~~0017.HK~~	
	~~0144.HK~~	~~1299.HK~~	~~0083.HK~~	
	~~0151.HK~~	~~1398.HK~~	0101.HK	
	0267.HK	~~2318.HK~~	~~0688.HK~~	
	0291.HK	2388.HK	~~1109.HK~~	
	0293.HK	~~2628.HK~~		
	~~0322.HK~~	~~3328.HK~~		
	~~0386.HK~~	~~3988.HK~~		
Symbol	0494.HK			
	~~0700.HK~~			
	0762.HK			
	~~0857.HK~~			
	~~0883.HK~~			
	0941.HK			
	~~0992.HK~~			
	~~1044.HK~~			
	~~1088.HK~~			
	1199.HK			
	~~1880.HK~~			
	~~1898.HK~~			
	~~1928.HK~~			
Total:	10	5	4	3

Table 3
Harvard IV-4 categories.

No.	Description
1	Positive vs. negative
2	"Osgood" semantic dimensions
3	Pleasure, pain, virtue and vice
4	Overstatement and understatement
5	Language of a particular "institution"
6	Roles, collectivities, rituals, and forms of interpersonal relations
7	Ascriptive social
8	Places, locations and routes
9	Objects
10	Communicating
11	Motivation-related
12	Process or change
13	Cognitive orientation
14	"I" vs. "we" vs. "you" orientation
15	"Yes", "No", negation and interjections

Table 4
Loughran-McDonald categories.

No.	Description	No. of words
1	Negative words	2349
2	Positive words	354
3	Uncertainty words	291
4	Litigious words	871
5	Modal words strong	19
6	Modal words weak	27

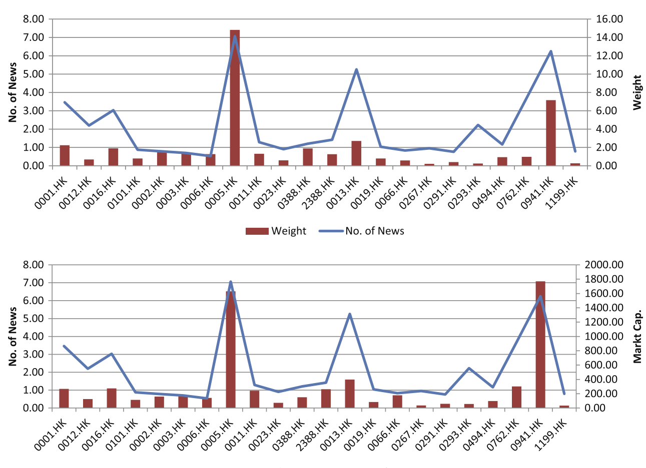

Fig. 2. Correlations between the average number of daily news pieces with stock weight and market capital respectively.

and the market capital of the stock become larger, the coverage of news on the stock will become bigger. This observation from our data sets is consistent with the results reported in Zhang and Skiena [44].

We obtain all the stocks' daily quotes (i.e., Open, High, Low, and Close prices) from Yahoo! Finance.[10] The daily OHLC data has the same period of data with the news. In Section 3.4, we will discuss about the preprocessing of the daily news and price data.

3.3. Sentiment dictionary

We use two manually constructed sentiment dictionaries, i.e., Harvard IV-4 sentiment dictionary[11] (HVD) and Loughran–McDonald financial sentiment dictionary (LMD). The manually constructed dictionaries are usually more accurate than the automatically constructed ones because of the careful selection by linguistic experts.[12]

HVD contains more than 10,000 words. The sentiment dimensions are categorized into 15 groups, which are described in Table 3. Each word is analyzed and projected onto 182 sentiment dimensions. LMD is constructed by Loughran and Bill McDonald [23]. We use the version 2012. This dictionary contains more than 3911 words and the detailed information about this dictionary is listed in Table 4.

3.4. The generic framework

Following the preprocessing steps in [22], we build up a generic framework to run our analysis processes, and the work flow is illustrated in Fig. 3. Section 3.4.1 shows the preprocessing of the prices and news articles. Section 3.4.2 shows how we map the news article onto sentiment space. Section 3.4.3 illustrates how we construct the instances and labels.

3.4.1. Preprocessing of daily prices and news

In the tuple of daily prices, we select *Open* and *Close* and considers daily *Open-to-Close* price return as a candidate of the prediction label, which is

$$R = \frac{Close - Open}{Open}. \quad (1)$$

The reasons why we choose *Open-to-Close* return instead of *Close-to-Close*[13] return have three folds from practitioner's point of view: (1) **Seasonality**. When considering the inter-day return, there

[10] http://finance.yahoo.com/.
[11] http://www.wjh.harvard.edu/ inquirer/homecat.htm.

[12] There have been some techniques that make progresses on improving the quality of the auto-generated dictionaries [2,32]. However, in this work, we mainly consider the manually generated dictionaries.
[13] Similar to *Open-to-Close* return, *Close-to-Close* return is the price return across two consecutive trading days, which is $R = \frac{Close_t - Close_{t-1}}{Close_{t-1}}$.

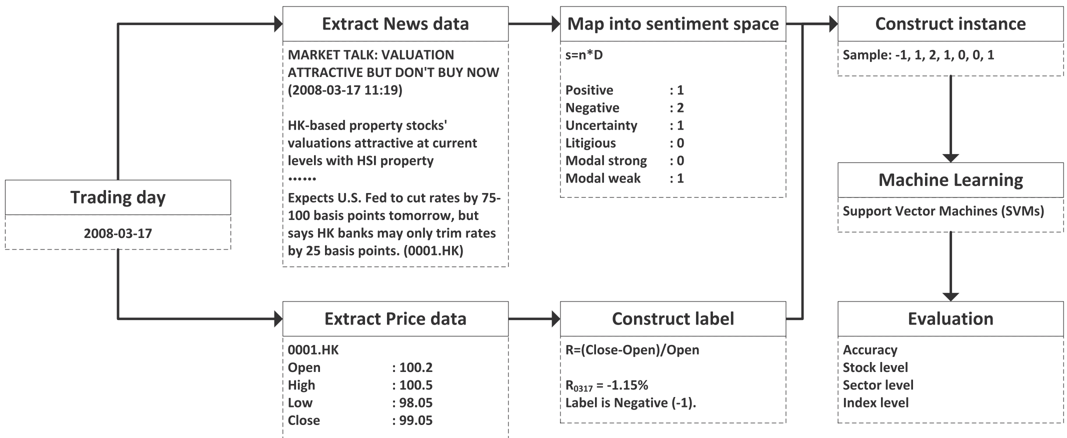

Fig. 3. The architecture of the generic framework.

is seasonality issue. For example, it is the busy trading season from every year's September to May next year, and less active during the summer vacation; the trading is more active at the beginning and the end of the week and less active in the middle; commodities are traded actively near the rolling days. These observed phenomena will bring seasonality patterns to the *Close-to-Close* return, and are hard to remove. (2) **Non-trading day gap**. The *Close-to-Close* return across weekends or holidays behaves different from the one in weekdays, since there is more time in the weekends and holidays, which brings risks to investors. To overcome the inventory carrying risk during the time gap, investors may turn stocks to cashes before the gap and re-enter market after the gap which has impacts on the *Close-to-Close* returns. (3) **$T+0$ market mechanism**. Hong Kong market uses $T+0$ instead of $T+N$.[14] Algorithmic traders would not like to carry the inventory after market close and take the overnight risk. Information has been reflected in the *Close* price by the intraday trading.

The daily news is first translated into a vector of terms. For each piece of news n_i, vector space model is applied and n_i is turned into a vector of term frequency values. Compiling all the news pieces together, we have a matrix N,

$$N = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_l \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1m} \\ t_{21} & t_{22} & \dots & t_{2m} \\ \vdots & & \ddots & \vdots \\ t_{l1} & t_{l2} & \dots & t_{lm} \end{bmatrix}, \tag{2}$$

where t stands for the term frequency, l is the index of the news and m is the index of the words. Among all the words that appear in the data set, we remove the stop words and keep the informational words. Since the dimension of the terms is large and not each dimension has non-zero value for every news, matrix N is large and sparse. Note that, for many stocks on some trading days, there are more than one piece of news. In this case, all the news about the same stock are first concatenated as one piece and then used for further preprocessing.

3.4.2. Mapping to sentiment space

We use matrix D to denote the sentiment dictionary. Each row in D corresponds to one word, and each column corresponds to one sentiment dimension. To map the matrix N to sentiment space S, we apply *inner-multiply* on matrix N and D, which is

$$S = N \odot D. \tag{3}$$

Since the columns in N do not match the rows in D, i.e., the words in N do not necessarily appear in D and words in D are not all used in news articles, we choose *inner-multiply* operator to mutually retrieve and match the columns and rows into two matrices.

3.4.3. Labeling

For each instance (row) in S, we label it with the return generated in Section 3.4.1. The daily *Open-to-Close* return is further divided into three categories by two thresholds which are symmetric around zero, denoted as th. And the prediction label is defined as follows:

$$L(x) = \begin{cases} \textit{positive} & \text{if } R(x) \geqslant th \\ \textit{neutral} & \text{if } -th < R(x) < th \\ \textit{negative} & \text{if } R(x) \leqslant -th \end{cases} \tag{4}$$

where x stands for a specific date. We will calibrate th and show the results under different th in Section 4.

3.4.4. Learning and evaluation

With the instances and the corresponding labels generated, we use support vector machines (SVMs), a successful text classification learning model employed by many researchers, to train on the preprocessed data sets.

We use classification accuracy (acc) to evaluate the prediction performance, which is defined as,

$$acc = \frac{t_{++} + t_{00} + t_{--}}{all}, \tag{5}$$

and

$$all = t_{++} + t_{00} + t_{--} + f_{0+} + f_{-+} + f_{+0} + f_{-0} + f_{+-} + f_{0-}, \tag{6}$$

where t_{++}, t_{00}, t_{--}, f_{0+}, f_{-+}, f_{+0}, f_{-0}, f_{+-} and f_{0-} are defined in Table 5.

3.5. A running case

In Table 6, we give a sample piece of news which is released on 2008-03-17, 11:19 am, talking about the asset valuation of

Table 5
The definition of terms in *acc*.

	Predict +	Predict 0	Predict −
True +	t_{++}	f_{+0}	f_{+-}
True 0	f_{0+}	t_{00}	f_{0-}
True −	f_{-+}	f_{-0}	t_{--}

[14] $T+0$ means that investors can buy and sell the same equity within the same trading day. In contrast, $T+N$ requires an extra N days to do the opposite trade action.

Table 6
A sample piece of news.

MARKET TALK: VALUATION ATTCIVE BUT DON'T BUY NOW (2008–03-17 11:19)
HK-based property stocks' valuations attractive at current levels with HSI property subindex down 4.9% at 25,763.53, says Castor Pang at Sun Hung Kai Financial, but advises investors not to buy now as sentiment remains uncertain. "The sector is generally trading at slight premium to par vs. their forecast NAVs, which is very attractive," but HK bourse is not going to be immune to more expected selloffs in US markets, he says. Tips HSI to test key 20,000 level in short term, now down 4.4% at 21,253.42. Expects US Fed to cut rates by 75–100 basis points tomorrow, but says HK banks may only trim rates by 25 basis points. CK (0001.HK) down 5.7% at HKD 99, NWD (0017.HK) down 8.9% at HKD 15.64 and Sino Land (0083.HK) down 8.7% at HKD 15.82

Table 7
News represented by a vector of term frequency value.

Term	Freq.	Term	Freq.	Term	Freq.
Market	2	Subindex	1	Expect	2
Talk	1	Down	5	Selloff	1
Valuation	2	Advise	1	Tip	1
Attractive	3	Investor	1	Test	1
But	4	Sentiment	1	Key	1
Not	3	Remain	1	Short	1
Buy	2	Uncertain	1	Term	1
Now	3	Sector	1	Cut	1
hk	3	Generally	1	Rate	2
Base	3	Trade	1	Point	2
Property	2	Slight	1	Tomorrow	1
Stock	1	Premium	1	Trim	1
Current	1	Forecast	1		
Level	2	Bourse	1		
hsi	2	Immune	1		

Table 8
News represented by a vector of sentiment value.

	Positive	Negative	Uncertainty	Litigious	Modal strong	Modal weak
Attractive	1	0	0	0	0	0
Uncertain	0	0	1	0	0	1
Short	0	1	0	0	0	0
Cut	0	1	0	0	0	0
s:	1	2	1	0	0	1

0001.HK. In this example, all the stop words are firstly removed, and we calculate the term frequency of each remaining word and list them in Table 7. The news is then represented by a vector of term frequencies, denoted by n. We use Loughran–McDonald financial sentiment dictionary which has 6 sentiment dimensions. We apply inner-multiply to n and D_{lm},

$$s = n \odot D_{lm}, \tag{7}$$

where vector s is the generated representation of news in sentiment space. The detailed information about s is shown in Table 8. The open price on 2008-03-17 is 100.2 and the close price is 99.05, so the *Open-to-Close* return is -1.15%. If we choose threshold th to be ±0.5%, then we label the news with *negative*. From the sentiment vector s, we could roughly observe that the *negative* dimension has larger weight ($weight = 2$) than the other dimensions ($weight \leq 1$). Although the observation is consistent with the true label generated from the prices and matches the interpretation of the news content, we cannot use this trivial approach to determine the news sentiment and make predictions just based on observations. In Section 4, we will present the analysis that uses SVMs to incorporate all the sentiment dimensions into generating predictions.

4. Experiment results and discussion

4.1. Experiment setup

In the experiment, we evaluate and compare six different approaches. We set up two approaches that use sentiment dictionaries, one approach that uses SenticNet and one approach that uses *bag-of-words*. Besides, we employ two traditional approaches that are based on the polarity asymmetry of the news sentiment. We illustrate the setup of each approach as follows.

We split the news and prices data sets into three parts along the time axis. The principles of the splitting method are (1) training data set should have relatively more trading days than validation and independent testing data sets; (2) validation data set should have almost the same trading days as independent testing set; (3) all the data sets should cover at least one trading year in order to reduce the yearly seasonality effect. We tried several splitting ways in our experiments, and find the results' change can be negligible. Therefore, we only report the result that uses the following splitting method: (1) from January 2003 to December 2005 is the training data set; (2) from January 2006 to December 2006 is the validation data set; and (3) from January 2007 to March 2008 is the independent testing data set.

During the training and validation phases, the parameters of SVMs are tuned, and the best combination of the parameters is kept for the independent testing phase. To be specific, we use RBF kernel SVMs. The parameters γ and C are tuned through a grid search method, which searches in a two-dimension space, $\gamma \in \{0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000\}$ and $C \in \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\}$, which in total $9 \times 10 = 90$ combinations.[15]

To evaluate the news impact and compare the prediction accuracy in both sentiment approach and non-sentiment approach, we set up two approaches using HVD and LMD respectively and one approach using SenticNet 3.0-beta (SN), and also a benchmark using only *bag-of-words* (BoW). In addition, Zhang and Skiena [44] proposed an approach based on sentiment polarity which is defined as

$$\text{Polarity} = \frac{positive - negative}{positive + negative}. \tag{8}$$

We also implement this approach using sentiment polarity (PO) except changing from binary classification to three-category classification, which is defined in Eq. (9),

$$\text{PO} = \begin{cases} positive & \text{if } \frac{f(pos) - f(neg)}{f(neg)} \geqslant th \\ negative & \text{if } \frac{f(neg) - f(pos)}{f(pos)} \geqslant th \\ neutral & \text{otherwise} \end{cases}, \tag{9}$$

Moreover, we consider the sentiment polarity issue in an adversarial environment, and add one more approach with the opposite scoring formula, which is,

$$\text{APO} = \begin{cases} positive & \text{if } \frac{f(neg) - f(pos)}{f(pos)} \geqslant th \\ negative & \text{if } \frac{f(pos) - f(neg)}{f(neg)} \geqslant th \\ neutral & \text{otherwise} \end{cases}. \tag{10}$$

4.2. Threshold in labeling method

As mentioned in Section 3.4.3, the *Open-to-Close* return is further changed to nominal label by comparing with threshold th. The determination of the th is explained in this section.

[15] It would be better not use cross-validation for time-series-like problem, which might bring *look-into-the-future* issue if the lagged correlation is high [4].

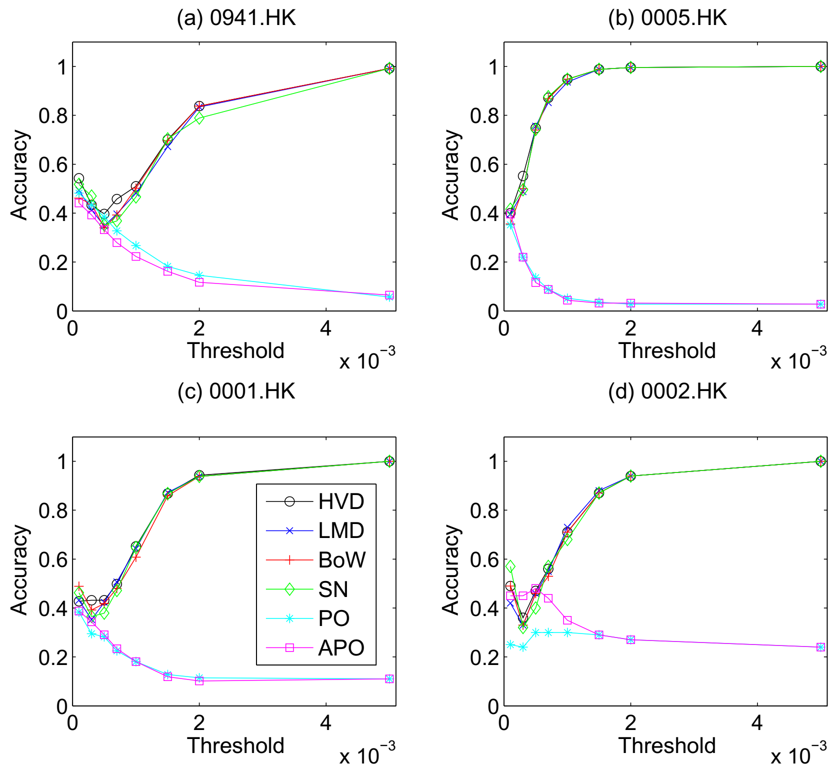

Fig. 4. Accuracy changes along with the threshold. (a) 0941.HK is from Commerce sector; (b) 0005.HK is from Finance sector; (c) 0001.HK is from Properties sector; (d) 0002.HK is from Utilities sector.

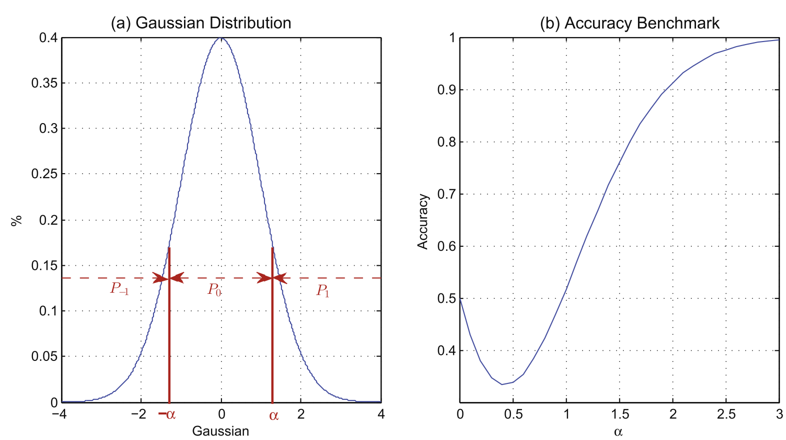

Fig. 5. Accuracy changes along with the threshold. (a) Gaussian distribution is assumed; (b) accuracy curve represents a up-side-down *Sine* function shape.

Different values of *th* are calibrated in the experiments. To illustrate our calibration logic, we first pick up four stocks which have the highest weight in each sector, and show their corresponding *acc* results along with the change of *th* in Fig. 4. We find that the curves of HVD, LMD, SN and BoW represent an up-side-down *Sine* function shape which consists of two parts: when the value of *th* is small, the curve represents a U-shape; when the value of *th* becomes great, the curve is similar to $y = \sqrt{x}$ with an upper bound 1.00. These patterns could be observed across different stocks.[16]

[16] We do not draw the curves for all the stocks, due to the limit of the paper length. The curve of 0005.HK (HSBC) seems not follow the pattern. That is because HSBC is a very liquid stock with small *bid-ask* spread, $th = 0.001$ is also big for the stock in comparison with the other ones. If we use *th* smaller than 0.001, the U-shape part will appear.

The reason for this observation is explained by Fig. 5. The distribution of stock price *Open-to-Close* return is assumed to be Gaussian distribution with *fat tail*,[17] as shown in Fig. 5a. Suppose we arbitrarily choose a value for *th* and the corresponding return value is denoted by α in the figure. Thus, the true label *negative*, *neutral* and *positive* have the following probabilities

$$P_{-1} = \int_{-\infty}^{-\alpha} \mathrm{pdf}_{Gaussian}(x)\,\mathrm{d}x, \tag{11}$$

$$P_0 = \int_{-\alpha}^{\alpha} \mathrm{pdf}_{Gaussian}(x)\,\mathrm{d}x, \tag{12}$$

$$P_{+1} = \int_{\alpha}^{+\infty} \mathrm{pdf}_{Gaussian}(x)\,\mathrm{d}x. \tag{13}$$

[17] In this explanation, *fat tail* could be neglected.

Table 9
Accuracy results in validation data set.

Symbol	HVD	LMD	BoW	SN	PO	APO	Symbol	HVD	LMD	BoW	SN	PO	APO
Commerce							Finance						
0013.HK	0.4327	0.4245	0.4408	0.4245	0.2898	0.2531	0005.HK	0.7480	0.7560	0.7400	0.7420	0.1360	0.1160
0019.HK	0.4421	0.5474	0.4421	0.4421	0.2737	0.2632	0011.HK	0.7400	0.8000	0.7400	0.8200	0.2700	0.2300
0066.HK	0.4035	0.3684	0.3684	0.4561	0.3860	0.2632	0023.HK	0.3776	0.3878	0.4184	0.4184	0.3571	0.3367
0267.HK	0.5000	0.4324	0.4324	0.4595	0.2973	0.2162	0388.HK	0.3089	0.3821	0.3659	0.3902	0.3333	0.3902
0291.HK	0.4677	0.4194	0.5000	0.5000	0.4032	0.3226	2388.HK	0.3953	0.3333	0.4186	0.4031	0.3488	0.3721
0293.HK	0.3810	0.4218	0.3673	0.4218	0.2177	0.4082							
0494.HK	0.4043	0.4043	0.4362	0.4787	0.4043	0.2872							
0762.HK	0.3395	0.3721	0.3767	0.3302	0.3721	0.2651							
0941.HK	0.3968	0.3482	0.3401	0.3522	0.3846	0.3320							
1199.HK	0.4568	0.4444	0.5185	0.4568	0.3210	0.4074							
Properties							Utilities						
0001.HK	0.4317	0.4229	0.4185	0.3789	0.2819	0.2907	0002.HK	0.4700	0.4600	0.4000	0.4141	0.3000	0.4800
0012.HK	0.4076	0.3567	0.3694	0.3694	0.3185	0.3185	0003.HK	0.3889	0.4333	0.4000	0.3533	0.3556	0.3333
0016.HK	0.4061	0.4670	0.4467	0.4162	0.2437	0.2741	0006.HK	0.4000	0.4462	0.4462	0.3997	0.2923	0.4154
0101.HK	0.4747	0.4949	0.4646	0.4848	0.3131	0.3333							

Given the label distribution without extra learning, people can conduct a random draw based on the *prior* distribution and make predictions. If the true label is the same as the random draw, the numerator of *acc* increases by 1. Therefore, the accuracy of this approach can be calculated by

$$acc = P_{-1}^2 + P_0^2 + P_{+1}^2. \tag{14}$$

Since α and $-\alpha$ are symmetric around 0,

$$P_{-1} = P_{+1} \tag{15}$$

$$P_0 = 1 - 2P_{+1} \tag{16}$$

substitute Eq. (15) and (16) into (14), and let $P_{+1} = P$, Eq. (14) becomes

$$acc = 6P^2 - 4P + 1. \tag{17}$$

If we draw the curve of *acc* along with the change of α, we get Fig. 5b. We can see that the curve in Fig. 5b has the up-side-down *Sine* function shape, with the minimum value at $P = \frac{1}{3}$, which is equivalent to flipping a fair coin with three outcomes for the prediction. On the other side, *acc* increases along with α on the right hand side of the minimum point, which means when the value of *th* is great, most of the samples are labeled with *neutral* and the probability of correct guess becomes high. In another word, the accuracy can be increased only by manipulating the labeling method without any change of the learning model.

PO and APO curves in Fig. 4 decrease as *th* increases, which is different from the other three approaches. This is because the predictions made by PO are determined by Eq. (9) without learning the *prior* label distribution in the training data set. As the *th* increases, the error rate on *negative* and *positive* classes increases, and the total *acc* decreases.

Based on the above explanations, the value of *th* cannot be too large as it will bring too much *acc* increment due to the effect explained in Fig. 5b. On the other hand, *th* cannot be too small, since *th* should be greater than the basic market transaction cost which is 0.003 (30 bps) in Hong Kong market. To balance both the constraints, we choose $th = 0.005$ (50 bps) which is near the minimum value point on the curve.

4.3. Individual stock level comparison

In Tables 9 and 10, we present the accuracy results in both validation and independent testing sets for all the stocks in the selected universe.

To compare the performance of each approach at stock level, we conduct an apple-to-apple comparison between each pair of approaches, and record the *wins vs. losses* in Tables 11 and 12. From the comparison in validation data set, we find that both HVD and LMD perform better than BoW (12 vs. 10, and 14 vs. 8, respectively). This result shows the performance improvement provided by changing from *bag-of-words* space to sentiment space. It can also be observed that HVD, LMD, SN and BoW perform much better than PO and APO approaches. This means only using the polarity of *pos* and *neg* sentiment dimensions cannot give good performance.

In the independent testing set, the performance of HVD, LMD, SN and BoW decreases a little bit. This is because the independent testing set is the hold-out unseen data to the models, so performance will usually be better on the validation data set than testing phase due to parameter calibration on the validation phase. It is also observed that LMD still performs better than BoW. The result that HVD, LMD, SN and BoW are better than PO and APO is still true in independent testing set.

4.4. Sector level comparison

Besides the comparison at stock level, we also conduct similar comparison at sector level. As shown in previous sections, the selected stocks come from 4 sectors. We calculate the prediction accuracy of each sector acc_{sector} by applying the market weight vector to the acc_{stock} vector, which is,

$$acc_{sector} = acc_{stock} \cdot weight_{sector}. \tag{18}$$

The results are shown in Tables 13 and 14. In the tables, numbers in bold font indicate the best performer, and numbers marked by underline denote the second best performer. We can observe that LMD achieves the best in 3 sectors in both validation and testing data sets; HVD has 1 best and 1 s best results in the validation set, and 2 s best in the testing set; SN has 3 s best in the validation set, and 1 s best in the testing set. At sector level, the performance results show a similar pattern to the stock level comparison, where LMD, HVD and SN outperform the other three approaches. In addition, we find that LMD performs better than HVD in 3 out of 4 sectors in both validation and testing data sets. The reason could be

Table 10
Accuracy results in independent testing data set.

Properties							Utilities						
Symbol	HVD	LMD	BoW	SN	PO	APO	Symbol	HVD	LMD	BoW	SN	PO	APO
0013.HK	0.3723	0.3628	0.4033	0.3461	0.2816	0.3793	0005.HK	0.6948	0.6968	0.6767	0.6929	0.2016	0.1895
0019.HK	0.4118	0.4902	0.4575	0.3856	0.3793	0.2759	0011.HK	0.5094	0.5472	0.5236	0.5660	0.3482	0.3929
0066.HK	0.4107	0.3839	0.3304	0.4196	0.3818	0.3455	0023.HK	0.3756	0.3503	0.3807	0.3503	0.3939	0.3030
0267.HK	0.4041	0.3630	0.3973	0.3904	0.3889	0.3333	0388.HK	0.3611	0.3785	0.3472	0.3750	0.3758	0.3576
0291.HK	0.3770	0.3525	0.4672	0.4426	0.3500	0.3500	2388.HK	0.3716	0.3257	0.3716	0.3807	0.3034	0.4045
0293.HK	0.3421	0.3872	0.3534	0.3759	0.3109	0.3361							
0494.HK	0.3764	0.3258	0.4045	0.4382	0.4048	0.3571							
0762.HK	0.3409	0.3750	0.3665	0.3523	0.4380	0.3504							
0941.HK	0.3644	0.3583	0.3381	0.3725	0.4170	0.3644							
1199.HK	0.4580	0.4733	0.4885	0.4504	0.4800	0.3400							
Properties							Utilities						
Symbol	HVD	LMD	BoW	SN	PO	APO	Symbol	HVD	LMD	BoW	SN	PO	APO
0001.HK	0.3883	0.3908	0.3714	0.3471	0.3622	0.3351	0002.HK	0.4037	0.4224	0.4348	0.3789	0.2459	0.3934
0012.HK	0.4028	0.3611	0.3715	0.3958	0.3740	0.2977	0003.HK	0.3457	0.4198	0.4012	0.3765	0.3056	0.3056
0016.HK	0.3462	0.3764	0.3874	0.4066	0.4251	0.3713	0006.HK	0.2963	0.3796	0.3611	0.3981	0.2558	0.4419
0101.HK	0.4673	0.4860	0.4766	0.4486	0.4087	0.2870							

Table 11
Stock level comparison in validation data set.

	HVD	LMD	BoW	SN	PO	APO
HVD	–	11 vs. 11	12 vs. 10	10 vs. 12	20 vs. 2	18 vs. 4
LMD	–	–	14 vs. 8	11 vs. 11	19 vs. 3	19 vs. 3
BoW	–	–	–	12 vs. 10	20 vs. 2	18 vs. 4
SN	–	–	–	–	20 vs. 2	20 vs. 2
PO	–	–	–	–	–	13 vs. 9
APO	–	–	–	–	–	–

Table 12
Stock level comparison in independent testing data set.

	HVD	LMD	BoW	SN	PO	APO
HVD	–	9 vs. 13	8 vs. 14	10 vs. 12	15 vs. 7	17 vs. 5
LMD	–	–	12 vs. 10	12 vs. 10	14 vs. 8	17 vs. 5
BoW	–	–	–	11 vs. 11	14 vs. 8	17 vs. 5
SN	–	–	–	–	15 vs. 7	18 vs. 4
PO	–	–	–	–	–	16 vs. 6
APO	–	–	–	–	–	–

Table 13
Sector level comparison in validation data set.

	HVD	LMD	BoW	SN	PO	APO
Commerce	**0.4086**	0.3915	0.3920	0.4000	0.3531	0.3154
Finance	0.6931	**0.7061**	0.6927	0.6976	0.1646	0.1532
Properties	0.4320	**0.4373**	0.4268	0.4063	0.2854	0.2996
Utilities	0.4080	**0.4442**	0.4134	0.4231	0.3147	0.3988

Table 14
Sector level comparison in independent testing data set.

	HVD	LMD	BoW	SN	PO	APO
Commerce	0.3758	0.3763	0.3776	0.3853	**0.3892**	0.3540
Finance	0.6483	**0.6509**	0.6318	0.6408	0.2245	0.2164
Properties	0.3976	**0.4030**	0.3956	0.3858	0.3857	0.3277
Utilities	0.3307	**0.4001**	0.3868	0.3877	0.2707	0.3884

Table 15
Index level comparison.

	HVD	LMD	BoW	SN	PO	APO
Validation	0.5892	**0.5976**	0.5858	0.5876	0.2230	0.2172
Independent testing	0.5460	**0.5527**	0.5391	0.5445	0.2789	0.2665

that LMD is designed specific for financial sentiment, and HVD is constructed for general purpose.

4.5. Index level comparison

Similar to the comparison at sector level, we also examine the performance of six approaches at the index level. We calculate the accuracy at index level by changing Eq. (18) to Eq. (19)

$$acc_{index} = acc_{stock} \cdot weight_{index}. \tag{19}$$

The results are shown in Table 15. Being consistent with the stock and sector level comparisons, LMD and HVD are the best among all six approaches (LMD achieves the best in both validation and testing sets; HVD achieves the second best in those two data sets). And also, LMD is better than HVD as observed in the sector level comparisons.

5. Conclusion and future work

Financial news articles are believed to have impacts on stock price return. In this paper, we analyze the news impact from sentiment dimensions. We first implement a generic stock price prediction framework. Secondly, we use Harvard psychological dictionary and Loughran–McDonald financial sentiment dictionary to construct the sentiment dimensions. News articles are then quantitatively projected onto sentiment space. We evaluate the models' prediction accuracy and empirically compare their performance at different market levels. To make it a fair comparison, instance labeling method is rigorously discussed and tested, and the threshold *th* is carefully chosen. Experiments, which are conducted on five years historical Hong Kong Stock Exchange prices and news articles, show that

1. Sentiment analysis does help improve the prediction accuracy. At stock, sector and index levels, the models with sentiment analysis outperform the *bag-of-words* model in both validation and independent testing data sets;
2. Simply focusing on positive and negative dimensions could not bring useful predictions. The models which use sentiment polarity do not perform well in all the tests;
3. There is a minor difference between the models using two different sentiment dictionaries.

For the future work, firstly, we consider shrinking the granularity of the prediction time horizon, which is to analyze the relationship between news impact and intra-day stock price return. As the trading speed becomes fast due to the adoption of algorithmic trading, capturing the intra-day news impact by sentiment analysis would be worth investigating. However, the intra-day noise is much higher than the daily level, which adds difficulty to the intra-day trading signal mining. Secondly, HVD and LMD are manually built in our current work. How to automatically expand the dictionaries as well as keeping the accuracy of the dictionaries will be an interesting direction to take in the future.

References

[1] A. Abbasi, H. Chen, A. Salem, Sentiment analysis in multiple languages: feature selection for opinion classification in web forums, ACM Trans. Inform. Syst. (TOIS) 26 (2008) 12.
[2] S. Bao, S. Xu, L. Zhang, R. Yan, Z. Su, D. Han, Y. Yu, Mining social emotions from affective text, IEEE Trans. Knowl. Data Eng. 24 (2012) 1658–1670.
[3] M. Bautin, L. Vijayarenu, S. Skiena, International sentiment analysis for news and blogs, in: Proceedings of the International Conference on Weblogs and Social Media, 2008.
[4] F. Bießmann, J.M. Papaioannou, M. Braun, A. Harth, Canonical trends: detecting trend setters in web data, in: International Conference on Machine Learning, 2012.
[5] E. Cambria, C. Havasi, A. Hussain, Senticnet 2: a semantic and affective resource for opinion mining and sentiment analysis, in: FLAIRS Conference, 2012, pp. 202–207.
[6] E. Cambria, T. Mazzocco, A. Hussain, Application of multi-dimensional scaling and artificial neural networks for biologically inspired opinion mining, Biol. Inspired Cogn. Architec. 4 (2013) 41–53.
[7] E. Cambria, B. Schuller, Y. Xia, C. Havasi, New avenues in opinion mining and sentiment analysis, IEEE Intell. Syst. (2013).
[8] E. Cambria, R. Speer, C. Havasi, A. Hussain, Senticnet: a publicly available semantic resource for opinion mining, Artif. Intell. (2010) 14–18.
[9] L. Chen, F. Wang, Preference-based clustering reviews for augmenting e-commerce recommendation, Knowl.-Based Syst. 50 (2013) 44–59.
[10] A.K. Davis, J.M. Piger, L.M. Sedor, Beyond the Numbers: An Analysis of Optimistic and Pessimistic Language in Earnings Press Releases, Federal Reserve Bank of St. Louis Working Paper Series, 2006.
[11] A. Devitt, K. Ahmad, Sentiment polarity identification in financial news: a cohesion-based approach, in: Annual Meeting of the Association of Computational Linguistics, 2007, p. 984.
[12] M.W. Dictionary, The Concept of Sentiment, 2013. <http://www.merriam-webster.com/dictionary/sentiment>.
[13] T. Fukuhara, H. Nakagawa, T. Nishida, Understanding sentiment of people from news articles: temporal sentiment analysis of social events, in: Proceedings of the International Conference on Weblogs and Social Media, 2007.
[14] G. Fung, J. Yu, W. Lam, News sensitive stock trend prediction, Adv. Knowl. Discov. Data Min. (2002) 481–493.
[15] N. Godbole, M. Srinivasaiah, S. Skiena, Large-scale sentiment analysis for news and blogs, in: Proceedings of the International Conference on Weblogs and Social Media, 2007.
[16] M. deHaaff, Sentiment Analysis, Hard but Worth It!, Customerthink, 2010.
[17] V. Hatzivassiloglou, K.R. McKeown, Predicting the semantic orientation of adjectives, in: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 1997, pp. 174–181.
[18] C. Havasi, E. Cambria, B. Schuller, B. Liu, H. Wang, Knowledge-based approaches to concept-level sentiment analysis, IEEE Intell. Syst. 28 (2013) 0012–14.
[19] C. Havasi, E. Cambria, B. Schuller, B. Liu, H. Wang, Statistical approaches to concept-level sentiment analysis, IEEE Intell. Syst. 28 (2013) 0006–9.
[20] C. Hung, H. Lin, Using objective words in sentiwordnet to improve sentiment classification for word of mouth, IEEE Intell. Syst. (2013).
[21] S.M. Kim, E. Hovy, Determining the sentiment of opinions, in: Proceedings of the 20th International Conference on Computational Linguistics, Association for Computational Linguistics, 2004, p. 1367.
[22] X. Li, C. Wang, J. Dong, F. Wang, X. Deng, S. Zhu, Improving stock market prediction by integrating both market news and stock prices, in: A. Hameurlain, S. Liddle, K.D. Schewe, X. Zhou (Eds.), Database and Expert Systems Applications of Lecture Notes in Computer Science, vol. 6861, Springer, Berlin/Heidelberg, 2011, pp. 279–293.
[23] T. Loughran, B. McDonald, When is a liability not a liability? Textual analysis, dictionaries, and 10-ks, J. Finance 66 (2011) 35–65.
[24] T. Mullen, N. Collier, Sentiment analysis using support vector machines with diverse information sources, in: Proceedings of EMNLP, 2004, pp. 412–418.
[25] T. Mullen, J. Wiebe, P. Hoffmann, Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis, Comput. Linguist. 35 (2009) 399–433.
[26] T. Nasukawa, J. Yi, Sentiment analysis: capturing favorability using natural language processing, in: Proceedings of the 2nd International Conference on Knowledge Capture, ACM, 2003, pp. 70–77.
[27] V. Niederhoffer, The analysis of world events and stock prices, J. Business 44 (1971) 193–219.
[28] A. Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, in: Proceedings of LREC, 2010.
[29] B. Pang, L. Lee, A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts, in: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2004, p. 271.
[30] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, in: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, vol. 10, 2002, pp. 79–86.
[31] S. Poria, A. Gelbukh, A. Hussain, D. Das, S. Bandyopadhyay, Enhanced senticnet with affective labels for concept-based opinion mining, IEEE Intell. Syst. (2013).
[32] Y. Rao, J. Lei, W. Liu, Q. Li, M. Chen, Building emotional dictionary for sentiment analysis of online news, World Wide Web (2013) 1–20.
[33] J. Ritter, Behavioral finance, Pacific-Basin Finance J. 11 (2003) 429–437.
[34] R. Schumaker, H. Chen, Textual analysis of stock market prediction using breaking financial news: the AZFin text system, ACM Trans. Inform. Syst. (TOIS) 27 (2009) 12.
[35] Y. Seo, J. Giampapa, K. Sycara, Financial News Analysis for Intelligent Portfolio Management, Robotics Institute, Carnegie Mellon University, 2004.
[36] P.C. Tetlock, Giving content to investor sentiment: the role of media in the stock market, J. Finance 62 (2007) 1139–1168.
[37] P.C. Tetlock, M. Saar-Tsechansky, S. Macskassy, More than words: quantifying language to measure firms' fundamentals, J. Finance 63 (2008) 1437–1467.
[38] A. Tsai, R.H. Tsai, J. Hsu, Building a concept-level sentiment dictionary based on commonsense knowledge, IEEE Intell. Syst. (2013).
[39] P.D. Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, in: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, 2002, pp. 417–424.
[40] J. Wiebe, Learning subjective adjectives from corpora, in: AAAI/IAAI, 2000, pp. 735–740.
[41] D. Wu, G. Fung, J. Yu, Z. Liu, Integrating multiple data sources for stock prediction, Web Inform. Syst. Eng. (WISE) 2008 (2008) 77–89.
[42] D. Wu, G. Fung, J. Yu, Q. Pan, Stock prediction: an event-driven approach based on bursty keywords, Frontiers Comput. Sci. China 3 (2009) 145–157.
[43] R. Xia, C. Zong, X. Hu, E. Cambria, Feature ensemble plus sample selection: a comprehensive approach to domain adaptation for sentiment classification, IEEE Intell. Syst. (2013).
[44] W. Zhang, S. Skiena, Trading strategies to exploit blog and news sentiment, in: Proceedings of the International Conference on Weblogs and Social Media, 2010.