Engineering Applications of Artificial Intelligence 85 (2019) 569–578

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Sentiment analysis on stock social media for stock price movement prediction☆

Ali Derakhshan, Hamid Beigy *

Sharif Intelligent Systems Laboratory, Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

ARTICLE INFO

Keywords:
Sentiment analysis
Opinion mining
Model-based opinion mining
Stock prediction

ABSTRACT

The opinions of other people are an essential piece of information for making informed decisions. With the increase in using the Internet, today the web becomes an excellent source of user's viewpoints in different domains. However, in one hand, the growing volume of opinionated text and on the other hand, complexity caused by contrast in user opinion, makes it almost impossible to read all of these reviews and make an informed decision. These requirements have encouraged a new line of research on mining user reviews, which is called *opinion mining*. User's viewpoints could change during the time, and this is an important issue for companies. One of the most challenging problems of opinion mining is *model-based opinion mining*, which aim to model the generation of words by modeling their probabilities. In this paper, we address the problem of model-based opinion mining by introducing a part-of-speech graphical model to extract user's opinions and test it in two different datasets in English and Persian where the Persian dataset is gathered in this paper from Iranian stock market social network. In the prediction of the stock market by this model, we achieved an accuracy better than methods that are using explicit sentiment labels for comments.

1. Introduction

The ability to predict stock prices is an essential issue concerning academia as well as business, but it is difficult to achieve a model that can do this. In the current stock markets, the feelings of stockholders whether positive or negative is an essential indicator of the future value of that stock. In recent years, the expansion of the internet and social networks make user opinions about the shares for various companies available in a large volume, and even social networks specifically for stockholders have emerged which people can express and discuss their ideas about future of each stock. Sentiment information for stocks in addition to historical prices of them could help to predict the future price of stocks better.

Stock prices are affected by many factors, including macroeconomics and various news. However, the focus of this research is solely on the feelings of users (using their comments). To achieve the best possible prediction model for stocks, we should aggregate all information about them including news and the company periodical reports, but our goal here is to achieve the best possible accuracy using only users opinions and comments in the social media.

To be able for extracting sentiments from these social networks we should do opinion mining in large quantities on the data which is a tough task, because texts in social networks are usually short, full of idioms, having unusual grammatical structures and so many other problems. Furthermore, literature in this context shows conflicting results in predicting the market. Although some recent researchers proclaimed weak to strong predictive capabilities (Nguyen et al., 2015; Bollen et al., 2011), earlier researchers concluded that mood information in social media has no predictive power (Antweiler and Frank, 2004; Tumarkin and Whitelaw, 2001). Still, Use of comments on social networks to predict stock prices have remained challenging.

The goal of this research is to develop a model using mood information in social media that could be used to predict stock market fluctuations (up or down) in the next day. In the proposed model, we use extracted features in one or two consecutive previous days and train a model in a supervised manner to predict the next day's prices.

One of our contributions is using part-of-speech in the LDA model separating words based on their part-of-speech tags. We call the proposed method *LDA-POS*, and this model achieved notable results on two datasets one in English and the other in Persian. The English dataset we used is from Nguyen paper and contains comments for 18 stocks for more than one year which is a massive dataset in this field of research where datasets are usually short (Nguyen et al., 2015). Another contribution of this paper is gathering a Persian dataset during this research which contains comments for five stocks in Iran's stock market

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.engappai.2019.07.002.

* Corresponding author.
E-mail addresses: aderakhshan@ce.sharif.edu (A. Derakhshan), beigy@sharif.edu (H. Beigy).

https://doi.org/10.1016/j.engappai.2019.07.002
Received 20 August 2018; Received in revised form 23 June 2019; Accepted 2 July 2019
Available online 29 July 2019

for about six months. The experimental results show the superiority of the proposed model in both datasets.

The remainder of the paper is organized as follows. Section 2 presents a literature review of the related research on predictive approaches. Section 3 describes both the English and Persian datasets in details including their statistics, their sources, and the way we used them in our task. Section 4 represents methods for stock movement prediction and Section 5 describes the evaluation methods we used and compares the proposed model to other related models. Section 6 concludes the paper and provides incentives for further work.

2. Related work

Predicting the stock market is an exciting field both for academics and industry. The central question in this area is that is it possible to predict stock price movements at all. Some of the researches were based on the random walk and Efficient Market Hypothesis (EMH) theory (Fama, 1991; Fama et al., 1969). They suggested that changes in stock prices are only because of news and events that could affect the costs and because it is not possible to predict events and news, it is impossible to predict stock price changes (Walczak, 2001). In practice, some researches were showing that stock prices do not follow random walk theory and they could be predicted in some degrees (Qian and Rasheed, 2007; Bollen et al., 2011; Vu et al., 2012). Some researches achieved descent accuracy results at 56% hit in predicting the direction of changes in stock prices (Schumaker and Chen, 2009a; Si et al., 2013; Tsibouris and Zeidenberg, 1995).

Further to discussed theories, there are two main philosophies in stock trading: fundamental analysis and technical analysis. In the fundamental analysis, the economic conditions of the company and economic indicators are used to assess the overall status of that company and to predict its stock prices. On the other hand, technical analysis is based on analyzing stocks price history during a time. Searching for repetitive patterns in prices to predict stock price fluctuations. Some researchers used only historical prices (Cervelló-Royo et al., 2015; Patel et al., 2015; Ticknor, 2013; Zuo and Kita, 2012a,b). To find patterns in the history of prices, many researchers implemented methods including Bayesian Networks (Zuo and Kita, 2012a,b), time-series and Auto Regressive models (Patel et al., 2015; Zuo and Kita, 2012a).

While the mentioned techniques did not use the sentiments on social media to predict the stock prices, it could be beneficial to include these data to improve prediction performance.

Most of the researchers in this field used only one stock (Bollen et al., 2011; Qian and Rasheed, 2007; Si et al., 2013), and in many of them, the number of test samples were insufficient (about 15 samples), which seems inadequate to reach a conclusion (Bollen et al., 2011; Vu et al., 2012). As we know, there is no research showing noteworthy results on several stocks in a long time. In this study, we used two datasets with more than 24 stocks in a relatively long interval.

Sentiment analysis has been widely used in product and restaurant reviews (Liu and Zhang, 2012; Pang and Lee, 2008). There have been some researchers trying to include textual data to improve stock market prediction. There are two primary sources of textual data that can be used in this task, the first one which is available for a longer time is economic news and the second one which is a new source is social media, especially social media that exclusively built for the stock market. These sentiments are aggregated in the model (Schumaker and Chen, 2009a,b; Sadeghi and Beigy, 2013).

The focus of sentiment analysis has changed over the time to aspect level opinion mining, while in many scenarios like product or restaurant reviews, it is crucial to separate aspects of different features and determine their polarity separately. Some initial researches in this domain, start their work base on the idea that most nouns could be aspects and their nearby adjective could contain the polarity for that aspect (Hu and Liu, 2004). While other prominent aspect-based opinion mining models, parse the sentences and extracted their syntax tree and find the coreferences of the nouns to address the problem successfully (Federici and Dragoni, 2016) .

Another approach to opinion mining is argumentation based opinion mining which uses argumentation theory to model and evaluates the pieces of information called arguments. An argument could support, contradict or explain a statement and by presenting the arguments relations in the form of a graph, one can make decisions based on them. A paper in this field introduced SMACk as argumentation based opinion mining framework that can analyze online social media texts (Dragoni et al., 2016). This framework is based on abstract bipolar argumentation theory that can extract relevant documents from textual documents (Dragoni et al., 2016). A recent paper in this field has tried to combine argumentation theory and natural language processing methods to find most debated arguments in online shopping framework and enable users to make more informed decisions (Dragoni et al., 2018). The introduced model combines Argumentation and Aspect-Based Opinion Mining (Dragoni et al., 2018).

Some researches in understanding emotions from the text emphasized on the broad applications of polarity detection in academia and also in industry (Cambria, 2016), they also suggest that although sentiment mining approaches are mainly using the bag-of-words model as the primary linguistic unit is a word in the first look, however determining specific and sentics needs multi-word expressions. Sentics, particularly determine the emotional information for real-world entities and are essential in sentiment polarity detection. This research highlights the importance of integrating semantic knowledge in addition to machine learning approaches. The writer also suggests that the next-generation models will include the commonsense knowledge data as well as brain-inspired reasoning methods (Cambria, 2016).

Based on a survey on applications of text mining in financial domain (Kumar and Ravi, 2016), about 70% of previous researches in this field have done using regular methods like decision trees, SVMs and regression analysis. Based on another recent survey the usage portion of regular methods is the same as other methods, as using complicated models gets poor performance in general (Xing et al., 2018b).

For predicting the market, most of the studies did not find the features extracted from the text sufficient, and they usually combined numeric economic data to their features or used ensemble methods whether in feature level or decision level to make their prediction robust (Xing et al., 2018b).

In recent years, by emerging deep neural networks, their usage in this field of research has dramatically increased. One of the prominent analyses has used Deep Belief Networks(DBN) in addition to Recurrent Neural Networks(RNN) to predict the market and has reduced the binary classification error rate to 40.05% from a baseline with 47.30% error rate (Yoshihara et al., 2015).

The most recent paper in this field has predicted the market well using ensemble learning of evolving clustering and LSTMs. This paper also emphasized that it is not sufficient for individuals to make investments solely based on public mood data as the public mood do not affect the market directly and other factors must also be considered (Xing et al., 2018a).

To predict stock market prices using twitter messages authors of Si et al. (2013) applied a non-parametric topic model. This model was a continuous Dirichlet Process Mixture(cDPM) to learn daily topics. Then, time series were created on the everyday topics. Using a non-parametric model and ability to estimate the number of topics automatically is the main advantage of this method. However, they used a small dataset, with messages only for three months.

3. Dataset

In this study we used two types of datasets as model inputs. The market social network datasets, which are used for comment analysis, and the market data that shows daily prices per share. We used the

Persian and English language social networks to explore comments and sentiments, where the users submit their comments for Iran and U.S. markets, respectively. Specifications of each dataset are described briefly in the rest of this section.

3.1. English dataset

This dataset was collected during a research on this field of study (Nguyen et al., 2015). The dataset consists of the English language users comments on 18 shares at Yahoo Message Board (Yahoo Finance Message Board), including their comments in a period of one year (from 32 July 2012 to 19 July 2013). The name of the stock companies and their stock abbreviation symbols are summarized in Table 1. We will use the stock summary symbols in the result tables. There are a total of 787,547 comments in the dataset. Some shares get more comments while the others get fewer comments per day. The number of comments per different shares in this data is also demonstrated in Table 1.

In order to prepare the training and testing sets, we divided the comment dataset into two time periods. For *AAPL* stock, 61 working days from 16 July 2012 to 01 Oct 2012 is used for the training set, and 83 working days from 12 Nov. 2012 to 13 March 2013 is used for the testing set. For all of the other stocks, the period from 23 July 2012 to 28 March 2013, containing 171 working days, is used for the training set and 78 working days from 01 April 2013 to 19 July 2013 is used for the testing set.

As mentioned in Nguyen et al. (2015), the comments for current transaction date is the set of all comments from 4 pm of previous transaction date to 4 pm of the current transaction date, because 4 pm is the when U.S. market gets closed.

Table 1 shows the statistics of English dataset stocks. For the *AAPL*, *Dell* and *KO* stocks the minimum number of messages in each transaction date is zero. It means that there are some transaction dates in this dataset which we have no messages for them. While we want to predict stock price movements using sentiments in users messages, having some days with no data in this dataset could add some random error to results for different methods. We want the results to be as accurate as possible (to reliably compare various methods), we decided to remove these three stocks from this dataset and use other 15 stocks with at least one message in a day, in our experiments.

Some previous studies have used Twitter as a major source of comments and sentiments to analyze market comments (Azar and Lo, 2016). However, this dataset has some advantages over Twitter which has inspired us to use it. Anyone can comment on any topic on Twitter. We can separate the market specific comments with hashtags and a series of distinguishing words, but there are still a large number of irrelevant comments on Twitter. We also do not have emotion labels on Twitter to use it as a basis for comparing the results(Human sentiment method). Although there are comments in the English social network datasets that are unrelated to the market, and even contradictory or false, they are much better than the comments on Twitter.

The U.S. stock market dataset contains the prices of stocks used in the prediction model as labels. The information in this dataset was extracted from the Yahoo finance website (Yahoo Finance). The U.S. stock markets are closed during weekends and major holidays but are open on weekdays. Daily stock prices are available for regular business days. There are daily opening prices, the highest and lowest prices, closing prices and adjusted closing prices. While we want to use a reasonable number of comments to predict stock prices for the next trading day, daily price fluctuation does not have importance to us; therefore, we use adjusted close prices per working day. U.S. markets use dollars for trading with the correct two decimal places (cents).

3.2. Persian dataset

This dataset is collected during this research. User comments gathered from SAHAMYAB website (Sahamyab stock twitter). There is a part in this website called tweets, where the users can post their comments on different stocks. The dataset contains user comments, spanning approximately six months (four months for train and two months for test). The name of the stock companies and their stock symbols are summarized in Table 2. We used these stock abbreviation symbols in the result tables.

There are a total of 21,205 comments from stock users in this dataset, among which 11,183 comments were directly sentiment labeled by users, indicating that a high percentage of comments are labeled emotions. The dataset's daily statics number of comments for different shares, given in Table 2, includes user comments on the shares mentioned above for 106 working days. In this dataset, 76 days from April 30, 2016 to August 28, 2016 are allocated to training data and 30 days from September 13, 2016 to November 02, 2016 are allocated to testing data.

The comments for the current transaction date in this dataset is the set of all comments from 12 pm of previous transaction date to 8 am of the current transaction date. Iran's stock market is open from 8 am to 12 pm in regular working days, hence for each transaction date we used users comments before the market gets opened, and the comments during the four open hours of the market (8 am to 12 pm) are discarded.

The comments in this dataset can have metadata like images and title, but we only used the comments texts. Also, we used only comments without sentiment labels for methods that are using the comments texts, and we used the labels in explicitly labeled comments in human sentiment method as a baseline.

Iran's stock market dataset contains the prices of stocks used in the prediction model as labels. The information in this dataset was extracted from Tehran Securities Exchange Technology Management Co. website (Tehran Securities Exchange). Iran's market is closed on weekends and other official holidays, with fewer days of activity compared to similar markets abroad. Daily opening prices, the highest and lowest prices, closing prices and final set prices per share can be accessed online on trading days. In this study, we use closing prices per share for each working day. The Iranian rial is the currency of Iran, with no decimal places.

3.3. The balancing of the labels

Our train datasets are not balanced, and the number of positive and negative samples in the training sets can be different. However, in the datasets we used, the range of ratio between positive and negative samples where about the same, so we are not highly unbalanced and the F-measure results also confirms this. In both our English and Persian datasets, the relative number of the labels in the test sets are also significant in addition to the provided statistics. Imagine that the positive labels (corresponding to the increase in prices) consist 90% of the labels, in this case, if a classifier always says that the prices will increase, it would be correct 90% of the times, so it is essential to investigate this property of the datasets. In our datasets, for the test set labels, we counted the positive and negative labels and calculated their relative abundance, after that, we calculated the average of the most frequent labels(always bigger than 0.5 in two class classification) over all the stocks. For the Persian dataset this indicator is equal to 54.00%, and for the English dataset, it is equal to 55.21%. This indicator is not a method to predict the price movements, while our classifier does not have access to test sets in the training phase and it is possible that for some stocks the most frequent class for test and train sets do not be the same.

Table 1
Statistics of the English dataset for each transaction date (Nguyen et al., 2015).

Stocks	Company name	The number of messages				Mean of the number of human sentiments
		Min	Median	Mean	Max	
AAPL	Apple Inc.	0	1093	1678	11 220	350
AMZN	Amazon.com Inc.	24	154	192	1 963	28
BA	The Boeing Company	46	173	203	1 053	16
BAC	Bank of America Corporation	94	282	343	1 366	49
CSCO	Cisco Systems Inc.	69	247	274	972	10
DELL	Dell Inc.	0	18	42	587	10
EBAY	eBay Inc.	1	17	29	267	3
ETFC	E Trade Financial Corporation	2	42	56	315	12
GOOG	Google Inc.	10	69	93	1 305	16
IBM	IBM Inc.	3	14	20	195	3
INTC	Intel Corporation	37	177	200	958	29
KO	The Coca-Cola Company	0	6	8	89	2
MSFT	Microsoft Corporation	27	139	172	815	53
NVDA	NVIDIA Corporation	10	65	80	410	11
ORCL	Oracle Corporation	5	67	79	372	6
T	AT&T Inc.	10	52	59	251	8
XOM	Exxon Mobil Corporation	10	37	44	202	4
YHOO	Yahoo! Inc.	22	121	141	860	27

Table 2
Statistics of the Persian dataset for each transaction date.

Stocks	Company name	The number of messages				Mean of the number of human sentiments
		Min	Median	Mean	Max	
KHZAMIA	Vehicle parts manufacturing	3	17	22	90	13
SHABENDAR	Bandar Abbas Oil Refining	4	28	42	323	21
SHAPNA	Esfahan Oil Refining	0	7	10	59	5
VNAFT	Oil, Gas industry investment	0	7	10	42	5
KHODRO	Iran Khodro Automobile	21	94	116	521	62

4. Methods for stock movement prediction

To predict stock prices, at first a set of features (using different methods) is extracted in a daily basis, then a Support Vector Machine (SVM) is used to predict the price movement by classifying the features in two categories up and down indicating increase and decrease in the stock price, respectively. As mentioned in Section 3, the English dataset used in this article is the same as the dataset used in Nguyen et al. (2015) and we want to compare our proposed method with methods presented in Nguyen et al. (2015). So we tried to keep the parameters and prediction model the same as the article mentioned to have a fair comparison. Therefore we chose the linear kernel for the SVM as the default kernel (same as the base article).

4.1. Price only method

In this method, we used only stock price history as our features to predict its price movement. This method is used as one of our base-line methods, and we want to see that by using only historical prices ow much is it possible to predict the market behavior.

For each transaction date, the price movement (up or down) for one previous date is shown by $price_{t-1}$ and for two previous dates by $price_{t-2}$. In this method $price_{t-1}$ and $price_{t-2}$ are the only features fed into the given classification model. Features used in each method are shown in Table 3, and the suffix En and Per denote the features for English and Persian dataset, respectively.

To follow the Ref. Nguyen et al. (2015), in our English dataset, we add *price only method* features to the feature set of other methods, but in our Persian dataset, we do not include features generated in Price only method to the feature set of other methods and we used only the features generated in each method only for that method. By this separation, we can determine and compare the predictive power of features in each method exclusively. When we want to create a practical stock prediction model, we can see each method as an expert and by aggregating each expert idea about price movement using ensemble learning methods we can build a model which incorporates all the available methods and outperforms all of them.

4.2. Human sentiment method

As we saw in Section 3, both English, and Persian datasets consist of some comments which users expressed their opinions. In some of the comments, users explicitly determined their optimistic or pessimistic overview of that stock by explicitly labeling their comment (*buy* label means that user believes that the price of that stock will rise and users must buy it and *sell* label means the opposite opinion). While we want to predict stock price movements using users opinions, this feature is the most prominent feature that shows the opinion of users about a specific stock.

In English dataset the labels are *strong sell*, *sell*, *hold*, *buy* and *strong buy* while in Persian dataset there are only two labels *buy* and *sell*.

In each transaction date, we count the number of messages with *buy* and *sell* labels in the Persian dataset. While we have only two labels here, by using the percentage of one label we can know the percentage of the other, so we use the percentage of *buy* for this dataset. For each transaction date, we calculate the percentage of *buy* label among explicitly labeled messages for current and the previous transaction dates and denote them by $HsentPer_{i,t}$ and $HsentPer_{i,t-1}$, respectively. Then we use these features in our prediction model. The features used for the English dataset is similar to what we have done for the Persian dataset. The percentage of each label for each transaction date is calculated, and these five percentages are the representative of the human sentiments in that day.

Features used in prediction model for this dataset are $HsentEn_{i,t}$ and $HsentEn_{i,t-1}$ in addition to $priceEn_{t-1}$ and $priceEn_{t-2}$. Table 3 shows the features used in Human sentiment method.

Table 3
Features of the prediction model.

Method	Features in English dataset	Features in Persian dataset
Price only method	$priceEn_{t-1}, priceEn_{t-2}$	$pricePer_{t-1}, pricePer_{t-2}$
Human sentiment	$HsentEn_{i,t}, HsentEn_{i,t-1}, priceEn_{t-1}, priceEn_{t-2}$	$HsentPer_{i,t}, HsentPer_{i,t-1}$
LDA-based method	$priceEn_{t-1}, priceEn_{t-2}, ldaEn_{i,t}, ldaEn_{i,t-1}$	$ldaPer_{i,t}$
Aspect-based sentiment	$priceEn_{t-1}, priceEn_{t-2}, Asent_{i;t}, Asent_{i;t-1}, I_{i,t}, I_{i,t-1}$	Unavailable
LDA-POS method	$PosAdjEn_{i,t}, PosNounEn_{i,t}, PosPrepEn_{i,t}, PosVerbEn_{i,t}$	$PosAdjPer_{i,t}, PosNounPer_{i,t}, PosPrepPer_{i,t}, PosVerbPer_{i,t}$
Neural network method	$ldaEn_{i,t}$	Unavailable

Table 4
Notations in LDA (Blei et al., 2003; Nguyen et al., 2015).

Notation	Definition
α, β	Hyperparameters
ϕ	The distribution over words
T	The number of topics
θ	The message specific topic distribution
z	A topic
w	A word in the message d
N_d	The number of words in the message d
D	The number of messages

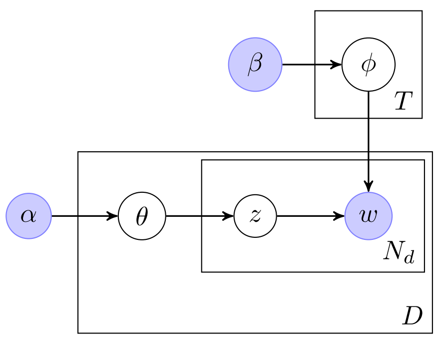

Fig. 1. Graphical model representation of LDA (Blei et al., 2003).

4.3. LDA-based method

This model is based on generative probabilistic model of Latent Dirichlet Allocation which considers every document(messages in our model) as a mixture of latent topics, and each topic is a probability distribution over words. We used this model as a baseline model and for comparison with Ref. Nguyen et al. (2015), we used the same parameter values. In Fig. 1, the graphical model representation of LDA are given and Table 4 shows the notations use in the LDA model.

For the English dataset, stop words are removed from the dataset, then the remaining words are lemmatized using Stanford CoreNLP (Manning et al., 2014). LDA is trained on the training dataset, then topics are inferred for messages in the test dataset. We used Gibbs sampling with 1000 iterations to infer the topics and choose 50 topics.[1] For each transaction date t in our test dataset, we fetch all messages and calculate the probability of each topic for every message and use their average as the input to the proposed prediction model.

For the Persian dataset, stop words are removed from the dataset, then the remaining words are lemmatized using JHazm library (A library developed specially for the Persian language). Because messages are short in this dataset, all messages in each transaction date are concatenated together, and we call them the document of that day. LDA is trained on the documents of the training dataset, then topics are inferred for documents in the test dataset. We used Gibbs sampling with 1000 iterations to infer the topics and choose 50 topics.[2] For each transaction date t in our test dataset, we calculated the probability of each topic for every document and used it as input (a vector of 50 probability numbers as we have 50 topics) to our prediction model.

Features used in prediction model in English dataset are $priceEn_{t-1}$, $priceEn_{t-2}, ldaEn_{i,t}$ and $ldaEn_{i,t-1}$ which subscripts t and $t-1$ indicate transaction date t and $t-1$ respectively. For Persian dataset the features used in prediction model is $ldaPer_{i,t}$. Table 3 shows the features used in LDA-based method.

4.4. Aspect-based sentiment

The aspect-based sentiment is the method proposed by Nguyen et al. (2015). We used the English dataset provided by this paper in our work, and we want to compare our proposed method results with it as well.

In this method, frequent consecutive nouns in all messages are detected (more than 10) in the dataset, then the sentiment polarity of these common consecutive nouns which are called topics is determined. To find the sentiment score of each topic, the opinion words of the sentences with that topic are detected using the SentiWordNet (opinion word list with their polarity Baccianella et al., 2010), and the topics are scored based on distance and polarity of each opinion word in a sentence containing that topic. The final score for each topic is the average score for that topic in all the sentences having that topic. To highlight the importance of more frequent topics, the percentage of messages with that topic is also included as a feature for this method.

The features used in the prediction model for this method are $Asent_{i;t}$, $Asent_{i;t-1}$, $I_{i,t}$, $I_{i,t-1}$, $priceEn_{t-1}$ and $priceEn_{t-2}$. $Asent$ denotes the polarity of topics and I indicates their corresponding importance in the prediction model. Subscripts t and $t-1$ indicate transaction date t and $t-1$, respectively. Table 3 shows the features used in Aspect-based sentiment and for the detailed explanation of this method you can refer to the Nguyen's paper (Nguyen et al., 2015).

To implement this method on the Persian dataset, we need an opinion word lexicon for the Persian language. Although there are some lexicons of opinion words and their polarity in Persian, their total lexicon size is about 1500 words (Dashtipour et al., 2016; Dehdarbehbahani et al., 2014), which is not comparable to SentiWordNet with more than 117 thousand entries. Because there are many synonyms to describe an opinion in Persian and detecting similarity and expanding these small Persian lexicons is out of this research focus, we decided not to implement this method in Persian. Lack of sufficient opinion words in the lexicon makes it impossible to score the topics in the sentences.

4.5. LDA-POS Method

In this section, we describe the proposed LDA-POS model. In the LDA-based method at first stop words are removed from the dataset, and then the distribution of different topics are assessed. In this method instead of removing stop words from the dataset, the part of speech of all words in the sentences are determined, then we create four categories of POS tags and the words with the same POS category tags are grouped as a document of that POS tag. Afterward, we assume each document with specific POS tag, a distribution of different topics

[1] We used mallet library to implement LDA, and in this library, English stop words have already been provided.

[2] We used Mallet library to implement LDA, and because mallet does not provide Persian stop words, we add the stop word from https://www.ranks.nl/stopwords/persian website which is the same source that English stop-words are acquired from in mallet.

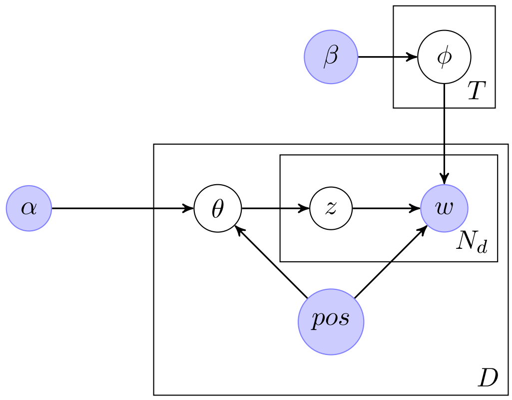

Fig. 2. Graphical model representation of LDA-POS.

Table 5
Notations in LDA-POS.

Symbol	Definition
α, β	Hyperparameters
ϕ	The distribution over words
T	The number of topics for all POS tags
θ	The document specific topic distribution
z	A topic
w	A word in the document d
N_d	The number of words in the document d
pos	The part of speech tag for the document
D	The number of documents

in that POS tag which each topic is a distribution of words. We used the training part of our datasets and some POS taggers(explained in the following paragraphs) to create documents of each POS category. Then we used these tagged documents to infer the topic distributions. Fig. 2 shows the graphical model representation of this method and the notations of this model are described in Table 5.

We implement this method on both English and Persian datasets. In each transaction date in the training days of the datasets, we determined the part-of-speech role of the words in the sentences of all messages (in the Persian language we used JHazm[3] (Nourian et al., 2015), and in the English language we used StanfordCoreNLP (Manning et al., 2014) for part-of-speech tagging and lemmatization). After that, we lemmatized the words; then we group the words with the same part-of-speech category in a document corresponding to that part-of-speech (Table 6 shows the part-of-speech categories and how different part-of-speech tags in the English and Persian taggers are mapped to them). For each transaction date in the train part of the datasets, we generate a document for each part-of-speech category, and in each part-of-speech category, we infer the topics using Gibbs sampling with 1000 number of iterations. We choose 50 topics for each part-of-speech category. Then the probability of each topic in different part-of-speech categories in the datasets for each transaction date is calculated. The feature generated by this method for each transaction date is a set of 50 topics probabilities for each *verb*, *adjective*, *preposition* and *noun* part-of-speech category summing up to a vector of 200 numbers.

While we separated the POS categories and created the feature vector with 200 dimensions for each transaction date containing the topics of all categories, some of the topics might have tiny probabilities for some stocks and therefore harm the performance of the SVM classifier. To prevent this, we perform a feature selection mechanism on the feature vector of each stock. We calculated the average of a feature in the training set, and if its square is bigger than 0.03, we kept that feature in the feature vector otherwise we simply ignored(omitted) that feature. The parameter 0.03 has been chosen experimentally. If it is bigger, it might clip some valuable feature, and if it is smaller, all the features will be selected and we find this amount appropriate. By using this preprocessing method, our results on English dataset improved over 1%, and the results on the Persian dataset remains the same.

Features used in prediction model for the English dataset are $priceEn_{t-1}$, $priceEn_{t-2}$, $PosAdjEn_{i,t}$, $PosNounEn_{i,t}$, $PosPrepEn_{i,t}$ and $PosVerbEn_{i,t}$ which the terms *Adj*, *Noun*, *Prep* and *Verb* indicate *adjective*, *noun*, *preposition* and *verb*, respectively in the topic probability of each part-of-speech category. Features used in prediction model for Persian datasets are $PosAdjPer_{i,t}$, $PosNounPer_{i,t}$, $PosPrepPer_{i,t}$ and $PosVerbPer_{i,t}$. Table 3 shows the features used in LDA-POS method.

4.6. Neural network method

Instead of using SVM with linear kernel we can use different classification methods to do this. Neural Network models gain much popularity because of the performance they show on various tasks, while deep models have too many parameters to learn and they usually need big training sets to train well, we used a two-layered neural network with one hidden layer. The activation function of neurons in the hidden layer is “tanh” and the output neuron activation function is “sigmoid”.

The inputs of this model(input layer) are the features generated by the LDA method. The number of topics in the LDA is 50, so the input layer size of this shallow neural network is equal to 50. For the output layer, we choose sigmoid as activation function so that we could see the output as the probability of a sample belonging to either of two classes 0 or 1 that corresponds to increase and decrease of the next day prices. If the output is bigger than 0.5, it belongs to class 1; else it belongs to class 0.

The number of hidden units are 40 in our implementation. We selected this number experimentally by testing different amounts and trying to maximize the results. Also, this amount seems a good heuristic. The error back propagation is used to update the weight of this model. The number of iteration is 10,000, and the learning rate is 0.1. To avoid overfitting, it is conventional to add regularization term(weight decay) to favor new weights, and it is proven to be helpful in practice. The regularization amount here is 0.995. Table 7 show the accuracy results for this model on the English dataset.

5. Experimental results

In this section, we use the computer experiments to evaluate the proposed model, and then we analyze the results.

5.1. Evaluation measure

Both Persian and English datasets are divided into two parts. One for training the prediction model and the other for testing. In English dataset, the period from July 23, 2012 to March 28, 2013 is for training and contains 171 transaction dates, and the period from April 1, 2013 to July 19, 2013 is for testing and contains 78 transaction dates as mentioned by Nguyen et al. (2015). In Persian dataset the period from April 30, 2016 to August 28, 2016 is used for training and contains 76 transaction dates, and the period from September 2, 2016 to November 2, 2016 is for testing and contains 30 transaction dates. While Iran stock market is closed on weekends and in normal working days, a stock could be frozen due to an administerial decision, the transaction days in our dataset might not be exactly the same for all stocks but in the mentioned period the number of transaction dates is the same.[4] We give each transaction date an up or down label based on the price increment or decrement compared to the previous transaction date. Accuracy is selected as the evaluation measure for this paper because it shows the proportion of true results in the test set and is suitable for our task and

[3] https://github.com/mojtaba-khallash/JHazm.

[4] Section 3.2 describes the Persian dataset in details and all transaction dates for each stock are available in dataset statistic files.

Table 6
List of word tags in each category.

Category name	POS Tag	POS tagger	On language	Explanation
Verb	VB	StanfordTagger	English	Verb, base form
Verb	VBD	StanfordTagger	English	Verb, past tense
Verb	VBG	StanfordTagger	English	Verb, gerund or present participle
Verb	VBN	StanfordTagger	English	Verb, past participle
Verb	VBP	StanfordTagger	English	Verb, non3rd person singular present
Verb	VBZ	StanfordTagger	English	Verb, 3rd person singular present
Verb	V	JHazm	Persian	Verb
Adjective	JJ	StanfordTagger	English	Adjective
Adjective	JJR	StanfordTagger	English	Adjective, comparative
Adjective	JJS	StanfordTagger	English	JJS Adjective, superlative
Adjective	RB	StanfordTagger	English	Adverb
Adjective	RBR	StanfordTagger	English	Adverb, comparative
Adjective	RBS	StanfordTagger	English	Adverb, superlative
Adjective	ADJ	JHazm	Persian	Adjective
Adjective	ADV	JHazm	Persian	Adverb
Preposition	CC	StanfordTagger	English	Coordinating conjunction
Preposition	DT	StanfordTagger	English	Determiner
Preposition	IN	StanfordTagger	English	Preposition or subordinating conjunction
Preposition	MD	StanfordTagger	English	Modal
Preposition	PDT	StanfordTagger	English	Predeterminer
Preposition	UH	StanfordTagger	English	Interjection
Preposition	PP	JHazm	Persian	Prepositional phrase
Preposition	PREP	JHazm	Persian	Preposition
Preposition	CONJ	JHazm	Persian	Conjunction
Preposition	PUNC	JHazm	Persian	Punctuation
Noun	NN	StanfordTagger	English	Noun, singular or mass
Noun	NNS	StanfordTagger	English	Noun, plural
Noun	NNP	StanfordTagger	English	Proper noun, singular
Noun	NNPS	StanfordTagger	English	Proper noun, plural
Noun	N	JHazm	Persian	Noun
Noun	NP	JHazm	Persian	Noun phrase

also to be able compare with the results given in Nguyen et al. (2015). Eq. (1) shows the *Accuracy* measure.

$$Accuracy = \frac{tp + tn}{tp + fp + fn + tn} \qquad (1)$$

where tp is the number of samples correctly categorized for positive samples, tn is the number of samples, correctly rejected for the negative samples, fp is the number of samples incorrectly categorized for the positive samples, and fn is the number of samples incorrectly rejected for the negative samples.

5.2. The results

Tables 7 and 8 show the accuracy results of different stocks for the English and Persian datasets. These tables also provide the average and variance of each method on all stocks. Using LDA-POS method we reached the overall average accuracy of 56.24% on 15 stocks in the English dataset and overall average accuracy of 55.33% on five stocks in the Persian dataset. As mentioned in Section 2 in some papers the accuracy of 56% is also reported as a pleasant result, but it should be considered that in most of them, the dataset is limited to a short period and their dataset usually consists of only one stock. In this paper, we implement our method on two datasets in different languages with a considerable number of stocks and a relatively protracted period (look at dataset details in Section 3). In some stocks, the accuracy of our method are outstanding. For example, VNAFT with 66.66%, YAHOO with 62.82% and so on.

The fundamental question in utilizing mood information in social networks to predict stock price movements is that whether this information has any predicting power. To answer this question, we compare the results of the human sentiment and the proposed LDA-POS method to the price only method. In the English dataset, the human sentiment method outperforms price only method by 0.26% and our LDA-POS method outperforms price only method by 2.38% and in Persian dataset, these improvement rates are 0.67% for human sentiment and 2.33% for our LDA-POS method. Based on these accuracy improvements we can infer that using mood information in social media can be helpful in predicting the stock market.

To be able to compare our results with the results reported in Nguyen et al. (2015), we add price only features to the feature vector of other methods in the prediction model. In Nguyen et al. (2015), the price only features were also concatenated to the feature vector of other methods in the English dataset(Table 3 show the features used in the prediction model for each method and dataset). We believe this approach is inefficient if not fallacious, so we do not include these features in the feature vector of different methods in the Persian dataset. If we add price only features to the LDA-POS feature vector in the Persian dataset the average accuracy results would be higher. Omitting other features from the feature vector of different methods help us to determine the predicting power of various methods precisely and therefore their comparison would be more meaningful. The features of price only method are just ones and negative ones, adding these features to the feature vector of other methods which most their elements are usually probability distribution with small amounts, makes the feature vector heterogeneous in it could reduce the performance of the prediction model. In order to aggregate features of different methods to improve the overall accuracy results, we can assume the outcome of each method as a belief of an expert and use techniques in ensemble learning to aggregate these beliefs and achieve an accuracy higher than all. In this paper, the focus is on proposing a method with the best possible accuracy, but in practical implementations integrating the results of different methods and approaches is necessary to achieve the best result.

The LDA-POS method outperforms human sentiment method by 2.12% in the English dataset and by 0.66% in the Persian dataset. Therefore we can claim that our method can capture sentiment information from messages in each transaction day in a way that it has more predicting power than explicitly labeled human sentiments. In the implementation of the LDA-POS method, we used only messages without human sentiment label, so we can apply LDA-POS method on messages in social networks that they do not have sentiment labels,

Table 7
Results of accuracies of 15 stocks.

Stocks	Baseline models				Our model		
	Price only	LDA-based method	Human sentiment	Aspect-based sentiment	LDA-POS method	Neural model on LDA	LDA-POS F-measure
AMZN	0.4605	0.5132	0.4868	0.7105	0.5769	0.3846	0.7179
BA	0.6316	0.5526	0.6053	0.5921	0.6154	0.5512	0.7619
BAC	0.5658	0.5526	0.5921	0.4474	0.5897	0.6153	0.7288
CSCO	0.5526	0.4737	0.4474	0.4605	0.5513	0.5384	0.7058
EBAY	0.5921	0.5658	0.4605	0.5789	0.5128	0.5384	0.6724
ETFC	0.5789	0.4868	0.5921	0.5526	0.5513	0.5512	0.7058
GOOG	0.5000	0.5658	0.5658	0.5263	0.5385	0.5512	0.6896
IBM	0.4868	0.5395	0.4737	0.5526	0.5513	0.4743	0.6956
INTC	0.4474	0.5000	0.4605	0.5263	0.5769	0.4102	0.7317
MSFT	0.5789	0.5526	0.6579	0.5263	0.5385	0.4743	0.6896
NVDA	0.6053	0.3947	0.5789	0.5395	0.5385	0.4871	0.6896
ORCL	0.4868	0.5921	0.5263	0.5395	0.5128	0.5000	0.6724
T	0.5526	0.5000	0.4737	0.5132	0.5513	0.5512	0.7008
XOM	0.4868	0.4342	0.6447	0.5395	0.6026	0.5769	0.7155
YAHOO	0.5526	0.5263	0.5526	0.5526	0.6282	0.5384	0.7289
AVERAGE	0.5386	0.5166	0.5412	0.5439	**0.5624**	**0.5162**	**0.7071**
VARIANCE	0.0032	0.0028	0.0050	0.0036	0.0012	0.0038	0.0006

Table 8
Results of accuracies of 5 Iranian market stocks.

Stocks	Baseline models			Our model	
	Price only	LDA-based method	Human sentiment	LDA-POS linear	LDA-POS F-measure
Khzamia	0.5333	0.4666	0.6000	0.5333	0.4166
Shabendar	0.5000	0.5333	0.5000	0.5000	0.5714
Shapna	0.5333	0.4666	0.5000	0.6000	0.6000
Vnaft	0.6333	0.4666	0.6300	0.6666	0.5833
Khodro	0.5000	0.4666	0.5000	0.4666	0.6363
AVERAGE	0.5400	0.4800	0.5467	**0.5533**	**0.5616**
VARIANCE	0.0030	0.0008	0.0042	0.0064	0.0072

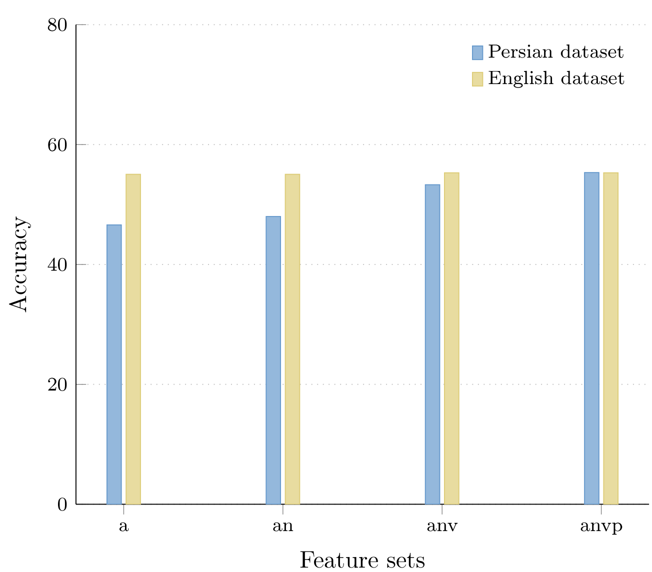

Fig. 3. Features used in the classification model. In this figure "a" stands for adjective, "n" stands for noun, "v" stands for verb, "p" stands for preposition, and their concatenation stands for their union.

and we can expect a better or comparable result than human annotated sentiments.

The LDA-POS method outperforms aspect-based sentiment, human sentiment and LDA-based method in English dataset by 1.85%, 2.12% and 4.58%, respectively. The proposed method also outperforms human sentiment and LDA-based method in Persian dataset by 0.66% and 7.33%, respectively.

The aspect-based sentiment needs a sentiment polarity lexicon in its implementation. The polarity and sentiment for words in various stocks and domains could be completely different, and use of a static sentiment polarity lexicon could harm the effectiveness of this method for several stocks. Also generating comprehensive sentiment polarity lexicon like the one used in this method for the English dataset is a grim task. Therefore implementation of this method on languages other than English depends on the availability of such lexicon. For example, in the Persian language, there are just extremely limited polarity lexicons with a maximum total size of 1500 words which makes implementation of aspect-based sentiment inefficient if not impossible. The advantage of the LDA-based method and LDA-POS method is that they do not need any further or language-dependent data to find different topic distributions and they have fewer limitations to be implemented in various languages. Another issue of aspect-based sentiment is its high variance. In the English dataset, the variance of this method is three times more than our LDA-POS method. Aspect-based sentiment works well on small number of stocks like AMZN and BA, but in some stocks, its accuracy is even less than 50%. The reason for this could be the domain sensitivity of the polarity lexicon, which makes use of aspect-based sentiment limited to only a few stocks. On the other hand in the LDA-POS method the accuracy are consistent, and for an unknown stock on an unknown language, use of LDA-POS method produces more trustable results.

The accuracy results for the LDA-based method were 52.63% in the English dataset and 48% in the Persian dataset. If we augment the price only features to feature vector of the LDA-based method in Persian, its accuracy would be equal to 52.66%. In the LDA-based method, the first step is removing the stop words from all document and then further procedures will be applied to what is remained to infer hidden topics. The LDA-based method looks at these remainder words as a bag of words, and it does not consider their part-of-speech roles in the sentences, While the meaning of each word can be assessed only in its context with having a specific part of speech. If we want to extract features to capture semantic changes in users opinions, it is better to seek this change in each part-of-speech. For example, to assess users opinions about a product or company, we should observe how the adjectives that users use change, how is the change in other categories of part-of-speech, to capture the general sentiment change accurately.

The LDA-POS method focus is on the part-of-speech role of words in the sentences, therefore at first in this method, the words with similar part-of-speech are grouped, and then we infer the topic distributions in each category. In the LDA-POS method, we divided the words in the datasets into four categories, which the detail of categories and implementation of this method is described in Section 4.5. In the Persian language, the way the words are used and their corresponding part-of-speech plays a crucial role in determining the meaning of a sentence, in such a level that a subtle change could alter the meaning of the whole sentence, while the English language is not as sensitive as Persian. To assess the importance of the features (topic distributions) for each part-of-speech in LDA-POS based method, we add the inferred features for each part-of-speech category one by one to the feature vector in both English and Persian dataset. Fig. 3 shows the changes in the accuracies while expanding the feature vector. As it can be observed in Fig. 3, the accuracy changes in the English dataset are negligible, while in the Persian language, in each step we can see a significant improvement, even when we added features for the preposition category (T_{prep}), a considerable increase in accuracy is observed, while in LDA-based method removing stop words almost removes all the prepositions in this language. In the English dataset, we could see no accuracy improvement by adding features from preposition category, and this shows that some part-of-speech roles do not play an essential role in the sentiment of the sentences in English and can be safely removed. These results illustrate the massive importance of part-of-speech roles in the Persian language.

The LDA-POS method outperforms the presented Neural Network model by 4.62%. The reason we used 2-layered neural network but not a deeper model is that neural networks have many parameters corresponding to the weights of their neurons which should be learned in the learning stage and that demands comparably more number of samples than simpler models to confidently learn their value. To predict the changes in the price well, the trend of the prices, social sentiments, and economic indicators are crucially important even more than their absolute value. To capture the patterns one could use recurrent neural models to obtain them, and the sentiments derived from the LDA-POS method could be used in that model which are reasonably beneficial based on their performance.

One of the weaknesses of the LDA-POS method is because it searches for topics in POS categories, so if the number of comments in a day or textual information were limited, this would prevent the method to capture the actual distribution of the topics. For example, if there are few comments in a day, consequently the number of words in the adjective category will be low, and this makes the method to overfit to the few samples the model observes in that day. In the datasets we used, on average there are plenty comments in each day, and this is not a problem in our case, but the limited number of days(samples) leads us to use simpler classification models.

Many factors are limiting the performance of our Persian dataset. The first restriction is because of the characteristics of the Persian language. In the formal written texts in this language, the vowels are omitted, so one word can have different pronunciations and different meanings. Also, Persian users usually write words based on their pronunciation and do not follow the formal language. So there is not even a universal slang dictionary for Persian, because in this manner a word could have many written forms which are not tractable.

The second factor hindering the performance comes from the scarcity of comments for different stocks. Not all the shares receive enough comments in daily basis to allow us to use them to predict daily price fluctuation, so we have a limited number of stocks to only which receive a reasonable amount of comments in regularly.

Further limitations come from the regulators of the market. In contrast to the global market, which trading the stocks of companies are most of the time possible; in Iran stock market, it is prevalent that the market administrator halts the exchange of a particular stock. It could last even for months due to new information or news. Afterward, they open it with a new price that could be dramatically different from the old price. To avoid these halt periods, we choose the time-line of our stocks carefully, and all the shares in our dataset are open to exchange during the dataset period. If we want to extend our dataset and add more working days to it, some of our stocks would be halted in the extended period. It is possible to extend one of our stocks to more elongated period, but we regard having the same time-line and number of samples for all the stocks a positive point for our dataset, and we do not want to make it fragmented.

6. Conclusion and future work

Recently predicting stock markets using machine learning methods has gained much attention. There are different approaches to satisfy this goal, including using sentiments in social media by assessing the human sentiments in user reviews. In this paper, we proposed a new method which incorporates part-of-speech tags into topic modeling methods, and we call our method "LDA-POS" method. The average accuracy results for this method on quite large datasets in both English and Persian languages reaches promising results of 56.24% and 55.33% respectively, and we outperform the related work which we used its English dataset. Usually, the sentiment extractions methods which perform flawlessly in the English language, do not perform well on the Persian language, while LDA-POS method results were similar in both languages. We show that some words with specific part-of-speech do not have significant in English and can usually be removed in the preprocessing step, while these words could have great significance in Persian. Also in this paper, we generated a dataset for the Persian language including five stocks, their user reviews and price movements which is a valuable resource and as our knowledge it is a first Persian stock dataset containing quite a protracted time.

There are some ideas to improve the proposed model and suggest them for future works. In our method, we select 50 topics, but in fact, the real number of topics for each stock could be different. We can solve this problem by using non-parametric methods or use some methods to guess the number of topics before applying the method. Also, we label price movements up and down. It is better to have more granular labels. Although the size of our Persian dataset seems to be sufficient, if we expand it to contain more stocks, the results derived from this dataset will be more reliable.

Furthermore, the general goal of research in this field is to predict the market and gave investors suggestions to buy specific stocks. To reach this goal, we should implement a market simulator which calculates the gain and loss by obeying the suggestions, and we should also integrate the suggestions (idea) of different methods together efficiently and produce suggestions which is more beneficial than each individual method.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions which improved the paper.

References

Antweiler, W., Frank, M.Z., 2004. Is all that talk just noise? the information content of internet stock message boards. J. Finance 59 (3), 1259–1294.

Azar, P., Lo, A.W., 2016. The wisdom of twitter crowds: Predicting stock market reactions to fomc meetings via twitter feeds. J. Portf. Manag..

Baccianella, S., Esuli, A., Sebastiani, F., 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Lrec, Vol. 10. pp. 2200–2204.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022, URL http://dl.acm.org/citation.cfm?id=944919.944937.

Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. J. Comput. Sci. 2 (1), 1–8.

Cambria, E., 2016. Affective computing and sentiment analysis. IEEE Intell. Syst. 31 (2), 102–107.

Cervelló-Royo, R., Guijarro, F., Michniuk, K., 2015. Stock market trading rule based on pattern recognition and technical analysis: Forecasting the djia index with intraday data. Expert Syst. Appl. 42 (14), 5963–5975.

Dashtipour, K., Hussain, A., Zhou, Q., Gelbukh, A., Hawalah, A.Y., Cambria, E., 2016. Persent: a freely available persian sentiment lexicon. In: International Conference on Brain Inspired Cognitive Systems. Springer, pp. 310–320.

Dehdarbehbahani, I., Shakery, A., Faili, H., 2014. Semi-supervised word polarity identification in resource-lean languages. Neural Netw. 58, 50–59.

Dragoni, M., da Costa Pereira, C., Tettamanzi, A.G., Villata, S., 2018. Combining argumentation and aspect-based opinion mining: The smack system. AI Commun. (Preprint), 1–21.

Dragoni, M., Pereira, C.D.C., Tettamanzi, A.G., Villata, S., 2016. Smack: An argumentation framework for opinion mining. In: International Joint Conference on Artificial Intelligence (IJCAI). IJCAI/AAAI Press, pp. 4242–4243.

Fama, E.F., 1991. Efficient capital markets: Ii. J. Financ. 46 (5), 1575–1617.

Fama, E.F., Fisher, L., Jensen, M.C., Roll, R., 1969. The adjustment of stock prices to new information. Int. Econ. Rev. 10 (1), 1–21.

Federici, M., Dragoni, M., 2016. A knowledge-based approach for aspect-based opinion mining. In: Semantic Web Evaluation Challenge. Springer, pp. 141–152.

Hu, M., Liu, B., 2004. Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 168–177.

Kumar, B.S., Ravi, V., 2016. A survey of the applications of text mining in financial domain. Knowl.-Based Syst. 114, 128–147.

Liu, B., Zhang, L., 2012. A survey of opinion mining and sentiment analysis. In: Mining Text Data. Springer, pp. 415–463.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, D., 2014. The stanford corenlp natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. pp. 55–60.

Nguyen, T.H., Shirai, K., Velcin, J., 2015. Sentiment analysis on social media for stock movement prediction. Expert Syst. Appl. 42 (24), 9603–9611.

Nourian, A., Rasooli, M.S., Imany, M., Faili, H., 2015. On the importance of ezafe construction in persian parsing. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Vol. 2. pp. 877–882.

Pang, B., Lee, L., 2008. Opinion Mining and Sentiment Analysis (Foundations and Trends (R) in Information Retrieval). Now Publishers Inc.

Patel, J., Shah, S., Thakkar, P., Kotecha, K., 2015. Predicting stock market index using fusion of machine learning techniques. Expert Syst. Appl. 42 (4), 2162–2172.

Qian, B., Rasheed, K., 2007. Stock market prediction with multiple classifiers. Appl. Intell. 26 (1), 25–33.

Sadeghi, S., Beigy, H., 2013. A new ensemble method for feature ranking in text mining. Int. J. Artif. Intell. Tools 22 (03), 1350010.

Sahamyab stock twitter. Oct 2016, https://www.sahamyab.com/stocktwits.

Schumaker, R.P., Chen, H., 2009a. A quantitative stock prediction system based on financial news. Inf. Process. Manage. 45 (5), 571–583.

Schumaker, R.P., Chen, H., 2009b. Textual analysis of stock market prediction using breaking financial news: The azfin text system. ACM Trans. Inf. Syst. (TOIS) 27 (2), 12.

Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., Deng, X., 2013. Exploiting topic based twitter sentiment for stock prediction. ACL (2) 2013, 24–29.

Tehran Securities Exchange Technology Management Co. Oct 2016, http://www.tsetmc.com/Loader.aspx?ParTree=15.

Ticknor, J.L., 2013. A bayesian regularized artificial neural network for stock market forecasting. Expert Syst. Appl. 40 (14), 5501–5506.

Tsibouris, G., Zeidenberg, M., 1995. Testing the efficient markets hypothesis with gradient descent algorithms. In: Neural Networks in the Capital Markets, Vol. 8. Wiley, Chichester, pp. 127–136.

Tumarkin, R., Whitelaw, R.F., 2001. News or noise? internet postings and stock prices. Financ. Anal. J. 57 (3), 41–51.

Vu, T.-T., Chang, S., Ha, Q.T., Collier, N., 2012. An experiment in integrating sentiment features for tech stock prediction in twitter. In: Proceedings of the Workshop on Information Extraction and Entity Analytics on Social Media Data. p. 2338.

Walczak, S., 2001. An empirical analysis of data requirements for financial forecasting with neural networks. J. Manag. Inf. Syst. 17 (4), 203–222.

Xing, F.Z., Cambria, E., Welsch, R.E., 2018a. Intelligent asset allocation via market sentiment views. IEEE Comput. Intell. Mag. 13 (4), 25–34.

Xing, F.Z., Cambria, E., Welsch, R.E., 2018b. Natural language based financial forecasting: a survey. Artif. Intell. Rev. 50 (1), 49–73.

Yahoo Finance. Oct 2016, https://finance.yahoo.com/.

Yahoo Finance Message Board. Oct 2016, https://finance.yahoo.com/quote/GE.

Yoshihara, A., Seki, K., Uehara, K., 2015. Leveraging temporal properties of news events for stock market prediction. Artif. Intell. Res. 5 (1), 103.

Zuo, Y., Kita, E., 2012a. Stock price forecast using Bayesian network. Expert Syst. Appl. 39 (8), 6729–6737.

Zuo, Y., Kita, E., 2012b. Up/down analysis of stock index by using Bayesian network. Eng. Manag. Res. 1 (2), 46.