

Web Based Sentiment Analysis for Stock Market

Ms. Vinisha K¹, Mrs. Mohana Priya S², Ms. Gajalakshmi S³, Ms. Gayathri D⁴

^{1,3,4} PG Scholar, ² Assistant Professor,

^{1,3,4} Dept of CSE, ² Dept of IT, PPG Institute of Technology, Coimbatore, 641035, India

Abstract- With the sudden growth of social media, sentiment analysis has developed fastly in recent years. Finding the opinion sites and monitoring them on the web is difficult task. So there is a need for automatic opinion discovery systems. Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. Sentimental analysis plays a major role in data mining and research. The goal of sentiment analysis is to find the sentiment polarity of texts or documents. This paper deals with finding sentiment polarity for stock market.

Keywords- Sentiment analysis, opinion mining, rule based, learning based.

I. INTRODUCTION

Sentiment analysis or opinion mining is the computational study of one's opinions, sentiments, and emotions from text. It is a language independent technology that understands the meaning of the text. Sentiment analysis uses NLP, statistics, or machine learning methods to extract, identify, or otherwise characterize the sentiment content of a text unit. For the past few years, it attracted the attentions from both academia and industry due to many challenging research problems and a wide range of applications. Opinions are important because whenever we want to take a decision we want to hear others' opinions. This is not only for individuals but also for organizations. In the past, when an individual needed to take a decision, he or she asks for opinions from friends and families.

With the sudden growth of social media, sentiment analysis has developed fastly in recent years. When an organization wanted to find opinions of the general public about its products and services, it conducted surveys and focus groups. People can now post reviews of products at merchant sites and express their views on almost anything like forums and blogs, and at social network sites. If anyone wants to buy a product, one is no longer limited to asking one's friends and families because there are many user reviews on the web. For a company, it may no longer need to conduct surveys or focus groups in order to gather consumer opinions about its products and those of its competitors because there is a plenty of such information publicly available.

The main issue of sentiment analysis is to identify how sentiments are expressed in texts. Also whether the expressions indicate positive or negative opinion towards the subject. Communication platforms, such as blogs, wikis, forums, and social-networking groups, have become a big data-mining source for the detection of public opinions. Sentiment analysis is important nowadays because for eg political parties may want to know whether people support their program or not, before investing in to a company one can leverage sentiment of the people for the company to find out where it stands, a company might want to find out the review of its products. It provides many benefits like monitor changes of opinion over time, lower cost and identify sources of feedback etc. The basic components of sentiment analysis are opinion holder- who is talking, object- item on which opinion is expressed, opinion- attitude or view of the opinion holder.

II. SYSTEM ANALYSIS

Existing work on sentiment analysis can be grouped into two main categories: rule and learning based approaches. Rule-based approach often needs an expert-defined dictionary of subjective words; this approach predicts the polarity of a sentence or document by analyzing the occurring patterns of such words in text [3]. Yuen et al. proposed an approach that derives the semantic polarity of words on the basis of morphemes [4]. In learning-based approaches, Hu and Liu [7] developed an approach to extract opinion features from product reviews based on linguistic patterns called class sequential rules, which can be derived from a group of labeled training sequences of words and part-of-speech tags.

A. Disadvantages

- Large training data sets with positive and negative examples are needed, which are often costly and time consuming
- Sentiment polarity results cannot be as exact as expected if the context of the text is not considered

III. PROPOSED WORK

Our proposed method comprises two stages: the semantic orientation and rule based method using latent semantic analysis. First is to extract the semantic words based on the learning algorithm, through labeling of subject and objects. Then second is to classify the opinions based on lexicons like adverbs and adjectives in the sentence. This paper deals with finding sentiment polarity for stock market. Experts would have given their comments for companies so that before investing in to a company one can leverage sentiment of the people for the company to find out where it stand. The technique used is parts of speech. The words in the text are tagged using a POS tagger so that it assigns label to each word. Then we extract the sentiment words using SND patterns. Then the average sentiment orientation of all phrases we gathered is computed. Finally sentiment polarity for each and every comment can be found. Experimental analysis proves the precision and recall of the opinion mining through sentimental analysis yields better results compared with state of approaches.

A. Advantages

- Learning-based approach is that it doesn't want expert knowledge to build the related bases
- Better accuracy

IV. ARCHITECTURE OF SENTIMENT ANALYSIS

A rule based approach is used in the stock sentiment analysis. The components of architectures are 1) collection of web data 2) data preprocessing 3) extraction of subjects and objects 4) developing sentimental analysis 5) sentiment calculation 6) sentiment polarity.

A. Collection of web data

Information regarding the share market on the web can be classified into three different categories. The first category consists of news, expert statement, discussions, etc., from the share market website. The second includes posts from the stock market sector in forums. These forums provide a platform through which users can exchange information about share market. The third includes real time information about stock in micro blogging, which can be found from the social media and ensuring that the conclusions are definitely based on public opinion or, at least, part of the public opinion. Data from the specific websites can be collected by the open application programming interface or correspondent crawler.

B. Data preprocessing

Data preprocessing is used to remove noisy and inconsistent data. Also it preprocesses stop words and negative words. In the pre-processing, the following steps are included: 1) partition of text, 2) tagging the words, and 3) the reinstatement of synonymous expressions. According to the material and the demand of the algorithms we propose to construct the "sentiment base" in the application of share market analysis.

C. Extraction of subjects and objects

Subjects and objects are mainly extracted by context mining. Context mining should obtain results as efficiently as possible to provide the necessary background knowledge for the subsequent steps. The second approach of extracting subjects and objects is text analysis, which extracts the opinion-oriented information through the pure text. The extractions of properties are based on the sentiment bases, modifier bases, and rule bases.

D. Developing sentiment analysis

Sentiment analysis consists of three bases namely sentiment, modifier and rule bases.

1) Sentiment base:

The sentiment analysis has two modules, i.e., the sentiment lexicon and its words' sentiment polarity. If there is no available sentiment lexicon in the stock market domain, then our primary task is to establish the sentimental lexicon [18]. Lexicon source of subjectivity includes verbs, adjectives and nouns with their polarity (i.e., positive, negative, or neutral) and strength (i.e., strong or weak) are noted. This lexicon is able to define the original polarity of a word only but the actual polarity of a word may be modified by its context in a sentence. The sentiment polarity of a word is determined by its morphemes. If the morphemes of a word occur in the positive lexicon more no of times than in the negative lexicon then the word is considered as positive; otherwise, the word is considered to be negative.

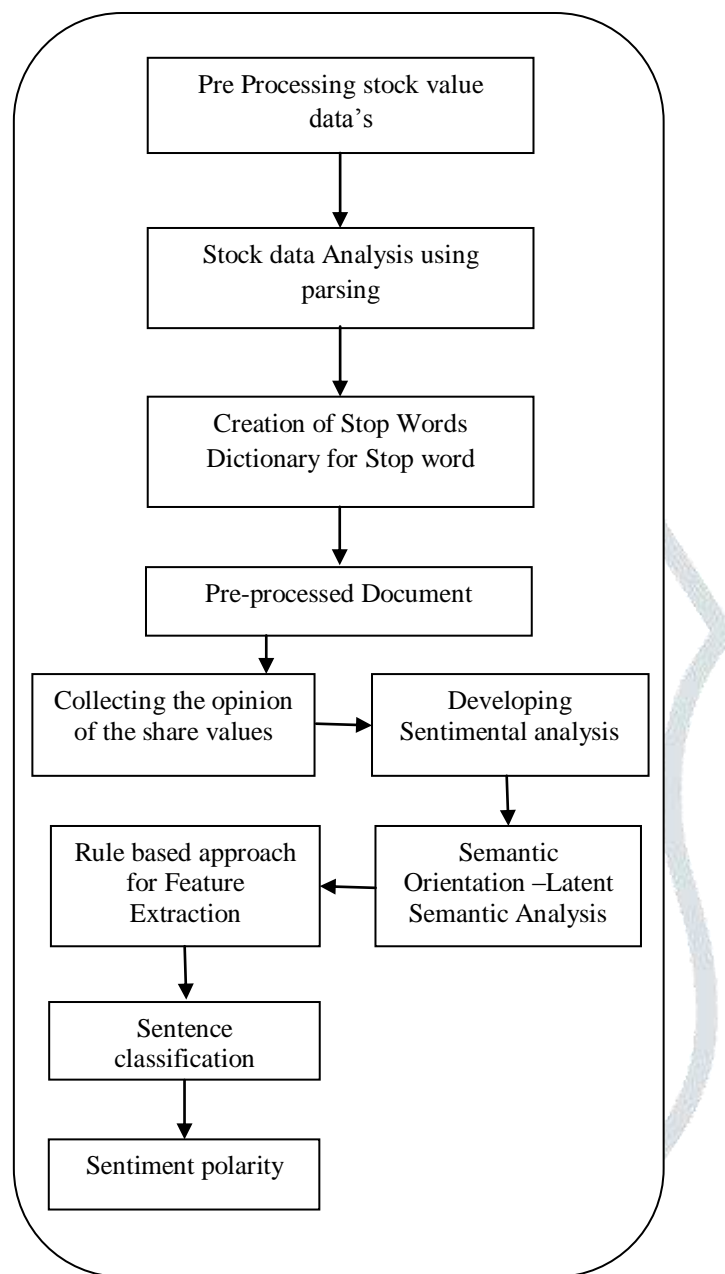


Fig. 1 Overall view of sentiment analysis

2) Modifier base:

In compliance with previous assumptions, the original sentiment of text is determined by the sentiment words. Also the sentiment is modified by adverbs. Negation adverbs reverse the sentiment polarity (e.g., “good” is positive, but it becomes negative if preceded by the word “not”). Similarly, degree adverbs that either strengthen or weaken the intensity of the sentiment polarity. In addition to that, sentence structure also affects the sentiment polarity value of a sentence.

3) Semantic rule base:

The construction of the semantic rule is important in the rule-based approach because the sentiment of share market depends on growth of the company and companies department. The semantic rule of sentiment is the SND pattern. S is the sentiment word, N is the negative word and D is the degree word. Each sentiment word has its unique modifiers. Among these three factors S is considered as the most important. The sentiment words are extracted using SND patterns. The main relationships between a sentiment word and its modifiers depend on their locations and classes in the sentence. In the N+D+S

rule, N is the modifier of D, and N+D is the modifier of the sentiment word (S). Therefore, the characteristics of N+D+S are the same as those of D+S. However, in the D+N+S rule, the negative word (N) is the modifier of the sentiment word (S), and the degree word (D) is the modifier of N+S. S usually represents a verb or a noun.

D) Sentiment calculation and polarity

The calculation of sentiment polarity of words is a basic step in the construction of the sentiment word base. In this paper, we use a sentiment polarity score to express the sentiment of a text whether it is positive or negative or neutral. The sentiment polarity score is calculated according to the rules defined by the SND patterns identified in the text. The sentiment polarity of a sentence is defined as p_s . If p_s greater than 0 then it is positive otherwise p_s is negative. We then calculate the polarity of sentences according to the rules defined in the rule bases. Therefore, in calculation of the polarity of a document in total, the position of the sentiment sentence is to be taken into consideration.

V. CONCLUSION

Share market data has been modeled as prediction analysis and monitored data to the server. The proposed solution has been modeled to analyze the share market of the web data to the server with respect to the opinion extraction in the data through proposing a latent semantic analysis algorithm which yields better performance in terms of scaling rate and accuracy against data classification. Sentiment analysis has been calculated through polarity scores – Positive, Negative and Neutral in the word occurrence.

REFERENCES

[1]	B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the Web," in Proc. 14th Int. Conf. World Wide Web, 2005, pp. 342–351.
[2]	B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends Inf. Retrieval, vol. 2, no. 1/2, pp. 1–135, Jan. 2008.
[3]	K. Bloom, N. Garg, and S. Argamon, "Extracting appraisal expressions," in Proc. HLT-NAACL, 2007, pp. 308–315.
[4]	J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin, "Learning subjective language," Comput. Linguist., vol. 30, no. 3, pp. 277–308, Sep. 2004.
[5]	F.-Y. Wang, "Agent-based control for networked traffic management systems," IEEE Intell. Syst., vol. 20, no. 5, pp. 92–96, Sep./Oct. 2005.
[6]	S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in Proc. 20th Int. Conf. Comput. Linguist., 2004, pp. 1367–1373.
[7]	M. Hu and B. Liu, "Opinion feature extraction using class sequential rules," presented at the AAI Spring Symposium Computational Approaches Analyzing Weblogs, Palo Alto, CA, USA, 2006, Paper AAI-CAAW-06.
[8]	V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in Proc. 8th Conf. Eur. Chapter Assoc. Comput. Linguist., 1997, pp. 174–181.
[9]	A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in Natural Language Processing and Text Mining. New York, NY, USA: Springer-Verlag, 2007, pp. 9–28.
[10]	E. Riloff, S. Patwardhan, and J. Wiebe, "Feature subsumption for opinion analysis," in Proc. Conf. Empirical Methods Natural Lang. Process., 2006, pp. 440–448.
[11]	C. L. Zhang, D. Zeng, J. X. Li, F. Y. Wang, and W. L. Zuo, "Sentiment analysis of Chinese documents: From sentence to document level," J. Amer. Soc. Inf. Sci. Technol., vol. 60, no. 12, pp. 2474–2487, Dec. 2009.
[12]	B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in Proc. ACL Conf. Empirical Methods Natural Lang. Process., 2002, vol. 10, pp. 79–86.
[13]	P. D. Turney, "Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews," in Proc. 40th Annu. Meet. Assoc. Comput. Linguist., 2002, pp. 417–424.
[14]	J. Wiebe, T. Wilson, and M. Bell, "Identifying collocations for recognizing opinions," in Proc. ACL Workshop Colloc., Comput. Extract., Anal., Exp., 2001, pp. 24–31.
[15]	L.-W. Ku, Y.-T. Liang, and H.-H. Chen, "Opinion extraction, summarization and tracking in news and blog corpora," in Proc. AAI Spring Symp., Comput. Approaches Anal. Weblogs, 2006, pp. 100–107.
[16]	R. W. Yuen, T. Y. Chan, T. B. Lai, O. Kwong, and B. K. T'sou, "Morpheme-based derivation of bipolar semantic orientation of Chinese words," in Proc. 20th Int. Conf. Comput. Linguist., 2004, pp. 1008–1014.
[17]	J. Kamps, M. Marx, R. J. Mokken, and M. De Rijke, "Using WordNet to measure semantic orientations of adjectives," in Proc. Int. Conf. Lang. Resourc. Eval., 2004, pp. 1115–1118.
[18]	D. Rao and D. Ravichandran, "Semi-supervised polarity lexicon induction," in Proc. 12th Conf. Eur. Chapter Assoc. Comput. Linguist., 2009, pp. 675–682.