An Improved Sentiment Classification using Lexicon into SVM

S.S.K. Rastogi
Student, M. Tech
IET Alwar, Rajasthan, India

Rohit Singhal
Associate Prof., CS Deptt.
IET Alwar, Rajasthan, India

Anil Kumar
Asst. Prof., CSE Deptt.
HCST,Mathura, India

ABSTRACT

With the emergence of web 2.0 and availability of huge amount of digital data on the social web, people always want to discover unknown, to predict events that could occur, and the procedure on how it works and change over time. Similarly, sentiment analysis is related with the automatic extraction of sentiment information from textual data available at various social webs. While most sentiment analysis deals commercial jobs like fetching opinions from product reviews, there is significant growth in social web and it becomes a source to promote various products. This is actual reason why most of the commercial web support login through social web like facebook, twitter. There are two approaches to sentiment analysis. First one is based on lexicon and second is machine learning. It has been proved that machine learning approach performs better than lexicon based approaches but it ignores the knowledge encoded in sentiment lexicons. This paper describes a method that includes sentiment lexicons as prior information to SVM approach, a machine learning techniques, to improve the accuracy of sentiment analysis. It also describes a technique to automatically create domain specific sentiment lexicons for this learning purpose. A result shows that the domain specific lexicons lead to a significant accuracy improvement for sentiment analysis.

General Terms

SVM; Lexicon; Sentiment Classification

Keywords

Lexicon Approach; Sentiment analysis; Polarity Detection

1. INTRODUCTION

Due to huge availability of information from online social websites like facebook and micro-blogging websites such as Twitter has attracted a lot of different types of researchers and practitioners for research. Several frameworks for detecting sentiments and opinions in social web have been developed for different application purposes, and Sentiment Analysis [1] is known as essential tool in social web monitoring platforms providing business services. In fact, social web posts, comments, reviews constitute a valuable asset for firms to directly find the customer's needs and preferences.

For example, tweets are a valuable [2] mine holding opinions of groups of people, possibly about a specific subject, issues or product. Generally, extracting sentiments given in tweets has been used for several purpose like to examine political sentiment by Tumasjan et al., 2010, to extract important information during times of mass emergency by Verma et al., 2011, to identify moods and happiness in known geographical location [3] from geotagged tweets by Mitchell et al., 2013, and in several social web monitoring services.

The main objectives of sentiment analysis are subjectivity, polarity and sentiment strength detection. Subjectivity detection deals with given text subjectivity, polarity detection [4] deals with whether given subjective text is positive, negative or neutral and finally sentiment strength detection finds the strength of positive or negative sentiment word. Social web, blogging and online forums have become the vast repository of review and comments on many topics which make it potential source of information for research. The social web [5] is being commercially used for automatically extracting consumer opinions about products or brands. An application could make use of large database from websites, use information retrieval methods to identify potentially relevant data, and then extract important information about given products or brands, such as which feature are liked or disliked.

The sentiment analysis, also called opinion mining, has developed many algorithms to categorize whether a comment, reviews, posts or tweets is subjective or objective, and whether any view expressed is positive or negative. Sentiment analysis can be done through various methods and applied on a large scale. One well known research focused on the average level of sentiment [6] showed in blogs in order to find overall trends in levels of happiness as well as age and geographic differences in the term of happiness.

A similar method used Facebook status updates to identify variation in mood over the year and to find "*the overall emotional mood of the nation*" and another project evaluated "*six dimensions of emotion in Twitter*". An important feature of public opinion and responses is sentiment whether people think positive or negative towards an occasion and how this varies over [7] time. There are many methods which combining social web data and other data. **S**entiment analysis algorithms have been able to give new understanding into human actions as a result and this approach shows promise for data journalism.

Actual application of sentiment analysis is different from its most general use for detecting the polarity [8] of opinions of customer products or music although it can use similar methods. Problem with most sentiment analysis is that it can be topic dependent in the sense that algorithms that predict sentiment exactly on texts from one specific field may be much less correct on another. Thus, while general sentiment analysis algorithms may not give overall good performance [9], it appears that there will be possibility for improving them by customizing them for specific topics.

For some specific topics, such changes may be required to give realistic performance if sentiment is normally expressed with a specialist or vocabulary. Hence, methods are required to either

develop topic-specific sentiment analysis algorithms for the social web or to modify general social web sentiment analysis [10] algorithms for specific areas. Most sentiment analysis algorithms are not appropriate for politics because they can exploit non-sentiment features of text that relate with sentiment, such as politicians' names. This can give more correct results overall if negative sentiment is expressed in ambiguous ways [11], such as with sarcasm and irony, so that the politician's name (or other non-sentiment feature) is the best clue for the presence of negativity. Sentiment analysis can depend on fixed sets of sentiment-related terms and ignore non sentiment aspects (i.e., a lexical approach).

2. RELATED WORK

There are two approaches to sentiment analysis are lexicon based and machine learning. A lexicon based approach [12] generally works with a lexicon of positive and negative words. The overall sentiment of a document text is determined by the sentiments of a set of words and words appearing in the text (Liu, 2007; Zhou and Chaovalit, 2008). However, a major challenge to this technique is that the polarity of several words is domain and context dependent. They either does not contain such expressions or bind them with an overall polarity score based on data collected from certain corpus. While excluding such words leads to poor coverage, simply binding them with a polarity words leads to poor precision.

Abbasi et al., 2008 studied that classification algorithms i.e. support vector machines (SVMs) [13] are commonly used because they perform better than other techniques in most machine learning contexts. Nevertheless, with a few exceptions (Read, 2005; Wilson et al., 2006), explicit comparisons with other methods have not been included in opinion mining publications.

There are two approaches to improve lexical sentiment analysis for the social web by allowing a general algorithm to be modified for a specific topic. The first method [14] considers the mood of the posts within the topic and the second method recognizes and adds topic-specific terms to the general sentiment lexicon. Although sentiment analysis often focuses on reviews of movies or consumer products these most likely form a small fraction of the social web.

An another polarity detection method given by Turney is to recognize average polarity of words within texts by approximating how often they occur with a set of seed words [15] of known and definite sentiment (e.g., good, terrible), typically using web search engines to estimate relative occurrence frequencies. It is supposed here that positive words will tend to come about with other positive words more than with negative words, and vice-versa. This approach requires relatively little lexical input knowledge and is flexible for different areas in the sense that a small set of initial general keywords can be used to generate a different lexicon for each application area. The seed words method appears to perform reasonably well in a variety of different contexts and learns area-specific sentiment-related words, such as 4G for mobile phones given by Zagibalov et al, 2010.

Turney et al has assessed the positive or negative polarity, rather than presence or absence, of sentimental content in text. Kim et al, 2004 have combined these two tasks, identifying subjectivity and detecting its sentiment polarity. The indicators of sentimental content have been extracted from different sources such as WWW and combined in a different ways which contain factor analysis and machine learning techniques, to determine when a text contains sentimental content and what is the polarity of that content. More recent studies have carried on in field of computer-aided

objectivity [16] which means determining linguistic features to be used to automatically categories text into positive or negative news. A method given by Hu and Liu, are used to calculating the semantic orientation [17] of sentiment words. In this method small set of seed of known sentiment words are defined first and develop the set automatically by incrementing synonyms and acronyms. This algorithm categorizes sentiments as either positive or negative.

Opinion mining can be divided into two or three parts, while more may be needed for some specific domains given by Balahur et al., 2010. In first steps, the documents are split into sections, such as sentences, and each sentences tested to see if it holds any sentiment if it is subjective or objective (Pang & Lee, 2004). In the second steps, the analysis of subjective sentences is done to detect their sentiment polarity. At last the object about which the opinion is told may be fetched which is given by Gamon et al, 2005. Opinion mining normally does not deal with discrete emotion [18] such as happiness, surprise rather it deals only with positive and negative sentiment and also does not detect sentiment strength but sometimes uses the strength of relationship of words with positive or negative sentiment, by Kaji et al 2007 and does not concurrently identify both positive and negative emotions.

Machine learning approach is used to find positive and negative sentiment by opinion mining algorithms. , where these sentimental word could be a subset of the words in the text, parts of speech or n-grams (i.e., the frequency of occurrence of all n consecutive words, where n is typically 1, 2, or 3 given by Abbasi et al, 2006. Another work on online movie review tried to include emotions but is more domain independent than words. Ng et al., 2006 studied about artificial features resulting from adjective polarity lists. The other features typically give small, but important increases in performance. Prabowo et al 2009 studied rules-based methods which identify structures in sentences related with.

3. DOMAIN SPECIFIC LEXICONS

Previous studied shows that sentiments of several texts are domain or context dependent. For instance long is positive if it is related with the camera feature of 'Battery Life'. However, if long is related with camera feature "Shutter Lag" then it carries negative sentiment. Therefore, it is important to know the context being discussed when one want to determine the related sentiment. Based on this study, contexts/domains based lexicon [19] is being made which indicate both specific domain as well as sentiment in that particular domain. For instance, lexicon about 'Camera Picture Quality' would comprise of two sub-lexicons. One includes texts such as mage, picture, photo, close up etc, which are good index for the aspect of 'Picture Quality' in the domain of digital cameras. The other one includes texts that convey positive or negative sentiments if the related aspect is camera picture quality. For instance, second sub-lexicon would show that while clear and sharp are positive, blurry is negative when they are related with camera picture quality. This objective can be obtained by combination of corpus filtering, web search with linguistic patterns and dictionary expansion. Detail descriptions of these techniques are given in the following subsections.

3.1 Corpus Filtering

To make a foundation for domain specific lexicons, training corpus are used, in which each camera review text is commented with a camera feature as well as the related sentiment. This approach compromise two steps as follows.

First steps extract all texts that occur in training sentences about any particular camera aspect and these may be nouns, adjectives, verbs, adverbs as well as their negated forms [20]. These steps create an initial list of lexicons for each camera feature.

In second steps, for texts in the list for each of the camera features, we make sure that if word or phrase also occurs in any other camera aspect lexicon. If yes, it is removed from the lexicon. After this step of filtering, list of lexicon are obtained for each camera feature, which contains only texts unique to that camera feature.

However, the drawback of this technique [21] is that the list of the lexicons would completely depend on the list of the corpus, and making a large training corpus list is time intensive, expensive and sometimes very complicated task due to complexity of natural language. This drawback can be overcome by enhancing initial domain specific lexicons obtained from training corpus through web search and filtering linguistic patterns as well as dictionary expansion. These two approaches are explained in the next two subsections.

3.2 Web Search and Filtering Using Linguistic Patterns

To improve the exposure of the domain specific lexicons, which is obtained from training corpus, two different linguistic patterns [22] are made and used them for searching queries to find more texts conceptually related with the camera feature. The two linguistic patterns are as follows.

Pattern 1: "Camera feature include(s) *"
Pattern 2: Camera feature + "Seed Word and *"

In these two patterns, 'Camera feature' refers to words such as camera accessories and camera price. 'Seed Word' denotes seed words for a specific camera feature. For instance, cheap and expensive can provide as seed words for camera feature price. Note that in Pattern 1, the camera feature name is incorporated as part of an exact search query, whereas in pattern 2, the camera feature name provides as the context for the search query. Based on semantic nature of a camera feature, we choose one of these two patterns to get expressions conceptually associated to that feature. For instance, while "camera accessories include *" is very useful for finding accessory expressions, 'camera picture + "clear and *"' is better for getting expressions associated to camera pictures.

3.3 Dictionary Extension

Although extension through referring synonyms and antonyms recorded in dictionaries is a commonly used technique when a general purpose sentiment lexicon is made (Hu and Liu, 2004), one found that this technique to be not always appropriate for making domain specific lexicons. The reason is that creating domain specific lexicons needs finding sentences that are conceptually correlated [23]; however sentences that are conceptually correlated are not necessarily synonyms or antonyms. For example, 'sharp' and 'clear' are conceptually correlated to camera picture qualities, but they are not exact synonyms from a linguistic point of view. However, in some cases, using dictionaries can still be very successful. When lexicons for camera price are made through web searching and filtering using pattern 2 then price lexicons are given as:
Price Lexicon: *[cheap, promote, expensive, worthy, lowest, discount, promo, coupon, value]* By binding the synonyms of

'cheap' and 'expensive' in WordNet (Fellbaum, 1998), Price Lexicon are further expanded.

3.4 Domain Specific Polarity Lexicon

This section deals with domain specific lexicons obtained from previous section that how to categories positive sentiment as well as negative sentiment in each domain lexicons. For example, one would like to make the following sub-lexicons for 'Picture Quality'.

Picture Quality Positive Lexicon: [*clear, bright, sharp, sunny, crisp, sober, stable, tidy, vivid*]

Picture Quality Negative Lexicon: [*dark, dim, gray, blurry, hazy, humid, fuzzy, blurred*]

For each term in the Picture Quality Lexicon that made through the combination of corpus filtering, web search and dictionary expansion [24], we need to find if it only appears in the training data labeled as showing a positive sentiment or a negative sentiment about the camera's picture quality. If it is the former case, terms are included into the Picture Quality Positive Lexicon, while if it is the latter case, terms are included into the Picture Quality Negative Lexicon. Next sections describe how lexicon knowledge is integrated into support vector machine to improve sentiment classification.

4. LEXICONS INTEGRATION INTO SVM

For each review text about cameras, we have to predict both the camera feature discussed in that text as well as the related sentiment regarding that camera feature. This goal can be achieved by performing a two step classification [25]. In first steps, classifiers are trained to predict the camera feature being discussed. In second step, classifiers are trained to predict the sentiment related with that camera feature. Finally, the two step prediction results are aggregated together to generate the final prediction. In both steps, lexicon information is included into conventional SVM learning. To illustrate this approach, sentence (1) is used as an example.
(1) Extra nice protections are given to camera due to rigid case.
Using nouns, adjectives, verbs, and adverbs as unigram feature words in a conventional SVM learning, this sentence can be represented as the following vector of words.

[Case, extra, nice rigid, camera, give, protection]

By including the information encoded in the lexicons, one automatically produces and inserts additional features into the above representation. For example, when the step 1 feature classification are performed, because the feature word '*case*' in the above representation is listed in domain specific lexicon about camera accessories, one would insert an extra feature word '*accessory*', and make the following new representation.

[Case, rigid, accessory, camera, extra, nice, protection]

The camera aspect is being promoted such as '*accessory*' if expressions of camera features occur in the sentence. In the next step of polarity prediction, domain specific sentiment lexicon and a common domain independent sentiment lexicon extracted from the MPQA [26] opinion corpus are incorporated.
For example, because '*nice*' signifies as a positive word in the MPQA lexicon, we would insert a feature word '***positive***'. In addition, if the first step prediction result for sentence (1) is '*accessory*', and '*rigid*' is also a positive word in our domain

specific lexicon regarding camera accessories, we would generate an extra feature word '*positive*' in final representation for sentence (1) for the second step polarity prediction as shown below.

[Case, rigid, camera, extra, nice, protection, positive]

We thus promote a '*positive*' prediction regarding the aspect of '*accessory*'. Experiments show that including lexicon information into SVM learning considerably improves the accuracy for classification task; compared to the common MPQA sentiment lexicon, the domain specific lexicon constructed is more effective.

5. EXPERIMENTAL EVALUATION

The sentiment analysis has been performed for classifying 45-way sentiment classification task. These 45 classes are resultant from 22 features related to camera purchases such as battery life, picture quality, customer support, LCD screen, and their related polarity values positive and negative, as well as neutral sentiment about any of the 22 features. The goal is to record each input sentence into one of the 45 classes. As stated in the previous section, two step classifications are performed for this task. Final joint classifiers consist of two classifiers. The first is a '*feature classifier*', which executes a 23-way camera feature classification. The second is a '*polarity classifier*', which executes a 3-way (positive, negative and neutral) classification. The final predictions are combined from the predictions made by these two classifiers. The classification accuracy is defined as follows.

$$Accuracy = \frac{No.\,of\ sentences\ correctly\ classified}{Total\ no.\,of\ sentences} \quad (1)$$

5.1 Dataset

Multi-Domain Sentiment Dataset created by Blitzer et al. has been selected for experiment purpose. Therefore, most of the classes in this dataset are not balanced, and the majority of the class has 15% of the sentences. Previous works such as Hu and Liu (2004) and Popescu and Etzioni (2005) only fetched explicitly expressed product features, and they do not recognize implicitly shown product features. Also, they do not further classify the extracted noun phrases. By contrast, both the explicitly and implicitly [27] shown product features should be extracted and further classify the semantically related words regarding product features. Another work by Zhao et al. extracted both implicitly and explicitly shown product features, and they also further classified the product features. However, in terms of sentiment extraction, they only extracted sentiment words related with product features, and did not further recognize the polarities of the sentiment words. By contrast, the polarities related [28] with the product features should be identified. Instead, the majority class (13%) as baseline has been used and incorporating lexicon information with SVM approach with a traditional SVM approach because the latter is the modern algorithm reported in the literature for sentiment analysis. Results show that both the traditional SVM approach and this given approach considerably work better the majority class baseline.

5.2 Results

There are total four experiments which are performed. In first experiment 1, the traditional SVM approach is performed and refers to this approach as SVM. In experiment 2, Information determined in the domain independent MPQA are included into sentiment dictionary into SVM approach and refer to this experiment as 'MPQA + SVM'. In experiment 3, information

encoded in the domain specific lexicons are included into SVM approach and refer to this experiment as 'Domain-Lexicons + SVM'. In experiment 4, both the information encoded in the MPQA and the domain specific lexicons are included into SVM approach and refer to this experiment as 'DomainLexicons + MPQA + SVM'. All of results are performed 10 times for stability, and they are summarized in Table 1.

The results in Table 1 and Fig. 1 show that it outperform and give overall best performance when both the domain independent MPQA lexicon and the domain specific lexicons are included. Incorporating the domain specific lexicons is more effective [29]. The improvement achieved by this approach is logically important with p <0.000001 according to paired t-test.

Table 1: Overall Performance Comparison

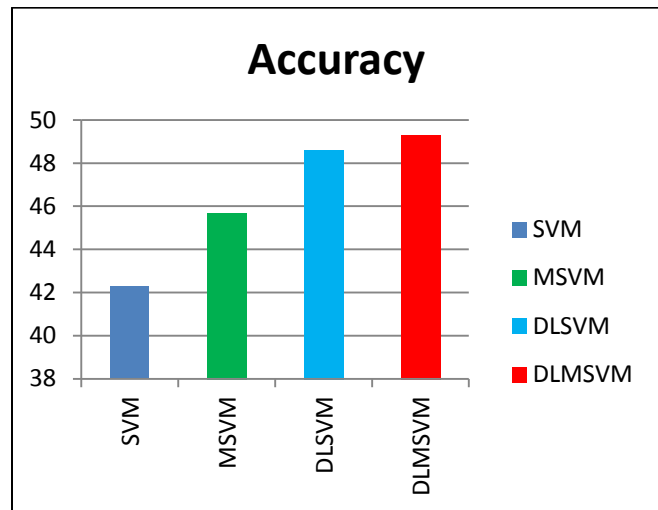| Approach	Accuracy
SVM	42.3%
MPQA+SVM	45.7%
Domain Lexicons+SVM	48.6%
Domain Lexicons+MPQA+SVM	49.3%

Fig. 1. Overall Accuracy Comparison

Results in Table 2 shows that including lexicon with SVM approach considerably improves both the accuracy for camera feature classification and the accuracy for polarity classification. Both improvements are logically important with p <0.000001 and p <0.05 respectively according to paired t-test.

Table 2: Separate Performance Comparison

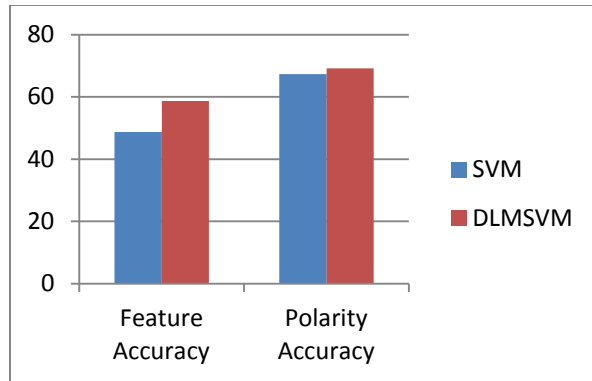| Approach	Feature Accuracy	Polarity Accuracy
SVM	48.7%	67.3%
Domain Lexicons+MPQA+SVM	58.7%	69.2%

Fig. 2. Individual performance comparison

6. CONCLUSIONS

Incorporating the information encoded in sentiment lexicons, mainly domain specific lexicons, can drastically improve the accuracy for proposed sentiment analysis. It also described how domain specific sentiment lexicons for the domain of camera reviews through a combination of corpus filtering, web searching and filtering and dictionary expansion are created. Finally, novel method has been developed to incorporate the lexicon information into machine learning technique such as SVM to improve sentiment analysis.

As the future work a multilingual corpus of facebook data and evaluate the uniqueness of the corpus across different languages can be done. Multilingual sentiment classifier for social web can be done.

7. REFERENCES

[1] G. Vural, B. B. Cambazoglu, P. Senkul, and O. Tokgoz. A framework for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish, Computer and Information Sciences III, pp 437-445, 2013.

[2] Cuneyd M. O. and Arzucan O. Word Polarity Detection Using a Multilingual Approach. Springer-Verlag Berlin Heidelberg, CICLing 2013, Part II, LNCS 7817, pp. 75–82, 2013.

[3] Thelwall, M., & Buckley, K. Topic-based sentiment analysis for the Social Web: The role of mood and issue-related words. Journal of the American Society for Information Science and Technology, 64(8), 1608–1617, 2013.

[4] Erik T., Mykola P. RBEM: A Rule Based Approach to Polarity Detection. In Proceedings of WISDOM'13, August 11, Chicago, USA, 2013.

[5] Rodrigo D. F., Gaston L. and Juan D. V. Sentiment Polarity of Trends on the Web Using Opinion Mining and Topic Modeling. In the proceeding of web intelligence constorium chile research center, USA, 2013.

[6] E. Cambria, B. Schuller, Y. Xia, and C. Havasi. New avenues in opinion mining and sentiment analysis. Intelligent Systems, IEEE, 28(2):15-21, 2013.

[7] D. Bespalov, Y. Qi, B. Bai, and A. Shokoufandeh. Sentiment classification with supervised sequence encoder. In Proceedings of European Conference on Machine Learning

and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), volume LNCS 7523, pages 159-174. Springer, 2012.

[8] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou, Sentiment Strength Detection for the Social Web. In the Journal of the american society for information science and technology, 63(1):163–173, 2012.

[9] Thelwall, M., Buckley, K., & Paltoglou, G. Sentiment strength detection for the social Web, Journal of the American Society for Information Science and Technology, 63(1), 163-173, 2012.

[10] G. Paltoglou and M. Thelwall. Twitter, myspace, digg: Unsupervised sentiment analysis in social media. Volume 3, pages 66:1-66:19, New York, USA,ACM, 2012.

[11] E. Cambria, D. Olsher, and K. Kwok. Sentic activation: A two-level affective common sense reasoning framework. In Proceedings of AAAI, pages 186-192, 2012.

[12] X. Glorot, A. Bordes, and Y. Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In Proceedings of the 28th International Conference on Machine Learning, ICML2011, pages 513-520, 2011.

[13] Hassan, A., Abu-Jbara, A., Jha, R., Radev, D.: Identifying the Semantic Orientation of Foreign Words. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, vol. 2, pp. 592–597, 2011.

[14] Thelwall, M., Buckley, K., & Paltoglou, G. Sentiment in Twitter events. Journal of the American Society for Information Science and Technology, 62(2), 406–418, 2011.

[15] K. Veselovská. Sentence-Level Polarity Detection in a Computer Corpus. In the Proceedings of WDS'11, Part I, 167–170, 2011.

[16] Taboada, M., Brooke, J.,Tofiloski, M.,Voll, K.,&Stede, M. Lexicon based methods for sentiment analysis. Computational Linguistics, 37(2), 267–307, 2011.

[17] E. Tromp and M. Pechenizkiy. Graph-based n-gram language identification on short texts. In Proceedings of the 20th Machine Learning conference of Belgium and The Netherlands, pages 27-34, 2011.

[18] E. Tromp and M. Pechenizkiy. Senticorr: Multilingual sentiment analysis of personal correspondence. In Proceedings of IEEE ICDM 2011 Workshops, pages 470-479. IEEE, 2011.

[19] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), 2544–2558, 2010.

[20] Taboada, M., Lu,Y.,Kong, X., Quan, X., Liu,W.,& Xu,Y. Exploring the sentiment strength of user reviews. Lecture Notes in Computer Science, 6184/2010, 471–482, 2010.

[21] A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), pages 1320-1326, 2010.

[22] Haccianella, S., Esuli, A., & Sebastiani, F. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of the Seventh conference on International Language Resources and Evaluation, 2010.

[23] M. Mathioudakis and N. Koudas, "Twittermonitor: trend detection over the twitter stream," in Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data, ser. SIGMOD '10. Indianapolis, Indiana, USA: ACM, pp. 1155–1158, 2010.

[24] Hassan, A., Qazvinian, V., Radev, D.: What's with the Attitude? Identifying Sentences with Attitude in Online Discussions. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 1245–1255, 2010.

[25] Hassan, A., Radev, D.: Identifying Text Polarity Using Random Walks. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 395–403, 2010.

[26] Balahur, A., Kozareva, Z., & Montoyo, A. Determining the polarity and source of opinions expressed in political debates. Lecture Notes in Computer Science, 5449, 468–480, 2009.

[27] V. Sindhwani and P. Melville. Document-word co-regularization for semi-supervised sentiment analysis. In Eighth IEEE International Conference on Data Mining (ICDM'08), pages 1025-1030, 2008.

[28] Ann Devitt and Khurshid Ahmad. Sentiment Polarity Identification in Financial News: A Cohesion-based Approach. In the Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 984–991, Prague, Czech Republic, June 2007.

[29] Turney, P.D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL '02) (pp. 417–424). Stroudsburg, PA: Association for Computational Linguistics. 2002.