

RICH SEMANTIC SENTIMENT ANALYSIS USING LEXICON BASED APPROACH

Hedayatullah Lodin¹ and Prem Balani²

Department of Information Technology, G.H. Patel College of Engineering and Technology, India

Abstract

Web is a huge repository of information, and a massive amount of data is generated everyday on online platforms. Information, can be facts and opinions, facts are objective statements about an event, and opinions are subjective statements that reflect the sentiments of a person towards an event. Research on sentiment analysis has increased tremendously in recent years due to its wide variety of applications. To analyze sentiments, certain methods have been proposed, which can be broadly categorized as supervised machine learning and lexicon based approaches. Supervised machine learning methods are giving high accuracy but these methods need training data and are domain dependent, while lexicon-based methods are not domain dependent. Although, building of lexicon is costly, but once constructed, it can be applied for a wide variety of domains, but still lexicon based methods are restricted to their dictionaries and are full-dependent on the presence of terms that explicitly reflect the sentiment, while in many cases the sentiment of a term is implicitly reflected by the semantics of its context. Therefore, we've proposed context aware, semantically rich (conceptual & contextual semantics) lexicon-based method which is different from traditional lexicon-based methods that assigns sentiment score and strength to terms in a dynamic way, and outperforms baselines.

Keywords:

Sentiment Analysis, Lexicon, Supervised Machine Learning, Contextual and Conceptual semantics

1. INTRODUCTION

Nowadays, the age of Internet has changed the way in which people express their sentiments, share their experiences, knowledge and perceptions about an event or an object. It is now mainly done through Websites, Forums, blogs, social networks, and content sharing systems. This means, on online platforms massive amount of data is generated every day. So, the web is a huge repository of information. The textual information which is available, can be basically categorized into two broad categories, facts and opinions. Facts are the objective statements about events or objects. Opinions (sentiments) are the subjective statements that reflect the sentiments and perceptions of a person towards an event or an object which form the core of Sentiment Analysis.

Sentiment analysis (SA) has grown to be one of the most active and fastest growing research areas in Natural Language Processing (NLP) field. Research on sentiments and opinions appeared much earlier [33], [42], [43], [44]. The term sentiment analysis was first mentioned in the work of Nasukawa and Yi [41], and the term opinion mining was first mentioned in the work of Dave, Lawrence and Pennock [40].

There are many definitions for sentiment analysis. Bing Liu [2], precisely defined sentiment analysis as: "Given a set of evaluative text documents D that contain opinions (or sentiments) about an object, opinion mining or sentiment analysis aims to extract attributes and components of the object that have been

commented on in each document d belongs to D and to determine whether the comments are positive, negative or neutral". So, sentiment analysis involves classifying opinions in text into categories like "positive", "negative" and "neutral".

Opinion can be mathematically represented as a quintuple (o, f, so, h, t) , where o = an object which can be person, event, product etc.; f = object's feature; so = sentiment orientation (positive, negative or neutral); h = opinion holder; t = time when the opinion is expressed [3].

Sentiment analysis research has increased tremendously in recent years due to its huge business applications and importance. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service. Businesses and Organizations, spend a huge amount of money to find consumer sentiments and opinions for brand analysis, new product perception and service benchmarking. Individuals are interested on other's opinions ("what other people think") to make decisions for purchasing a product or using a service.

According to [2], since 2000, a large number of methods and enhancements have been proposed for the problem of Sentiment analysis. These methods can be broadly classified as Supervised Machine Learning, Lexicon based, and Hybrid methods.

This paper presents a lexicon based method that overcomes the limitations of traditional lexicon based and supervised machine learning methods, with following characteristics:

- Domain independent
- Context aware
- Dynamic (assigning score and strength of terms)

The rest of the paper is organized as follows. section 2, contains related work. Section 3, contains the proposed work. Section 4, contains the experimental results and section 5 contains the concluding remarks.

2. RELATED WORK

Sentiment analysis is a subfield of Natural Language Processing that subtends with the subjectivity and determination of opinion in a text. According to [4], Sentiment analysis can be considered as a very restricted NLP problem, where it is necessary to understand the positive or negative sentiments concerning each sentence, entity or topic. As it is mentioned in (section 1), it aims to analyze people's sentiments, opinions, emotions, etc., towards objects such as topics, products, individuals, organizations, services, etc. Sentiment analysis can be performed at different levels:

- Document Level
- Sentence Level
- Entity/Aspect Level.

Document level considers whole document as an opinion on an entity or aspect of it, thus the entire document is classified as positive, negative or objective [5], [6], [35]. Sentence level is fine-grained than document level sentiment classification, in which the sentence can be classified as positive, negative or neutral [7], [8], [6]. And finally, when more precise information is necessary, then the entity/aspect level arises. It is the finest-grained level, it considers a target on which the opinion holder expresses a positive or negative opinion [7], [6], [5].

The literature survey indicates that, the major techniques for sentiment analysis are:

- Supervised Machine Learning Approaches.
- Lexicon-Based Approaches
- Hybrid Approaches (Using the best of above two approaches) [18], [19], [20].

2.1 SUPERVISED MACHINE LEARNING APPROACHES

Supervised Machine Learning approaches need training data. These methods are based on training classifiers (e.g., Naïve Bayes, SVM, Maximum Entropy etc.), from various combinations of features, such as: word n-grams [9], [10], [11], POS tags [12], syntax features and syntactic dependency (e.g., hashtags, punctuations, parsing etc.) that can determine the meaning of a sentence [13], [28], [29], [49], with/without words' prior sentiment and semantic concepts [14].

Supervised machine learning methods have achieved the accuracies in the range of 80%-84% as reported in [21], [14]. However, the training data is difficult to obtain [3], especially for continuously changing and evolving data (e.g., Twitter data), The distant supervision approach [9] overcomes this limitation and makes use of automatically generated training data, where emoticons such as “:(” and “:-)” are typically used to label tweets as negative or positive. But, it has been observed that the automatic labelling of training data affects the performance of the classifiers and introduces errors [23]. Another limitation of supervised machine learning methods, is their domain dependency, e.g., classifiers trained on data from one domain will produce unsatisfactory performance when applied to data from another domain, (e.g., trained classifiers on tweets relating to health reform, applied on tweets relating to products) [24], [25]. So, the fastest emerging varieties of topics from twitter, affect the applicability of supervised machine learning approaches [25].

2.2 HYBRID APPROACHES

Hybrid of the Machine learning approaches and Lexicon-based approaches, has also been tried and tested to obtain the best of both the worlds [18], [19], [20].

Appel et al. [1], proposed a hybrid method, that uses NLP essential techniques and a sentiment lexicon, enhanced with SentiWordNet and fuzzy sets for semantic orientation of polarity and sentiment strength for sentence level sentiment analysis. They evaluated their proposed method on three different datasets, and achieved high accuracy and precision compared to Naïve Bayes and Maximum Entropy techniques. But this method is ill-suited for twitter like data (which contains slangs, jargons, idioms, poor grammar etc.).

Besides hybrid approaches, the combination of supervised and unsupervised techniques, or semi-supervised techniques, can also be used for sentiment classification [26], [27].

2.3 LEXICON-BASED APPROACHES

Lexicon based approaches are working based on dictionaries of words, where each word is associated with a specific sentiment in order to calculate the overall sentiment. These approaches rely on sentiment lexicons of words and phrases with their sentiment orientations, such as pre-built dictionaries SentiWordNet [15], LIWC lexicon [17] or the MPQA subjectivity lexicon [16].

These methods try to overcome the limitations of supervised machine learning methods, as they do not need training data. Lexicons are costly to obtain, but once constructed they are applicable to a wide variety of domains (not domain dependent) [25]. SentiWordNet [30] is a lexicon that is widely used for sentiment analysis, it is based on an English lexical dictionary called WordNet. SentiWordNet group adjectives, nouns, verbs, and other grammatical classes into synonym sets called synsets.

Lexicon-based approaches are working effectively on conventional text [3], and they tend to be ill-suited for informal text (e.g., twitter data, which contains a large number colloquial expressions of malformed words like “loov”, “idk”, “gr8”). Besides this, many lexicon-based methods are using lexical structures for sentiment determination, which becomes problematic in traditional text which contains ungrammatical sentences [25].

In [5], Thelwall et al. proposed an algorithm (SentiStrength) to overcome the aforementioned limitations of lexicon based approaches. SentiStrength is a human-coded lexicon specifically built for social web. This method applies several lexical rules (e.g., existence of emoticons, negations, intensifiers, booster words), to determine the sentiment strength of an online post. SentiStrength, as well as [31], focuses not only on polarity (positive/negative) detection, but also on identifying sentiment strength. In the case of [31], sentiment strength varies from -5 to +5 (very negative to very positive).

Lexicon-based methods are restricted by their lexicons, and particularly by the static prior sentiment values of words or terms regardless of their contexts. However in [5], authors have proposed an algorithm to assign an updated sentiment strength to words in the lexicon, but still it needs to be trained from manually annotated corpora. Another problem with lexicon-based methods is their full-dependency on the presence of terms that explicitly reflect sentiment, while in many cases the sentiment of a term is implicitly reflected by the semantics of its context [32].

To overcome the limitations of lexicon-based methods, many methods have been proposed to explore the semantics for sentiment analysis that implicitly reflect the sentiment. Semantic methods of sentiment analysis can be broadly classified into contextual semantic and conceptual semantic approaches [25].

Contextual semantic (also called statistical semantics) methods are determining semantics from the co-occurrence patterns of words [50], that have been used for sentiment analysis [33], [34], [22]. In [34], the authors used PMI (pointwise mutual information) to find the statistical correlation between a given word and a balanced set of 14 polar (positive and negative) words. If the word has a stronger degree of correlation to negative words than positive

ones, the word has negative orientation, and if the degree of correlation is stronger to positive words' side, the word has positive orientation. However, large lexical knowledge is not required, but still the identification speed is very limited [36], because for retrieving the co-occurrence frequencies of words it uses search engines. According to [37], due to the use of entire web as a corpus and limited choice of paradigm words, the contextual semantic approach is unable to work well with words, which have domain specific orientations in order to assign sentiments. For example, it is unable to differentiate the word "Long" as a positive word (to describe the battery life) and as negative (to describe a program's running time).

"The battery of this camera lasts very long."

"This program takes a long time to run."

Conceptual semantic methods are those, which are using external semantic knowledge bases with the techniques of NLP, for capturing the conceptual representations of words that implicitly reflect sentiment [25]. In [14], Saif Hassan et al. showed that the combination of general conceptual semantics with supervised classifiers is giving a higher sentiment accuracy. In [38], Cambria et al. proposed a concept-based lexicon called SenticNet, which contains 50k fine grained concepts with their sentiment orientations. SenticNet works well with conventional text [39], unlike SentiStrength [5], ill-suited for Twitter and the like. However, conceptual semantic methods have been shown that outperform syntactical methods [32], but they are restricted by the boundary of their underlying knowledge bases. Especially when processing Twitter data (with rapid language deformation), it will be problematic.

3. PROPOSED WORK

From literature review as in (section 2), it can be concluded that, there is need for domain independent, context aware, semantically rich (conceptual & contextual semantics) methods to analyze explicitly and implicitly reflected sentiment. Therefore, a lexicon based method has been proposed, that is different from typical lexicon based methods.

The proposed approach has been designed with the following features:

- Lexicon based, that can be applied for different domains, and updating sentiment score and strength dynamically.
- Context aware (Able to capture semantics with regard to context and concept).

The systematic workflow of the proposed approach as shown in (Fig.1) can be summarized in the following steps:

3.1 CREATION OF TERM-INDEX

In this step, a term-index is created and several text pre-processing procedures such as filtering out non-English terms, POS tagging are performed.

3.2 GENERATION OF VECTOR (TERM CONTEXT VECTOR)

In this step, each term m is represented as a vector of all its context terms (i.e., words that occur with a word m in the same context) in the text. Term context vector of a term m is a vector c

that occurs with m term. $c = (c_1, c_2, c_3, \dots, c_n)$ and the semantic relation is found between term m and its context term c_i .

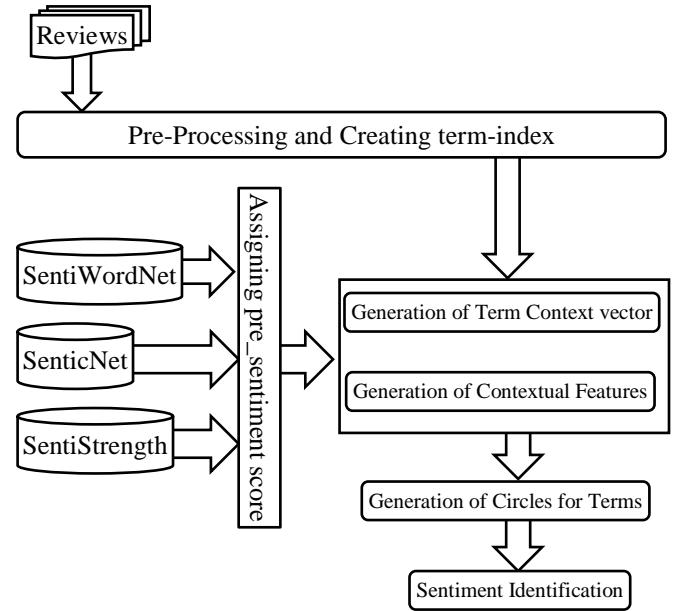


Fig.1. Systematic workflow of the proposed work

3.3 GENERATION OF CONTEXTUAL FEATURES

Here, for each term; its degree of correlation to all its context terms has been calculated and also an initial score has been assigned to these context terms using lexicons (SentiWordNet, SentiStrength, and SenticNet).

- Pre-sentiment Score: Based on POS tag(s), each c_i is assigned sentimental score from three lexicons.
- Term Degree of Correlation: Inspired by TF-IDF scheme, the term degree of correlation will be calculated as:

$$TDC(m, c_i) = f(c_i, m) \times \log(N/Nc_i) \quad (1)$$

where, $f(c_i, m)$ is the total number of times context-term(c_i) occurs with m in the review, N is the number of terms in the review, and Nc_i is the number of terms that occur with context-term (c_i).

3.4 GENERATION OF CIRCLE

This step, converts the term-context vector of m into a two dimensional geometric circle. Each context term is located in the circle based on: Angle θ_i (defined by its pre-sentiment score $(c_i) * PI$), $\theta_i = \text{Pre-sentiment}(c_i) * \pi$ and Radius r_i (defined by its degree of correlation with the term m), where $r_i = \text{Term degree of correlation}(m, c_i)$

3.5 SENTIMENT IDENTIFICATION

To identify the sentiment orientation and strength of terms, polar coordinate system is converted to Cartesian coordinate system. Therefore, the trigonometric functions (properties of circle) have been used to find the coordinates of terms:

$$x_i = r_i \cos \theta_i \text{ and } y_i = r_i \sin \theta_i$$

where, y value represents the sentiment and x value represents the sentiment strength.

Now, for contextual semantic of term m is circled, that is composed by the set of (x, y) .

Based on the contextual features, terms will take place in the plane of two dimensional geometric circle. Words that take place in two upper quadrants will have positive sentiment ($\sin\theta > 0$). The upper left quadrant representing stronger positive sentiment because it has larger angle than upper right quadrant.

Similarly terms which take place in two lower quadrants will have negative sentiment ($\sin\theta < 0$). The larger the radius (term degree of correlation), the important the context term to term- m .

One of the effective way to approximate the sentiment of the term m is to calculate the geometric median of all its points which are located on the circle plane. Geometric median for n given points (p_1, p_2, \dots, p_n) is calculated as:

$$g = \arg \min_{g \in R^2} \sum_{i=0}^n \|p_i - g\|^2 \tag{2}$$

For overall sentiment and strength of a review, using Eq.(2) the geometric median is found for g_m of $g_1, g_2, g_3, \dots, g_n$.

The main concept here is that, the sentiment of a word or term is not fixed, it depends on the context where the word has been used or simply, each term depends on its contextual semantics. For conceptual semantics of terms SenticNet has been used. To capture the term's contextual semantics distributional hypothesis has been followed, terms tend to have similar meanings would occur in similar contexts [10]. So, the co-occurrence pattern of a word with other words for contextual semantics has been calculated.

4. EXPERIMENTAL STUDIES

4.1 EXPERIMENTAL SETUP

The baselines and the proposed approach RSSA (Rich-Semantic Sentiment Analysis using lexicon based approach), have been implemented using java technology. They have been evaluated on two different datasets (twitter and SaaS datasets):

- Stanford Twitter Sentiment (STS-Gold) dataset [14], which contains total 2032 tweets, out of total, 632 are positive and 1400 are negative.
- Software as a Service (SaaS) dataset [49], which contains total 6258 reviews of customers, out of total, 5397 are positive, and 264 are negative reviews.

In order to analyze the comparative results accurately, the same datasets have been used to evaluate both proposed approach and baselines.

4.2 EXPERIMENTAL RESULTS

The proposed method is a lexicon based approach, therefore those lexicon-based that are recent, widely used and giving higher accuracy as baselines (SentiWordNet, SenticNet, and SentiStrength) have been used.

Table.1. Confusion Matrix

	Class Negative	Class Positive
Predicted Negative	tn	fp
Predicted Positive	fn	tp

Normally accuracy, precision, recall and f-measure are used for evaluating performance of sentiment analysis classification

algorithms [46][47]. They are usually computed from a confusion Matrix [45][48], as in Table.1.

In Table.1, tn (true negative) is the number of correctly predicted negative classes, tp (true positive) is the number of correctly predicted positive classes, fp (false positive) is negative classes predicted as positive, and fn (false negative) is positive classes predicted as negative. Those evaluation metrics are computed based on the values in the confusion matrix as follows:

$$Accuracy = (tp+tn)/(tp+tn+fp+fn) \tag{3}$$

$$Precision = tp/(tp+fp) \tag{4}$$

$$Recall = tp/(tp+fn) \tag{5}$$

$$F - measure = 2 \frac{Precision * recall}{Precision + recall} \tag{6}$$

The baselines and proposed method have been implemented and evaluated on both datasets (STS-Gold and SaaS datasets), and their performance (accuracy) has been calculated as shown in Table.2.

Table.2. Performances of proposed method and baselines

Methods	Accuracy for STS-Gold dataset	Accuracy for SaaS dataset
SentiWordNet	60%	72.1%
SenticNet	58.1%	87.3%
SentiStrength	67.3%	75.7%
RSSA	72.8%	89.69%

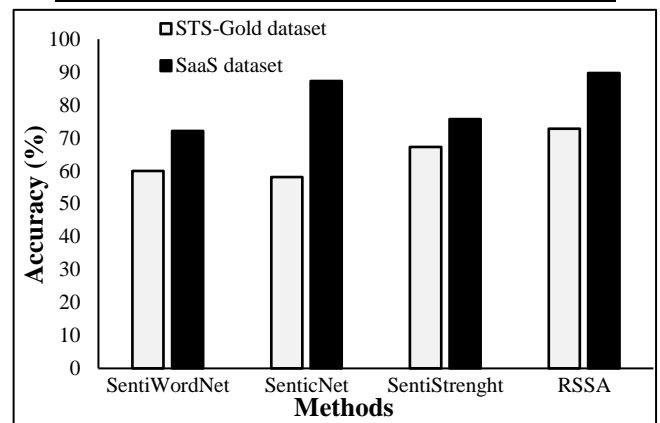


Fig.2. Comparative results of RSSA with baselines

For STS-Gold dataset, the accuracy obtained is 72.8% accuracy; the highest accuracy for this dataset among all three baselines and for SaaS dataset as it contains huge amount of conventional text, the accuracy is 89.69% accuracy, which is a very satisfactory.

5. CONCLUSION

Sentiment analysis is one of the challenging research areas in NLP, which its research has increased recently due to its huge business and individual applications. In literature, certain methods have been proposed for sentiment analysis. Supervised machine learning methods are giving high accuracy but these methods need training data and are domain dependent. Classifiers, trained on one domain may not be applicable for another domain. Whereas,

lexicon-based methods are not domain dependent. Although, building of lexicon is costly, but once constructed, it can be applied for a wide variety of domains, but still lexicon based methods are restricted to their dictionaries. Another problem with lexicon-based methods is their full-dependency on the presence of terms that explicitly reflect the sentiment, while in many cases the sentiment of a term is implicitly reflected by the semantics of its context. Therefore, a lexicon based method that is different from typical lexicons has been proposed and it outperforms baselines and the accuracy obtained is 72.8% for STS-Gold dataset, and 89.69% for SaaS dataset. The proposed approach offers dynamic sentiment score and strength according to the context and concept of terms (contextual and conceptual semantics). It is a context aware method that is applicable for wide variety of domains.

REFERENCES

- [1] Orestes Appel et al., "A Hybrid Approach to the Sentiment Analysis Problem at the Sentence Level", *Knowledge-Based Systems*, Vol. 108, pp. 110-124, 2016.
- [2] Bing Liu, "Sentiment Analysis and Opinion Mining, Synthesis Lectures on Human Language Technologies", Morgan and Claypool Publishers, 2012.
- [3] Bing Liu, "Sentiment Analysis and Subjectivity", Handbook of Natural Language Processing, 2010
- [4] Erik Cambria, Bjorn Schuller, Yunqing Xia and Catherine Havasi, "New Avenues in Opinion Mining and Sentiment Analysis", *IEEE Intelligent Systems*, Vol. 28, No. 2, pp. 15-21, 2013.
- [5] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai and Arvid Kappas, "Sentiment Strength Detection in Short Informal Text", *Journal of the American Society for Information Science and Technology*, Vol. 61, No. 12, pp. 2544-2558, 2010.
- [6] Noura Farra, Elie Challita, Rawad Abou Assi and Hazem Hajj, "Sentence-Level and Document-Level Sentiment Mining for Arabic Texts", *Proceedings of IEEE International Conference on Data Mining Workshops*, pp. 1114-1119, 2010.
- [7] A. Agarwal, F. Biadys and K.R. McKeown, "Contextual Phrase-Level Polarity Analysis using Lexical Affect Scoring and Syntactic N-Grams", *Proceedings of 12th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 24-32, 2009.
- [8] Robert Remus and Christian Hanig, "Towards Well-Grounded Phrase-Level Polarity Analysis", *Proceedings of 12th International Conference on Computational Linguistics and Intelligent Text Processing*, pp. 380-392, 2011.
- [9] Alec Go, Richa Bhayani and Lei Huang, "Twitter Sentiment Classification using Distant Supervision", CS224N Project Report, Stanford, pp. 1-12, 2009.
- [10] Albert Bifet and Eibe Frank. "Sentiment Knowledge Discovery in Twitter Streaming Data", *Proceedings of International Conference on Discovery Science*, pp. 1-15, 2010.
- [11] Alexander Pak and Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", *Proceedings of LREC Conference*, Vol. 10, pp. 1320-1326, 2010.
- [12] Apoorv Agarwal et al., "Sentiment Analysis of Twitter Data", *Proceedings of Workshop on Languages in Social Media*, pp. 30-38, 2011.
- [13] Efthymios Kouloumpis, Theresa Wilson and Johanna D. Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG!", *Proceedings of 5th International AAAI Conference on Weblogs and Social Media*, pp. 538-541, 2011.
- [14] Hassan Saif, Yulan He and Harith Alani, "Semantic sentiment analysis of twitter", *Proceedings of International Semantic Web Conference*, 2012.
- [15] Stefano Baccianella, Andrea Esuli and Fabrizio Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining", *Proceedings of LREC Conference*, Vol. 10, pp. 2200-2204, 2010.
- [16] Theresa Wilson, Janyce Wiebe and Paul Hoffmann. "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis", *Proceedings of Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 347-354, 2005.
- [17] James W. Pennebaker, Matthias R. Mehl, and Kate G. Niederhoffer, "Psychological Aspects of Natural Language use: Our Words, Our Selves", *Annual Review of Psychology*, Vol. 54, No. 1, pp. 547-577, 2003.
- [18] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining Lexicon-based and Learning based Methods for Twitter Sentiment Analysis", Technical Report, HP Laboratories, pp. 1-7, 2011.
- [19] Andrius Mudinas, Dell Zhang, and Mark Levene, "Combining Lexicon and Learning based Approaches for Concept-Level Sentiment Analysis", *Proceedings of 1st International Workshop on Issues of Sentiment Discovery and Opinion Mining*, pp. 1-7, 2012.
- [20] Ji Fang and Bi Chen, "Incorporating Lexicon Knowledge into SVM Learning to Improve Sentiment Classification", *Proceedings of Workshop on Sentiment Analysis where AI meets Psychology*, pp. 94-100, 2011.
- [21] Hassan Saif, Yulan He and Harith Alani, "Alleviating Data Sparsity for Twitter Sentiment Analysis", *Proceedings 2nd Workshop on Making Sense of Microposts*, pp. 2-9, 2012.
- [22] Hiroya Takamura, Takashi Inui and Manabu Okumura, "Extracting Semantic Orientations of Words using Spin Model", *Proceedings of 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics*, pp. 133-140, 2005.
- [23] Michael Speriosu et al., "Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph", *Proceedings of 1st Workshop on Unsupervised Learning in NLP Association for Computational Linguistics*, pp. 53-63, 2011.
- [24] Anthony Aue and Michael Gamon, "Customizing Sentiment Classifiers to New Domains: A Case Study", *Proceedings of Recent Advances in Natural Language Processing*, Vol. 1. No. 3, pp. 1-7, 2005.
- [25] Hassan Saif, Yulan He, Miriam Fernandez and Harith Alani, "Contextual Semantics for Sentiment Analysis of Twitter", *Information Processing and Management*, Vol. 52, No. 1, pp. 5-19, 2016.
- [26] Kyoungok Kim and Jaewook Lee, "Sentiment Visualization and Classification via Semi-Supervised Nonlinear

- Dimensionality Reduction”, *Pattern Recognition*, Vol. 47, No. 2, pp. 758-768, 2014.
- [27] Arnd Christian Konig and Eric Brill, “Reducing the Human Overhead in Text Categorization”, *Proceedings of 12th ACM International Conference on Knowledge Discovery and Data Mining*, pp. 598-603, 2006.
- [28] Bing Liu and Lei Zhang, “A Survey of Opinion Mining and Sentiment Analysis”, *Mining Text Data*, pp. 415-463, 2012.
- [29] Jose M.Chenlo and David E. Losada, “An Empirical Study of Sentence Features for Subjectivity and Polarity Classification”, *Information Sciences*, Vol. 280, pp. 275-288, 2014.
- [30] Andrea Esuli and Fabrizio Sebastiani, “Sentiwordnet: A Publicly Available Lexical Resource for Opinion Mining”, *Proceedings of LREC Conference*, pp. 1-5, 2006.
- [31] Maite Taboada et al., “Lexicon-based methods for Sentiment Analysis”, *Computational Linguistics*, Vol. 37, No. 2, pp. 267-307, 2011.
- [32] Erik Cambria, “An Introduction to Concept-Level Sentiment Analysis”, *Proceedings of Mexican International Conference on Artificial Intelligence*, pp. 1-6, 2013.
- [33] Peter D. Turney, “Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews”, *Proceedings of 40th Annual Meeting on Association for Computational Linguistics*, pp. 417-424, 2002.
- [34] Peter D.Turney and Michael L. Littman, “Measuring Praise and Criticism: Inference of Semantic Orientation from Association”, *ACM Transactions on Information Systems*, Vol. 21, No. 4, pp. 315-346, 2003.
- [35] Rodrigo Moraes, Joao Francisco Valiati and Wilson P. Gavião Neto, “Document-Level Sentiment Classification: An Empirical Comparison between SVM and ANN”, *Expert Systems with Applications*, Vol. 40, No. 2, pp. 621-633, 2013.
- [36] Tao Xu, Qinke Peng and Yinzhao Cheng, “Identifying the Semantic Orientation of Terms using S-HAL for Sentiment Analysis”, *Knowledge-Based Systems*, Vol. 35, pp. 279-289, 2012.
- [37] Xiaowen Ding, Bing Liu and Philip S. Yu, “A Holistic Lexicon-based Approach to Opinion Mining”, *Proceedings of International Conference on Web Search and Data Mining*, pp. 231-240, 2008.
- [38] Erik Cambria, Catherine Havasi and Amir Hussain, “SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis”, *Proceedings of FLAIRS Conference*, pp. 202-207, 2012.
- [39] Lisette Garcia Moya, Henry Anaya Sanchez and Rafael Berlanga Llavori, “Retrieving product Features and Opinions from Customer Reviews”, *IEEE Intelligent Systems*, Vol. 28, No. 3, pp. 19-27, 2013.
- [40] Kushal Dave, Steve Lawrence and David M. Pennock, “Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews”, *Proceedings of 12th International Conference on World Wide Web*, pp. 519-528, 2003.
- [41] J. Yi, T. Nasukawa, R. Bunescu and W. Niblack, “Sentiment Analyzer: Extracting Sentiments about a given Topic using Natural Language Processing Techniques”, *Proceedings of 3rd IEEE International Conference on Data Mining*, pp. 1-8, 2003.
- [42] Satoshi Morinaga et al., “Mining Product Reputations on the Web”, *Proceedings of 8th ACM International Conference on Knowledge Discovery and Data Mining*, pp. 341-349, 2002.
- [43] Simon Tong and Daphne Koller, “Support Vector Machine Active Learning with Applications to Text Classification”, *Journal of Machine Learning Research*, Vol. 2, pp. 45-66, 2001.
- [44] Janyce Wiebe, “Learning Subjective Adjectives from Corpora”, *Proceedings of International Conference on American Association for Artificial Intelligence*, pp. 1-6, 2000.
- [45] Fabrizio Sebastiani, “Machine Learning in Automated Text Categorization”, *ACM Computing Surveys*, Vol. 34, No. 1, pp. 1-47, 2002.
- [46] Rudy Prabowo and Mike Thelwall, “Sentiment Analysis: A Combined Approach”, *Journal of Informetrics*, Vol. 3, No. 2, pp. 143-157, 2009.
- [47] Arman Khadjeh Nassirtoussi et al., “Text Mining for Market Prediction: A Systematic Review”, *Expert Systems with Applications*, Vol. 41, No. 16, pp. 7653-7670, 2014.
- [48] Basant Agarwal and Namita Mittal, “Prominent Feature Extraction for Review Analysis: An Empirical Study”, *Journal of Experimental and Theoretical Artificial Intelligence*, Vol. 28, No. 3, pp. 485-498, 2016.
- [49] Asma Musabah Alkalbani et al., “Sentiment Analysis and Classification for Software as a Service Reviews”, *Proceedings of IEEE 30th International Conference on Advanced Information Networking and Applications*, pp. 53-58, 2016.
- [50] Peter D. Turney and Patrick Pantel, “From Frequency to Meaning: Vector Space Models of Semantics”, *Journal of Artificial Intelligence Research*, Vol. 37, No. 1, pp. 141-188, 2010.