IJCSE International Journal of Computer Sciences and Engineering Open Access

Survey Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Sentimental Analysis: A Survey

[1]Akanksha Mrinali, [2*]Sanjeev Kumar Sharma

[1] Oriental Institute of Science and Technology, Bhopal, India
[2] Oriental Institute of Science and Technology, Bhopal, India

[*]*Corresponding Author: sanjeevsharma@oriental.ac.in*

Available online at: www.ijcseonline.org

Accepted: 13/Jul/2018, Published: 31/Jul/2018

Abstract: Sentiment analysis (SA) is an intellectual and extracting process of the user's feelings and emotions. It is one of the promising fields of Natural Language Processing (NLP) such as text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study effective states and subjective information. Sentiment analysis is widely applied to a voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. In this paper, the latest algorithms of sentiment analysis applications are investigated and presented briefly. This paper also introduces a survey on the different techniques and challenges of sentiment analysis.

Keywords: Sentiment Analysis, Opinion Mining, Product Review, Data Review

I. Introduction

Sentiment Analysis is sometimes known as Opinion Mining (OM). It refers to the use of Natural Language Processing (NLP), Text Analysis, Computational Linguistics, and biometrics to systematically identify, extract, quantify, and study effective states and subjective information. Sentiment analysis is widely applied to a voice of the customer materials such as reviews and survey responses, online, and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. Sentiment analysis uses data mining processes and techniques to extract and capture data for analysis in order to distinguish the subjective opinion of a document or collection of documents, like blog posts, reviews, news articles, and social media feeds like tweets and status updates. Sentiment Analysis (SA) is the computational study of people's opinions, attitudes, and emotions toward an entity. The entity can represent individuals, events or topics. These topics are likely to be covered by reviews. The target of SA is to find opinions, identify the sentiments they express, and then classify their polarity as shown in Fig. 1.1.

Opinionative words or phrases

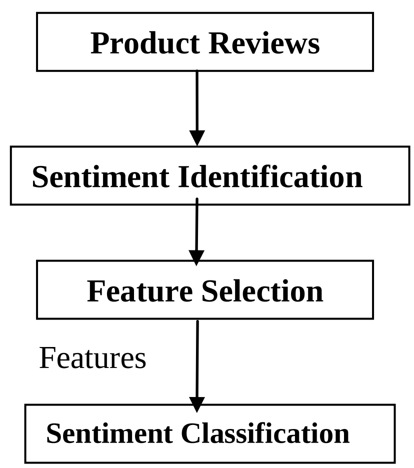

Fig.1.1 The process of Sentiment analysis on product reviews

Sentiment analysis is the computational study of people's opinions, appraisals, and emotions toward entities, events and their attributes. In the past few years, it attracted a lot of attention from both academia and industry due to many challenging research problems and a wide range of applications. Opinions are important because whenever we need to make a decision we want to hear others' opinions. This is not only true for individuals but also true for organizations. However, there was almost no computational study on opinions before the Web because there was little-opinionated text available. In the past, when an individual needed to make a decision, he/she typically asked for opinions from friends and families in the following manner.

a. Is this product review positive or negative?

b. Is this customer email satisfied or dissatisfied?
c. Based on a sample of tweets, how are people responding to this ad campaign/product release/news item?
d. How have bloggers' attitudes about the president changed since the election?

When an organization wants to find an opinion of the general public about its products and services, it conducted surveys and focus groups. However, with the explosive growth of the social media content on the Web in the past few years, the world has been transformed. People can now post reviews of products at merchant sites and express their views on almost anything in discussion forums and blogs and at social network sites. Now if one 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. Section I contains the introduction of sentimental analysis or opinion mining. Section II contains the related work already done in the field of sentimental analysis. Section III comprised of different approaches of sentiment analysis, Section IV contains the applications and challenges of Sentiment Analysis. Section V explains the various advantages of sentimental analysis. Section VI describes the limitations of Sentiment Analysis. Section VII concludes research work with future directions.

II. Background and Related Work

A lot of research work has been done in the field of Sentiment Analysis and it is briefly explained in subsequent paragraphs.

Somprasertsri and Lalitrojwong suggested a procedure for mining thing attributes and what's more decisions in light of reasoning about syntactic and what's more semantic information [1]. Through the usage of dependence associations and furthermore ontological data with probabilistic based models, consequences of examination exhibited that the proposed method was more versatile and furthermore viable.

Li et al., proposed a sentiment mining framework which mines accommodating supposition data from camera surveys by used Semantic Part Naming (SRL) and in addition extremity figuring strategies [2]. Highlights vocabulary and slant dictionary were built for mining traits and additionally full of feeling elements. At last, the complexities between positive and in addition negative conclusions were exhibited outwardly. The trial comes about demonstrated the framework is doable and compelling.

L. Recommended new item includes mining which used syntactic dependence information recently through separation of ostensible and in addition non-nominal terms [2]. Ostensible semantic structures would be parsed based on reliance trees together with the proposed display treating non-ostensible terms like the semantic neighbours of related ostensible terms. Semantic structure parsing would yield obstinate match stream with a couple of ostensible terms and in addition, their semantic neighbours, based on fine-grained item properties might be procured through co-bunching strategy through factorization. Assessments by and large group entropy, perplexity and also manual assessment demonstrated the advantages of the proposed system. Item characteristics to a great degree durable in fine-grain were removed in a programmed way.

A novel collective fluffy set hypothesis based separating model which incorporated subjective and additionally target information was proposed by Cheng and Wang [3]. The procedure offered exhaustive outcomes and in addition fathomed issues of customary Shared Sifting (CF) frameworks, new client and also a new thing. The trials showed that the novel system yielded outstanding proposals. CF was utilized for a few business frameworks, for example, IMDB, Netflix, and others with awesome achievement. CF framework's central idea is the age of suggestions based on comparable clients' past encounters. Client conclusions are arranged into objective and also subjective parts. The previous is acquired from lay clients while the last is requested from specialists.

A semi-robotized technique for making opinion word references in a few dialects was recommended by Steinberger et al., and it yielded abnormal state supposition lexicons for two dialects and programmed interpretation into a third dialect[4]. Words found in target dialect word list are ordinarily used like the word sense in that of the two source dialects.

A Senti Frame Net, an extension of Frame Net, as a reflection for SA was proposed by Ruppenhofer and Rehbein [5]. SA is featured by down to business centred strategies that use shallow techniques for strength yet rely upon datasets and procedures ad-hoc creation. Progression toward profound investigations depends onthe advancement of shallow delineations with semantically propelled rich information and also centred around different research branches and additionally mix of assets for making collaborations with related work in common dialect preparing.

Treatment of item includes extraction as a succession naming undertaking using discriminative learning models called Contingent Irregular Fields (CRFs) for handling it was proposed by Huang et al., [6]. It joined parts-of-discourse includes and also sentence structure highlights into a CRF learning system. It presented semantic information based and additionally distributional setting based likeness measurements for computing comparability between item property articulations for item highlight arrangement. An effective diagram pruning based order convention for characterizing the component articulations gathering into different semantic gatherings is proposed.

Examinations affirmed the adequacy of the approach as opposed to different strategies.

A generative Probabilistic Viewpoint Mining Model (PAMM) to recognize angles/subjects identified with class marks or clear-cut meta-data of a corpus was created by Cheng et al.,[7]. It lessened having angles framed by blending ideas of various classes; thus distinguished perspectives are simpler to be translated. Perspectives discovered additionally can recognize class: They can recognize a class from others. A proficient Desire Augmentation (EM) - calculation for parameter estimation is created. Results on audits of 4 drugs demonstrated that PAMM discovers preferred angles over different methodologies when estimated with mean point-wise shared data and grouping exactness.

Novel roads in OM and also SA that have profitable and tremendous measures of unstructured information with respect to general suppositions were recommended by Cambria et al., (where in history, current use, and in addition the eventual fate of Sentiment Mining and also SA was nitty-gritty[8].

Fundamental examinations systems using Korean Twitter information for mining worldly and additionally spatial patterns of brand pictures was demonstrated by Cho et al.,[9]. Freely accessible Korean morpheme analyser examined Korean tweets linguistically, and manufactured Korean extremity word references having a thing, descriptive word, verb, as well as root for breaking down ever y tweet's notion. Opinion arrangement is done by an SVM and additionally multi-ostensible Guileless Bayes classifier.

Veeraselvi and Saranya exhibited sentiment recognition and association subsystem, which had just been incorporated into the proposed bigger inquiry noting framework [10]. The subjectivity order framework utilized Hereditary Based Machine Learning (GBML) method that thought about subjectivity as a semantic issue. The grouping of a survey was evaluated through the normal semantic introduction of expressions in the audit which contain modifiers or intensifiers. Exploratory after the effects of the proposed methods was productive and created prominent assessments.

A novel assessment mining technique which misuses the semantic web-guided answers for upgrading results acquired by means of traditional NLP strategies and additionally SA methods were proposed by Peñalver-Martinez et al.,[11]. The proposed method's points are the change of highlights based assessment mining using ontology's at the element determination stage and also the arrangement of the novel vector butt-centric sis-based method for SA. This technique was executed and assessed in a certifiable motion picture re-see the setting and it yielded great results rather than customary strategies.

III. Approaches of Sentiment Analysis

There are various approaches of Sentiment Analysis [12]. The classification of these approaches of Sentiment analysis is shown in Fig. 1.2.

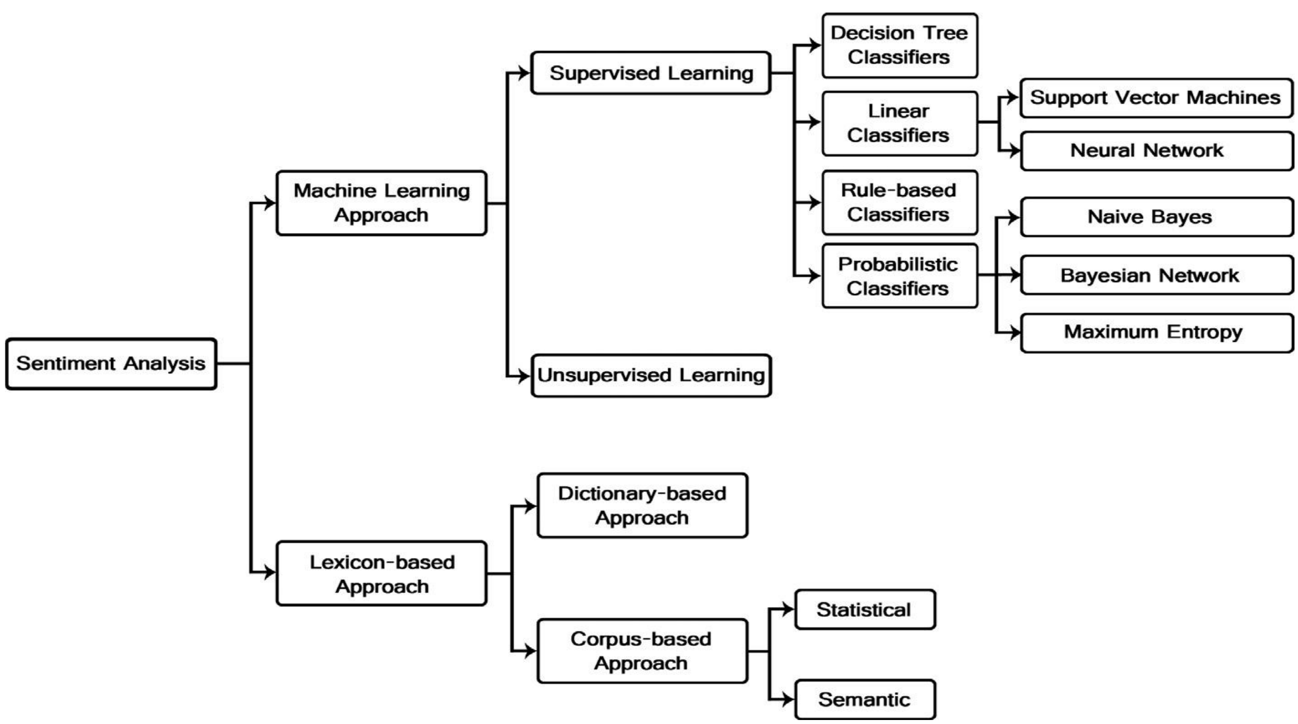

Fig. 1.2 Sentiment classification techniques

Sentiment analysis plays a great role in the area of researches done by many methods; there are many methods to carry out sentiment analysis. Still, many types of research are going on to find better alternatives due to its importance scenario. Some of the methods are discussed below.

A. Machine learning approach

Machine learning approach relies on the famous ML algorithms to solve the SA as a regular text classification problem that makes use of syntactic and/or linguistic features. Text Classification Problem Definition: We have a set of training records D = {X1, X2 ,..., Xn} where each record is labelled to a class. The classification model is related to the features in the underlying record to one of the class labels. Then for a given instance of the unknown class, the model is used to predict a class label for it. [13,14].

A.1. Supervised learning

The supervised learning methods depend on the existence of labelled training documents. There are many kinds of supervised classifiers in literature. In the next subsections, we present in brief details some of the most frequently used classifiers in SA.

a) Decision tree classifiers

The decision tree implementations in text classification tend to be small variations on the standard in order to mine the content structures of topical terms in sentence-level contexts by using the Maximum Spanning Tree (MST) structure to discover the links among the topical term ‘‘t’’ and its context words. Decision tree classifier provides a hierarchical decomposition of the training data space in which a condition on the attribute value is used to divide the data. The condition or predicate is the presence or absence of one or more words. The division of the data space is done recursively until the leaf nodes contain certain minimum numbers of records which are used for the purpose of classification. These are other kinds of predicates which depend on the similarity of documents to compare sets of terms which may be used to further partitioning of documents. The different kinds of splits are single attribute split which is use the presence or absence of particular words or phrases at a particular node of the tree in order to perform the split. Similarity-based multi-attribute split uses documents or frequent words clusters and the similarity of the documents to these words clusters in order to perform the split.

b)Linear Classifiers

Given X ¼ fx1 . . . . . . :xng is the normalized document word frequency, vector A ¼ fa1 . . . . . . and is a vector of linear coefficients with the same dimensionality as the feature space, and b is a scalar; the output of the linear predictor is defined as p ¼ A:X þ b, which is the output of the linear classifier. The predictor p is a separating hyper plane between different classes. There are many kinds of linear classifiers; among them is Support Vector Machines (SVM) which is a form of classifiers that attempt to determine good linear separators between different classes. Two of the most famous linear classifiers are discussed in the following subsections.

a. Support Vector Machines Classifiers (SVM)

It is a non-probabilistic classifier in which a large amount of training set is required. It has done by classifying points using a (d-1)-dimensional hyper plane. SVM finds a hyper plane with the largest possible margins. Support Vector Machines make uses of the concept of decision planes that define in such decision boundaries. A decision plane is one of the separates between a set of objects having a different class membership. An illustration is given in Fig. 1.3(a). It describes the objects belong to either class red or green, and the separating line defines the boundary. Here the original objects are Fig. 1.3(b) mapped or rearranged using a mathematical function known as a kernel and it is known as mapping or transformation. After transformation, the mapped objects are linearly separable and as a result, the complex structures having a curve to separate the objects can be completely avoided[15,16].

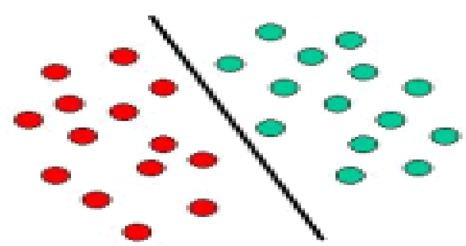

Fig 1.3 (a) Linear-classifier

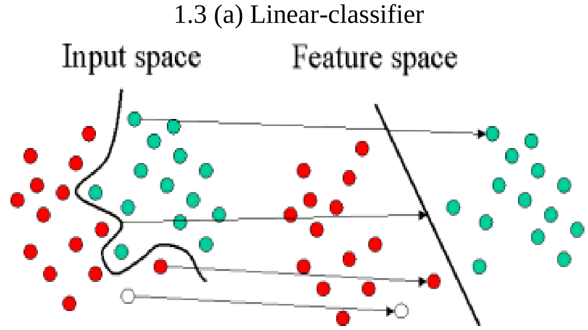

Fig. 1.3 (b). SVM illustration.

b. Neural Network (NN)

Neural Network consists of many neurons where the neuron is its basic unit. The inputs of the neurons are denoted by the vector where the word frequencies in the i^{th} document. There is a set of weights A which are associated with each neuron used in order to compute a function of its inputs f(_). The linear function of the neural network is pi ¼ A _ Xi. In a binary classification problem, it is assumed that the class label of X_i is denoted by Y_i and the sign of the predicted function pi yields the class label. Multilayer neural networks are used for non-linear boundaries. These multiple layers are used to induce multiple piecewise linear boundaries, which are used to approximate enclosed regions belonging to a particular class. The outputs of the neurons in the earlier layers feed into the neurons in the later layers. The training process is an additional complex because the errors need to be back-propagated over different layers. There are implementations of NNs for text data which are found.

c) Rule-Based classifier

of rules that represents a condition on the feature set expressed in disjunctive normal form while the other hand side is the class label. The conditions are on the term presence [17]. Term absence is rarely used because it is not revealing in sparse data. There are numbers of criteria in order to generate rules, the training phase constructs all the rules depending on these criteria. The most two common criteria are support and confidence. The support is the absolute number of instances in the training data set which are relevant to the rule. The Confidence refers to the conditional probability that the right-hand side of the rule is satisfied if the left-hand side is satisfied. Some combined rule algorithms were proposed in. Both decision trees and decision rules tend to encode rules on the feature space, but the decision tree tends to achieve this goal with a hierarchical approach. The main difference between the decision trees and the decision rules is that DT is a strict hierarchical partitioning of the data space, while rule-based classifiers allow for overlaps in the decision space.

d) Probabilistic classifiers

Probabilistic classifiers use by mixture models for classification. The mixture model assumes that each class is a component of the mixture. Each mixture component is a generative model that provides the probability of sampling a particular term for that component. These kinds of classifiers are also called generative classifiers. Three of the most famous probabilistic classifiers are discussed in the next subsections.

i. Naive Bayes Classifier (NB)

The Naive Bayes classifier is the simplest and most commonly used classifier. Naive Bayes classification model computes the posterior probability of a class, based on the distribution of the words in the document. The model works with the BOWs feature extraction which ignores the position of the word in the document. It uses the Bayes Theorem to predict the probability that a given feature set belongs to a particular label.

$$P(label|features) = \frac{\mathrm{P(label)} * \mathrm{P(features|label)}}{\mathrm{P(features)}}$$

P (label) is the prior probability of a label or the likelihood that a random feature set the label. P(features label) is the prior probability that a given feature set is being classified as a label. P (features) is the prior probability that a given feature set has occurred. Given the Naive assumption which states that all features are independent, the equation could be rewritten as follows:

$$P(label|features = \frac{\mathrm{P(label)} * \mathrm{P(f1|label)} * \ldots\ldots * \mathrm{P(fn|label)}}{\mathrm{P(features)}}$$

An improved NB classifier was proposed by Kang and Yoo to solve the problem of the tendency for the positive classification accuracy to appear up to approximately 10% higher than the negative classification accuracy. This creates a problem of decreasing the average accuracy when the accuracies of the two classes are expressed as an average value. They showed that using this algorithm with restaurant reviews narrowed the gap between the positive accuracy and the negative accuracy compared to NB and SVM. The accuracy is improved in recall and precision compared to both NB and SVM.

ii. Bayesian Network (BN)

The main assumption of the NB classifier is the independence of the features. The other extreme assumption is to assume that all the features are fully dependent. This leads to the Bayesian Network model which is a directed acyclic graph whose nodes represent random variables, and edges represent conditional dependencies. BN is considered a complete model for the variables and their relationships. Therefore, a complete joint probability distribution (JPD) over all the variables is specified for a model. In Text mining, the computation complexity of BN is very expensive; that is why it is not frequently. They proposed the use of multi-dimensional Bayesian network classifiers. It joined by different target variables in the same classification task in order to exploit the potential relationships between them. They extended the multi-dimensional classification framework to the semi-supervised domain in order to take advantage of the huge amount of unlabelled information available in the context. They showed their semi-supervised multi-dimensional approach outperforms the most common SA approaches, and that their classifier is the best solution in a semi-supervised framework because it matches the actual underlying domain structure.

iii. Maximum Entropy Classifier (ME)

The Maximum Classifier (known as a conditional exponential classifier) converts labelled feature sets to vectors using encoding. This encoded vector is then used to calculate weights for each feature that can then be combined to determine the most likely label for a feature set. This classifier is parameterized by a set of X {weights}, which is used to combine the joint features that are generated from a feature-set by an X {encoding}. In particular, the encoding maps each C {(feature set, label)} pair to a vector. The probability of each label is then computed using the following equation:

$$P(fs|label = \frac{\text{dot prod(weights. encode(rs, label))}}{\text{sum(dot prod(weights, encode(rs, l)forlinlablels)}}$$

ME classifier was used to detect parallel sentences between any language pairs with small amounts of training data. The other tools that were developed to automatically extract parallel data from non-parallel corpora use language specific techniques or require large amounts of training data. Their results showed that ME classifiers can produceuseful results for almost any language pair. This can be allowed to create the parallel corpora for many new languages.

A.2. Unsupervised learning

The unsupervised learning methods overcome these difficulties. Many research works were presented in this field including the work presented by Ko and Seo. They proposed a method that divides the documents into sentences and categorized each sentence using keyword lists of each category and sentence similarity measured. The unsupervised approach was used to automatically discover the aspects discussed in Chinese social reviews and also the sentiments expressed in different aspects. They used the model to discover multi-aspect global topics of social reviews, than they extracted the local topic and associated sentiment based on a sliding window context over the review text. They worked on social reviews that were extracted from a blog data set and a lexicon. They showed their approach obtained good topic partitioning results and helped to improve SA accuracy. It helped to discover multi-aspect fine-grained topics and associated sentiment.

B. Lexicon-based approach

Opinion words are employed in different sentiment classification tasks. Positive opinion words are used to express some desired states, while negative opinion words are used to express some undesired states. There is also opinion on phrases and idioms which together are called the opinion lexicon. There are three main approaches in order to compile or collect the opinion word list. The manual approach is very time consuming and it is not used alone. It is usually combined with the other two automated approaches as a final check to avoid the mistakes that resulted from automated methods. There are two automated approaches are presented in the following subsections.

a) Dictionary-based approach

The dictionary-based approach presented the main strategy of a small set of opinion words is collected manually with known orientations. Then, this set is grown by searching in the well-known corpora Word Netor thesaurusfor their synonyms and antonyms. The newly founded words are added to the seed list then the next iteration starts. The iterative process stops when no new words are found. After the process is completed, manually inspection can be carried out to remove or correct errors. The dictionary-based approach has a major disadvantage which is the inability to find opinion words with domain and context specific orientations. Qiu and He used a dictionary-based approach to identify sentiment sentences in contextual advertising. They proposed an advertising strategy to improve ad relevance and user experience. They used syntactic parsing and sentiment dictionary and proposed a rule-based approach to tackle topic word extraction and consumers' attitude identification in advertising keyword extraction. They worked on web forums from automotvieforums.com. These results demonstrated the effectiveness of the proposed approach on the advertising keyword extraction and ad selection.

b) Corpus-based approach

The Corpus-based approach helps to solve the problem to find opinion words with context-specific orientations. These methods depend on the syntactic patterns or patterns that occur together along with a seed list of opinion words to find other opinion words in a large corpus. They started with a list of seed opinion adjectives and used them along with a set of linguistic constraints to identify additional adjective opinion words and their orientations. The constraints are connectives like AND, OR, BUT, EITHER-OR. .....;the conjunction AND for example says that conjoined adjectives usually have the same orientation. This idea is called sentiment consistency, which is not always consistent practically. There are also hostile expressions such as but, however which are indicated as opinion changes. In order to determine if two conjoined adjectives are of the same or different orientations, learning is applied to a large corpus. Then, the links between adjectives form a graph and clustering is performed on the graph to produce two sets of words: positive and negative. Using the corpus-based approach alone is not as effective as the dictionary-based approach because it is hard to prepare a huge corpus to cover all English words, but this approach has a major advantage that can help to find domain and context specific opinion words and their orientations using a domain corpus. The corpus-based approach is performed using a statistical approach or semantic approach as illustrated in the following subsections:

a) Statistical approach

The statistical approach can be used to find the co-occurrence patterns or opinion words. This could be done by deriving posterior polarities using the co-occurrence of adjectives in a corpus, as proposed by Fahrni and Klenner. This technique resolves the problem of the unavailability of some words if the used corpus is not large enough. The polarity of a word can be identified by studying the occurrence frequency of the word in a large annotated corpus of texts. If the word occurs more frequently among positive texts, then its polarity is positive. If it occurs more frequently among negative texts, then its polarity is negative. If it has equal frequencies, then it is a neutral word. The similar opinion words frequently appear together in a corpus. Therefore, if two words appear together frequently within the same context, they are likely to have the same polarity. Therefore, the polarity of an unknown word can be determined by calculating the relative frequency of co-occurrence with another word. Statistical methods are used in many applications related to SA.

b) Semantic approach

The Semantic approach gives sentiment values directly and relies on different principles for computing the similarities between words. This principle gives similar sentiment values to semantically close words for example; Word Net provides different kinds of semantic relationships between the words used to calculate sentiment polarities. Word Net could be used for obtaining a list of sentiment words by iteratively expanding the initial set with synonyms and antonyms and then determining the sentiment polarity for an unknown word by the relative count of positive and negative synonyms of the word.The Semantic approach is used in many applications to build a lexicon model for the description of verbs, nouns, and adjectives to be used in SA as the work presented by Maks and Vossen. Their model described the detailed subjectivity relations among the actors in a sentence expressing separate attitudes for each actor. These subjectivity relations are labelled with information concerning both the identity of the attitude holder and the orientation (positive vs. negative) of the attitude. Their model included a categorization into semantic categories relevant to SA. It provided means for the identification of the attitude holder, the polarity of the attitude and also the description of the emotions and sentiments of the different actors involved in the text.

IV. Applications and Challenges

There are various applications of Sentiment Analysis discussed in subsequent paragraphs.

A. Recommender System: A recommender system or a recommendation system is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item. The goal of a Recommender System (RS) is to generate meaningful recommendations to users about items or products that might be of interest to them. This new area of research is gaining more importance mainly due to the effects of widespread use of social media.

B. Business Intelligence: Business needs to have a complete understanding of their customer's opinions and need to their products or services they offer, but they face the challenges of dealing of unstructured text from the source of customer's opinions and need. Consumer products and services sentiments as explained by a source of customer review and references, but a source for customer services, business intelligence, and product brand reputation management.

a) Are the customers satisfied with services, products, and support?
b) What customers think of products and services offered by competitors
c) What influences the market and how opinions propagate
d) What do the customers like
e) What problems do customers have
f) And, what additional features would the customer like to have and are willing to pay for.

Sentiment Analysis thus finds its use in Consumer Market for Product reviews, marketing for knowing consumer attitudes and trends, Social Media for finding a general opinion about recent hot topics in town, Movie to find whether a recently released movie is a hit.

a) Applications to Review-Related Websites Movie Reviews, Product Reviews etc.
b) Applications as a Sub-Component Technology Detecting antagonistic, the heated language in emails, spam detection, context-sensitive information detection etc.
c) Applications in Business and Government Intelligence Knowing Consumer attitudes and trends
d) Applications across Different Domains Knowing public opinions for political leaders or their notions about rules and regulations in place etc.

C. Sentiment Analysis for survey

A Revinate Surveys captures organic guest feedback with an open text format, which is then evaluated with the same technology as an online review. With the added volume of feedback, your decisions will be informed and strengthened by an even wider audience sample.

D. Challenges of Sentiment analysis

The challenges of Sentiment Analysis have been analyzed and divide into many levels such as document level, sentence level, word/term level or aspect level. It also divides the challenges into two types to ease to deal with them and focus on the degree of accurate meaning. This research discusses these sentiment challenges, the factors affecting them, and their importance. This research is based on two comparisons among the ten previous types of research in sentiment analysis to choose the suitable challenge for each research and to show their effects on the sentiment accuracy. The first comparison discusses the relationship between the sentiment analysis challenges and review structure. The second comparison examines a significance of solving the sentiment challenges to improve accuracy.

The first comparison is between the ten research papers. The target of this comparison is recognizing the relationship between the sentiment challenges and review structure and how to affect the sentiment results. Sentiment reviews structure becomes an essential factor which effects on the selecting the important challenges should the researchers face in their research by assuming the types of review format as in the following:

a) Structured Sentiments are found in formal sentiment reviews, but it targets the formal issues such as books or research. Because the writers are professional and Sentiments are notices about the scientific or factual issues.

b)Semi-Structured Sentiments lie on the range of between the formally structured sentiments and unstructured sentiments. These require understanding several issues about reviews. This type which depends on Pros and Cons is listed separately by the writer and the contents of Pros and Cons are usually short phrases.

c) Unstructured Sentiments are an informal and free text format, the writer does not follow any constraint. There is no formal separation of Pros and Cons and the content may consist of several sentences, where each sentence contains features and/or opinions. For the example below the unstructured reviews have the potential to provide more abundant and detailed opinion information than its counterpart).

Table 1 illustrates that the comparison of various strategies of sentiment analysis. The first factor is domain oriented, that requires having an orientation of the topic domain and its features or keywords to determine the fitting challenge for the research or application. The comparison relies on the relationship between the domains and the structure of the review. Another result is the negation that the most important challenge which has the greatest impact in any sentiment analysis and evaluation whether structured semi-structured or unstructured review. But the comparison shortcoming requires updatable research constantly to reach the suitable challenges easily and quickly. The second comparison explains the summary of sentiment challenges and how to improve the accuracy of each one based on the previous works. Its goal is identifying the most significant challenges in sentiment and how to improve its results relevant to the used techniques and explains the proposition of using the techniques with respect to the sentiment analysis (SA) challenge types (Theoretical or technical).

Table 1.1: Study of the sentiment challenges relevant to review structure.

Ref. No	Domain-oriented	Challenge-type	SA challenge	Review-structure
Yanfang et al. [18]	N	Theoretical	NLP overheads (Ambiguity	Semi-structured
Duyu et al. [19]	Y, social media	Theoretical	NLP overheads (Emotions	Unstructured
Marina et al. [20]	Y, the game on Amazon mechanical Turk	Theoretical	World knowledge	Unstructured
Svetlana et al. [21]	Y, tweets	Theoretical	NLP overheads (Short Abbreviations)	Unstructured
Qingxi and Ming [22]	N, online customers reviews	Theoretical	Spam and fake detection	Unstructured
Bing and Liang [23]	Y	Theoretical	Domain dependence	Un-structured, twitter
Robert [24]	Y	Theoretical and Technical	Negation +bipolar words	Semi-structured, sentences or topics documents
Stanislav[25]	Yes	Theoretical	Domain dependence	Unstructured conjunction with a predefined taxonomy of emotional terms
Alexandra et al. [26]	Y	Theoretical	Domain dependence	Structured, news articles
Christine et al. [27]	Y, tweets	Theoretical	NLP overheads (Sarcasm) +negation	Unstructured

Nathan and Ruihong [28]	Y, tweets	Theoretical	NLP overheads (Sarcasm)	Unstructured
Subhabrata and Pushpak [29]	Y, products	Technical	Extracting features or keyword	Semi-structured
Gizem et al. [30]	Y, trip advisor	Technical	Extracting features or keyword	Semi-structured

Table 2identifies the usage of each technique. The theoretical challenges use lots of techniques to improve the results by solving the selective sentiment challenges. The highest techniques usage in the theoretical type is parts-of-speech (POS) tagging and lexicon-based techniques. Bag-of-words (BOW) technique is the second technique. And the last one is Maximum entropy (ME) technique. But the results are different in technical sentiment challenge type the highest usage technique is n-gram technique since it is based on phrases and expressions. And the least technique used here is a lexicon-based technique. The results of this comparison are very important in choosing the suitable technique to solve the sentiment challenges to reach the highest accuracy.

Table 2examines several parameters relevant to the sentiment analysis challenges. These parameters are the lexicon type, domain-oriented, dataset, the technique used and the accuracy results. This comparison summarizes the effect of sentiment challenge solutions in analyzing and evaluating sentiment analysis accurately. The lexicon type in comparison in Table 2 refers to the language of the dataset and the size of the dataset. There are several available lexicons as Sentiment, How- Net, and Word Net. The used lexicon has the sentiment word and polarity. The polarity differs in the sentiment classification polarity level. This classification of polarity is divided into several class levels such as two levels (Positive, and Negative polarities), three levels as in the hierarchical level, or four level (_, Neutral, +, Mixed), and more specified classification into five levels (Very Negative, Negative, Neutral, Positive, Very Positive polarities).

Table 2: Study of several parameters effects on the sentiment challenges.

Ref. No.	**SA challenge Type**	**SA challenge**	**Technique used**	**Domain Oriented**	**Lexicon type**	**Dataset**	**Accuracy**
Lucie et al.[31]	Technical	Bi-polar Words	n-gram (uni and bi-grams	Y	HL and MPQA lexicon	Dataset of 1,600 Facebook messages	70%
Qingxi and Ming [32]	Theoretical	Spam and fake reviews	Combine lexicon and use shallow dependency parser	N, online customers reviews	SentiWordNet and MPQA	Store#364,	85.7% for sentiment method but word counting approach 76.7%
Ouyang et al. [33]	Theoretical	Domain Dependence	Emotion Dependency Tuple (EDT-improved (BOW) TF-IDF and cross entropy, space vector model	N	Chinese	COAE2014 dataset	60%
Duyu et al. [34]	Technical	Technical Nlp overheads	Fine-grained emotions	Y	Chinese lexicon	35,000 tweets about the Sichuan	80%,

		(emotions)				earthquake	
Svetlana et al. [35]	Theoretical	Domain Dependence	SemEval-2013	Y	Tweets and MPQA English	2000 positive words and 4700 negative words, also the popular MPQA	Improve accuracy and F-measure about 13% from baseline to reach 69%
Chetan and Atul [36]	Technical	Huge lexicon	Lexicon based technique	Y	6,74,412 tweets	The polarities of the words in the dictionary are set according to a specific domain,	73.5%
Ivan et al. [37]	Technical	Bi-polar Words	Combination of features (n-grams) and preprocessing techniques (unsupervised stemming and phonetic transcription	Y	English Facebook	Facebook dataset containing 10,000 posts	69%
Alexandra et al. [38]	Theoretical	Domain Dependence	Word Net-lexicon based	Y	News reviews	Newspaper articles (the set of 1292 quotes	82% improve the base line 21%
Andrius et al. [39]	Technical	Huge lexicon	Bag-of-word11s SVM.	Y, CNET, IMDB movie reviews	Santi	The first dataset Software Review, second dataset Movie Reviews	82.30%

V. Advantages

a) By listening to and analyzing comments on Facebook and Twitter, local government departments can guess the public sentiment towards their department and the services they provide and use the results to improve services such as parking and leisure facilities, local policing, and the condition of roads.

b) Universities can use sentiment analysis to analyze the student feedback and comments garnered either from their own surveys or from online sources such as social media. They can then use the results to identify and address any areas of student dissatisfaction, as well as identify and build on those areas where students are expressing positive sentiments [40].

c) Sentiment analyzes the customer reviews on sites like Trip Advisor and Yelp, hotels, and restaurants, which can not only manage their reputations by improving the services offered but can also; guess the general customer attitude to their business or brand. Businesses can compare their results with their competitors to better understand people's attitude to their business. They can identify where they may be excellent, or identify where there's room for improvement compared to the competitor. They can also conduct market research into general sentiment around key issues, topics, products, and services, before developing and launching their own new services, products or features.

d) The success of your marketing campaign is not measured only by the increase in the number of followers, likes, or comments but also lies in how many positive discussions you are able to help facilitate amongst your customers. By doing sentiment analysis, you can see how much positive or negative discussions have occurred amongst your audience. By combining the quantitative and the qualitative measurements, you can measure the true Rate of Interest (ROI) of your marketing campaign.

e) Sentiment Analysis helps us complete our market research by getting to know "what customers' opinions are about our products/services and how we can align our products/services quality and features with their tastes?"

f) Once our customer purchases our product, we want to keep them loyal to our brand as long as possible and can be an evangelist for our brand. Customers can essentially become more effective for us. That is why it is of utmost importance to have the best customer service in place and keep our current customers happy. There are many factors that contribute to great customer service, such as on-time delivery, being responsive in social media, and adequate compensation for product's errors. Sentiment analysis can pick up negative discussions, and give real-time alerts so that can respond quickly. If customers complain about something related to the brand, the faster react, the more likely customers will forget about being annoyed in the first place, and be satisfied with the great customer service. Sentiment analysis is a part of social listening to manage complaints helps to avoid leaving customers feeling ignored and angry.

g) Constant monitoring of what is currently happening in social media conversations also helps to prevent or at least lighten the damage of online communication crisis. The crisis may start from the product's bad quality, unacceptable customer service, or other serious social issues such as environmental harm, animal cruelty or child labour usage in emerging markets. If it doesn't manage customer complaints fast enough, the conversation can go viral and lead to a huge crisis that might not be able to cope back from. Uber (cab) is a great example of a recent crisis that viral in social media[41].

h) Another advantage of sentiment analysis is to boost sales revenue. Since the increase in sales revenue is the final outcome of successful marketing campaigns, improved products/service quality, and customer service, which can be achieved with sentiment analysis.

i) As of June of 2017, Facebook achieved a high-flying status with its two billion users, while Instagram has reached 700 million users and Twitter about 317 million. Even though Twitter had a relatively small audience base, tweets provide a nicely controlled domain to be targeted for online research.

j) While 20,000 mentions for your brand on Twitter last week is a spirit lifter, you're in big trouble if these mentions are all about complaints. You can determine the state of your marketing strategy on social channels through sentiment analysis.

VI. Limitations

There are some limitations of sentiment analysis, which are given below.

a) Sentiment analysis tools can identify and analyze many pieces of text automatically and quickly. But computer programs have problems recognizing things like sarcasm and irony, negations, jokes, and exaggerations - the sorts of things a person would have little trouble identifying. And failing to recognize these can skew the results [42].

b) 'Disappointed' may be classified as a negative word for the purposes of sentiment analysis, but within the phrase "it wasn't disappointed", it should be classified as positive.

c) We would find it easy to recognize as sarcasm the statement "It' really loving the enormous pool at my hotel!", if this statement is accompanied by a photo of a tiny swimming pool; whereas an automated sentiment analysis tool probably would not, and would most likely classify it as an example of positive sentiment.

d) With short sentences and pieces of text, for example like those you find on Twitter especially, and sometimes on Facebook, there might not be enough contexts for a reliable sentiment analysis. However, in general, Twitter has a reputation for being a good source of information for sentiment analysis, and with the new increased word count for tweets, it's likely it will become even more useful.

e) So, automated sentiment analysis tools do a really great job of analyzing text for opinion and attitude, but they're not perfect.

f) When we're using a tool like typically to analyze our text to see if it conveys the sentiment want for our readers/audience, combine the results it gives us with our human judgment to identify anything the tool may not be able to easily determine.

g) Types highlight phrases in our text by positive and negative sentiment, making it super easy for to see where our document is either expressing exactly the sentiments we want it to or where we may need to make some changes.

VII. Conclusion

Sentiment analysis is the field of study that analyzes people's sentiments, attitudes, or emotions towards certain entities. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Sentiment Analysis handles the evaluation of opinions of various kinds, broadly categorized into positive or negative opinions where the reviews show that various attributes, as well as classification protocols, combine effectively for overcoming individual shortcoming and benefits. Future work is required for improving performance metrics. The most important difficulty is utilizing other languages, handling negations as well as providing opinion summaries on the basis of product attributes, handling implicit product attributes, the complexity of sentences and so on. In our research work, the aspect of feature selection and classification of opinions is focused on techniques optimizing and the classifier are proposed in this work. Various sentiment analysis methods and its different levels of analysing sentiments have been studied in this paper.

Within the world of internet, the majority of people depend on social networking sites to get their valued information, analysing the reviews from these blogs will yield a better understanding and help in their decision-making that can be found by the relationship between the proportion of sentiment techniques usage in theoretical and technical types to solve sentiment challenges.

References

[1] Som Prasertsri, G & Lalitrojwong, P , 'Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization', J. UCS, vol. 16, no. 6, pp. 938-955,2010.

[2] Li, X, Dai, L & Shi, H ,"Opinion mining of camera reviews based on Semantic Role Labelling", in Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on, vol. 5, pp. 2372-2375,2010.

[3]Wang, D, Zhu, S & Li, T, Sum View "A Web-based engine for summarizing product reviews and customer opinions", Expert Systems with Applications, vol. 40, no. 1, pp. 27-33, 2013.

[4] Lin, C, He, Y, Everson, R & Rüger , S, "Weakly supervised joint sentiment-topic detection from text", Knowledge and Data Engineering, IEEE Transactions on, vol. 24, no. 6, pp. 1134-1145,2012

[5] Ruppenhofer, Josef and Rehbein, Ines," Semantic frames as an anchor representation for sentiment analysis", Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis,2012.

[6] Cleland-Huang, J & Mobasher, B, "Using data mining and recommender systems to scale up the requirements process", in Proceedings of the 2nd international workshop on Ultra-large-scale software-intensive systems, pp. 3-6, 2008.

[7] Hai, Z, Chang, K & Cong, G, "One seed to find them all: mining opinion features via association", in Proceedings of the 21st ACM international conference on Information and knowledge management, pp. 255-264, 2012.

[8] Cambria, E, Scholler, B, Xia, Y & Havasi, C, "New avenues in opinion mining and sentiment analysis", IEEE Intelligent Systems, no. 2, pp. 15-21, 2013.

[9] Cho, KS, Jung, NR & Kim, UM, "Using word map and score based weight in opinion mining with map reduce", in Service-Oriented Computing and Applications (SOCA), 2010 IEEE International Conference on, pp. 1-4, 2010.

[10] Veeraselvi, S & Saranya, C," Semantic orientation approach for sentiment classification", in Green Computing Communication and Electrical Engineering (ICGCCEE), 2014 International Conference on, pp. 1-6, 2014

[11] Peñalver-Martinez, I, Garcia-Sanchez, F, Valencia-Garcia, R, Rodríguez-García, MÁ, Moreno, V, Fraga, A & Sánchez-Cervantes, JL, "Feature-based opinion mining through Ontology's", Expert Systems with Applications, vol. 41, no. 13, pp. 5995-6008, 2014.

[12] Angulakshmi, G & ManickaChezian, R, "An analysis on opinion mining: techniques and tools", International Journal of Advanced Research in Computer Communication Engineering, vol. 3, no. 7, pp. 7483-7487, 2014.

[13] Ortigosa-Herna´ ndez Jonathan et al. "Approaching sentiment analysis by using semi-supervised learning of multi-dimensional classifiers", Neuron computing, pp.98–115,2012.

[14] Kaufmann JM. JMaxAlign, "A Maximum Entropy Parallel Sentence Alignment Tool", In the Proceedings of COLING'12: Demonstration Papers, Mumbai. pp. 277–88,2012.

[15] Ankush Sharma, Aakanksha, Assistant Professor, Department of C.S.E, Chandigarh University Gharuan, India, International journal of Advanced Research in Computer and Communication Engineering, " A Comparative Study Of Sentiments Analysis Using Rule Based and Support Vector Machine " volume 3,2014.

[16] Walaa Meddhat , Ahmed Hassan ,Hoda Korashy "Sentiment analysis algorithms and applications: A survey, Ain Sham University, Faculty of Engineering, Computer & Systems Department, Egypt 19 April 2014.

[17] Chin-Shrng Yang, Hsiao-Ping Shih, Department of Information Management, Yuan Ze University, ChangLi, Taiwan," A Rule-Based Approach For Effective Sentiment Analysis" PACIS 2012.

[18] Yanfang, C., Pu, Z., Anping, X., "Sentiment analysis based on an expanded aspect and polarity-ambiguous word lexicon", Int. J. Adv. Comput. Sci. Appl, Vol. 6 pp.2,2015.

[19] Duyu, T., Bing, Q., Ting, L., Qiuhui, S., "Emotion analysisplatform on Chinese microblog", CoRR J,2014.

[20] Marina, B., Claudiu, C.M., Boi, F,"Acquiring commonsense knowledge for sentiment analysis using human computation. In:Proceeding", WWW'14 Companion, Seoul, Korea,2014.

[21] Svetlana, K., Xiaodan, Z., Saif, M.M., "Sentiment analysis ofshort informal texts", J. Artif. Intell. Res. Vol.50,2014.

[22] Qingxi, P., Ming, Z., "Detecting spam review through sentiment analysis", J. Software, Vol.9, pp.8,2014.

[23] Bing, X., Liang, Z., "Improving twitter sentiment analysis with topic-based mixture modeling and semi-supervised training", In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers). Association for Computational Linguistics, Baltimore, Maryland, USA,2014.

[24] Robert, R., "Modelling and Representing Negation in Data-driven Machine Learning-based Sentiment Analysis", ESSEM@ AI* IA,2013.

[25] Stanislav, B., "An Approach to Feature Extraction for Sentiment Analysis of News Texts",2013.

[26] Alexandra, B., Ralf, S., Mijail, K., Vanni, Z., Erik, V.D.G., Matina, H., Bruno, P., Jenya, B., "Sentiment analysis in the news", In: Proceedings of the Seventh International Conference on Language Resources and Evaluation LREC,Vol.10,2013.

[27] Christine, L., Florian, K., Antal, V.D.B., "The perfect solution for detecting sarcasm in tweets #not", In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis. Association for Computational Linguistics, Atlanta, Georgia, pp. 29–37,2013.

[28] Nathan, G., Ruihong, H., "Sarcasm as the contrast between a positive sentiment and negative situation", In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing EMNLP, 2013.

[29] Subhabrata, M., Pushpak, B., "Feature Specific Sentiment Analysis for Product Reviews", CICLing, part I, LNCS 78181, Springer-Verlag, Berlin Heidelberg.

[30] Gizem, G., Berrin, Y., Dilek, T., Yucel, S., "New features for sentiment analysis: do sentences matter?",In: SDAD 2012 The 1st International Workshop on Sentiment Discovery from Affective Data, pp.5,2012.

[31] Lucie, F., Eugen, R., Daniel, P., "Analysing domain suitability of a sentiment lexicon by identifying distributional bipolar words", In Proceedings of the Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis. EMNLP,2015.

[32] Qingxi, P., Ming, Z., "Detecting spam review through sentiment analysis", J. Software,Vol.9,pp.8,2014.

[33] Ouyang, C., Zhou, W., Yu, Y., Liu, Z., Yang, X., Topic "sentiment analysis in Chinese news" Int. J. Multimedia Ubiquitous Eng., Vol.9 pp.11- 385,2014.

[34] Duyu, T., Bing, Q., Ting, L., Qiuhui, S., "Emotion analysis platform on Chinese micro blog", Co RR J, 2014.

[35] Svetlana, K., Xiaodan, Z., Saif, M.M., "Sentiment analysis ofshort informal texts" J. Artif. Intell. Res, Vol.50, 2014.

[36] Chetan, K., Atul, M., "A scalable lexicon based technique for sentiment analysis", Int. J. Foundations Computer Sci. Technology, Vol.4, pp.5, 2014.

[37] Ivan, H., Tomas, P., Josef, S., "Sentiment analysis in Czech social media using supervised machine learning", In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis. Association for Computational Linguistics, Atlanta, Georgia,pp.65–74,2013.

[38] Alexandra, B., Ralf, S., Mijail, K., Vanni, Z., Erik, V.D.G., Matina, H., Bruno, P., Jenya, B., "Sentiment analysis in the news", In Proceedings of the Seventh International Conference on Language Resources and Evaluation LREC, Vol.10,2013.

[39] Andrius, M., Dell, Z., Mark, L., "Combining lexicon and learning based approaches for concept-level sentiment analysis", In WISDOM' Beijing, China, pp.12, 2012

[40] Ivan, H., Tomas, P., Josef, S., "Sentiment analysis in Czech social media using supervised machine learning", In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis. Association for Computational Linguistics, Atlanta, Georgia,pp.65–74,2013.

[41] Tsytsarau Mikalai, Palpanas Themis, "Survey on mining subjective data on the web", Data Min Knowledge Discovery, Vol.24, pp.478–514,2012.

[42] Hatzivassiloglou V, McKeown K. "Predicting the semantic orientation of adjectives", In Proceedings of the annual meeting of the Association for Computational Linguistics ACL,pp.97, 1997.