

Sentiment Detection, Recognition and Aspect Identification

Salma Faiz Alasmari
Computer Science Department
King Khalid University
Abha, Saudi Arabia

Mohammed Dahab
Computer Science Department
King Abdulaziz University
Jeddah, Saudi Arabia

ABSTRACT

Sentiment analysis is the study of people's opinions, attitudes, feelings, and emotions discuss any object such as entities, events, topics, product, issues, services, etc. are respected for extraction of useful subjective information out of the text. The task is challenging and considered for customers and producers, for instance, Customers need to have general ideas about products, and companies need to know customers' needs and the opportunity to investigate and analyze those opinions towards their products and services. The increasing growth of the social network sites generating a big resource of social data. So, researchers considering about harnessing the effectiveness of social media and sentiment analysis. These recent studies classified according to their contributions in the different SA methods. The main target of this paper is to give the illustration of the current trend of research in the sentiment analysis Detection, Recognition, Aspect Identification.

Keywords

Sentiment Analysis, Sentiment Detection, Sentiment Recognition, Aspect identification, Text Mining

1. INTRODUCTION

Sentiment analysis is a kind of text classification that classifying documents based on the direction of the object they contain. It thus plays an important part of Natural Language Processing. It is applications' particularly have spread to every conceivable domain, from company's products, services, healthcare, and financial activities to social media events. It is known as the process of detecting the contextual polarity of the text. It determines whether given text is positive, negative or neutral. It is known as opinion mining too. It is using to derives the appraisal, emotion or attitude of the speaker. The social networks act as a medium where the users can post many opinions a day, and these blogs can use for classification. A lot of research work held in the field of sentiment analysis due to its significance in the marketing level competition and the changing needs of the people. Sentiment analysis requires the usage of a training dataset for evaluating its

In general, sentiment analysis can consider a classification process as illustrated in Fig. 1. and has been investigated basically at three levels [1]. In document-level, the primary task is to classify whether the entire opinion document represents a positive or negative sentiment. This level of analysis supposes that each document represents opinions of a single entity. In sentence-level, the primary task is to check whether each sentence represented a positive, negative, or neutral opinion. This level of analysis is approximately related to subjectivity classification, which distinguishes objective sentences that express factual information from subjective sentences that express an opinion. Document-level and the sentence-level studies do not discover what exactly people liked and did not like. Aspect level performs the fine-grained

analysis. As an alternative to looking at language constructs (documents, paragraphs, sentences, clauses or phrases), aspect-level directly looks at the opinion itself [2][40].

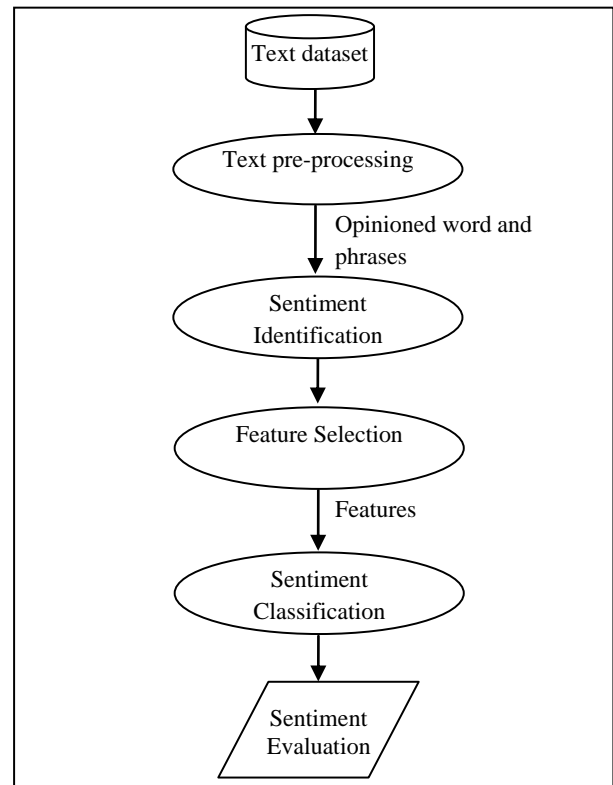


Fig 1: Sentiment Analysis process [41]

One characteristic of the sentiment analysis generation process is that the overall sentiment is not determined by a few individuals but by an aggregation of all the available sentiments. Therefore, it is necessary to be able to analyze user-generated content automatically. This paper, is a survey include a literature review of some existing studies in Sentiment Detection, Recognition and Aspect Identification. Sentiment detection is a subfield of natural language processing and a significant process of text analysis which the primary target is identifying the presence of emotions within the specific text. The ability to distinguish what type of thing included in that sentence called the sentiment recognition. Finally, the sentiment aspect identification considered in details of the object expresses opinions in the text.

The remainder of this paper organized as follows. Section 2 discusses some existing work in sentiment detection, Section 3 identifies sentiment recognition and discusses some current work in, Section 4 identifies aspect identification and discusses some existing studies, finally, Section 5 concludes the paper.

2. THE PROBLEM OF SENTIMENT ANALYSIS

In this section, it discusses the Sentiment Analysis problem, which allows us to identify a structure from the unstructured text and to decide a unified context for the current study that consists of three parts: Sentiment Detection, Recognition and Aspect identification.

In general, sentiment analysis is the process of identifying, extracting opinions, sentiments and emotions expressed in a specific text, and study subjective information, especially for determining whether the people's feedbacks towards a particular product, a service or a person are positive, negative, or neutral.

By use this review on Amazon's item to present the problem "I bought a Camera last month and was fabulous.". It observed that the sentence expressed positive sentiment or emotion and the target of the sentiment is the camera. Also, the holder in a sentence is the writer of the review "I." With this example, It helps to define the sentiment analysis problem. In the web, it described a prototype that publishes the precise definition of sentiments discussed by users:

A post identified as a text created by the people on social networks (e.g.. YouTube, Facebook, LinkedIn, Twitter), generating an opinion on a particular object. A post appears with reviews, comments, likes or location for classification purposes.

A sentiment is a feeling, thought or idea relied on a user expression about an object. It is composed of an objective, a polarity, and emotions. Sentiments expressed by posts of the user regarding a particular object [3]. An opinion is the general essential sentiment of published text. It aggregates the sentiments (document, sentence- or feature-level) of the post. The characteristics of an opinion generate those of the sentiment and consist of reviews, likes or comments [3] [4].

Formally, the terminology *object* will be used to indicate the entity that has discussed in the text. An object *o* definition [5] is an entity which can be a product, person, event, organization, or issue. It related to a pair, $o: (C, T)$, where *C* represented a set of *components*, *sub-components*, and so on, and *T* is a set of *attributes* of *o*. Also, every single component has its individual set of sub-components and attributes. An example for an entity, A Canon is the product of digital Camera. It has a set of components, e.g., lens, battery, and modes. Also, a set of attributes, e.g., photo resolution, filtering, and zooming. The memory component also has its set of attributes, e.g., memory capacity, and memory speed.

The opinion defined as a positive or negative sentiment or emotion (called sentiment orientations) expressed about an entity from an opinion holder [6]. Liu and Zhang [7] identified opinion as essential five elements ($E_i, A_{ij}, OO_{ijkl}, H_k, T_1$). E_i represents the name of an entity, A_{ij} is an attribute of E_i , the orientation of the opinion is OO_{ijkl} about attribute A_{ij} of entity E_i , H_k represents the opinion holder, and T_1 is the time when the opinion is expressed by the opinion holder H_k . This definition converting unstructured text into structured data and provides the accurate information for opinions analysis [7]. Dahab and Al Asmari [8] discussed two approaches of sentiments which are called regular sentiment and comparative sentiment. Then categorized sentiments based on how they included in text into explicit sentiment and implicit sentiment.

The Objective of sentiment analysis presented a collection of opinionated documents *D*, discover all sentiment essential five elements ($E_i, A_{ij}, OO_{ijkl}, H_k, T_1$) in document *D*. Based on the previous review,

Next target is to discuss some main research topics of sentiment analysis, and this study aims to discuss the following tasks:

- Task 1 (Sentiment Detection): Extract sentiment within a particular text in document *D*.
- Task 2 (Sentiment Recognition): Extract the entity name and the entity types.
- Task 3 (Sentiment Aspect identification): Extract all features of the entity includes of a sentiment polarity (positive or negative).

The SA aspect identification focuses more on feature-level sentiment analysis, which relied on the idea that an opinion contains a sentiment polarity (positive or negative) and a feature of products.

3. SENTIMENT DETECTION

Nowadays the interesting in investigating social media data has been increasing because it offers a platform that allows users to express their opinions towards things. Sentiment detection is a significant process of text analysis which the primary target is detected sentiment within a particular text. The notion of sentiment detection was introduced first by [9] that date back to 1998. Sentiment detection has applied to a variety of social media, especially to reviews of products or services, tweets, etc. This section is focusing on existing research on sentiment detection overall.

3.1 Sentiment Detection Approaches

In general, Sentiment detection considering machine learning algorithms that divided into a supervised method or an unsupervised (statistically) method. Supervised methods aim at generating sentiment classification models. Whereas, in unsupervised methods the sentiment derived from the document using document's statistical features or employing available lexicons including emotional words [10]. Existing research on sentiment detection is mainly carried out in online reviews. Next, will focus on some studies considered in sentiment detection based on a machine learning or lexicon.

3.1.1 Sentiment Detection Based on a Machine Learning

Machine learning out of a text mining can consider as the process of developing programs with the ability to search through data to look for patterns and can change when exposed to new data .Therefore sentiment detection classification based on a machine learning approach can clearly express as a supervised learning technique with three categories, positive, negative and neutral. Training and testing data used in the existing research are commonly product reviews [7].

The currently supervised learning classification techniques as shown in fig.2 can perform to sentiment detection, e.g., naive Bayesian, nearest neighbor and support vector machines. Hernández-Farías, et al. [11] presented an irony detection model capable of detecting irony sentiments through ironic and non-ironic sentences. It considered a set of structural features that associate both text characteristics and knowledge. It applied a set classifiers such as Naive Bayes, Decision Tree, Neural Network, Maximum Entropy, Support

Vector Machine for irony terms detection task. The model presented the capability to distinguish between ironic and non-ironic sentences.

Subsequent studies in this survey used many more features and techniques in learning [10]. As most machine learning applications, the primary task of sentiment detection is to study a practical set of features. Some of the current features are listed below.

- Term frequency: these features are specific words or n-grams words and their frequency. For example, Hartmann et al. [1] provided a sentiment detection system based on character n-grams with a Naive Bayes classification to determine sentiments of texts. Their study used the movie review database as a test. They demonstrated that character n-grams achieving a high level and capable results comparable to highly advanced methods because it required a minimal amount of knowledge and gave fewer assumptions about the data.
- Part of speech (POS) tagging: It found in many types of research that adjectives are important indicators of opinions. Thus, adjectives have treated as special features. Maynard and Funk [12] developed a system able to determine which higher quality opinions from a set of pre-election tweets. The system could detect sentiments by use extra-linguistic contextual information and different advanced NLP methods such as tokenization, part-of-speech tagging, morphological analysis, sentence splitting, and so on.
- Opinion word-based features: In general, the users discuss their opinions by using diverse positive with

negative polarities or set of emotions in their word. Also, some opinion words are adjectives, adverbs, nouns or verbs can express opinions too [7]. Specifying the overall sentiment of text usually used lexicon-based features, but they couldn't detect more accurate features that are inherent in the public language appeared on social websites. Giatsoglou, et al. [10] proposed word embedding-based methods and aim to detect semantic and syntactic features of words out of text collections in a language independent method.

- Negations: The existing of negation words are regularly modify the opinion direction, but some cases do not represent negation. For example, of negative sentence "I didn't like this phone." In this example "not" is not modify the opinion direction "not only but also" [7].

The applying of unsupervised learning classification techniques on opinion words, phrases or documents would be completely natural for sentiment detection. For example, Lin, et al. [13] developed three Bayesian models for unsupervised sentiment classification based on the document, named the latent sentiment model, the joint sentiment topic model, and the Reverse-JST model. The result shows that both JST and Reverse-JST

always performs worse than JST meaning that the JST model is more suitable for sentiment topic detection. The sentiment detection based on word embeddings and the use of a Deep Convolutional Neural Network developed with the help of an unsupervised neural language model, such as in Severyn and Moschitti [14].

Whereas Turney [15] introduced an unsupervised learning classification algorithm for detect sentiments based on the phrases in the review that include adjectives or adverbs.

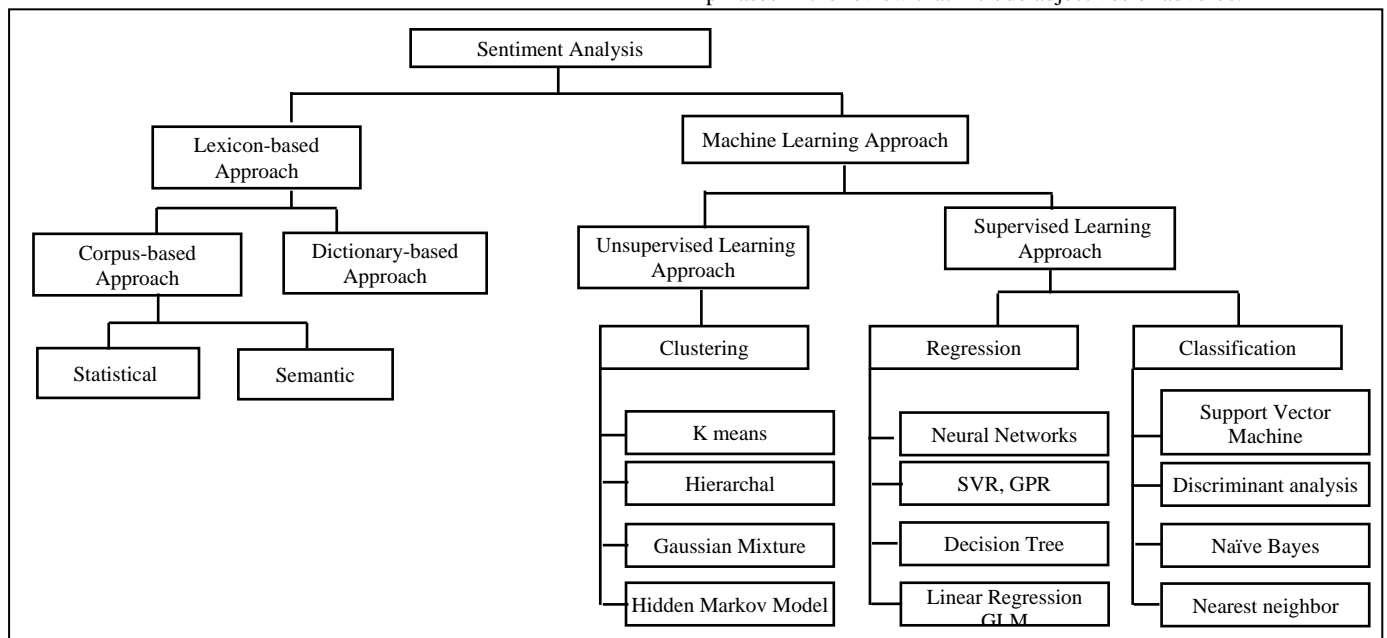


Fig 2: Sentiment Analysis Classification Techniques

The algorithm first extracts phrases including adjectives or adverbs as indicators of opinions. Next, estimates the semantic orientation SO of the extracted phrases calculated as

the mutual information between a specific phrase and the term “excellent” minus the mutual information between a particular phrase and the term “poor.” The review is classified as

recommended if the average SO of all phrases in the review is positive, otherwise not recommended. Concluding classification achieves accuracies on reviews from a set of domains range from 84% for automobile reviews to 66% for movie reviews.

To summarize, we can see that the benefit of sentiment detection especially based on opinion document is that it provides a general opinion on an entity or topic. Whereas the drawbacks are that it does not provide enough details of people positive and negative opinions and it is not simply applicable on forum and blogs which contain multiple entities.

3.1.2 Sentiment Detection Based on Lexicon

The lexicon-based feature sentiment detection approach relies on the co-occurrence of a sentiment lexicon that might include a list of emotional orientations. Usually, a sentiment lexicon consisted of a set of words in a special language, containing emotion weight and annotated with some orientations or particular rating score. For example, Mohammad & Turney [16] developed the NRC Word-Emotion Association Lexicon called EmoLex is the large size in terms along with the emotional orientations. The lexicon includes 14,182 words annotated in a binary method (0 = not existing; 1 = existing) with respect the sentiment polarity (positive, negative). In this context, when the lexicon richer in words besides using a multi-orientation emotion spectrum, compared to an easy binary polarity method, can enhance the classification accuracy in sentiment detection [10]. One of the key ideas is the one demonstrated by Carrillo-de-Albornoz and Plaza [17] the model detecting emotions in the text by using an effective lexicon that attaches emotional meanings to text

and processing negation, intensifiers, and modality as polarity modifiers. The emotion-based representing text as a set of emotions improved the accuracy in different sentiment classification tasks. Within the scope of the lexicon-based problem, Allison [18] suggested the study of supervised classification methods using lexical based-features on three tasks: the first two are the movie reviews, and the collection of political speeches these datasets are binary (positive and negative) classification tasks. The third task, using a corpus of the political speeches but includes five classes, the research proved that lexical-based classifiers of the third task achieve well results on the sentiment detection.

In the case of social media on the Web, Thelwall [19] discussed SentiStrength as a lexicon-based sentiment detection program to detect the strength of sentiments available, especially in social web texts. Although SentiStrength shows the capability of creating variants for different languages, It also employs a lexicon in uses a positive/negative sentiment terms and their polarities together with grammatical rules (e.g. negation) to deal the social web methods (as emoticons, links to other websites, punctuations, and misspellings) for sentiment detection. For example, "I am not busy" is a negative sentence because the bigram "not happy" with the lexical approach coded as negative because "happy" is a positive term and "not" is negating term. It looks advanced in pick up indirect indicators of sentiment that will generate fake sentiment patterns. The SentiStrength gives superior accuracy on a range of human-level on general short social web texts except when the texts usually include sarcasm terms.

3.1.3 Sentiment Detection Based on Dictionary

The problem with the dictionary-based approach is the incapability to find relevance sentiment words with specific

field and context directions. Qiu and He [20] proposed a Dissatisfaction-oriented Advertising approach based on SA. They identified sentiment sentences in contextual advertising based on using a dictionary. They proposed a novel advertising approach to enhance ad relevance and user experience through applied syntactic parsing and sentiment dictionary. They propose a rule-based strategy to extract specific words of sentiment sentences related to consumers' opinions identification, which considered as the advertising keywords. Their experimental results proved the efficiency of the proposed approach on extracting advertising keyword and demonstrated selection.

3.2 Opinion Spam Detection

Opinion spam indicates to bogus or fake opinions that attempt to aim to misguide readers or systems through showing unworthy positive opinions for a specific thing to elevate their reputations or showing negative opinions to other objects to harm promote of the objects. Opinion spam detection, in fact, is significant for applications [6]. In the web, a spammer may be a person or may work as groups by using several identities. Manually dividing information as spam and non-spam is not useful for spammer can show the product like any other person [21].

Web and Email spams are a little frequent to most persons. Email spam indicates to unwanted economic emails marketing commercial products. The Webspam indicates to the usage of illegal resources to growing the search rank positions of specific Website. The main reason for spread spams for economic reasons. For example, in the Web search, the commercial or public value of a website resulted by a search engine is of considerable importance based on the rank position of the websites. In the other side, If the website of a product page rank position is very low. And a user searches for a particular product through use a search engine, then the opportunity that the user will access to the product page is enormously small. In fact, undesirable for the commerce that forces the sellers to contact Search Engine Optimization agencies which create spam to helping others improve their page ranking [6].

In general, there are three types of spam documents as identified by Jindal and Liu in [22]. The first type: untruthful (fake) opinions indicates to reviews that mislead readers or opinion mining systems by giving unworthy positive reviews or harm negative reviews. The second type: opinions target brands only means reviews that do not focus on the explicit products that they are considered to review, but concentrated on the comment of the brands, the producers or the sellers. The third type: non-opinions means not reviews towards particular thing although they look as reviews. There are two basic types: advertisements, and unrelated reviews not including opinions [6].

Chen and Chen [23] introduced research relied on a set of interior records of opinion spams collected from a shady marketing campaign. They discovered the properties of opinion spams and spammers on a website to get some insights. Then, they proposed features that could be possibly useful in detecting spam opinions in threads. The experimental results show of spam detection on first posts: the promotions topics focused on the user experiences' in some spam posts. These promotions topics also try to excite people through use specific words and pictures. It considered to submitted on work time, and the threads that spam initiate are more dynamic to locate at striking positions.

Their study primary intention is to preserve and continue the discussion in a thread to interest more attention on it.

4. SENTIMENT RECOGNITION

There are many other challenges in sentiment analysis besides the approaches that discussed in the previous section. In some studies, it attempts to push sentiment analysis boundaries further and aim to obtain an advanced understanding of user's opinion expressed through text. By extract, the entity name and the entity type refer to holders, foundations, locations, times, measures, percentage and so on. This extraction task is known as Named Entity Recognition (NER). For example, in social media websites in extracting entity name, in a regular opinion mining models, the people need to find opinions around specific entities, e.g., trending products or brands. Also, people usually can only provide some names because there are abundant various brands and models. Also, people using social websites write names of the same product brands in different ways. For example, "Macintosh" may be written as "Mac" or "MacBook," and "Hewlett Packard" may be written as "HP." This mean product names may have more differences. Therefore, it is significant for a system to automatically recognition them from an appropriate Corpus [6].

Much existing sentiment recognition models consider only a specific target in each tweet.

Wang, et al. [24] developed a corpus includes UK election tweets, with the percentage of 3.09 entities for each tweet. It involves a model for multi-target specific sentiment recognition. They developed through using the context about a target as syntactic dependency requiring the target. They presented results of the model on a benchmark corpus of single targets and multi-target election corpus, are outperforming than state-of-the-art performance in previous approaches to multi-target sentiment task as deep learning models for single target sentiment.

Santosh and Vardhan [25] introduced a framework called Resource Description Framework (RDF) that supports in structuring the unstructured reviews which include features and sentiments resulted from traditional preprocessing and extraction methods. For word sense clarification, they are taking the support of WordNet lexical database. They used SentiWordNet for sentiment word direction. Several existing RDF terminologies are supportive in the creation of machine data. Also, they used SPARQL queries are executed on RDF data to study the opportunity for classifying the reviews using feature data. The results indicate processed by the interface as a feature, sentiment pair so that reviews are filtered apparently and support the user-specific feature set.

Bharti, et al. [13] proposed two methods for sarcasm recognition in the tweet of Twitter data. The first one is a parsing-based lexicon generation algorithm (PBLGA). The second one is for sarcasm recognition based on the existence of the interjection term. The two methods together are examined and compared with the previous state-of-the-art method to sarcasm recognition. The first method earns a 0.89, 0.81 and 0.84 precision, recall and f - score correspondingly. The second approach earns 0.85, 0.96 and 0.90 precision, recall and f - score in tweets.

The unstructured data consist of large quantities of redundant features, which will obviously influence the effectiveness of machine learning. Liu, et al. [27] performed a multi-swarm particle swarm optimization (MSPSO) model to choose the influence emotional features, which generates multi different

particle swarms on many cross-training subsets. These swarms are investigated to get the preferable features by the F-Measure value. The experimental results on the dataset indicate that MSPSO can efficiently decrease redundancy of text features and detect distinctive features., MSPSO can earn the improved implementation when selecting the same dimensions compared with conventional feature selection methods. Also, the experimental result shows that 72.19% of subjects confirm of the usability of the recognition method and efficiency of the feature selection.

5. SENTIMENT ASPECT IDENTIFICATION

In recent years, sentiment analysis has become a motivating area of research and has developed a wide range of domains, such as products commerce, social recommendation and data prediction [2]. Sentiment identification identified as describe the object expresses opinions in the text in more details.

In text analysis, sentiments can hold at various levels: word level, sentence level, and document level. In word level sentiment classification, is used to extract the object features (e.g. product quality, size, price, etc.) which are the opinion subjects' summary of reviews or comments. Sentence-level sentiment classification studies each sentence as a particular unit and identifies if that sentence includes only one opinion and whether the opinion is positive, negative, or neutral. Document-level sentiment classification, identify if the document (e.g. product comments, blogs, or posts) expresses sentiments and specify whether the general sentiment direction of the document that relies on classes are positive, negative or neutral [28]. However, Sentence-level and document-level sentiment classification are interesting to study often considered mainly by microblogs sentiment analysis, which differentiates objective documents that express real information not influenced by the speaker's previous experiences, feelings or tastes. The subjective documents assume whether a sentence expresses an opinion on a single entity or not that reflects the scene through with the speaker views reality [29].

It can extract users' experience for a specific feature of product from reviews and post comments. This example from reviews of the book "Secret" clarify:

"Beginning I have read this once it was fabulous and great things have happened. But, I am feeling rejected to think of a person thinking "the world is a lovely" and not being vigilant to dark corners.". In the first sentence are obviously positive, with positive subjective words such as "fabulous" and "great" supporting this book. The last sentence is negative. A phrase like "feeling rejected" express that the author did not agree with some topics in the conclusion of the book. The sentiment does not only happen at the entire document level, and not restricted to a single object [30].

Sentiment analysis identification considered in extract users' opinion about the different feature of reviews. Features of specific entities that can rate by users, such as product components, advantages, and disadvantages, etc. It means SA identification focuses more on feature-level sentiment analysis, which relied on the idea that an opinion includes of a sentiment polarity (positive or negative) and features of the product. The objective of sentiment analysis identification is to support person's knowledge to know sentiment approval for the different attributes of a product and get an overall data of specific things, such as camera, phone, movie, and so on.

Liang, et al. [31] proposed a matrix factorization model called CONR for sentiment word identification (SWI). The model constructed of the document-word matrix and utilized two significant factors: sentiment matching and sentiment consistency. The model is non-seed words because the problem of extracting the polarities of sentiment words from the labels of seed words, the seed words with it labels are specified manually, and any lost keyword could result in weak performance.

In this context, Chunping, et al. [32] considered by identifying the user opinions discussed on the Web about movies a different feature, then classify them. They specific devoted to feature-level SA of short reviews. First, they identified opinion objects of sentiment sentence. Next, they used syntax parsing on sentiment dictionary to study author's experience about the objects. Their model designed for three levels: data pre-processing, the opinion objects identification and feature-level SA. They edited the second task as a sequence tagging process and resolved it using a conditional random field. Also, they proposed word embedding features model for improving the accuracy based on deep learning technique. The results show that opinion aspect identification created on CRFs achieve well in short movie reviews and the method is summarized and supportive for the user in decision making.

4.1 Sentiment Aspect Identification Based on a Machine Learning

With the appearance of Web 2.0 shows an important increase in studies correlated to information extraction from web social networks. These social websites such as Facebook, Twitter, YouTube, and LinkedIn enable their users to share their products or citizens' opinions that service both research and commercial perspectives [33]. Poria, et al. [34] shifting forward from analysis text only to fuse multimodal content such as audio and videos. They proposed a novel temporal deep convolutional neural network (CNN) for textual feature extraction. Also, applied multiple kernel learning model (MKL) for extracting heterogeneous features from various types of modalities like audio, video, and text. They performed the integration of CNN with RNN for overlaps through features learned during temporal convolution. They showed that MKL further improved organizing features into groups and each group has its specific kernel function, to gathering data from diverse modalities efficiently and outperform the state of the art of multimodal emotion SA on different datasets.

The statistical machine learning techniques used to select features and decrease dimensionality for sentiment classification on web reviews. Yao, et al. [35] taking adjectives, adverbs, and verbs simultaneously as the feature. The obtained results indicate the sentiment classification of web reviews show the highest accuracy. Also, the adjective has the best classification effect to the emotional text than the other two. They used document frequency reduction technique, Chi-square Statistic, Mutual Information and Information Gain algorithms (DF, CHI, MI, and IG). The (MI) technique is not appropriate for the sentiment classification, especially for Chinese web reviews. The remaining DF, CHI feature selection Techniques has the best classification effect, and the latter one IG method is the worst.

In sentiment analysis, the task of extracting implicit opinions of objects through using emotions or discussions. The SA identification involves fine-grained analysis methods and significant efforts to extract features for classification approaches. For sentiment analysis identification of tweets

collected from Twitter microblogging, Dash, et al. [36] was conducted three classification algorithms Naive Bayes, Maximum Entropy and Support Vector Machines. The research focused on the measure the efficiency of different text feature together. The experiments results show applying unigram, and Part-of-Speech Tagging (POS) as text features the classification algorithms accuracies enhanced compared to other alternative features conducted. The performance of NB with Laplace smoothing employing a set of features such as unigram, Part-of-Speech (POS) and SVM with unigram too are effectiveness in classifying the tweets.

In the same context, native researchers devoted to enhancing the text features performance. Xie, et al. [37] utilized API from Sina which a Chinese microblogging for sentiment analysis identification. They studied implementations of three approaches, first relied on the emoticon, second used the sentiment lexicon and last by the hybrid approach using SVM. The experiments indicate the SVM relied on hybrid approach accomplishes the best accuracy is 66.467% and 67.283%, with applying of target-independent feature and target-dependent features separately.

4.2 Sentiment Aspect Identification Based on a Semantic

For discovering semantics for SA, there are Several approaches have been proposed and can be classified into contextual semantic, and conceptual semantic approaches. The contextual semantic approaches are extracting semantics from the relevance patterns of words called (statistical semantic). Whether the conceptual semantic approaches conduct external semantic data principles with NLP methods to detect the conceptual representations of words that implicitly carry sentiment.

Saif, et al. [38] proposed a semantic sentiment identification of words named SentiCircle. It a lexicon-based approach for SA detects the contextual semantic of words on Twitter from their relevance and changes the sentiment of words depends on their contextual semantics. The sentiment detection at both entity-level and tweet-level evaluated via three different sentiment lexicons to derive word prior sentiments. The Results show that the approach effectively outperforms the state-of-the-art lexicon-based approach SentiStrength in accuracy and F-measure for entity-level subjectivity and polarity detections.

In general, the sentiment target affected by the context of the text for example, the users almost use context polarity terms to express their sentiment like happy, sad, good and poor and so on. The sentiment target identification relies on the sentiment word that has different polarity in a different context. It means the sentiment target couldn't identify based on the sentiment word only becomes more complex especially when this word contained negations and comparing words. The context determined by the sentiment of the sentence and the semantic consistency. Therefore, the existing studies are considered only on some conjunctions which can't show the complete use of the context. Chen, et al. [39] utilized context more effectively and proposed a model rely on Long Short-Term Memory and Convolutional Neural Network to identify the sentiment polarity of the sentence, also use the context when it is not easy to extract then derives the sentiment target from the sentiment of the sentence.

6. CONCLUSION

This paper introduced and surveyed the domain of sentiment analysis, and due to several challenging research problems and a wide variety of practical applications, it has been a very active research area in recent years. First, it presented an abstract of the problem of Sentiment Analysis and illustrated the basic terminologies in this topic. It then discussed the most widely studied topic of sentiment detection, which considered in detect sentiment within a document or sentence. Also, it discussed opinion spam detection, which is more and more becoming a significant issue as people are depending on opinions existing on the Web for decision making. Then, it described sentiment analysis recognition which briefly introduced the issue that considered with extracting sentiment in the document and recognizing whether it conveys a positive or negative opinion. Last but not least, it discussed the sentiment aspect identification which considered in describe the object expresses opinions in the text in more details. Finally, the paper concluded by saying that all the sentiment analysis approaches are highly challenging. There is a real and huge need in the foundations for such services because every business wants to know how consumer's opinions towards their products and services and those of their competitors. These real desires and the technical challenges will keep the field vibrant and lively for many years forwards. When thinking about the future of this survey first, it will discuss the aspect-level sentiment analysis more. Also, by discussing the semantic and information retrieval techniques that usually investigated to make improvement in sentiment analysis models.

7. REFERENCES

- [1] T. Hartmann, S. Klenk, A. Burkovski, and G. Heidemann, "Sentiment Detection with Character n-Grams," in Proceedings of the 7th International Conference on Data Mining (DMIN'11), 2011, pp. 364-368.
- [2] M. Devika, C. Sunitha, and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," *Procedia Computer Science*, vol. 87, pp. 44-49, 2016.
- [3] H. Tang, S. Tan, and X. Cheng, "A survey on sentiment detection of reviews," *Expert Systems with Applications*, vol. 36, pp. 10760-10773, 2009.
- [4] F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, *Sentiment Analysis in Social Networks*: Morgan Kaufmann, 2016.
- [5] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Springer, 2006.
- [6] Liu, B. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing*, Second Edition, N. Indurkha and F.J. Damerau, Editors. 2010.
- [7] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining text data*, ed: Springer, 2012, pp. 415-463.
- [8] M. Dahab and S. Al Asmari, "Towards Discovering the Sentiment Holder in Arabic Text," *International Journal of Computer Applications*, vol. 111, 2015.
- [9] Tan, S., Y. Wang., and X. Cheng. Combining learn-based and lexicon-based techniques for sentiment detection without using labeled examples. In Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-2008), 2008.
- [10] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Systems with Applications*, vol. 69, pp. 214-224, 2017.
- [11] I. Hernández-Farías, J.-M. Benedí, and P. Rosso, "Applying basic features from sentiment analysis for automatic irony detection," in *Iberian Conference on Pattern Recognition and Image Analysis*, 2015, pp. 337-344.
- [12] D. Maynard and A. Funk, "Automatic detection of political opinions in tweets," in *Extended Semantic Web Conference*, 2011, pp. 88-99.
- [13] C. Lin, Y. He, and R. Everson, "A comparative study of Bayesian models for unsupervised sentiment detection," in *Proceedings of the fourteenth conference on computational natural language learning*, 2010, pp. 144-152.
- [14] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2015, pp. 959-962.
- [15] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th annual meeting on association for computational linguistics*, 2002, pp. 417-424.
- [16] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," *Computational Intelligence*, vol. 29, pp. 436-465, 2013.
- [17] J. Carrillo-de-Albornoz and L. Plaza, "An emotion-based model of negation, intensifiers, and modality for polarity and intensity classification," *Journal of the American Society for Information Science and Technology*, vol. 64, pp. 1618-1633, 2013.
- [18] B. Allison, "Sentiment detection using lexically-based classifiers," in *International Conference on Text, Speech and Dialogue*, 2008, pp. 21-28.
- [19] M. Thelwall, "The Heart and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength," in *Cyberemotions*, ed: Springer, 2017, pp. 119-134.
- [20] G. Qiu, X. He, F. Zhang, Y. Shi, J. Bu, and C. Chen, "DASA: dissatisfaction-oriented advertising based on sentiment analysis," *Expert Systems with Applications*, vol. 37, pp. 6182-6191, 2010.
- [21] V. Patel, G. Prabhu, and K. Bhowmick, "A Survey of Opinion Mining and Sentiment Analysis," *International Journal of Computer Applications*, vol. 131, pp. 24-27, 2015.
- [22] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 2008, pp. 219-230.
- [23] Y.-R. Chen and H.-H. Chen, "Opinion spam detection in web forum: a real case study," in *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 173-183.

- [24] B. Wang, M. Liakata, A. Zubiaga, and R. Procter, "TDParse-multi-target-specific sentiment recognition on Twitter," 2016.
- [25] D. T. Santosh And B. V. Vardhan, "Automatic Machine Recognition Of Features And Sentiments From Online Reviews," In *Iaeng Transactions On Engineering Sciences: Special Issue For The International Association Of Engineers Conferences 2015, 2017*, pp. 264-280.
- [26] S. K. Bharti, K. S. Babu, and S. K. Jena, "Parsing-based sarcasm sentiment recognition in Twitter data," in *Advances in Social Networks Analysis and Mining (ASONAM), 2015 IEEE/ACM International Conference on*, 2015, pp. 1373-1380.
- [27] Z. Liu, S. Liu, L. Liu, J. Sun, X. Peng, and T. Wang, "Sentiment recognition of online course reviews using multi-swarm optimization-based selected features," *Neurocomputing*, vol. 185, pp. 11-20, 2016.
- [28] V. Jagtap and K. Pawar, "Analysis of different approaches to sentence-level sentiment classification," *International Journal of Scientific Engineering and Technology*, vol. 2, pp. 164-170, 2013.
- [29] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, pp. 1-167, 2012.
- [30] C. R. Fink, D. S. Chou, J. J. Kopecky, and A. J. Llorens, "Coarse- and Fine-Grained Sentiment Analysis of Social Media Text," *Johns Hopkins APL Technical Digest*, vol. 30, pp. 22-30, 2011.
- [31] J. Liang, X. Zhou, Y. Hu, L. Guo, and S. Bai, "Conr: A novel method for sentiment word identification," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, 2014, pp. 1943-1946.
- [32] O. Chunping, L. Yongbin, Z. Shuqing, and Y. Xiaohua, "Opinion Objects Identification and Sentiment Analysis," *International Journal of Database Theory and Application*, vol. 8, pp. 1-12, 2015.
- [33] R. T. Khasawneh, H. A. Wahsheh, I. M. Alsmadi, and M. N. AI-Kabi, "Arabic sentiment polarity identification using a hybrid approach," in *Information and Communication Systems (ICICS), 2015 6th International Conference on*, 2015, pp. 148-153.
- [34] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, "Convolutional MKL based multimodal emotion recognition and sentiment analysis," in *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, 2016, pp. 439-448.
- [35] J. Yao, H. Wang, and P. Yin, "Sentiment feature identification from Chinese online reviews," in *Advances in Information Technology and Education*, ed: Springer, 2011, pp. 315-322.
- [36] A. K. Dash, J. K. Rout, and S. K. Jena, "Harnessing Twitter for Automatic Sentiment Identification Using Machine Learning Techniques," in *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*, 2016, pp. 507-514.
- [37] L. Xie, M. Zhou, and M. Sun, "Hierarchical structure based hybrid approach to sentiment analysis of chinese micro blog and its feature extraction," *Journal of Chinese information processing*, vol. 26, pp. 73-82, 2012.
- [38] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of Twitter," *Information Processing & Management*, vol. 52, pp. 5-19, 2016.
- [39] P. Chen, B. Xu, M. Yang, and S. Li, "Clause sentiment identification based on convolutional neural network with context embedding," in *Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2016 12th International Conference on*, 2016, pp. 1532-1538.
- [40] Argamon, S., Koppel, M., & Avneri, G. (1998). *Routing documents according to style*.
- [41] In *First international workshop on innovative information systems*.
- [42] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, pp. 1093-1113, 2014.