Sentiment Analysis: A Survey of Current Research and Techniques

Jeevanandam Jotheeswaran, Dr. S. Koteeswaran

Research Scholar, VelTech Dr. RR & Dr. SR Technical University, Chennai, India

Department of Computer Science and Engineering, VelTech Dr. RR &Dr. SR Technical University, Chennai, India

ABSTRACT: Opinions and reviews on products and services are expressed in the Web through blogs, feedback forms; it is essential to develop methods to automatically classify and gauge them to identify the underlying sentiment about the product. Analyzing the polarity of sentiment expressed in data is Opinion Mining (OM). It is a system that identifies and classifies opinion/sentiment as represented in electronic text. Economic and marketing researches depend heavily on accurate method to predict sentiments of opinions extracted from internet and predict online customer's preferences. OM has many steps, and techniques for each step. This study ensures an overall survey about OM related to product reviews, and classification algorithms used for sentiment classification.

KEYWORDS: Opinion Mining (OM), Sentiment Analysis, Semantics, Machine Learning

I. INTRODUCTION

An Opinion is a judgment or belief a majority of people formed about a specific thing, not necessarily based on fact/knowledge. Opinion generally refers to what a person thinks about something or opinion is a subjective belief, and the result of emotion or facts interpretation [1]. Opinion Mining (OM), also called as Sentiment analysis, is a natural language processing type to find public mood about a product or topic. OM and Sentiment Analysis tool "process a set of search results for a given item, generating product attributes (quality, features etc.) and aggregating opinion". OM is automatic extraction of knowledge from others opinions on a particular topic/problem. It involves collecting and examining opinions about services or product in blog posts, tweets, reviews andcomments. Sentiment analysis is useful for economic and marketing strategizing such as in marketing where success of a new product launch can be judged, determines which product or service version is popular and also identifies demographics like particular features.

Some of the challenges faced in Sentiment analysis are that an opinion word which is positive in one situation can be negative in another situation and opinions are not expressed similarly by different people. Most reviews have positive and negative comments and are analyzed sentence by sentence. But, in more informal media like twitter or blogs, people are more likely to combine different opinions in same sentence which may or may not be easy to comprehend, but difficult for an algorithm to analyze [2].

With the growth of social media (forum discussions, reviews, blogs, comments and postings in social network sites, micro-blogs, Twitter) on the Web, organizations and individuals are using content in such media to make decisions. Generally, overall contextual polarity or writer sentiment about some aspect is determined using sentiment analysis. The challenge in sentiment classification is sentiment may be judgment, mood or evaluation of an object like a film, book or a product which can be a document or sentence or feature that is labeled positive or negative [3]. But, finding and monitoring opinion web sites and distilling information in them are a formidable task due to the proliferation of diverse sites. Each site has a huge volume of opinion text not always easily deciphered in long blogs and forum postings. An average human reader has difficulty identifying relevant sites and extracting and summarizing opinions in them. Hence, automated sentiment analysis systems are required [4].

1.1 FEATURE EXTRACTION AND SELECTION
Feature selection methods provide a criterion to eliminate terms from document corpus to lessen vocabulary space. Feature selection is done in literature as follows [8]:

1: Information gain (based on presence/absence of a term in a document, a threshold is set and terms with less information gain removed).

2: Odd Ratio (suitable for binary class domain which has one positive and one negative class to classify. The algorithm is run in each class and top- n features from the sorted list are taken).

3: Document Frequency (Measures appearances of a term in available corpus documents and based on threshold, computed terms are removed).

4: Mutual Information (words with frequent association in a document are chosen.

Features weighting mechanism is two types. They are:

1: Term Presence and Term Frequency- word occurring occasionally has more information than frequently occurring words.

2: Term frequency and inverse document frequency (TFIDF) - Documents are rated with highest rating for words appearing regularly in few documents and lowest rating for words appearing regularly in all documents.

Discovering regularities is easy and consumes less time when data suits machine learning which is achieved through removal of irrelevant and redundant data features and this process is termed feature selection. Feature selection is automatic and computationally tractable, unlike constructing new input data. Feature selection benefits learning by reducing data required to achieve learning, improves predictive accuracy, compacts learned knowledge and is easily understood. It also has lower execution time.

Current machine learning feature selection methods are in two divisions —wrappers: evaluationof features using learning algorithm, and filters: evaluation of feature through heuristics based on data's general characteristics. As feature selection is optimized for a specific learning algorithm, wrappers ensure better results regarding final predictive learning algorithm accuracy than filters. But, as a learning algorithm evaluates all features sets, wrappers are costly to run and intractable for large databases with many features [09].

Feature selection methods are,

- Correlation based feature selector (CFS),
- Information Gain,
- Support Vector Machine (SVM),
- Principal component analysis (PCA)

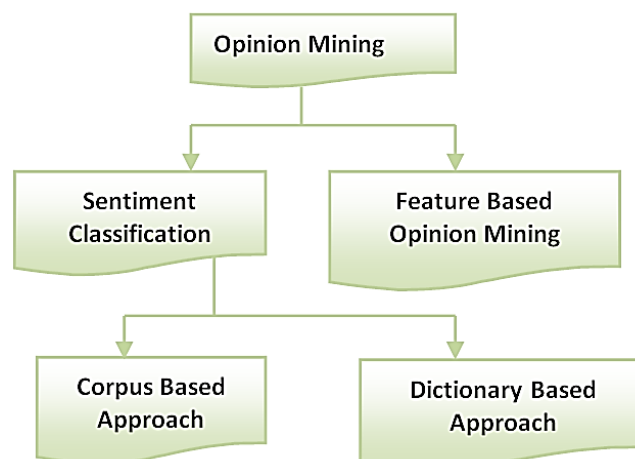### 1.2 CLASSIFICATION METHODS IN OPINION MINING

Figure 1.1 Classification of approaches of Semantic Orientation

Problems in extracting a text's Semantic Orientation (SO) (whether text is positive or negative to specific subject matter) starts from determining SO for individual words. The hypothesis is that, in a SO of relevant words in a text, SO for entire text is determined. SO approach to Sentiment analysis is unsupervised learning as it needs no prior training to mine data. Figure1.1 details classification of SO approaches.

Corpus Based Approach

A popular corpus-driven method is determining words emotional affinity to learn their probabilistic affective scores from a large corpus. The method is to assign a happiness factor to words based on their occurrence frequency in happy-labelled blog posts compared to total frequency in a corpus of blog posts labeled with "happy" and "sad" mood annotations. They compare happiness factor of scores of words with scores in the list.

Dictionary Based Approach

Dictionary based approach uses lexical resources like Word Net to automatically obtain emotion-related words for emotion classification. Beginning from primary emotion adjectives set, they retrieve similar words from Word Net using senses of all words in synsets with emotion adjectives. The process exploits synonym and hyponym relations in Word Net to manually locate words similar to nominal emotion words. Affective weights are acquired automatically from a very large text corpus in an unsupervised way.

Feature based opinion mining

Using OM, a review is evaluated at 3 levels- document, sentence and feature levels. Evaluating a review at document level, the entire review is classified positive or negative based on opinions expressed in that review. When evaluated at sentence level, then every sentence is classified as positive or negative while feature level or feature based OM gives summary where a product feature is liked or disliked by reviewers. The major feature based OM tasks include - (1) identifying products features in review, (2) determining opinion expressed by reviewer (positive, negative or neutral), (3) summarizing discovered information.

1.3 DATA FOR OPINION MINING

Data from sources mentioned below are used to locate opinions and ensure good recommendation for specific applications. Most common sources are blogs and review sites.

Blogs

With increasing internet usage, blogging and blog pages are increasing. Blog pages are a means to express one's personal opinions. Bloggers record daily events in their lives and express feelings, opinions and emotions. Many blogs have reviews on products, issues, etc. Blogs are a source of opinion in many sentiment analysis related studies.

Review sites

A factor considered to make decisions by purchaser before buying is to know comments by previous buyers. Opinions of others are an important factor for users in making purchasing decisions. User-generated reviews is available on the net of product or services in unstructured format. Reviewer's data in most sentiment classification studies are collected from e-commerce websites like www.amazon.com, www.yelp.com, [11] www.CNET download.com and so on.

Datasets

Raw datasets are available and a widely used review dataset for Movie domain, the Multi-Domain Sentiment (MDS) dataset, (http:// www.cs.jhu.edu/mdredze/datasets/sentiment) which has 4 types of product reviews from popular websites like Amazon.com including Books, DVDs, Electronics and Kitchen appliances [12] with 1000 positive and 1000 negative reviews for every domain. Most work in the field uses movie review data for classification.

Micro-blogging

Micro-blogging is a popular communication tool among Internet users. Millions of messages appear daily in common web-sites for micro-blogging like Twitter, Tumblr and Facebook. Twitter messages express opinions which are data source to classify sentiment [13]. In Twitter, information is represented as short text message called "tweet". Opinions about topics are expressed in tweets and considered for OM.

II. RELATED WORK

2.1 MACHINE LEARNING

A survey covering techniques and methods to get opinion oriented information from text presented by Khan et al., [14] deals with techniques related to OM. It followed a systematic literature review process to conduct the survey. It focused on machine learning techniques based on usage and importance of OM. It tried to identify commonly used classification techniques for opinionated documents to aid future research.

A work on Chinese OM, which emphasized mining opinions on online reviews, was presented by Zhang et al., [15]. This paper was based on machine learning methods. Using a real-world Amazon CN dataset on opinions, it conducted comparative experimental studies and concluded that the approach was effective. Though machine learning based method outperformed its alternatives (especially SVM based method), such methods need large labeled training instances, which are time consuming and labor intensive to get.

A classification scheme of pre-release movie popularity using C4.5 and PART classifier algorithm was proposed by Asad et al., [16] which defined relation between post release movies attributes using correlation coefficient. Movie data across the internet makes it a good candidate for machine learning and knowledge discovery. But, most research is directed to bi-polar classification of movie or a movie recommendation system based on reviews by viewers on various internet sites.

A review's helpfulness depends on 3 important factors: reviewer's expertise, review writing style and its timeliness was shown by Liu et al., [17]. The volume of available reviews and variations in review quality are an impediment to effective reviews use, as most helpful reviews are buried under voluminous low quality reviews. Based on these factors analysis, a nonlinear regression model for helpfulness prediction is presented. This empirical study on IMDb movie reviews dataset proves that the new approach is very effective.

A new, collaborative fuzzy set theory based filtering framework that integrates subjective and objective information was proposed by Cheng and Wang [18]. This method provides a comprehensive result and also solves problems of conventional Collaborative Filtering (CF) systems, new user and new item. The experiment indicates that the new method produces high-quality recommendations. CF was applied to many commercial systems like IDDB, Netflix and others successfully. A CF system's basic idea is generating recommendations based on similar users past experiences. Users' options are categorized into objective and subject information. The former was furnished by common users and latter represents solicited opinions by experts (film critics).

A semi-automatic approach to create sentiment dictionaries in many languages presented by Steinberger et al., [19] produced high-level gold standard sentiment dictionaries for two languages translating them automatically to third languages. Words found in target language word lists are likely to be used as their word senses are similar to that of two source languages.

Three supervised machine learning algorithms like Naïve Bayes, SVM and character based N-gram model was compared for sentiment classification of reviews on travel blogs for 7 popular travel destinations in the US and Europe by Ye et al., [20]. Empirical findings revealed that SVM and N-gram approaches outperformed Naïve Bayes approach, and when training datasets had many reviews, all 3 approaches had at least 80%accuracy.Internet applications growth in tourism leads to enormous amount of personal reviews for travel-related information on the Web. These reviews appear in forms like BBS, blogs, Wiki or forum websites.Query functions in search engines like Yahoo and Google help users find reviews they needed on specific destinations. Returned pages from search engines are beyond humanvisual capacity.

How features based on syntactic dependency relations improve OM performance was explored by Joshi and Penstein-Rosé [21]. Using a of dependency relation triples transformation, it converts them to "composite back-off features" that generalize better than regular lexicalized dependency relation features. A novel approach to mine opinions from product reviews, converting OM task to identify product features, expressions of opinions and relations between them was presented by Wu et al., [22]. Taking advantage of the observation that most product features are phrases, a phrase dependency parsing concept, which extends traditional dependency parsing to phrase level, is introduced. This is then implemented to extract relations between product features and expressions of opinions. Evaluation showed mining task benefitted from phrase dependency parsing.

Machine learning experiments regarding sentiment analysis in forum texts, blog, and reviewsseen on the World Wide Web and written in French, English, and Dutch were presented by Boiy and Moens [23]. This paper trains from a set of example sentences or statements that are annotated manually as positive, negative or neutral regarding a certain

entity. It is interested in feelings expressed by people regarding consumed products. Also, it learns and evaluates many classification models configured in a cascaded pipeline. It deals with several problems, being the input texts noisy character, sentiment attribution to a particular entity and the training set's small size.

2.2 SEMANTICS BASED

A novel space designed feature to classify polarity and strength of relationships from biomedical abstracts was described by Swaminathan et al., [24]. Semantics-based sequential features was explored and constructed at three levels: entity, phrase, and sentencein addition to conventional syntactic features like unigrams and bigrams. A wrapper-based method then selects optimal feature sets for polarity and strength prediction. A multi-stage SVM classifier and an SVR predictor are built for polarity and strength prediction, respectively. Two different schemas, namely, (1 vs. all) and (2 vs. 2), build multi-stage SVM. Finally, 3 different kernel functions are considered at different stages of SVM classifier.

A semantic user modeling based on Twitter posts was investigated by Abel et al., [25]. It introduced and analyzed methods to link Twitter posts with related news articles to contextualize Twitter activities. It proposed and compared strategies exploiting semantics from tweets and related news articles to represent individual Twitter activities in a meaningful way, semantically. A large-scale evaluation validates this approach's benefits and showed that these methods relate tweets to news articles with high precision and coverage, enrich tweets semantics clearly and strongly impact semantic user profiles construction for the Social Web.

A SentiFrameNet, an extension to FrameNet, as a representation for sentiment analysis was proposed by Ruppenhofer and Rehbein [26]. Sentiment analysis is characterized by pragmatic focused approaches which use shallow techniques for robustness but rely on data sets and methods ad-hoc creation. Progress towards deep analysis depends on a) enriching shallow depictions with linguistically motivated, rich information, and b) focusing different research branches and combining resources to create synergies with related work in NLP.

Argument mining as a task to build a representation for an argumentative piece of text was considered by Peldszus and Stede [27] with the goal of providing a survey of literature on resulting representations (argument diagramming techniques) and on various automatic analysis process aspects. The authors provide a synthesized proposal for a scheme combining advantages from many earlier approaches for representation; the authors also discuss relationship between representing argument structure and rhetorical structure of texts. Then, for argument mining, the authors cover literature on closely-related tasks that were tackled in Computational Linguistics, as they think that these contribute to powerful argument mining systems than the prototypes in recent years.

An OM system that mined useful opinion information from camera reviews by using Semantic Role Labeling (SRL) and polarity computing method was proposed by Li et al., [28]. In the proposed method, feature and sentiment lexicon were used to mine features and emotional items respectively. Finally, the comparison between positive and negative opinion are visuallypresented.

A new OM method underpinned by context-sensitive text mining and inferential language modeling to improve OMeffectiveness was presented by Lau et al., [29]. Initial experiments reveal the new inferential OM method outperforms purely lexicon-based opinion finding method regarding many benchmark measures. The research helps the development of effective OM methods to discover business intelligence from Web knowledge base.

A new approach of Opinion Feature Extraction based on Sentiment Patterns (OFESP) proposed by Zhai et al., [34] considers reviews structure characteristics for higher precision and recall values. OFESP, with a self-constructed database of sentiment patterns, matches each review sentence to obtain features, filtering redundant features regarding domain, statistics and semantic similarity relevance. Experimental studies demonstrated that OFESP outperforms on precision, recall and F-score for real-world dataset. Meanwhile, compared to the syntactic analysis based approach, OFESP performs better on recall and F-score.

The role of Binary Grammatical relation or Dependency (BGD) in the words of a sentence for OM was discussed by Srivastava et al., [30]. In a sentence, words are arranged in sequence to communicate information. A sentence's

complete meaning is not only determined by meaning of words, but also by the pattern in which they are arranged. Essentially every word in a sentence possesses grammatical links with other words to ensure correct meaning and they are called BGD.

Treating product feature extraction as a sequence labeling task employing a discriminative learning model named Conditional Random Fields (CRFs) to tackle it was proposed by Huang et al., [31]. An innovative OM methodology that took advantage of the new Semantic Web-guided solutions to enhance results got through traditional natural language processing techniques and sentiment analysis processes was proposed by Peñalver-Martinez et al., [32]. The proposed methodology's goals are: (1) to improve feature-based OM using ontologies at feature selection stage, and (2) provide a new vector analysis-based method for sentiment analysis. The methodology was implemented and tested in a real-world movie review-themed scenario, yielding good results compared to conventional approaches.

2.3 OTHER TECHNIQUES

An OM framework that extracted opinions and views of consumers/customers, and analyzed them to provide concrete market flow with proven statistical data was demonstrated by Shandilya and Jain [33]. The software used classification, clustering and lingual knowledge-based OM to provide these features. A new approach that enabled to predict and explain collective opinion formation by social swarming was introduced by Kaiser et al., [34]. A computational model to simulate collective opinion formation is derived from ant colony meta-heuristic and applied to an exemplary online community where members' opinions are identified through text mining. Web 2.0 platforms ensure more power to people over sharing information and exchanging opinions. Increased social interactivity leads to emergence of self-organized communities where members form opinions through social swarming. This approach is compared with 3 approaches for validation.

A systematic analysis framework using Korean Twitter data to mine temporal and spatial trends of brand images was demonstrated by Cho et al., [35]. A publicly available Korean morpheme analyzeranalyzed Korean tweets grammatically, and constructed Korean polarity dictionaries having a noun, adjective, verb, and/or root to analyze each tweet message's sentiment. Sentiment classification is performed by a SVM and multinomial Naive Bayes classifier.
Related works about market prediction were reviewed by Nassirtoussi et al., [36] based on online-text-mining to produce a picture of generic components that all have. It compared every system with the rest and identified the main differentiating factors. This comparative systems analysis expands onto theoretical and technical foundations. This work helps research to structure this field and identify exact aspects needing more research.

Data collected from Twitter was analyzed by Aldahawi et al., [37] who investigated variance from using automated sentiment analysis tool versus human classification. This interest is in understanding how users' motivation to post messages affects classification quality. The data set uses Tweets from two world's leading oil companies, BP America and Saudi Aramco, and other users that follow and mention them, representing the West and Middle East. Results revealed that both methods yield different positive, natural and negative classifications based on culture and relationship of the poster to two companies, questioning automated sentiment analysis tools reliability for some users.
The construction of a word in English and Chinese from letter (alphabet) and character to words and phrases respectively was described by Chen et al., [38]. The keyless research system obtains deferential number of characters in a phrase, from many Chinese text documents. Synonyms are found from a synonymous process. About 50 top key words and 50 bottom key words described quality of class teaching by students in an open-ended writing part in the questionnaire was used for evaluation. It is concluded that key words are from text documents when key words are not defined beforehand, frequency of key word appearance in text documents is obtained, and model is applicable to other fields.

Liang et al., [39] suggested mining for users' opinions on items based on item taxonomy by experts and folksonomy contributed by users. It explored personalized item recommendations based on users' opinions. Experiments on real word datasets from Amazon.com and CiteULike proved the proposed approaches effectiveness. Item folksonomy or tag information is web 2.0 information. Item folksonomy has rich opinion information of users on item classifications and descriptions and is used as another information source for OM.
.

III. SUMMARY

The papers reviewed in the sections above are tabulated.

Table 1: Summary of Reviews

Year	Authors	Title of the Paper	Algorithm Used/Method
2009	Khan et al., [15]	Mining opinion from text documents: A survey	Machine learning techniques based on Opinion Mining (OM)
2008	Zhang et al., [16].	Sentiment classification for chinese reviews using machine learning methods based on string kernel	Real-world Amazon CN dataset used
2012	Asad et al., [18]	Movie popularity classification based on inherent movie attributes using C4. 5, PART and correlation coefficient	C4.5 and PART classifier algorithm
2008	Liu et al., [19]	Modeling and predicting the helpfulness of online reviews	IMDb movie reviews dataset
2011	Cheng and Wang [20]	A novel fuzzy recommendation system integrated the experts' opinion	Collaborative fuzzy set theory based filtering framework used
2012	Steinberger et al., [21]	Creating sentiment dictionaries via triangulation	High-level gold standard sentiment dictionaries
2009	Ye et al., [23]	Sentiment classification of online reviews to travel destinations by supervised machine learning approaches	Comparison made between Naïve Bayes, SVM and character based N-gram model
2009	Joshi and Penstein-Rosé [24]	Generalizing dependency features for opinion mining	Dependency relation triples transformation
2009	Boiy and Moens [27]	A machine learning approach to sentiment analysis in multilingual Web texts	Sentiment analysis
2010	Swaminathan et al., [28]	Opinion mining for biomedical text data: Feature space design and feature selection	Semantics-based sequential features
2011	Abel et al., [29]	Semantic enrichment of twitter posts for user profile construction on the social web	Semantic user modeling based on Twitter posts
2012	Ruppenhofer and Rehbein [30]	Semantic frames as an anchor representation for sentiment analysis	SentiFrameNet
2013	Peldszus and Stede [31]	From argument diagrams to argumentation mining in texts: a survey	Argument mining
2010	Li et al., [32]	Opinion mining of camera reviews based on Semantic Role Labeling	Semantic Role Labeling (SRL) and polarity computing method
2010	Zhai et al., [34]	Extracting Opinion Features in Sentiment Patterns	Opinion Feature Extraction based on Sentiment Patterns (OFESP)
2010	Srivastava et al., [35]	Exploiting grammatical dependencies for fine-grained opinion mining	Binary Grammatical relation or Dependency (BGD)
2012	Huang et al., [36]	Fine-grained product features extraction and categorization in reviews opinion mining	Conditional Random Fields (CRFs)
2014	Peñalver-Martinez et al., [37]	Feature-based opinion mining through ontologies	Semantic Web-guided solutions
2009	Shandilya and Jain [38]	Opinion Extraction & Classification of Reviews from Web Documents	Classification, clustering and lingual knowledge-based OM used
2011	Kaiser et al., [39]	Are Humans Like Ants?--Analyzing Collective Opinion Formation in Online Discussions	Web 2.0 platforms

IV. CONCLUSION

Opinion Mining (OM) is natural language processing dealing with tracking the mood of people regarding a product or topic. OM combines information retrieval and computational linguistic techniques handling a document's opinions. The field's main goal is solving problems related to opinions on products, politics in newsgroup posts and review sites. It provides automatic opinion extraction, emotions and sentiments in text tracking attitudes and feelings on the web. People express views by writing blog posts, comments, reviews and tweets on various topics. Tracking products and brands and determining if they are viewed positively or negatively is done on the web. OM has slightly different tasks and many names like opinion extraction, sentiment analysis, sentiment mining, affect analysis, subjectivity analysis, emotion analysis and review mining. But, they all come under sentiment analysis or OM.

This study defined the concept of opinion in a sentiment analysis context, the main tasks being a framework of OM. Sentiment analysis deals with evaluation opinions or opinions type implying positive or negative sentiments. Reviews reveal that different features and classification algorithms combine efficiently to overcome individual drawbacks and benefit from each other's merits. Finally they enhance sentiment classification performance. More work in future is needed to improve performance measures. The main challenge is in using other languages, dealing with negation expressions and producing an opinions summary based on product features/attributes, handling of implicit product features, complexity of sentence/ document, etc. Future research could be dedicated for the challenges.

REFERENCES

1. Padmaja, S., & Fatima, S. S. (2013). Opinion Mining and Sentiment Analysis–An Assessment of Peoples' Belief: A Survey. *International Journal*.
2. Vinodhini, G., &Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: a survey. *International Journal*, *2*(6).
3. Buche, A., Chandak, D., &Zadgaonkar, A. (2013). Opinion Mining and Analysis: A survey. *arXiv preprint arXiv:1307.3336*.
4. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1-167.
5. Patel, F. N., &Soni, N. R. (2012). Text mining: A Brief survey. *International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-2 Number-4 Issue-6 December-2012*, 243.
6. Basari, A. S. H., Hussin, B., Ananta, I., &Zeniarja, J. (2013). Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization. *Procedia Engineering*, *53*, 453-462.
7. Zhang, W., Yoshida, T., & Tang, X. (2007, October). Text classification using multi-word features. In *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on* (pp. 3519-3524). IEEE.
8. Selvam, B., &Abirami, S (2013). A SURVEY ON OPINION MINING FRAMEWORK.
9. Isabella, J. (2012). Analysis and evaluation of Feature selectors in opinion mining. *Indian Journal of Computer Science and Engineering (IJCSE)*, *3*(6).
10. Somprasertsri, G., &Lalitrojwong, P. (2010). Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization. *J. UCS*, *16*(6)
11. Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets & classification algorithms for sentiment classification. *ISciences*, *181*(6), 1138-1152.
12. Pak, A., &Paroubek, P. (2010, May). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *LREC*.
13. Khan, K., Baharudin, B. B., & Khan, A. (2009, June). Mining opinion from text documents: A survey. In *Digital Ecosystems and Technologies, 2009. DEST'09. 3rd IEEE International Conference on* (pp. 217-222). IEEE.
14. Zhang, C., Zuo, W., Peng, T., & He, F. (2008, November). Sentiment classification for chinese reviews using machine learning methods based on string kernel. In *Proceedings of the 2008 Third International Conference on Convergence and Hybrid Information Technology* (Vol. 2, pp. 909-914).
15. Balahur, A., &Montoyo, A. (2008, October). A feature dependent method for opinion mining and classification. In *Natural Language Processing and Knowledge Engineering, 2008. NLP-KE'08. International Conference on* (pp. 1-7). IEEE.
16. Asad, K. I., Ahmed, T., & Rahman, M. S. (2012, May). Movie popularity classification based on inherent movie attributes using C4. 5, PART and correlation coefficient. In *Informatics, Electronics & Vision (ICIEV), 2012 International Conference on* (pp. 747-752). IEEE.
17. Cheng, L. C., & Wang, H. A. (2011, June). A novel fuzzy recommendation system integrated the experts' opinion. In *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on* (pp. 2060-2065). IEEE.
18. Steinberger, J., Ebrahim, M., Ehrmann, M., Hurriyetoglu, A., Kabadjov, M., Lenkova, P., ...&Zavarella, V. (2012). Creating sentiment dictionaries via triangulation. *Decision Support Systems*, *53*(4), 689-694.
19. Jin, W., Ho, H. H., &Srihari, R. K. (2009, June). OpinionMiner: a novel machine learning system for web opinion mining and extraction. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1195-1204). ACM.
20. Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, *36*(3), 6527-6535.
21. Wu, Y., Zhang, Q., Huang, X., & Wu, L. (2009, August). Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3* (pp. 1533-1541). Association for Computational Linguistics.
22. Binali, H., Potdar, V., & Wu, C. (2009, February). A state of the art opinion mining and its application domains. In *Industrial Technology, 2009. ICIT 2009. IEEE International Conference on* (pp. 1-6). IEEE.
23. Boiy, E., &Moens, M. F. (2009). A machine learning approach to sentiment analysis in multilingual Web texts. *Information retrieval*, *12*(5), 526-55
24. Abel, F., Gao, Q., Houben, G. J., & Tao, K. (2011). Semantic enrichment of twitter posts for user profile construction on the social web. In *The Semantic Web: Research and Applications* (pp. 375-389). Springer Berlin Heidelberg.

25. Ruppenhofer, J., &Rehbein, I. (2012, July). Semantic frames as an anchor representation for sentiment analysis. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis* (pp. 104-109). Association for Computational Linguistics.

26. Peldszus, A., &Stede, M. (2013). From argument diagrams to argumentation mining in texts: a survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, *7*(1), 1-31.

27. Li, X., Dai, L., & Shi, H. (2010, August). Opinion mining of camera reviews based on Semantic Role Labeling. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on* (Vol. 5, pp. 2372-2375). IEEE.

28. Lau, R. Y., Lai, C. L., & Li, Y. (2009, August). Leveraging the web context for context-sensitive opinion mining. In *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on* (pp. 467-471). IEEE.

29. Zhai, Y., Chen, Y., Hu, X., Li, P., & Wu, X. (2010, October). Extracting Opinion Features in Sentiment Patterns. In *Information Networking and Automation (ICINA), 2010 International Conference on* (Vol. 1, pp. V1-115). IEEE.

30. Srivastava, R., Bhatia, M. P. S., Srivastava, H. K., &Sahu, C. P. (2010, September). Exploiting grammatical dependencies for fine-grained opinion mining. In *Computer and Communication Technology (ICCCT), 2010 International Conference on* (pp. 768-775). IEEE.

31. Huang, S., Liu, X., Peng, X., &Niu, Z. (2012, December). Fine-grained product features extraction and categorization in reviews opinion mining. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on* (pp. 680-686). IEEE.

32. Peñalver-Martinez, I., Garcia-Sanchez, F., Valencia-Garcia, R., Rodríguez-García, M. Á., Moreno, V., Fraga, A., & Sánchez-Cervantes, J. L. (2014). Feature-based opinion mining through ontologies. *Expert Systems with Applications*, *41*(13), 5995-6008.

33. Shandilya, S. K., & Jain, S. (2009, March). Opinion Extraction & Classification of Reviews from Web Documents. In *Advance Computing Conference, 2009. IACC 2009. IEEE International* (pp. 924-927). IEEE.

34. Kaiser, C., Piazza, A., Krockel, J., &Bodendorf, F. (2011, October). Are Humans Like Ants?--Analyzing Collective Opinion Formation in Online Discussions. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)* (pp. 266-273). IEEE.

35. Cho, S. W., Cha, M. S., Kim, S. Y., Song, J. C., &Sohn, K. A. (2014, May). Investigating Temporal and Spatial Trends of Brand Images Using Twitter Opinion Mining. In *Information Science and Applications (ICISA), 2014 International Conference on* (pp. 1-4). IEEE.

36. Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications*, *41*(16), 7653-7670.

37. Aldahawi, H. A., & Allen, S. M. (2013, September). Twitter Mining in the Oil Business: A Sentiment Analysis Approach. In *Cloud and Green Computing (CGC), 2013 Third International Conference on* (pp. 581-586). IEEE.

38. Chen, H. S., & Su, J. S. (2009, December). The Keyless Technique of Text Data Mining and Its Application. In *Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference on* (pp. 834-840). IEEE.

39. Liang, H., Xu, Y., & Li, Y. (2010, December). Mining Users' Opinions Based on Item Folksonomy and Taxonomy for Personalized Recommender Systems. In*Data Mining Workshops (ICDMW), 2010 IEEE International Conference on* (pp. 1128-1135). IEEE.

Cheng, V., Leung, C., Liu, J., &Milani, A. (2013). Probabilistic Aspect Mining Model for Drug Reviews.