

A SURVEY ON OPINION MINING AND SENTIMENT POLARITY CLASSIFICATION

Sindhu C¹, Dr. S. ChandraKala²

¹PG Student, ²Professor, Computer Science and Engineering, Velammal Engineering College, Chennai, India.

csindhucse@gmail.com

Abstract

In recent years, the exponential increase in the Internet usage and exchange of user's opinions has become the motivation for Opinion Mining. Due to overwhelming amount of user's opinions, views, feedbacks and suggestions available through the web resources, it's very much essential to explore, analyze and organize their views for better decision making. Opinion Mining or Sentiment Analysis is a Natural Language Processing task that identifies the user's opinions expressed in the form of positive, negative or neutral comments underlying the text. This survey gives an overview of the efficient techniques, recent advancements and the future research directions in the field of Opinion Mining and Sentiment Polarity Classification.

Keywords-- Opinion Mining; Sentiment Analysis; Sentiment Classification; Information Extraction

I. INTRODUCTION

Opinion Mining or Sentiment Analysis refers to computational techniques for analyzing the opinions that are extracted from various sources like the blog posts, comments on forums, reviews about products, policies or any topic on social networking sites or tweets. It aims at determining the attitude of a user about some topic. The Web is a huge repository of structured and unstructured data. The analysis of this data to extract underlying user's opinion and sentiment is a challenging task. An opinion can be described as a quadruple consisting of a Topic, Holder, Claim and Sentiment [56]. Here the Holder believes a Claim about the Topic and expresses it through an associated Sentiment. To a machine, opinion is a "quintuple", an object made up of 5 different components: [Bing Liu in NLP Handbook] ($O_j, f_{jk}, SO_{ijkl}, h_i, t_l$), where O_j = the object on which the opinion is given, f_{jk} = a feature of O_j , SO_{ijkl} = the sentiment value of the opinion, h_i = Opinion holder, t_l = the time at which the opinion is given.

There are numerous challenges in the field of Opinion Mining. The most common challenges are given here. First, **Word Sense Disambiguation** (WSD), a classical NLP problem is often encountered, which is the task of selecting the appropriate senses of a word in a given context. For example, "an unpredictable plot in the movie" is a positive phrase, while "an unpredictable steering wheel" is a negative one. The opinion word *unpredictable* is used in different senses. Second, addressing the problem of **sudden deviation** from positive to negative polarity, as in "The movie has a great cast, superb storyline and spectacular photography; the director has managed to make a mess of the whole thing".

Third, **negations**, unless handled properly can completely mislead. "Not only do I not approve Sony W8, but also hesitate to call it a phone" has a positive polarity word *approve*; but its effect is negated by many negations. Fourth, keeping the target in focus (**entity identification**) can be a challenge as in "my camera compares nothing to Ricky's camera which is sleek and light, produces life like pictures and is inexpensive". All the positive words about Ricky's camera being the constituents of the document vector will produce an overall decision of positive polarity, which is wrong [7].

The importance and popularity of Opinion Mining have led to several papers which describes and implements it's variety of tasks using several different techniques, some of them are listed in **Table 1**, together with the years of publication and the tasks involved. **Fig 1** depicts the major important steps in order to achieve an opinion impact. The web users post their views, comments and feedbacks about a particular product or a thing through various blogs, forums, social networking sites, etc. Data is collected from such opinion rich sources in such a way that only the reviews related to the topic, that is searched is selected, which is considered to be the input document. This input document contains both factual and opinionated sentences. In order to mine the polarity of the user's opinion it becomes important to focus only on the opinionated sentences. The process of selecting the opinionated sentences and ignoring the factual sentence is called *Subjectivity Detection*, in opinion mining, which is then pre-processed. Preprocessing consists of tokenizing, stop words filtering and stemming. Then, the process of extracting relevant features is done.

Feature selection can potentially improve classification accuracy [58], narrow in on a key feature subset of sentiment discriminators, and provide greater insight into important class attributes. The extracted features contribute to a document vector upon which various machine learning techniques can be applied in order to classify the polarity (positive and negative opinions) using the obtained document vectors and finally the opinion impact is obtained based on the sentiment of the web users.

Table.1
Recent Papers on the related tasks of Opinion Mining

Tasks	Paper	Year
Sentiment Analysis	[66]	2012
	[70]	2012
	[73]	2011
Subjectivity Analysis	[47]	2011
Sentiment Detection	[48]	2009
Feature Selection for Opinion Mining	[49]	2011
	[62]	2008
	[63]	2011
Review Aggregation	[61]	2008
Supervised Machine Learning Approaches for Opinion Mining	[64]	2009
Sentiment Classification	[76]	2012
	[71]	2012
Active learning for Opinion Mining	[65]	2012

The organization of this paper includes Sentiment Analysis in section 2 under which subjectivity detection, negation and feature based sentiment classification are briefed in the sections 2.1 to 2.3. Feature Extraction and Feature Reduction are explained under 2.3.1 and 2.3.2. The Sentiment Classification is explained under the section 3, under which the polarity and intensity assignment is discussed under 3.1 and 3.2 respectively. The Machine Learning Approaches are discussed in the section 4, under which the Naïve Bayes Classification, Maximum Entropy and Support Vector Machines are briefed. Finally the applications and future challenges of Opinion Mining and Sentiment Classification are elaborated under the sections 5 and 6 respectively.

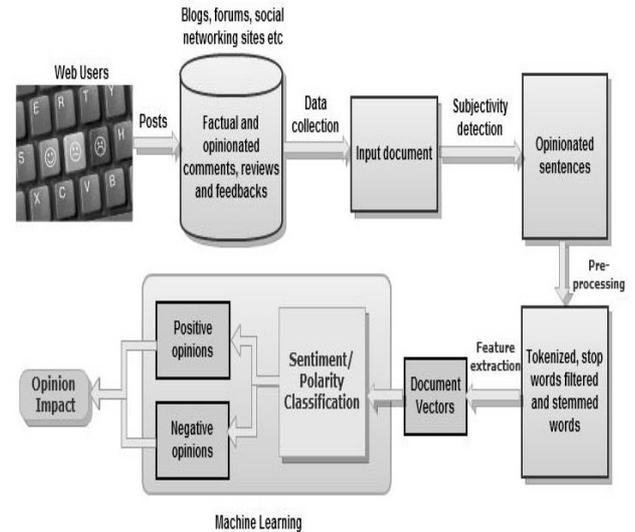


Fig.1. Systematic Work Flow of Opinion Mining

II. SENTIMENT ANALYSIS

Text categorization generally classifies the documents by topic. Such an ordinary keyword search will not be suitable for mining all kinds of opinions. Hence it becomes necessary to use that the sophisticated opinion extraction methods. Sentiment analysis is a natural language processing technique that helps to identify and extract subjective information in source materials. Sentiment analysis aims to determine the attitude of the writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state, or the intended emotional communication. A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level — whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. Beyond polarity sentiment classification, the emotional states such as "angry", "sad" and "happy" can also be identified. One of the challenges of Sentiment Analysis is to define the opinions and subjectivity of the study [7]. Subjectivity is highly context-sensitive, and its expression is often peculiar to each person. Subjectivity Detection and Negation are the most important preprocessing steps in order to achieve efficient opinion impact. They are discussed in the following sections.

2.2 Subjectivity Detection

Subjectivity detection, in opinion mining can be defined as a process of selecting opinion containing sentences [7]. (e.g.,) “India’s economy is heavily dependent on tourism and IT industry. It is an excellent place to live in.” The first sentence is a factual one and does not convey any sentiment towards India. Hence such a sentence does not play any role in deciding on the polarity of the review, and should be filtered out. Here, the polarity classifier assumes that the incoming documents are opinionated. Joint Topic-Sentiment Analysis is done by collecting only on-topic documents (e.g., by executing the topic-based query using a standard search engine). In Information extraction, both topic-based text filtering and subjectivity filtering are complementary as in [8]. If a document contains information on a variety of topics that may attract the attention of the user, then it will be useful to classify the topics and its related opinions. This type of analysis can be useful for comparative search analysis of related items and also to discuss on the texts that contains various features and attributes.

The political orientation of the websites can be done by classifying the concatenation of all the documents found on that particular site as in [9]. Analyzing sentiment and opinions in political oriented text, generally focuses on the attitude expressed via texts which are not targeted at a specific issue. In order to mine opinion, the main concentration is on non-factual information in text. There are various affect types; in general the concentration is on the six “universal” emotions as in [10]: anger, disgust, fear, happiness, sadness and surprise. These emotions could be easily associated with an interesting application of a human-computer interaction, where when a system identifies that the user is upset or annoyed, the system could change the user interface to a different mode of interaction as in [11].

2.3 Negation

Negation is a very common linguistic construction that affects polarity and, therefore, needs to be taken into consideration in sentiment analysis. When treating negation, one must be able to correctly determine what part of the meaning expressed is modified by the presence of the negation. Most of the times, its expression is far from being simple, and does not only contain obvious negation words, such as *not*, *neither* or *nor*.

Research in the field has shown that there are many other words that invert the polarity of an opinion expressed [50], such as *diminishers* / *valence shifters* (e.g., I find the functionality of the new phone *less* practical), *connectives* (Perhaps it is a great phone, *but* I fail to see why), or even *modals* (In theory, the phone *should* have worked even under water). As can be seen from these examples, modeling negation is a difficult yet an important aspect of sentiment analysis.

2.4 Feature Based Sentiment Classification

Feature engineering is an extremely basic and essential task for Opinion Mining. Converting a piece of text into a feature vector is the basic step in any data driven approach to Opinion Mining. It is important to convert a piece of text into a feature vector, so as to process text in a much efficient manner. In text domain, effective feature selection is a must in order to make the learning task effective and accurate. In text classification, with the bag of words model, each position in the input feature vector corresponds to a given word or phrase. In the bag of words framework, the documents are often converted into vectors based on predefined feature presentation including feature type and features weighting mechanism, which is critical to classification accuracy. The major feature types contain unigrams, bigrams and the mixtures of them, etc. The features weighting mechanism mainly includes presence, frequency, *tf*idf* and its variants [57]. The commonly used features used in Sentiment Analysis and their critiques [70] are Term Presence, Term Frequency, Term Position, Subsequence Kernels, Parts of Speech, Adjectives, Adjective-Adverb Combination, *n*-gram features etc.

2.3.1 Feature Extraction

Let us consider the *n*-gram features for feature extraction. An *n*-gram is a contiguous sequence of *n* items from a given sequence of text or speech. An *n*-gram could be any combination of letters [49] (syllables, letters, words, part-of-speech (POS), characters, syntactics, and semantic *n*-grams). The *n*-grams typically are collected from a text or speech corpus and *n*-gram features captures sentiment cues in text. *n*-gram features can be classified into two categories: 1) Fixed *n*-grams are exact sequences occurring at either the character or token level. 2) Variable *n*-grams are extraction patterns capable of representing more sophisticated linguistic phenomena.

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

A plethora of fixed and variable n -grams have been used for opinion mining [50]. Documents are often converted into vectors according to predefined features together with weighting mechanisms [57]. Correlation is a commonly used method for feature selection [58], [59]. The process of obtaining n -gram can be given as in the steps below:

1. *Filtering* - removing URL Links
2. *Tokenization* - Segmenting text by splitting it by spaces and punctuation marks, and forming bag of words
3. *Removing Stop Words* - Removing articles("a", "an", "the")
4. *Constructing n -grams* - from consecutive words
After the extraction of the features, they are reduced.

2.3.2 Feature Reduction

Feature reduction is an important part of optimizing the performance of a classifier by reducing the feature vector to a size that does not exceed the number of training cases as a starting point. Further reduction of vector size can lead to more improvements if the features are noisy or redundant. Reducing the number of features in the feature vector can be done in two different ways [41]:

- 1) reduction to the top ranking n features based on some criterion of "predictiveness",
- 2) reduction by elimination of set of features (e.g. elimination of linguistic analysis features etc.)

Now that the extracted features are reduced, the classification of the sentiment based on the polarity and intensity of the text using any of the machine learning approaches is to be done.

III. SENTIMENT CLASSIFICATION

Sentiment Classification mainly consists of two important tasks, including sentiment polarity assignment and sentiment intensity assignment [49]. Sentiment polarity assignment deals with analyzing, whether a text has a positive, negative, or neutral semantic orientation. Sentiment intensity assignment deals with analyzing, whether the positive or negative sentiments are mild or strong. There are several tasks in order to achieve the goals of Sentiment Analysis. These tasks include sentiment or opinion detection, polarity classification and discovery of the opinion's target. Sentiment Classification broadly refers to binary categorization, multi-class categorization, regression and ranking.

The various methodologies used in order to achieve Sentiment Classification are 1) Classification with respect to term frequency, n -grams, negations or parts of speech, 2) Identification of the semantic orientation of words using lexicon, statistical techniques and training documents, 3) Identification of the semantic orientation of the sentences and phrases, 4) Identification semantic orientation of the documents 4) Object feature extraction, 5) Comparative sentence identification.

3.1 Polarity Assignment

Sentiment polarity assignment deals with analyzing, whether a text has a positive, negative, or neutral semantic orientation. The Sentiment Polarity Classification is a binary classification task where an opinionated document is labeled with an overall positive or negative sentiment. Sentiment Polarity Classification can also be termed as a binary decision task. When a news article is given as an input, analyzing and classifying it as a good or bad news is considered to be a text categorization task as in [5]. Furthermore, this piece of information can be good or bad news, but not necessarily subjective (i.e., without expressing the view of the author). Summarizing reviews in order to collect information on to why the reviewers liked or disliked the product is another way of mining opinion. In order to determine the polarity of the outcomes as described in medical texts is yet another type of categorization related to the degree of positivity as in [6]. Few other problems related to the determination of the degree of positivity are the analysis of comparative sentences. Automated opinion mining often uses machine learning, a component of artificial intelligence.

3.2 Intensity Assignment

While Sentiment polarity assignment deals with analyzing, whether a text has a positive, negative, or neutral semantic orientation, Sentiment intensity assignment deals with analyzing, whether the positive or negative sentiments are mild or strong. Consider the two phrases "I don't like you" and "I hate you", where, both the sentences would be assigned a negative semantic orientation but the latter would be considered more intense than the first [49]. Effectively classifying sentiment polarities and intensities entails the use of classification methods applied to linguistic features. While several classification methods have been employed for opinion mining, Support Vector Machine (SVM) has outperformed various techniques including Naive Bayes, Decision Trees, Winnow, etc. [67], [68] and [69].

IV. MACHINE LEARNING APPROACHES

The aim of Machine Learning is to develop an algorithm so as to optimize the performance of the system using example data or past experience. The Machine Learning provides a solution to the classification problem that involves two steps: 1) Learning the model from a corpus of training data 2) Classifying the unseen data based on the trained model. In general, classification tasks are often divided into several sub-tasks:

1. Data preprocessing
2. Feature selection and/or feature reduction.
3. Representation
4. Classification.
5. Post processing

Feature selection and feature reduction attempt to reduce the dimensionality (i.e. the number of features) for the remaining steps of the task. The classification phase of the process finds the actual mapping between patterns and labels (or targets). Active learning, a kind of machine learning is a promising way for sentiment classification to reduce the annotation cost [65]. The following are some of the Machine Learning approaches commonly used for Sentiment Classification.

4.1 Naive Bayes Classification

A naive Bayes classifier is a simple probabilistic classifier based on Bayes' theorem and is particularly suited when the dimensionality of the inputs are high. Naïve Bayes classification is an approach to text classification that assigns the class $c^* = \arg \max_c P(c | d)$, to a given document d . Its underlying probability model can be described as an "independent feature model". The *Naive Bayes* (NB) classifier uses the Bayes' rule (1),

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)}, \quad (1)$$

Where $P(d)$ plays no role in selecting c^* . To estimate the term $P(d | c)$, Naive Bayes decomposes it by assuming the f_i 's are conditionally independent given d 's class as in (2).

$$P_{NB}(c|d) := \frac{P(c)(\prod_{i=1}^m P(f_i|c)^{n_i(d)})}{P(d)} \quad (2)$$

Where m is the no of features and f_i is the feature vector.

Consider a training method consisting of a relative-frequency estimation $P(c)$ and $P(f_i | c)$. Despite its simplicity and the fact that its conditional independence assumption clearly does not hold in real-world situations, Naive Bayes-based text categorization still tends to perform surprisingly well [13]; indeed, Naive Bayes is optimal for certain problem classes with highly dependent features[29].

4.2 Maximum Entropy

Maximum Entropy (ME) classification is yet another technique, which has proven effective in a number of natural language processing applications [26]. Sometimes, it outperforms Naive Bayes at standard text classification [27]. Its estimate of $P(c | d)$ takes the exponential form as in (3).

$$P_{ME}(c|d) := \frac{1}{Z(d)} \exp(\sum_i \lambda_{i,c} F_{i,c}(d, c)) \quad (3)$$

Where $Z(d)$ is a normalization function. $F_{i,c}$ is a *feature/class function* for feature f_i and class c , as in (4).

$$F_{i,c}(d, \hat{c}) := \begin{cases} 1, & n_i(d) > 0 \text{ and } \hat{c} = c \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

For instance, a particular feature/class function might fire if and only if the bigram "still hate" appears and the document's sentiment is hypothesized to be negative. Importantly, unlike Naive Bayes, Maximum Entropy makes no assumptions about the relationships between features and so might potentially perform better when conditional independence assumptions are not met.

4.3 Support Vector Machines

Support vector machines (SVMs) have been shown to be highly effective at traditional text categorization, generally outperforming Naive Bayes [40]. They are *large-margin*, rather than probabilistic, classifiers, in contrast to Naive Bayes and Maximum Entropy. In the two-category case, the basic idea behind the training procedure is to find a maximum margin hyper plane, represented by vector $\vec{\omega}$, that not only separates the document vectors in one class from those in the other, but for which the separation, or *margin*, is as large as possible. This corresponds to a constrained optimization problem; letting $c_j \in \{1, -1\}$ (corresponding to positive and negative) be the correct class of document d_j , the solution can be written as in (5).

$$\vec{\omega} := \sum_j \alpha_j c_j \vec{d}_j, \quad \alpha_j \geq 0, \quad (5)$$

Where the α_j 's (Lagrange multipliers) are obtained by solving a dual optimization problem. Those \vec{d}_j for which α_j are greater than zero are called *support vectors*, since they are the only document vectors contributing to $\vec{\omega}$. Classification of test instances consists simply of determining which side of $\vec{\omega}$'s hyperplane they fall on.

V. APPLICATIONS

Sentiment analysis and opinion mining systems also have a potential role in imparting sub-component technology for other systems.

Specifically, sentiment analysis system is an augmentation to *recommendation systems* [14, 15]; since it might recommend such a system not to suggest items that receive a lot of negative feedback. In online systems that display ads as sidebars is sometimes helpful to detect web pages that contain sensitive content inappropriate for ads placement [16]; for more sophisticated systems, it could be useful to bring up product ads when relevant positive sentiments are detected. It has also been argued that information extraction can be improved by discarding information found in subjective sentences [17]. *Detection of "flames"* (overly-heated opposition) in email or other types of communication is another possible use of subjectivity detection [4].

There are quite a large number of companies, big and small, that have opinion mining and sentiment analysis as part of their mission. Review-oriented search engines basically use sentiment classification techniques. Opinion Mining proves itself to be an important part of search engines. Topics need not be restricted to product reviews, but could include opinions about candidates running for office, political issues, and so forth. Summarizing user reviews is an important problem. One could also imagine that errors in user ratings could be fixed: there are cases where users have clearly accidentally selected a low rating when their review indicates a positive evaluation [3].

Opinion-oriented questions may require different treatment, hence *question and answering* is another area where opinion mining can prove useful [18, 19, and 20]. For definitional questions, providing an answer that includes more information about how an entity is viewed may better inform the user [18]. Summarization may also benefit from accounting for multiple viewpoints [21]. One effort seeks to use semantic orientation to track literary reputation [22]. In general, the computational treatment of affect has been motivated in part by the desire to improve human-computer interaction [23, 54, and 55]. It's the breadth of opportunities – promising ways text analytics can be applied to extract and analyze attitudinal information from sources as varied as articles, blog postings, e-mail, call-center notes and survey responses – and the difficulty of the technical challenges that make existing and emerging applications so interesting. Three other applications include influence networks, assessment of marketing response and customer experience management/enterprise feedback management.

VI. FUTURE CHALLENGES

People don't always express opinions the same way. Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In opinion mining, however, "the movie was great" is very different from "the movie was not great"[50].

Analyzing the sentiment of the web user reviews brings in several challenges. First, a word can have positive sense in one situation and negative in another. Consider the word "*long*" for instance. If a customer said a laptop's battery life was long, that would be a positive opinion. If the customer said that the laptop's start-up time was long, however, that would be a negative opinion [49]. These differences mean that an opinion system trained to gather opinions on one type of product or product feature may not perform very well on another.

People can be contradictory in their statements. Most reviews will have both positive and negative comments, which is somewhat manageable by analyzing sentences one at a time. However, the more informal the medium (twitter or blogs for example), the more likely people are to combine different opinions in the same sentence. For example: "the movie bombed even though the lead actor rocked it" is easy for a human to understand, but more difficult for a computer to parse. Sometimes even other people have difficulty understanding what someone thought based on a short piece of text because it lacks context. For example, "That movie was as good as his last one" is entirely dependent on what the person expressing the opinion thought of the previous film.

Pragmatics is a subfield of linguistics which studies the ways in which context contributes to meaning. It is important to detect the pragmatics of user opinion which may change the sentiment thoroughly. Capitalization can be used with subtlety to denote sentiment. In the examples given below, the first example denotes a positive sentiment whereas the second denotes a negative sentiment.

*I just finished watching THE DESTROY.
That completely destroyed me.*

Another challenge is the *entity identification*. A text or sentence may have multiple entities associated with it. It is extremely important to find out the entity towards which the opinion is directed. Consider the following examples [70].

Nokia is better than Sony.

Mani defeated Kumar in football.

The examples are positive for Nokia and Mani respectively but negative for Sony and Kumar.

VII. CONCLUSION

This survey discusses various approaches to Opinion Mining and Sentiment Classification. It provides a detailed view of different applications and potential challenges of Sentiment Classification that makes it a difficult task. Some of the machine learning techniques like Naïve Bayes, Maximum Entropy and Support Vector Machines has been discussed. Many of the applications of Opinion Mining are based on bag-of-words, which do not capture context which is essential for Opinion Mining. The recent developments in Opinion Mining and its related sub-tasks are also presented. The state of the art of existing approaches has been described with the focus on the following tasks: Subjectivity detection, Word Sense Disambiguation, Feature Extraction and Sentiment Classification using various Machine learning techniques. Finally, the future challenges and directions so as to further enhance the research in the field of Opinion Mining and Sentiment Classification are discussed.

REFERENCES

- [1] Alexander Pak and Patrick Paroubek, "Twitter as a corpus for Sentiment Analysis and Opinion Mining", Proceedings of the Seventh conference on International Language Resources and Evaluation LREC'10 Valletta, Malta: European Language Resources Association ELRA (May 2010).
- [2] Bo Pang, Lillian Lee and Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", 2012.
- [3] Lu's Cabral and Ali Hortac, "The dynamics of seller reputation: Theory and evidence from eBay, 2006". URL <http://pages.stern.nyu>.
- [4] Ellen Spertus. Smokey, "Automatic recognition of hostile messages", in Proceedings of Innovative Applications of Artificial Intelligence (IAAI), pages 1058–1065, 1997.
- [5] SanjivRanjan Das, Peter Tufano, and Francisco de Asis Martinez-Jerez. e, "Information: A clinical study of investor discussion and sentiment", Financial Management, 34(3):103–137, 2005.
- [6] Yun Niu, Xiaodan Zhu, Jianhua Li, and Graeme Hirst, "Analysis of polarity information in medical text", in Proceedings of the American Medical Informatics Association 2005, Annual Symposium, 2005.
- [7] ShitanshuVerma, and Pushpak Bhattacharyya, "Incorporating Semantic Knowledge for Sentiment Analysis", in Proceedings of ICON-2008
- [8] Ellen Riloff and JanyceWiebe, "Learning extraction patterns for subjective expressions", in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2003.
- [9] Gregory Grefenstette, Yan Qu, James G. Shanahan, and David A. Evans, "Coupling niche browsers and affect analysis for an opinion mining application", in Proceedings of Recherched'InformationAssist'ee par Ordinateur (RIAO), 2004.
- [10] Paul Ekman, "Emotion in the Human Face". Cambridge University Press, second edition, 1982.
- [11] Lisa Hankin, "The effects of user reviews on online purchasing behavior across multiple product categories", Master's final project report, UC Berkeley School of Information, May 2007. http://www.ischool.berkeley.edu/files/lhankin_report.pdf.
- [12] George Forman, "An Extensive Empirical study of feature selection Metrics for Text Classification", in Journal of Machine Learning Research 3 (2003).
- [13] David D. Lewis, "Naive (Bayes) at forty: The independence assumption in information retrieval", in Proceedings of the European Conference on Machine Learning (ECML), pages 4–15, 1998.
- [14] Junichi Tatemura, "Virtual reviewers for collaborative exploration of movie reviews", in Proceedings of Intelligent User Interfaces (IUI), pages 272–275, 2000.
- [15] Loren Terveen, Will Hill, Brian Amento, David McDonald, and Josh Creter, "PHOAKS: A system for sharing recommendations", Communications of the Association for Computing Machinery (CACM), 40(3):59–62, 1997.
- [16] Xin Jin, Ying Li, Teresa Mah, and Jie Tong, "Sensitive webpage classification for content advertising", in Proceedings of the International Workshop on Data Mining and Audience Intelligence for Advertising, 2007.
- [17] Ellen Riloff, JanyceWiebe, and William Phillips, "Exploiting subjectivity classification to improve information extraction", in Proceedings of AAAI, pages 1106–1111, 2005.
- [18] Lucian VladLita, Andrew Hazen Schlaikjer, WeiChang Hong, and Eric Nyberg, "Qualitative dimensions in question answering: Extending the definitional QA task", in Proceedings of AAAI, pages 1616–1617, 2005.
- [19] SwapnaSomasundaran, Theresa Wilson, JanyceWiebe, and VeselinStoyanov, "QA with attitude: Exploiting opinion type analysis for improving question answering in on-line discussions and the news", in Proceedings of the International Conference on Weblogs and Social Media (ICWSM), 2007.
- [20] VeselinStoyanov, Claire Cardie, and JanyceWiebe, "Multi-perspective question answering using the OpQA corpus", in Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pages 923–930, October 2005.
- [21] Yohei Seki, Koji Eguchi, Noriko Kando, and Masaki Aono, "Multi-document summarization with subjectivity analysis" in Proceedings of the Document Understanding Conference (DUC), 2005.
- [22] MaiteTaboada, Mary Ann Gillies, and Paul McFetridge, "Sentiment classification techniques for tracking literary reputation", in LRECWorkshop: Towards Computational Models of Literary Analysis, pages 36–43, 2006.
- [23] Jackson Liscombe, Giuseppe Riccardi, and DilekHakkani-Tur, "Using context to improve emotion detection in spoken dialog systems", in Interspeech, pages 1845–1848, 2005.
- [24] Hugo Liu, Henry Lieberman, and Ted Selker, "A model of textual affect sensing using real-world knowledge", in Proceedings of Intelligent User Interfaces (IUI), pages 125–132, 2003.

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

- [25] Matt Thomas, Bo Pang, and Lillian Lee, "Get out the vote: Determining support or opposition from Congressional floor-debate transcripts", In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 327–335, 2006.
- [26] Adam L. Berger, Stephen A. Della Pietra, and Vincent J. Della Pietra, "A maximum entropy approach to natural language processing", Computational Linguistics, 22(1):39–71, 1996.
- [27] Kamal Nigam, John Lafferty, and Andrew McCallum, "Using maximum entropy for text classification", in Proceedings of the IJCAI-99 Workshop on Machine Learning for Information Filtering, pages 61–67, 1999.
- [28] Andrew McCallum and Kamal Nigam, "A comparison of event models for Naive Bayes text classification". In Proceedings of the AAAI-98 Workshop on Learning for Text Categorization, pages 41–48, 1998.
- [29] Pedro Domingos and Michael J. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss", Machine Learning, 29(2-3):103–130, 1997.
- [30] Stephen Della Pietra, Vincent Della Pietra, and John Lafferty, "Inducing features of random fields". IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(4):380–393, 1997.
- [31] Stanley Chen and Ronald Rosenfeld, "A survey of smoothing techniques for ME models". IEEE Transactions Speech and Audio Processing, 8(1):37–50, 2000.
- [32] Sanjiv Das and Mike Chen, "Yahoo! for Amazon: Extracting market sentiment from stock message boards", in Proceedings of the 8th Asia Pacific Finance Association Annual Conference (APFA2001), 2001.
- [33] Douglas Biber, "Variation across Speech and Writing", Cambridge University Press, 1988.
- [34] Shlomo Argamon-Engelson, Moshe Koppel, and Galit Avneri, "Style-based text categorization: What newspaper am I reading?" in Proceedings of the AAAI Workshop on Text Categorization, pages 1–4, 1998.
- [35] Aidan Finn, Nicholas Kushmerick, and Barry Smyth, "Genre classification and domain transfer for information filtering", in Proceedings of the European Colloquium on Information Retrieval Research, pages 353–362, Glasgow, 2002.
- [36] Vasileios Hatzivassiloglou and Kathleen McKeown, "Predicting the semantic orientation of adjectives", in Proceedings of the 35th ACL/8th EACL, pages 174–181, 1997.
- [37] Vasileios Hatzivassiloglou and Janyce Wiebe, "Effects of adjective orientation and gradability on sentence subjectivity", in Proceedings of COLING, 2000.
- [38] Marti Hearst, "Direction-based text interpretation as an information access refinement", in Paul Jacobs, editor, Text-Based Intelligent Systems, Lawrence Erlbaum Associates, 1992.
- [39] Alison Huettner and Pero Subasic, "Fuzzy typing for document management", in ACL 2000 Companion Volume: Tutorial Abstracts and Demonstration Notes, pages 26–27.
- [40] Thorsten Joachims. "Text categorization with support vector machines: Learning with many relevant features", in Proceedings of the European Conference on Machine Learning (ECML), pages 137–142, 1998.
- [41] Michael Gamon, "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis", Proceeding COLING '04 Proceedings of the 20th international conference on Computational Linguistics Article No. 841, 2004.
- [42] Hsinchun Chen and David Zimbra, "AI and Opinion Mining", Published by the IEEE Computer Society, 2010.
- [43] Hsinchun Chen, "AI and Opinion Mining, Part 2", Published by the IEEE Computer Society, 2010.
- [44] S. Shivashankar and B. Ravindran, "Multi Grain Sentiment Analysis using Collective Classification", ECAI-2010. 823-828.
- [45] George Stylios, Dimitris Christodoulakis, Jeries Besharat, Maria-Alexandra Vonitsanou, Ioanis Kotrotsos, Athanasia Koumpouri and Sofia Stamou, "Public Opinion Mining for Governmental Decisions", Electronic Journal of e-Government Volume 8 Issue 2 2010 (pp203-214).
- [46] Anindya Ghose and Panagiotis G. Ipeirotis, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics". IEEE Transactions On Knowledge And Data Engineering, Vol. 23, No. 10, October 2011.
- [47] Janyce Wiebe and Ellen Riloff, "Finding Mutual Benefit between Subjectivity Analysis and Information Extraction", IEEE Transactions On Affective Computing, Vol. 2, No. 4, October-December 2011.
- [48] Huifeng Tang, Songbo Tan *, Xueqi Cheng, "A survey on sentiment detection of reviews", Expert Systems with Applications 36 (2009) 10760–10773.
- [49] Ahmed Abbasi, Stephen France, Zhu Zhang, and Hsinchun Chen, "Selecting Attributes for Sentiment Classification Using Feature Relation Networks", in IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No. 3, March 2011.
- [50] Michael Wiegand and Alexandra Balahur, "A Survey on the Role of Negation in Sentiment Analysis".
- [51] Thorsten Joachims, "Making large-scale SVM learning practical", in Bernhard Schölkopf and Alexander Smola, editors, Advances in Kernel Methods - Support Vector Learning, pages 44–56. MIT Press, 1999.
- [52] Jussi Karlgren and Douglass Cutting, "Recognizing text genres with simple metrics using discriminant analysis", in Proceedings of COLING, 1994.
- [53] Brett Kessler, Geoffrey Nunberg, and Hinrich Schütze, "Automatic detection of text genre", in Proc. of the 35th ACL/8th EACL, pages 32–38, 1997.
- [54] Hugo Liu, Henry Lieberman, and Ted Selker, "A model of textual affect sensing using real-world knowledge", in Proceedings of Intelligent User Interfaces (IUI), pages 125–132, 2003.
- [55] Junichi Tatemura, "Virtual reviewers for collaborative exploration of movie reviews" in Proceedings of Intelligent User Interfaces (IUI), pages 272–275, 2000.
- [56] Soo-Min Kim and Eduard Hovy, "Determining the Sentiment of Opinions". Proceedings of the COLING conference, Geneva, 2004.
- [57] Yuming Lin, Jingwei Zhang, Xiaoling Wang, Aoying Zhou, "Sentiment Classification via Integrating Multiple Feature Presentations". WWW 2012 – Poster Presentation, April 16–20, 2012, Lyon, France.

International Conference on Information Systems and Computing (ICISC-2013), INDIA.

- [58] M. Hall and L.A. Smith, "Feature Subset Selection: A Correlation Based Filter Approach," Proc. Fourth Int'l Conf. Neural Information Processing and Intelligent Information Systems, pp. 855-858, 1997.
- [59] G. Forman, "An Extensive Empirical Study of Feature Selection Metrics for Text Classification", J. Machine Learning Research, vol. 3, pp. 1289-1305, 2004.
- [60] Bo Pang¹ and Lillian Lee², "Opinion mining and sentiment analysis". Foundations and Trends in Information Retrieval, Vol. 2, No 1-2 (2008) 1-135, 2008.
- [61] Z. Zhang, "Weighing Stars: Aggregating Online Product Reviews for Intelligent E-Commerce Applications," IEEE Intelligent Systems, vol. 23, no. 5, pp. 42-49, Sept. 2008.
- [62] A. Abbasi, H. Chen, and A. Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," ACM Trans. Information Systems, vol. 26, no. 3, article no. 12, 2008.
- [63] SugeWang ,Deyu Li , Xiaolei Song , Yingjie Wei and Hongxia Li, "A feature selection method based on improved fisher's discriminant ratio for text Sentiment Classification" in Expert Systems with Applications 38, 8696-8702, 2011
- [64] Ye, Q., Zhang, Z. Q., & Rob, L, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches". Expert System with Application, 36(3), 6527-6535, 2009
- [65] Shoushan Li, ShengfengJu, Guodong Zhou and Xiaojun Li, "Active Learning for Imbalanced Sentiment Classification" in Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 139-148, Jeju Island, Korea, 12-14 July 2012.
- [66] Claudiu-CristianMusat THISONE, AlirezaGhasemi and BoiFaltings, "Sentiment Analysis Using a Novel Human Computation Game" in Proceedings of the 3rd Workshop on the People's Web Meets NLP, ACL 2012, pages 1-9.
- [67] A. Abbasi and H. Chen, "CyberGate: A System and Design Framework for Text Analysis of Computer Mediated Communication," MIS Quarterly, vol. 32, no. 4, pp. 811-837, 2008.
- [68] A. Abbasi, H. Chen, S. Thoms, and T. Fu, "Affect Analysis of Web Forums and Blogs Using Correlation Ensembles," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 9, pp. 1168-1180, Sept. 2008.
- [69] B. Pang and L. Lee, "A Sentimental Education: Sentimental Analysis Using Subjectivity Summarization Based on Minimum Cuts," Proc. 42nd Ann. Meeting of the Assoc. Computational Linguistics, pp. 271-278, 2004.
- [70] Subhabrata Mukherjee, "Sentiment Analysis-A Literature Survey", Indian Institute of Technology, Bombay, Department of Computer Science and Engineering, June 29, 2012.
- [71] Naradhipa, A.R.; Purwarianti, "A. Sentiment classification for Indonesian message in social media Cloud Computing and Social Networking" in International Conference on Digital Object Identifier: 10.1109/ICCCSN.2012.6215730, 2012.
- [72] Chenghua Lin; Yulan He; Everson, R.; Ruger, S, "Weakly Supervised Joint Sentiment-Topic Detection from Text", IEEE Transactions on Knowledge and Data Engineering, Volume: 24, Issue: 6, 2012.
- [73] Neviarouskaya, A.; Prendinger, H.; Ishizuka, M, "SentiFul: A Lexicon for Sentiment Analysis". IEEE Transactions on Affective Computing, Volume: 2, Issue: 1, 2011.
- [74] Xiaohui Yu; Yang Liu; Xiangji Huang; Aijun, "Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain", IEEE Transactions on Knowledge and Data Engineering, Volume: 24, Issue: 4, 2012.
- [75] Chien-Liang Liu; Wen-Hoar Hsaio; Chia-Hoang Lee; Gen-Chi Lu; Jou, E, "Movie Rating and Review Summarization in Mobile Environment", Part C: Applications and Reviews, IEEE Transactions on Systems, Man, and Cybernetics, Volume: 42, Issue: 3, 2012.
- [76] Bollegala, D.; Weir, D.; Carroll, J, "Cross-Domain Sentiment Classification using a Sentiment Sensitive Thesaurus". IEEE Transactions on Knowledge and Data Engineering, Volume: PP, Issue: 99, 2012.