



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Issue 2, February 2017

## A Survey on Research Issues in Opinion Mining

Dr.B.Radha, V.Meera

Assistant Professor, Department of Computer Science, Sree Saraswathi Thyagaraja College, Pollachi,  
Tamilnadu, India

Assistant Professor, Department of Commerce, Sree Saraswathi Thyagaraja College, Pollachi, Tamilnadu, India

**ABSTRACT:** Natural language processing is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human-computer interaction. Many challenges in NLP involve: natural language understanding, enabling computers to derive meaning from human or natural language input; and others involve natural language generation.

**KEYWORDS:** Opinion mining, Sentiment Analysis, Computational Linguistics, Text mining

### I. INTRODUCTION

The field of study that focuses on the interactions between human language and computers is called Natural Language Processing, or NLP for short. It sits at the intersection of computer science, artificial intelligence, and computational linguistics. NLP is a way for computers to analyze, understand, and derive meaning from human language in a smart and useful way. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation. NLP algorithms are typically based on machine learning algorithms. Instead of hand-coding large sets of rules, NLP can rely on machine learning to automatically learn these rules by analyzing a set of examples (i.e. a large corpus, like a book, down to a collection of sentences), and making a statical inference. In general, the more data analyzed, the more accurate the model will be.

### II. LITERATURE SURVEY

Guohong Fu et al. presented a fuzzy set theory based approach to Chinese sentence level sentiment classification. They define three fuzzy set to represent the respective sentiment polarity classes namely positive, negative and neutral sentiments. Here the authors extract a dictionary of sentiment morphemes from a sentiment lexicon, and compute their opinions scores using a modified chi-square technique. After that they develop two rule based strategies for word level and phrase level polarity identification, respectively. Then they calculate the final sentiment intensity of an opinioned sentence by summing the opinion score of all phrases within it. In the experiment the system involve three main modules, namely a lexical analysis module, subjectivity detection module and a sentiment classification module. The experiment results show that their system outperform the best system for Chinese opinion analysis (COP) pilot task at NTCIR-6 under the lenient evaluation standard.

N. Mohanda et al. focuses on tagging the appropriate mood in Malayalam text. Tagging is used to specify whether a sentence indicates a sad, happy or angry mood of the person involved or if the sentence contains just facts, devoid of emotions. The first step of proposed method is to manually collect the corpus from the Malayalam novels. Second step is to manually tag the corpus. The part-of speech of each word of each sentence will be manually tagged appropriately. The next step, semantic orientation is calculated with extracted words and modified formulas. Then the result of the calculation will help in classifying the sentence into one of the four classes- joy, sorrow,



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirce.com](http://www.ijirce.com)

Vol. 5, Issue 2, February 2017

Anger or Neural. Andrew L. Maas et al. presented a model that uses a mix of unsupervised and supervised techniques to learn word vectors capturing semantic term-document information as well as rich sentiment content. This model capture both semantic and sentiment similarities among words. Authors evaluate this model with document level and sentence level categorization task in the domain of online movie reviews. For document and sentence level authors compare this model's word representations with several bags of words weighting method, and alternative approaches to word vector induction. For experiment authors used IMDB review dataset. They evaluate classifier performance after cross validating classifier parameters on the training set using a linear SVM in all cases. However their model showed superior performance to other approaches, and performed best when concatenated with bag of words representation. Andrew et al also performed sentence level subjectivity classification. For this task classifier is trained to decide whether a given sentence is subjective, expressing the writer's opinions, or objective, expressing the writer's opinions, or objective, expressing purely facts. Author's uses dataset of Pang and Lee(2004), which contains subjective sentences from movie review summaries and objective sentences from movie plot summaries. Authors randomly split the 10,000 examples into 10 folds and report 10 fold cross validation accuracy using the SVM training protocol of Pang and Le(2004). However author find that their model provided superior feature when compared against others SVM.

Gizem et al. proposed and evaluate new feature to be used in a word polarity based approach to sentiment classification. They used dataset TripAdvisor corpus consists of around 250,000 customer-supplied reviews of 1850 hotels and each review is associated with a hotel and a 1-star-rating to 5-star. According to performances authors uses Support Vector Machine and Logistics regression. The SVM is trained using a radial basis function kernel as provided by Lib SVM. The authors consider reviews with star rating bigger than 2 are positive reviews and rest are negative reviews. However for results authors used grid search on validation set. After those optimum parameters, trained their system on training set and tested it on testing set. They find that using sentence level features bring improvements over the best result, albeit small.

Asad et al. proposed a suffix tree data structure to represent syntactic relationships between opinions targets and words in a sentence that are opinion-bearing. Data source are Sentiment corpora with sub-sentential annotations, such as the Multi-Perspective Question-Answering(MPQA) Corpus(Wilson and WIEBE, 2005) and the J.D. Power and Associates(JDPA) blog post corpus(Kessler et al.,2010.). Their baseline system is the initial setting of the labels for the sampler and next system involve combination of our SRT factors with the observed linguistic features. Authors also experiment with including and excluding combinations of POS, role, and word features. The accuracy measure does show overall improvement with the inclusion of more feature factor combinations.

## A. METHODOLOGIES PROPOSED FOR OPINION MINING

Opinion features such as reviews on a particular product are typically domain-specific. The feature appears frequently in the given review domain, and which are outside the domain is domain-independent corpus about product. Domain-specific opinion features are mentioned more frequently in the domain corpus of reviews, as compared to a domain-independent corpus. A domain-dependent review corpus and a domain-independent corpus is observed. Figure 1 shows that, first extract a list of candidate features from the review corpus by defining manually syntactic rules. Each extracted candidate feature, will estimate its IDR, which represents the statistical association of the candidate to the given domain corpus, and extrinsic-domain relevance, will reflects the statistical relevance of the candidate to the domain-independent corpus. Only candidates with IDR scores more exceeding a predefined intrinsic relevance threshold and EDR scores less than another extrinsic relevance threshold are extracted as valid opinion features. In short, this paper identifies opinion features that are domain-specific and at the same time domain-independent corpus are removed and ignored

## II. TEXT MINING

Also referred to as text data mining, roughly equivalent to text analytics, is the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a database), deriving patterns within the structured data, and finally evaluation and interpretation of the output. 'High quality' in text mining usually refers to some combination of relevance, novelty, and interestingness. Typical text



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Issue 2, February 2017

mining tasks include text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling (i.e., learning relations between named entities). Text analysis involves information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques including link and association analysis, visualization, and predictive analytics. The overarching goal is, essentially, to turn text into data for analysis, via application of natural language processing (NLP) and analytical methods. A typical application is to scan a set of documents written in a natural language and either model the document set for predictive classification purposes or populate a database or search index with the information extracted.

### III.COMPUTATIONAL LINGUISTICS

It is an interdisciplinary field concerned with the statistical or rule-based modeling of natural language from a computational perspective.

Traditionally, computational linguistics was performed by computer scientists who had specialized in the application of computers to the processing of a natural language. But little if any success was made. Computational linguists often work as members of interdisciplinary teams, which can include regular linguists, experts in the target language, and computer scientists. In general, computational linguistics draws upon the involvement of linguists, computer scientists, experts in artificial intelligence, mathematicians, Quantum Physicists, logicians, philosophers, cognitive scientists, cognitive psychologists, psycholinguists, anthropologists and neuroscientists, among others. Computational linguistics has theoretical and applied components. Theoretical computational linguistics focuses on issues in theoretical linguistics and cognitive science, and applied computational linguistics focuses on the practical outcome of modeling human language use.<sup>[1]</sup> The Association for Computational Linguistics defines computational linguistics as: "...the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models of various kinds of linguistic phenomena.

Modern computational linguistics is often a combination of studies in computer science and programming, math, particularly statistics, language structures, and natural language processing. Combined, these fields most often lead to the development of systems that can recognize speech and perform some task based on that speech. Examples include speech recognition software, such as Apple's Siri feature, spellcheck tools, speech synthesis programs, which are often used to demonstrate pronunciation or help the disabled, and machine translation programs and websites, such as Google Translate and Word Reference.<sup>[34]</sup>

Computational linguistics can be especially helpful in situations involving social media and the Internet. For example, filters in chatrooms or on website searches require computational linguistics. Chat operators often use filters to identify certain words or phrases and deem them inappropriate so that users cannot submit them.[34] Another example of using filters is on websites. Schools use filters so that websites with certain keywords are blocked from children to view. There are also many programs in which parents use Parental controls to put content filters in place. Computational linguists can also develop programs that group and organize content through Social media mining. An example of this is Twitter, in which programs can group tweets by subject or keywords.

### IV.OPINION MINING

Opinion mining is a type of natural language processing for tracking the mood of the public about a particular product. Opinion mining, which is also called sentiment analysis, involves building a system to collect and categorize opinions about a product. In this paper we focus on what are the research issues in opinion mining and how we can use it for future research.

A very basic step of opinion mining and sentiment analysis is feature extraction. Figure 1 shows the process of opinion mining and sentiment analysis.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Issue 2, February 2017

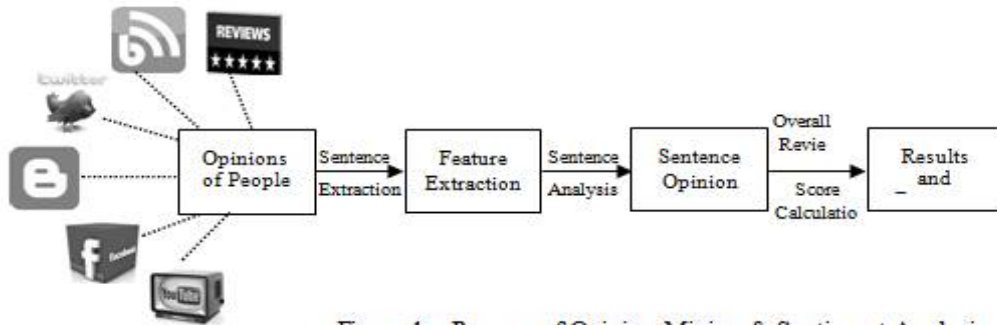


Figure 1. Process of Opinion Mining & Sentiment Analysis

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service. Opinion mining can be useful in several ways. It can help marketers evaluate the success of an ad campaign or new product launch, determine which versions of a product or service are popular and identify which demographics like or dislike particular product features. For example, a review on a website might be broadly positive about a digital camera, but be specifically negative about how heavy it is. Being able to identify this kind of information in a systematic way gives the vendor a much clearer picture of public opinion than surveys or focus groups do, because the data is created by the customer.

## V. VARIOUS METHODOLOGIES USED IN OPINION MINING

There are various methods used for opinion mining and sentiment analysis among which following are the important ones:

1) Naïve Bays Classifier. 2) Support Vector Machine (SVM). 3) Multilayer Perceptron. 4) Clustering.

## VI. BASIC COMPONENTS OF AN OPINION

◦ Opinion holder: The person or organization that holds a specific opinion on a particular object.

◦ Object: on which an opinion is expressed

◦ Opinion: a view, attitude, or appraisal on an object from an opinion holder.

Definitions: refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract

### Motivations of Opinion Mining

There is a lot of information to discover in online forum and discussions, news reports, client emails or blogs for

- Market research
- Media monitoring and
- Public opinion research

Opinion mining is a relevant technology to recognize opinions, emotional attitudes about products, services, persons and other topics.

## VII. RESEARCH TOPICS IN OPINION MINING

- Development of linguistic resources for opinion mining
  - Automatically build lexicons of subjective terms
- At the document/sentence level
  - Simple opinion extraction (a holder, an object, an opinion)
  - Subjective / objective classification



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Issue 2, February 2017

- Sentiment classification: positive, negative and neutral
- At the feature level
- Identify and extract commented features
- Group feature synonyms
- Determine the sentiments towards these features
- Comparative opinion mining
- Identify comparative sentences
- Extract comparative relations from these sentences

## VIII.CONCLUSIONS

Opinion mining is an emerging domain of data mining applied to summary the knowledge from large volume of data that may be customer comments, feedback and reviews on any product or topic etc.

Sentiment detection has a broad diversity of applications in information systems, containing classifying reviews, summarizing review and other real time applications. There are possible to be many other applications that is not discussed. It is constructed that sentiment classifiers are severely dependent on fields or topics. From the above work it is visible that neither classification model consistently outperforms the other, various types of face have distinguished distributions.

It is also point that various kinds of features and classification algorithms are united in an efficient way in order to overcome their singular drawbacks and benefit from each other's merits, and at least enhance the sentiment classification performance. This paper discusses about an overview of Opinion Mining and sentiment analysis in detail with the techniques.

## REFERENCES

1. Ion SMEUREANU, Cristian BUCUR, Applying Supervised Opinion Mining Techniques on Online User Reviews, Informatica Economică vol. 16, no. 2/2012.
2. Bo Pang and Lillian Lee, "Opinion Mining and Sentiment Analysis", Foundations and Trends<sub>R</sub> in Information Retrieval Vol. 2, Nos. 1–2 (2008).
3. Abbasi, "Affect intensity analysis of dark web forums," in Proceedings of Intelligence and Security Informatics (ISI), pp. 282–288, 2007.
4. K. Dave, S. Lawrence & D. Pennock. "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews." Proceedings of the 12th International Conference on World Wide Web, pp. 519–528, 2003.
5. B. Liu. "Web Data Mining: Exploring hyperlinks, contents, and usage data," Opinion Mining, Springer, 2007.
6. B. Pang & L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales." Proceedings of the Association for Computational Linguistics (ACL), pp.15124,2005.
7. Nilesh M. Shelke, Shriniwas Deshpande, Vilas Thakre, Survey of Techniques for Opinion Mining, International Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012.
8. Sannella, M. J. . M. Kim and E. Hovy, "Extracting Opinions Holders, and Topics Expressed in Online News Media Text," Proc. ACL/COLING Workshop Sentiment and Subjectivity in Text, 2006.
9. Forman, B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," Proc. 42nd Ann. Meeting on Assoc. for Computational Linguistics.
10. Brown, L. D., Hua, M. Hu and B. Liu, "Mining and Summarizing Customer Reviews, Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 168-177, 2004
11. B. Liu, "Sentiment Analysis and Opinion Mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167, May 2012 .
12. Y. Jo and A.H. Oh, "Aspect and Sentiment Unification Model for Online Review Analysis," Proc. Fourth ACM Int'l Conf. Web Search and Data Mining, pp. 815-824, 2011.
13. G. Qiu, C. Wang, J. Bu, K. Liu, and C. Chen, "Incorporate the Syntactic Knowledge in Opinion Mining in User-Generated Content," Proc. WWW 2008 Workshop NLP Challenges in the Information Explosion Era, 2008.
14. A. Popescu and O. Etzioni, "Extracting Product Features and Opinions from Reviews," Proc. Human Language Technology Conf. and Conf. Empirical Methods in Natural Language Processing, pp. 339- 346, 2005.
15. V. Hatzivassiloglou and J.M. Wiebe, "Effects of Adjective Orientation and Gradability on Sentence Subjectivity," Proc. 18th Conf. Computational Linguistics, pp. 299-305, 2000.
16. Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, and Z. Su, "Hidden Sentiment Association in Chinese Web Opinion Mining," Proc. 17th Int'l Conf. World Wide Web, pp. 959-968, 2008

## BIOGRAPHY

Dr.B.Radha received B.Sc Degree in Chemistry in 2001, MCA Degree in Computer Applications in 2004. She is currently working as a Assistant Professor in the Department of Computer Science at Sree Saraswathi Thyagaraja College. She received her Ph.D from Anna University in the year 2015 .Her research interest includes Resource allocation in Grid Computing and Simulation. She has published two technical papers in National conferences. Ten paper in the International journal and International Conferences she is a life time member of ISTE.She has acted as reviewer in many National and International conferences.