

# Statistical Modelling for Sentiment Classification: A Survey

Mineeta Khanuja

M.Tech Scholar, Department of Computer Science & Engg.  
Bharti College of Engineering & Technology, Durg(CG)

Asst.prof. Leelkanth Dewangan

Asst. Prof, Department of Computer Science & Engg.  
Bharti College of Engineering & Technology, Durg(CG)

**Abstract**— Persons express sentiments as part of day-by-day communiqué. Sentiments can be umpired by a mish mash of indications such as facial emotions, prosodies, shrugs, and actions. Emotions are also enunciated by written texts or from their handwriting style. Linguistic is a prevailing tool to converse and pass on the information. Automatic sentiment classification and examination methods are useful in numerous applications with psychosomatic basis. For instance, it can be magnificently applied to acquire user favorites and comforts from users' personal written text and vocalizations. These methods are frequently considered in the field of the domain of behavior or mood modeling and customer response analysis. In this paper we will discuss different machine learning approach for sentiment classification. We will conclude our survey based on accuracy (TPR and FNR) among different sentiment classification methods.

**Keywords**— TPR,FNR,NLP,IR

## I. INTRODUCTION

Sentiment Analysis (SA) or Mood Classification or Mood Classification is the mathematical modelling approach of person's sentiments, insolences and mood toward an object. The object can signify events, individuals, or issues. It is also a way to express sentiment. NLP (Natural Language Processing) procedures have long been functional to automatically classify the information content in written text. Applications such as topic based text classification, summarization, problem answering systems, and (IR) information retrieval systems naturally emphasis on the information contained in written text.

There are several fields where automatic sentiment classification is worthwhile for example there are several e-commerce websites in which they ask for feedback, product review etc. The user or customer can give diverse thoughts for different aspects of the identical entity like this sentence "the voice feature of this phone is not good, but the battery life is long". These criticisms/review are significant to the business owners as they can take commercial resolutions or decisions agreeing to the scrutiny results of users' sentiments about their products. The reviews sources are primarily review sites. Sentiment analysis (SA) is not only efficient on merchandise reviews but can also be pragmatic on stock markets, news tutelages, or political discussions.

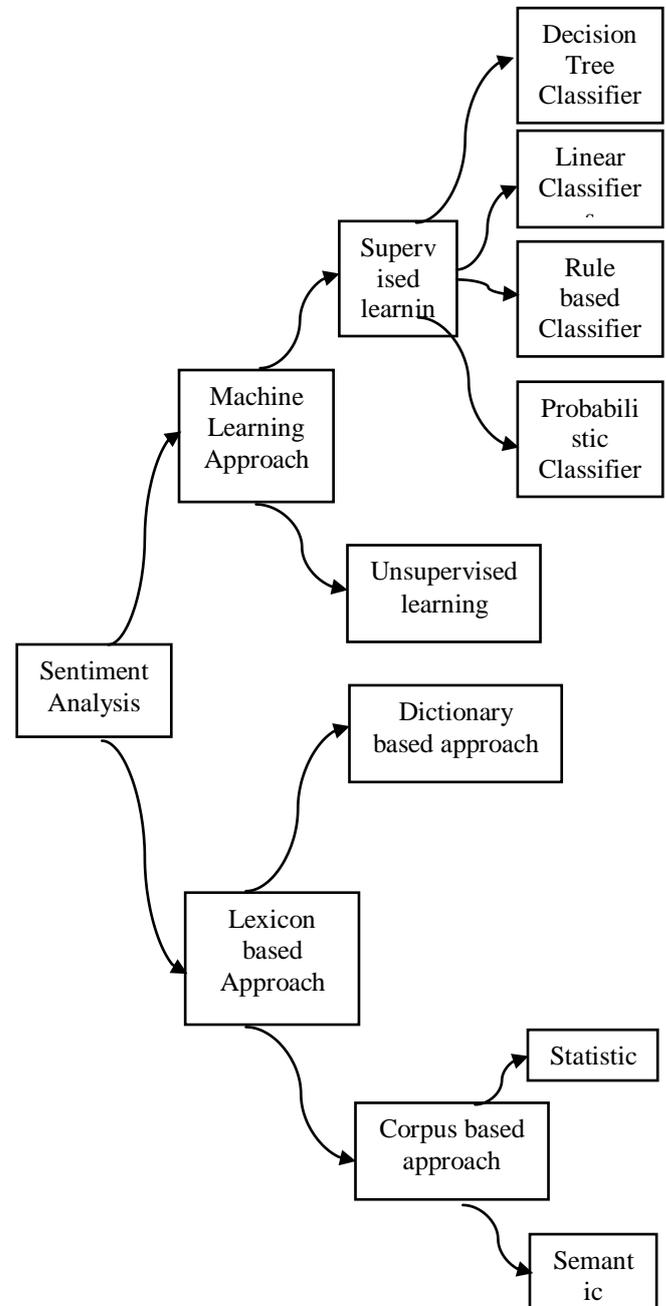


Fig.-1 Sentiment Classification Techniques

In political discussions for example, we could access out commons sentiments on a convinced voting contenders or political parties. The voting outcomes can also be forecast from party-political posts. The micro-blogging and social network sites are painstaking a very good source of information because individuals share and converse their thoughts about a certain topic spontaneously. They are also cast-off as data sources in the SA progression.

Further in next sections of this paper, in section II we will discuss different literatures, in section III we will discuss how we have motivated towards this research, in section IV we will give tabular comparison among different literature based on accuracy at last we will conclude our review.

## II. LITERATUE SURVEY

**Erik Cambria** said that emotions play an important role in successful and effective human-human communication. In fact, in many situations, emotional intelligence is more important than IQ for successful interaction. There is also signif cant evidence that rational learning in humans is dependent on emotions. Affective computing and sentiment analysis, hence, are key for the advancement of AI and all the research f elds that stem from it. Moreover, they f nd applications in various scenarios and companies, large and small, that include the analysis of emotions and sentiments as part of their mission. Sentiment-mining techniques can be exploited for the creation and automated upkeep of review and opinion aggregation websites, in which opinionated text and videos are continuously gathered from the Web and not restricted to just product reviews, but also to wider topics such as political issues and brand perception [IEEE 2016].

**Vivek Narayanan et. al.** explored different methods of improving the accuracy of a Naive Bayes classifier for sentiment analysis. We observed that a combination of methods like effective negation handling, word n-grams and feature selection by mutual information results in a significant improvement in accuracy. This implies that a highly accurate and fast sentiment classifier can be built using a simple Naive Bayes model that has linear training and testing time complexities. We achieved an accuracy of 88.80% on the popular IMDB movie reviews dataset. The proposed method can be generalized to a number of text categorization problems for improving speed and accuracy [Springer 2015].

**David Zimbra et. al.** present an approach to brand-related Twitter sentiment analysis using feature engineering and the Dynamic Architecture for Artificial Neural Networks (DAN2). The approach addresses challenges associated with the unique characteristics of the Twitter language, and the recall of mild sentiment expressions that are of interest to brand management practitioners. Author demonstrate the effectiveness of the approach on a Starbucks brand-related Twitter data set. The feature engineering produced a final tweet feature representation consisting of only seven dimensions, with greater feature density. Two sets of experiments were conducted in three-class and fiveclass tweet sentiment classification. Author compare the proposed

approach to the performances of two stateof-the-art Twitter sentiment analysis systems from the academic and commercial domains. The results indicate that the approach outperforms these state-ofthe-art systems in both three-class and five-class tweet sentiment classification by wide margins, with classification accuracies above 80% and excellent recall of mild sentiment tweets [IEEE 2016].

**Xilun Chen et. al.** presented ADAN, an adversarial deep averaging network for cross-lingual sentiment classification, which, for the first time, applies adversarial training to cross-lingual NLP. ADAN leverages the abundant resources on English to help sentiment analysis on other languages where little or no annotated data exist. We validate our hypothesis by empirical experiments on Chinese and Arabic sentiment classification, where we have labeled English data and only unlabeled data in the target language. Experiments show that ADAN outperforms several baselines including domain adaptation models and a highly competitive MT baseline. We further show that even without any bilingual resources, ADAN trained with random initialized embeddings can still achieve meaningful cross-lingual performance. In addition, we show that in the presence of labeled data in the target language, ADAN can naturally incorporate this additional supervision and yields even more competitive results [arXiv 2017].

**Sufal Das et. al.** concluded that Sentiment analysis is technically very challenging but more promising techniques are available, and it will become increasingly important as more people are buying and expressing their opinions on the web. Summarizing the reviews is not only useful to common shoppers, but also crucial to product manufacturers and has wide applications. Since people are interacting through internet, a huge data is being generated every second. Thus, a distributed parallel computing environment is very much needed to perform sentiment analysis efficiently [IJARCS 2015].

**Basant Agarwal et. al.** emphasis of paper is to discuss the research involved in applying machine learning methods mostly for sentiment classification at document level. Machine learning-based approaches work in the following phases, which are discussed in detail in this chapter for sentiment classification: (1) feature extraction, (2) feature weighting schemes, (3) feature selection, and (4) machine-learning methods [Speinger 2017].

**Walaa Medhat et. al.** said that Sentiment Analysis (SA) is an ongoing field of research in text mining field. SA is the computational treatment of opinions, sentiments and subjectivity of text. This survey paper tackles a comprehensive overview of the last update in this field. Many recently proposed algorithms' enhancements and various SA applications are investigated and presented briefly in this survey. These articles are categorized according to their contributions in the various SA techniques. The related fields to SA (transfer learning, emotion detection, and building resources) that attracted researchers recently are discussed. The main target of this survey is to give nearly full image of SA techniques and the related fields with brief details. The

main contributions of this paper include the sophisticated categorizations of a large number of recent articles and the illustration of the recent trend of research in the sentiment analysis and its related areas [Elsevier 2014].

**I.Hemalatha et. al.** proposed an efficient method for preprocessing. Where it has to be done before applying any classification algorithm. We have performed three preprocessing tasks. One task to remove URLs from the input file next one to remove special characters, here we can also remove repeated letters from a word, the last task is to remove question words. Now the preprocessed document can be given as input to any Machine Learning algorithms. Author focuses on the preprocessing techniques implemented on a specially designed algorithms in order to perform sentiment analysis [IJETTCS 2012].

**Giulio Angiani et. al.** said that Sentiment Analysis has become one of the most interesting topics in AI research due to its promising commercial benefits. An important step in a Sentiment Analysis system for text mining is the preprocessing phase, but it is often underestimated and not extensively covered in literature. In this work, our aim is to highlight the importance of preprocessing techniques and show how they can improve system accuracy. In particular, some different preprocessing methods are presented and the accuracy of each of them is compared with the others.

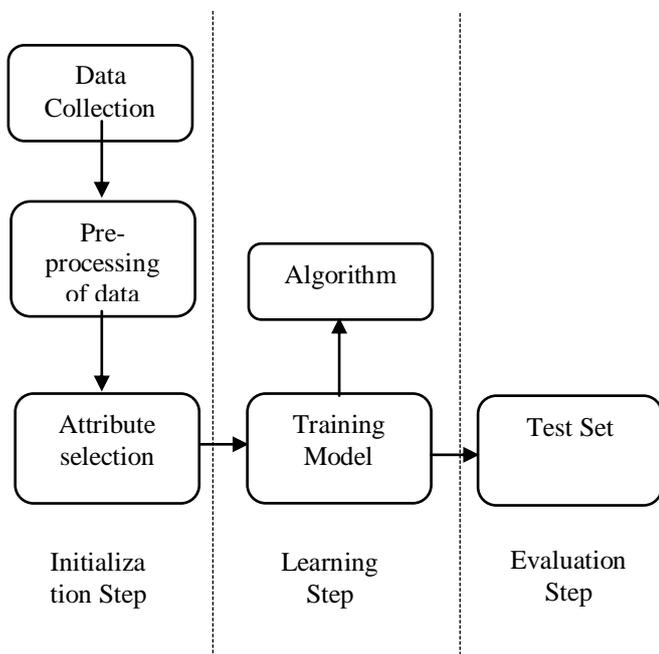


Fig.-2 Steps in SA

**D. Teja Santosh et. al.** said that extraction of appropriate product features using domain independent FOT on LDA topic clusters is carried out successfully. Opinion orientation of the identified features was carried out successfully using SentiWordNet. The LDA parameters “ $\alpha$ ” and “ $\beta$ ” are clearly specified in topic clusters generation process. The obtained features and their opinions help the new customer to read feature specific reviews. In the future, the

traditional LDA model is modified to identify the opinion bearing words and classifies the orientations of their opinions of the extracted features and can compare with the sentiwordnet calculated opinion orientations [MECS 2016].

### III. MOTIVATION

Long range interpersonal communication (Social networking) on the web has developed drastically in the course of the most recent decade. Interpersonal organizations have moved toward becoming widely used and well known mediums for data conveyance as well concerning social cooperation’s. Client exercises on social organizing sites give profitable understanding into singular conduct, encounters, feelings and interests. It is very comparative how a man communicates socially with the human instinct and conduct. In today’s situation, it is required to direct an identity content to precisely quantify clients’ identity; in this manner, it is unreasonable to utilize identity investigation models in numerous web-based social networking areas. A person’s prosperity depends to a great extent on the impression made on others. Accomplishment at work advertise, discovering sentimental accomplices, and picking up help and positive consideration from one’s social foundation intensely rely upon what others consider you.

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. It’s also known as opinion mining, deriving the opinion or attitude of a speaker.

Sentiment = feelings

- Attitudes
- Emotions
- Opinions

There are several application of sentiment analysis which motivated us towards this research.

- Based on a sample of tweets, how are people responding to this ad campaign/product release/news item?
- How have bloggers attitudes about the president changed since the election?
- Identifying child-suitability of videos based on comments.
- Identifying (in) appropriate content for ad placement.
- Use SA to

- A) Search the web for opinions and reviews of this and competing laptops. Blogs, Opinions, amazon, tweets, etc.
- B) Create condensed versions or a digest of consensus points

Messages frequently reflect different parts of the creator’s identity. In this research, we introduce a technique to extricate identity qualities from composed content. The most widely accepted of these traits are as follows:

- Openness (OPE)
- Conscientiousness (CON)
- Extraversion (EXT)
- Agreeableness (AGR)
- Neuroticism (NEU)

IV. COMPARISON

S.No.	Author/Paper title/Year	Description
1.	Erik Cambria/Affective Computing and Sentiment Analysis/IEEE 2016	Erik Cambria said that emotions play an important role in successful and effective human-human communication. Sentiment-mining techniques can be exploited for the creation and automated upkeep of review and opinion aggregation websites, in which opinionated text and videos are continuously gathered from the Web and not restricted to just product reviews, but also to wider topics such as political issues and brand perception.
2.	Basant Agarwal et. al./Machine Learning Approaches for Sentiment Analysis/Speinger 2017	The main emphasis of paper is to discuss the research involved in applying machine learning methods mostly for sentiment classification at document level. Machine learning-based approaches work in the following phases, which are discussed in detail in this chapter for sentiment classification: (1) feature extraction, (2) feature weighting schemes, (3) feature selection, and (4) machine-learning methods.
3.	Vivek Narayanan et. al./Fast and accurate sentiment classification using an enhanced Naive Bayes model/ ARXIV 2014	Author said that accuracy of algorithm can be increased by applying efficient feature selection algorithm. Proposed method can be generalized to a number of text categorization problems for improving speed and accuracy.
4.	David Zimbra et. al./Brand-Related Twitter Sentiment Analysis using Feature Engineering and the Dynamic Architecture for Artificial Neural	Author compare the proposed approach to the performances of two state-of-the-art Twitter sentiment analysis systems from the academic and commercial domains. The

	Networks/IEEE 2016	results indicate that the approach outperforms these state-of-the-art systems in both three-class and five-class tweet sentiment classification by wide margins, with classification accuracies above 80% and excellent recall of mild sentiment tweets.
5.	Vivek Narayanan et. al./Fast and accurate sentiment classification using an enhanced Naive Bayes model/Springer 2016	Naive Bayes model that has linear training and testing time complexities. We achieved an accuracy of 88.80% on the popular IMDB movie reviews dataset. The proposed method can be generalized to a number of text categorization problems for improving speed and accuracy.
6	Daoud Clarke et. al./Developing Robust Models for Favourability Analysis/ACLWEB 2012	Explore the effect of using five different types of features, the robustness of the models when tested on data taken from a later time period, and the effect of balancing the input data by undersampling. Author found varying choices for the optimum classifier, feature set and training strategy depending on the task and dataset.
7.	Sufal Das et. al./Sentiment Analysis for Web-based Big Data: A Survey/IJARCS 2017	Sentiment analysis is a very challenging and promising discipline which uses both intersection of information retrieval and computational linguistic techniques to deal with the reviews expressed in a source material. This work talks about the sentiment analysis process and focus on some machine learning techniques for sentiment classification and future challenges in opinion mining for big data.

V. CONCLUSION

In the context of sentiment analysis, for example, the products and services recommended to a person should be

those that have been positively evaluated by other users with a similar personality type. From comparison section of this paper we saw that accuracy of existing sentiment classification accuracy is not up to mark there is still need of improvement. By applying efficient preprocessing and machine learning classification technique we can improve the accuracy.

#### REFERENCES

- [1] Navonil Majumder et. al. Deep Learning-Based Document Modeling for Personality Detection from Text IEEE 2017.
- [2] Erik Cambria et.al. Affective Computing and Sentiment Analysis IEEE 2016
- [3] Giulio Angiani et. al. A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter Springer 2016
- [4] I.Hemalatha et. al. Preprocessing the Informal Text for efficient Sentiment Analysis IJETTCS 2012
- [5] Ahmed Hassan et. al. Sentiment analysis algorithms and applications: A survey Elsevier 2014
- [6] Daoud Clarke et. al. Developing Robust Models for Favourability Analysis ACLWEB 2012
- [7] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passoneau, R.: Sentiment Analysis of Twitter Data. Computer Science - Columbia University (New York, USA) (2011).
- [8] Balahur, A.: Sentiment Analysis in Social Media Texts. European Commission Joint Research Center (Varese, Italy) (2013).
- [9] Bao, Y., Quan, C., Wang, L., Ren, F.: The role of pre-processing in twitter sentiment analysis. In: International Conference on Intelligent Computing. pp. 615{624. Springer (2014).
- [10] Cambria, E., Olsher, D., Rajagopal, D.: Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In: Twenty-eighth AAAI conference on artificial intelligence (2014) [5] Duncan.
- [11] B., Zhang, Y.: Neural networks for sentiment analysis on twitter. In: Cognitive Informatics & Cognitive Computing (ICCI\* CC), 2015 IEEE 14th International Conference on. pp. 275{278. IEEE (2015)
- [12] Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: Proceedings of LREC. vol. 6, pp. 417{422. Citeseer (2006)
- [13] Fornaciari, P., Mordonini, M., Tomaiuolo, M.: A case-study for sentiment analysis on twitter. In: Proceedings of the 16th Workshop "From Objects to Agents"-WOA (2015).
- [14] E. Cambria, "Affective Computing and Sentiment Analysis," IEEE Intelligent Systems, vol. 31, no. 2, 2016, pp. 102–107.