

## Comparative Analysis of Techniques For Sentiment Classification

Prajakta Gosavi<sup>1</sup>  
Department of Computer Engineering,  
Shree L.R. Tiwari College of Engineering,  
Mira Road, Thane, India  
prajakta777@yahoo.co.in

Vaishali Shirsath<sup>2</sup>  
Department of Information Technology,  
Vidyavardhini College of Engineering and Technology,  
Vasai, India  
vaishalishirsath1@gmail.com

**Abstract** - The growing popularity of E-commerce, social medias, blogs etc. created a new platform where anyone can discuss and exchange his/her views, ideas and experience about any product or services. This trend accumulated a huge amount of user generated data on the web. If this content can be extracted and analyzed properly then it can act as a key factor in decision making. But manual extraction and analysis of this content is an difficult task, as the content is unstructured in nature and it is written in natural language. This situation opened a new area of research called Sentiment Analysis(SA) or Opinion Mining(OM). OM and SA is an extension of Data Mining that extracts and analyzes the unstructured data automatically. The main motive of this paper is to discuss the key concept used in Opinion Mining and Sentiment Analysis and also presents a comparative analysis of various techniques used in this area.

**Index Terms** - Opinion Mining, Sentiment Analysis, Natural Language Processing, Sentiment Lexicon, Sentiment Classification.

\*\*\*\*\*

### 1 INTRODUCTION

Sentiment is anyone's feelings, Attitudes, Emotions, and Opinions. Subjective impressions, not facts. Sentiment can refer to activity of five material senses (hearing, sight, touch, smell, and taste) associating them with or as something considered transcendental: for example Feelings and emotions, News sentiment, automatic detection of opinions embodied in news, Market sentiment, optimism or pessimism in financial and commodity markets.

A fact is something that has really occurred or is actually the case. Scientific facts are verified by repeatable experiments. Fact is sometimes used synonymously with truth, as distinct from opinion. "fact creates norms, and truth illumination".

SA (or OM) is defined as the task of finding the opinions of authors about specific entities. The decision-making process of people is affected by the opinions formed by thought leaders and ordinary people. When a anyone wants to buy a product online he or she will typically start by searching for reviews and opinions written by other people on the various offerings[1]. Sentiment analysis is one of the hottest research areas in computer science. Sentiment analysis is the procedure by which information is extracted from the opinions, appraisals and emotions of people in regards to entities, events and their

attributes. Sentiment analysis has now become the dominant approach used for extracting sentiments from online sources.

There is a huge explosion today of 'sentiments' available from social media including Twitter, Facebook, blogs, and user forums. These reviews of text are a gold mine for companies and individuals that want to monitor their reputation and get timely feedback about their products and actions.

Sentiment classification is a way to analyze the subjective information in the text and then mine the opinion. In decision making, the opinions of others have a significant effect on customers ease, making choices with regards to online shopping, choosing events, products, entities. The approaches of text sentiment classification typically work at a particular level like phrase, sentence or document level. There are different approaches to classify the sentiment which includes the machine learning methods. These machine learning methods include Naive Bayes, Maximum Entropy classification, and Support Vector Machines (SVM). Sentiment classification is a new field of Natural Language Processing that classifies subjectivity text into positive or negative.

Subjectivity analysis focuses on dividing text units into two categories: objective and subjective, whereas sentiment analysis attempts to divide the text units into three categories; negative, positive and neutral. With the advent of time and a need for better understanding

and extraction, it is slowly increased towards sentiment classification.

## 2 RELATED WORK

There are number of articles presented every year in the Sentiment analysis fields. These articles is increasing through years.[2][9][11][12][13] we have to study a product aspect ranking framework, which automatically identifies the important aspects of products from online consumer reviews, aiming at improving the usability of the numerous reviews.

In paper[3], ProBase , the largest existing taxonomy of common knowledge, is blended with ConceptNet , a natural-language-based semantic network of commonsense knowledge, and multi-dimensional scaling (MDS) is applied on the resulting knowledge base for sentiment analysis. The main aim of this paper is to build possibly the most comprehensive resource of common and commonsense knowledge and apply MDS on it, in order to perform a domain-independent concept-level analysis of opinion and sentiments on the Web.

In [5] it is expected that by fusing text-based sentiment classification with audio and video features. Thus, building on, on this paper they now going to introduce multi-modal sentiment analysis in on-line review videos, which can be immediately applied in multimedia retrieval and tagging of large on-line video archives. In this they have used ICT-MMMO: Multi-Modal Movie Opinion Database and introduce a real-life collection of review videos obtained from the YouTube and Expo TV platforms containing movie review videos by non-professional users.

[6] this paper, they explore a variation of the inference rating problem[4], assigning a text segment to one of several categories, each related to a different affective intensity. In contrast to previous research, researchers do not explore one dimension of affective expression, such as the number of stars of a review or the level of positiveness or negativeness in a document, but they have studied the two affective dimensions of valence and arousal, which have been shown to be key aspects of emotion, overall exploring the capability to accurately capture the affective state of online, socially active users.

In this[8] they refer to the problem of discovering and mining connections between social emotions and online documents as social affective text mining, including predicting emotions from online documents, associating emotions with latent topics, and so on. Straight forward method is to manually build a

dictionary of affective terms for each emotion, e.g., SentiWordNet and WordNetAffect .

## 3 SENTIMENT CLASSIFICATION

Sentiment classification is usually formulated as a two-class classification problem, positive and negative. Training and testing data used are normally product reviews. Since online reviews have rating scores assigned by their reviewers, e.g., 1-5 stars, the positive and negative classes are determined using the ratings. For example, a review with 4 or 5 stars is considered a positive review, and a review with 1 to 2 stars is considered a negative review. Sentiment classification is essentially a text classification problem. Traditional text classification mainly classifies documents of different topics, e.g., politics, sciences, and sports. In such classifications, topic related words are the key features[10].

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. Advanced, "beyond polarity" sentiment classification looks, for instance, at emotional states such as "angry", "sad", and "happy".

### 3.1 Type of Sentiment Classification

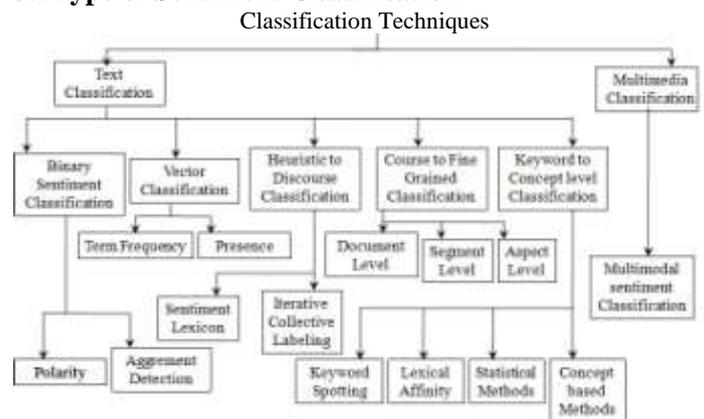


Figure 1. Sentiment Classification

#### 3.1.1 Text Classification

Text Classification is the task to classify documents into predefined classes. Text Classification is also called Text Categorization, Document Classification, Document Categorization. Two approaches manual classification and automatic classification. So based on literature survey we decide whether in the particular research paper they have used text classification or not.

#### 3.1.2 Multimedia Classification

Some years ago the content of the World Wide Web consisted mainly of text. Since then, also due to the development of the Web 2.0, it became populated

more and more with the other media types: images, audio files and videos. Multimedia classification is task to classify images, audio and videos into predefined classes. So based on literature survey we decide whether in the particular research paper they have used text classification or not. Multimedia retrieval and classification has received a lot of interest in recent years.

### 3.2 Sentiment Classification task

#### 3.2.1 Polarity Classification

The basic task of opinion mining is polarity classification. Polarity classification occurs when a piece of text stating an opinion on a single issue is classified as one of two opposing sentiments. Reviews such as “thumbs up” versus “thumbs down,” or “like” versus “dislike” are examples of polarity classification. Polarity classifications also identify pro and con expressions in online reviews and help make the product evaluations more credible.

#### 3.2.2 Agreement Detection

Agreement detection is another form of binary sentiment classification. Agreement detection determines whether a pair of text documents should receive the same or different sentiment-related labels. After the system identifies the polarity classification, it might assign degrees of positivity to the polarity that is, it might locate the opinion on a continuum between positive and negative. Also, it can classify multimedia resources according to mood and emotional content for purposes such as affective human-machine interaction, troll filtering, and cyber-issue detection.

### 3.3 Sentiment classification using Text features

#### 3.3.1 Term Presence

Presence is a binary-valued feature vector in which the entries indicate only whether a term occurs (value 1) or doesn't (value 0). Presence forms a more effective basis to review polarity classification and reveals an interesting difference.

#### 3.3.2 Term Frequency

Term frequency means how many times that term occurred in the given data set. The amount of times something happens within a certain period of time.

#### 3.3.3 Term Position

Term Position refers to how a token's position in a text unit might affect the text's sentiment.

#### 3.3.4 N-gram

The  $n$ -grams typically are collected from a text or speech corpus. When the items are words,  $n$ -grams may also be called shingles

An  $n$ -gram of size 1 is referred to as a "unigram"; size 2 is a "bigram" (or, less commonly, a "digram"); size 3 is a "trigram". Larger sizes are sometimes referred to by the value of  $n$ , e.g., "four-gram", "five-gram", and so on.

#### 3.3.5 Polarity label to terms

In this method polarity label that is positive, negative or neutral are given to the document or the data set given to us. Or we will assign the scale of positivity or scale of negativity to the given data.

#### 3.3.6 Part of Speech (POS)

General textual analysis uses part of speech (POS) information (for example, nouns, adjectives, adverbs, and verbs) as a basic form of word-sense disambiguation. Certain adjectives are good indicators of sentiment and guide feature selection to classify the sentiment. Also, selected phrases chosen by pre-specified POS patterns, usually including an adjective or adverb, help detect sentiments.

#### 3.3.7 Lexicon

Lexicons are the dictionaries in which the positive and negative polarities are already defined. Affective lexicons, in which lemmas are annotated with affective semantics such as Affective Norms for English Words (ANEW), provide validated ratings of valence and arousal of specific words.

### 3.4 Sentiment Classification Techniques

#### 3.4.1 Supervised Technique

This type of technique to be more accurate because the classifier is trained on a set of representative data called a corpus. sentiment classification also known as polarity classification or opinion mining attempt to address this problem by automatically classifying a piece of text as expressing positive or negative sentiment. Supervised sentiment classification the classifier is used such as Naïve bayes, Maximum entropy. And statistical methods are also used.

#### 3.4.2 Unsupervised Technique

In contrast, semantic orientation approach to sentiment classification is unsupervised technique because it does not require prior training in order to mine the data. Instead it uses the direction of a word's semantic orientation and the strength of semantic orientation to determine review sentiment. Lexicon i. e Dictionaries such as SenticWordNet is used in unsupervised sentiment classification.

### 3.5 Sentiment Classification Levels

#### 3.5.1 Document-Level

Opinions and sentiments that occur only at the document level they are called as document level classification. We classify the document on either positive, negative sentiment. Opinions and sentiments don't occur only at the document level, nor are they limited to a single valence or target. One document might contain positive and negative opinions toward one or more topics. So the next level classification has take place[4].

The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment. For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product. This task is commonly Sentiment Analysis and Opinion Mining known as document-level sentiment classification. This level of analysis assumes that each document expresses opinions on a single entity (e.g., a single product). Thus, it is not applicable to documents which evaluate or compare multiple entities.

#### 3.5.2 Segment Level

One document might contain positive and negative opinions toward one or more topics. Hence, later work adopted a segment-level opinion analysis that used graph-based techniques to distinguish sentimental from unsentimental sections.

#### 3.5.3 Phrase-Level

In this the sentiment classification is at the phrase level so we will get the better result compare to above to methods. classified items based on fixed, syntactic phrases used for expressing opinions.

#### 3.5.4 Sentence-Level

Here the sentiment classification is at the sentence level reduced text-analysis granularity to the sentence level by using the presence of opinion-bearing lexical items to detect subjective sentences. Sentence level that is given a document in that we will classify the sentiment at sentence level

#### 3.5.5 Aspect/Feature-Level

Even sentence-level approaches often fail to discover sentiments about an entity and/or its aspects. In an aspect-level approach, wherein an opinion consists of targets and the sentiments associated with them. For example, the sentence "the new iPhone 5's screen size is amazing, but its battery life is short" evaluates two aspects (opinion targets): the screen size and battery life of the same entity[7]. The sentiment about the iPhone 5's screen size is positive, but the sentiment about its battery life is negative. Based on this level of

analysis, we can produce a structured opinion summary about an entity and its aspects, and can draw more accurate statistics about those aspects.

### 3.6 Sentiment Classification Approaches

#### 3.6.1 Keyword Spotting

Keyword spotting is the most naive approach and probably also the most popular because of its accessibility and economy. Text is classified into affect categories based on the presence of fairly unambiguous affect words like 'happy', 'sad', 'afraid', and 'bored'.

#### 3.6.2 Lexical Affinity

Lexical affinity is slightly more sophisticated than keyword spotting as, rather than simply detecting obvious affect words, it assigns arbitrary words a probabilistic 'affinity' for a particular emotion. For example, 'accident' might be assigned a 75% probability of being indicating a negative affect, as in 'car accident' or 'hurt by accident'. These probabilities are usually trained from linguistic corpora.

#### 3.6.3 Statistical Method

Statistical methods, such as Bayesian inference and support vector machines, have been popular for affect classification of texts. By feeding a machine learning algorithm a large training corpus of affectively annotated texts, it is possible for the system to not only learn the affective valence of affect keywords (as in the keyword spotting approach), but also to take into account the valence of other arbitrary keywords (like exical affinity), punctuation, and word co-occurrence frequencies.

#### 3.6.4 Concept Based Method

Concept-based approaches to sentiment analysis focus on a semantic analysis of text through the use of web ontologies or semantic networks, which allow the aggregation of conceptual and affective information associated with natural language opinions. relying on large semantic knowledge bases, such approaches step away from blind use of keywords and word co-occurrence count, but rather rely on the implicit features Associated with natural language concepts. Unlike purely syntactical techniques, concept based approaches are able to detect also sentiments that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey any emotion, but which are implicitly linked to other concepts that do so.

#### 3.7 Language Dependency

In this we are checking that whether the studied research paper is Language dependent or Language

Independent .By knowing whether the particular sentiment classification technique can be used in other language means if we change the prior language then there would be no change in experimental result.

### 3.8 Domain dependency

In this we are checking that whether the studied research paper is Domain Dependent or Domain Independent.By knowing whether a particular sentiment classification technique can be used in other domain for eg .if you performing sentiment classification on product then if we change the domain from product to other that is hotel then also it has to work[11].

## 4 COMPARATIVE ANALYSIS

Research paper/ Parameters	Type of Classification	Technique	Coarse to Fine Grained	Approach	Domain Dependency	Language Dependency
Product Aspect Ranking & its Application [2]	Text	Supervised Learning	Feature Based	Statistical	Domain Independent	English Language Dependent
Semantic MDS for open domain Sentiment Analysis [3]	Text	Supervised And Unsupervised technique	Document-Level	Concept-based	Domain Independent	English Language Dependent
YouTube Movie Reviews: Audio-Visual Context [5]	Multimodal	Supervised Technique	N/A	Statistical Method	Cross Domain	English Language Dependent
Seeing Stars of valence & Arousal in Blog Posts [6]	Text	Supervised Technique	Document-Level	Keyword Spotting	Domain Dependent	English Language Dependent
Predicting Emotional Responses For long Term Informal Text [7]	Text	Unsupervised technique	Sentiment Level	Keyword Spotting	Domain Dependent	English Language Dependent
Mining Social Emotion From Affective Text [8]	Text	Supervised Technique	Document-Level	Statistical Method / Keyword Spotting	Domain Dependent	Chinese Language Independent

Table 1. Comparative Analysis

## 5 CONCLUSION

This survey paper presented overview of different sentiment classification techniques.After studying

various papers we can say that most of the sentiment classification techniques are Domain Dependent, Language dependent, Text based and have Document level specifications.Sentiment analysis used for various real world applications.

## REFERENCES

- [1] Pang, Bo, and Lillian Lee. "Opinion mining and sentimentanalysis." Foundations and Trends® in Information Retrieval 2.1–2 (2008): 1-135.
- [2] Zha, Zheng-Jun, et al. "Product aspect ranking and its applications." IEEE Transactions on Knowledge and Data Engineering 26.5 (2014): 1211-1224.
- [3] Cambria, Erik, et al. "Semantic multidimensional scaling for open-domain sentiment analysis." IEEE Intelligent Systems 29.2 (2014): 44-51.
- [4] Cambria, Erik, et al. "New avenues in opinion mining and sentiment analysis." IEEE Intelligent Systems 28.2 (2013): 15-21.
- [5] Wöllmer, Martin, et al. "Youtube movie reviews: Sentiment analysis in an audio-visual context." IEEE Intelligent Systems 28.3 (2013): 46-53.
- [6] Paltoglou, Georgios, and Michael Thelwall. "Seeing stars of valence and arousal in blog posts." IEEE Transactions on Affective Computing 4.1 (2013): 116-123.
- [7] Paltoglou, Georgios, et al. "Predicting emotional responses to long informal text." IEEE Transactions on Affective Computing 4.1 (2013): 106-115.
- [8] Bao, Shenghua, et al. "Mining social emotions from affective text." IEEE transactions on knowledge and data engineering 24.9 (2012): 1658-1670.
- [9] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002.
- [10] Read, Jonathon. "Using emoticons to reduce dependency in machine learning techniques for sentiment classification." Proceedings of the ACL student research workshop. Association for Computational Linguistics, 2005.
- [11] Glorot, Xavier, Antoine Bordes, and Yoshua Bengio. "Domain adaptation for large-scale sentiment classification: A deep learning approach." Proceedings of the 28th international conference on machine learning (ICML-11). 2011.
- [12] Bhonde, Swati B., and Jayashree R. Prasad. "Sentiment Analysis-Methods,Applications & Challenges." International Journal of Electronics Communication and Computer Engineering 6.6 (2015): 634.
- [13] Wang, Sida, and Christopher D. Manning. "Baselines and bigrams: Simple, good sentiment and topic classification." Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. Association for Computational Linguistics, 2012.
- [14] Fahrni, Angela, and Manfred Klenner. "Old wine or warm beer:Target-specific sentiment analysis of adjectives." Proc. of the Symposium on Affective Language in Human and Machine, AISB. 2008.