# Sentiment Analysis Approaches for Social Media Monitoring

Bisma Shah
Hamdard University , Department of Computer Science New Delhi -62
Shahbisma2009@gmail.com

Farheen Siddiqui
Hamdard university Department of Computer Science New Delhi -62

*Abstract:* **Social media has become an indispensable part of social life. It influences the beliefs, values, and attitudes of people, as well as their intentions and behaviours. Social media acts not only as a platform for personal communication, but also, as a means for communicating opinions about products and services or even political and general events among its users. Due to its widespread and popularity, a huge amount of user reviews or opinions are being generated and shared daily on social media. Therefore, converting social media content into information, key concepts, and themes is crucial for generating knowledge and formulating strategies to support operational, managerial, and strategic decision making. Opinion mining and Sentiment analysis are the formalization for studying and constructing opinions and sentiments. The current paper aims at studying and providing the comparison of different methods of sentiment analysis used for extracting the polarity (positive, negative or neutral) of social media dataset.**

**Keywords**: Sentiment analysis, Naive Bayes Classifier, Social media, Lexical knowledge base, Decision tree.

## 1. Introduction

With the explosion of internet there is an abundance of data available on-line, they can be numerical or text file and they can be structured, semi-structured or non-structured. The internet is replete with reviews, comments and ratings by virtue of the developing prominence of sites like Amazon.com and Epinion.com, where individuals can express their sentiment on various items and rate them. Many researchers and practitioners aim attention at the approaches and techniques to extract useful information from these data lately. Advancement in computer technology along with many retrieval techniques and tools have been proposed according to different data types. In addition to data and text mining, there has seen a growing interest in non-topical text analysis in recent years. **Sentiment Analysis** is one of them. Sentiment analysis, also known as **Opinion Mining** helps in identifying and extracting subjective information in source materials and categorizes them as **positive**, **neutral**, or **negative**.

In recent years, more attention has been paid to the problem of sentiment classification [1]. Using appropriate mechanisms and techniques, the vast amount of data generated online can be processed into information to support **operational**, **managerial**, and **strategic** decision making [2].Sentiment analysis aims to identify and extract opinions and attitudes from a given piece of text towards a specific subject [3].

Social media measurement or 'social media monitoring' is concerned with the active monitoring of social media channels for extraction of useful information about a company or organization, usually tracking of various social media content as a way to determine the volume and sentiment of online conversation about a brand or topic. Methodologies which can speedup processing and reduce latency are required for real time analysis of social media application. The task of mining user opinions from social media data is not straight forward; it can be accomplished in different ways.

Sentiment analysis provides new techniques of classification. The objective could be a basic **polarity classification (positive or negative)**, or **a multi-class one (like the 5 star classification)**. Sentiment analysis can be done on a **global topic level** or on a **more specific level**. Global topic level analysis provides general opinion on a particular product being taken into account and a specific level analysis fetches opinions based on the product's aspects.

The sentiment found inside remarks, feedback or critiques give helpful pointers to a wide range of purposes. These sentiments can be categorised either into **two** categories: **positive** and **negative**; or into an **n-point scale**, e.g., very good, good, satisfactory, bad, very bad. In this respect, a sentiment analysis task can be interpreted as a classification task where each category represents a sentiment. Opinions obtained from sentiment analysis enables companies to estimate the extent of product acceptance, judge the success of a new product launched and to determine strategies to improve product quality. It also facilitates policy makers or politicians to analyse public sentiments with respect to policies, public services or political issues.

Figure 1 shows the process of analyzing sentiments in case of product reviews.

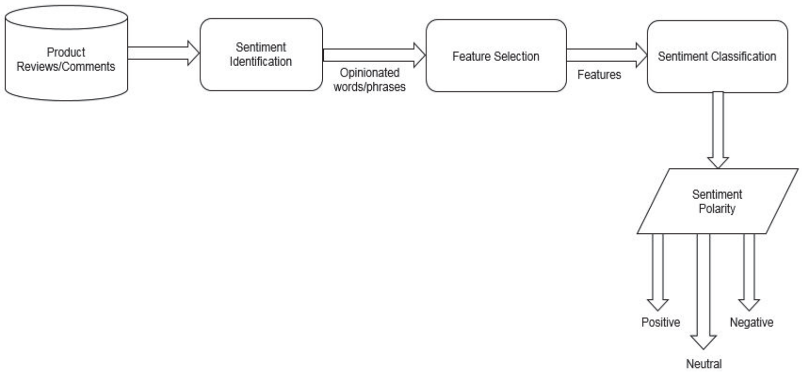


Figure 1. Analysis of sentiments in product reviews.

## 2. Different Classification Approaches for Sentiment Analysis

A large number of approaches for opinion mining tasks were used in history. A few of them are listed here. The data of review from automobiles, bank, movies, and travel destinations was used by Janice M.Wiebi [4]. Classification of words was based on two classes (positive or negative) and overall positive or negative score for the text was use to obtain the count. A document is acknowledges as positive document on the off chance that the reports contain more positive than negative terms, else it is negative. Document and sentence level classification is used here for classification. Likes and dislikes of the opinion holder on each feature cannot be found by these classifications, although they are useful and improve the effectiveness of a sentiment classification.

Pang et al [1], Mukras R.J [5] used several **statistical** feature selection methods and directly applied machine learning techniques on the data of movie review, customer feedback review and product review. They concluded from their work that in sentiment classification not only does machine learning algorithm perform well. They demonstrated that the presence or absence of a word is by all accounts more characteristic of the content instead of the frequency of a word.

Morie Rimon [6] classified the sentiments using the approach based on keywords. Under this approach, sentiment indicators used are terms, mainly adjectives (e.g. awesome, awful). The list of indicators can be prepared manually, composed semi automatically using sources such as WordNet or acquired by machine learning algorithms that infer the best indicators from tagged samples in the domain of interest.

Different machine learning classifiers and feature extractors were used by Alec co [7] to classify sentiments. Naive Bayes, Maximum Entropy and Support Vector Machines (SVM) are some of the machine learning classifiers. Part of speech (POS) tags are used with unigrams, bigrams, unigrams and bigrams, and unigrams as feature extractors.

Machine learning and semantic orientation approaches were combined into one frame of reference by Yan Dang, Yulei Zhang and Hsinchun Chen [8] by proposing a lexicon enhanced method for sentiment classification. In particular, they used additional dimension of features for the machine learning classifiers as the words with semantic orientations.

Changli Zhang [9], used Decision learning method for sentiment classification and performed the analysis on customer feedback review and product review data. In Decision tree learning, approximations of discrete valued target functions are done and the learned function is represented by a decision tree. To improve human readability, sets of if-then rules can be used to represent learned trees.

Sub-trees of word dependency trees as features for sentence wise sentiment polarity classification were used by Kudo et al. [10]. Boosting algorithm with the sub tree-based decision stamps as weak learners was utilized.

**Sentiment Analysis** refers to the process of identifying subjective information in source materials [3, 11].

Typical tasks of Sentiment Analysis include:

1) Classifying textual documents into positive and negative polarity categories, [12, 13].
2) Identifying textual topics and their associated opinions [14, 15].
3) Opinion summarization (Hu and Liu, 2004; Ku et al., 2006) [16, 17].

The paper by Hatzivassiloglou and McKeown (1997) [18] describes a basic sentiment classification framework which inspired a number of later works on this topic.

We have classified the various Sentiment Analysis approaches based on the exhaustive study of previous work done. Figure 2 shows this Classification for Sentiment Analysis approaches.

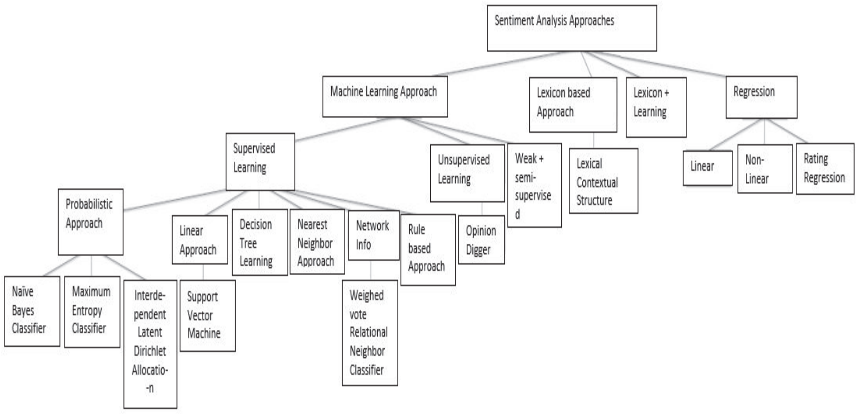


Figure 2. Sentiment classification approaches.

### 2.1 Machine Learning Approach

To solve the Semantic Analysis as a regular text classification problem, we can use machine learning approach that relies on the famous machine learning algorithms and make use of syntactic and/or linguistic features. An example of machine learning approach is given below:

**Blending Feature Mining and Sentiment Analysis for Product Review Rating**

This solution was introduced in 2011 by de Albornoz et al. [19]. In this method three categories of product review rating are produced globally by measuring the polarity and strength of the expressed opinion. This is a Document-level sentiment classification method based on machine learning. Because this method tries to identify the strength of the opinion as well as the relevance of the feature the opinion is about, it is chosen as a representative of the global rating solutions. The **mechanism** of Automatic Product Review Rating [19] is straightforward and is summarized as:

(i) Detecting salient product features, that is, the important features are identified;

(ii) Extracting the user opinion on each product feature from the sentences;

(iii) Quantifying the user opinions, that is, finding polarity and strength of opinion;

(iv) Predicting the rating of a review, that is, a global score is computed.

**Advantages**

Both the strength of an opinion as well as the relevance of the feature the opinion is about are considered.

**Disadvantages**

- Dependent on WordNet database.
- Is specific to reviews written in English only.

Further, we can categorize machine learning approaches as follows:

#### 2.1.1 Supervised Learning

The supervised learning methods make use of labelled training data for classifying data. A large number of supervised classifiers are available in literature, some of the most frequently used classifiers in Semantic Analysis are presented below in brief detail:

##### 2.1.1.1 Probabilistic Approach

Probabilistic classifiers, for the purpose of classification, employ mixture models. As per the assumption of mixture model, each class accounts for a separate component of the mixture. Each mixture component provides the probability of sampling a particular term for that component and hence, these kinds of classifiers are also called **generative classifiers**, since each component here is a generative model.

##### 2.1.1.1.1 Naive Bayes Classifier

The Naive Bayes classifier (NBC) is an extremely simple classifier that relies on Bayesian probability and the assumption that feature probabilities are independent of one another (Vachaspati, P and Wu, C., 2012) [20]. The posterior probability of a class is determined in Naive Bayes classification model, given the distribution of the words in the document. The probability of a given feature set belonging to a particular label is predicted by using Bayes Theorem. Finding the probabilities of categories given a text document by using the joint probabilities of words and categories is the basic idea of this approach.

As per Bayes theorem, for a given data point x and class C, the conditional probability is

$$P(C / x) = \frac{P\,(x/C).P(C)}{P(x)} \qquad (1)$$

Now assuming that the probability of each of the attributes of a data point x = {x1, x2… xj}, occurring in a given class is independent, we can estimate the probability of x as follows:

$$P(C/x) = P(C). \prod P(x_j / \mathrm{C}) \qquad (2)$$

Thus, the conditional probabilities of each attributes occurring on the predicted classes are required to be estimated from the training data set to train a Naive Bayes classifier.

**Advantages**

- Model is easy to interpret.
- Efficient computation.

**Disadvantages**

Assumptions of attributes being independent, which may not be necessarily valid.

**2.1.1.1.2 Maximum Entropy Classifier (ME)**

The Max Entropy classifier is a probabilistic classifier that belongs to the class of exponential models. Unlike the Naive Bayes classifier, the Max Entropy does not assume that the features are conditionally independent of each other. The MaxEnt is based on the Principle of Maximum Entropy. The MaxEnt classifier selects from all the models that fit the training data, the one which has the largest entropy. Text classification problems such as language detection, topic classification, sentiment analysis and more can be solve by Max Entropy classifier.

The MaxEnt Classifier (known as a conditional exponential classifier) encodes labelled feature sets to vectors. This encoded vector calculates the weights for each feature and the most likely label for a feature set is determined using these weights. The parameter of this classifier is a set of X {weights}, which joins the common features generated from a feature-set by an X {encoding}. In particular, the encoding maps each C {(featureset, label)} pair to a vector. The probability of each label is then computed using the following equation:

$$P(fs \mid label) = \frac{dotprod(weights, encode(fs, label))}{sum(dotprod(weights, encode(fs, l)) for l in labels)} \qquad (3)$$

Gibbs, loglinear, exponential and multinomial logic models are other names for Maximum entropy (ME) model [21].

**Advantages**

- Maximum Entropy classifier is used when the prior distributions are unknown.
- Maximum Entropy classifier is used when the conditional independence of the features cannot be assumed.

**Disadvantages**

More time is required to train Max Entropy classifier compared to Naive Bayes. This is because of the optimization problem that needs to be solved in order to estimate the parameters of the model.

**2.1.1.1.3 Interdependent Latent Dirichlet Allocation (ILDA)**

Interdependent Latent Dirichlet Allocation (ILDA), introduced in 2011 by Moghaddam and Ester [22] is based on the probabilistic assumption that an aspect and its corresponding rating is interdependent. ILDA is a probabilistic graphical model in which each review is represented as a blend of latent aspects and ratings. ILDA is performed at document-level. It assumes that aspects and their ratings are represented by multinomial distributions and it groups head terms into aspects and sentiments into ratings. ILDA relies on a concept introduced in 2003 by Blei et al.: Latent Dirichlet Allocation (LDA) - a generative probabilistic model for collections of discrete data such as text corpora [23], where each item of a collection is displayed as a finite mixture over a fundamental set of latent variables.

**Advantages**

- Assumes that a sentiment and its rating are mutually dependent.
- Overcomes a notable issue of different strategies where a positive assumption communicated with a word having negative meaning is misconstrued (e.g., low price).

**Disadvantages**

The relation of extracted clusters with the actual aspects or ratings is not specific (general drawback for unsupervised models).

### 2.1.1.2 Linear Approach

The output of the linear classifier is defined as

$$p = A'.X' + b \qquad (4)$$

Where **X'** $\{x_1. \; . \; . \; xn\}$ is the normalized document word frequency, vector **A'** $\{a_1. \; . \; . \; an\}$ is a vector of linear coefficients with the same dimensionality as the feature space, and b is a scalar; the predictor p is a separating hyperplane between different classes. Support Vector Machines (SVM) is one of the kinds of linear classifiers [24], [25] which attempts to determine good linear separators between different classes.

**Support Vector Machines (SVM) Classifiers**

Support Vector Machine (SVM) is used for performing two-class data classification of linear data. Given a set of input vectors, an SVM will attempt to find a separating hyperplane that gives the largest decision boundary between the two classes. This approach has the distinct advantage that small document variations will not cause misclassifications [26].Here the training data is mapped non-linearly to a higher dimension. Then a linear optimal hyperplane separating different classes is searched within the higher dimension using support vectors. Vladimir vapnik, Bernhard Boser and Isabelle Guyan has developed this approach in 1992. For traditional text classification purposes, a support vector machine (SVM) is considered the most efficient classifier. Illustration of dividing hyperplanes for a sample of points belonging to two classes is given in Figure 3.

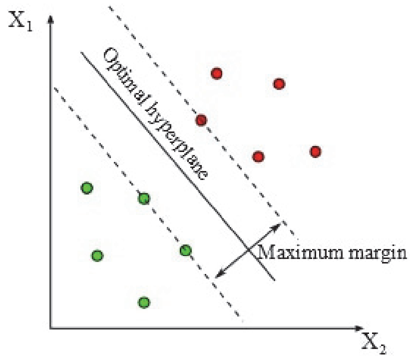


Figure 3. Hyperplane Separating Two Classes in SVM.

Availability of unlimited number of separating hyper planes dividing the two classes in the above illustration can be seen. To pick the most ideal hyperplane, a natural thought is pick the one that has the biggest separation between any points from either class, in this manner making the widest possible margin between points from the distinctive classes. The understanding behind this technique is that a hyper plane with a huge margin would be a "more secure" classification boundary, less inclined to make prediction errors by being near the boundaries of one of the classes.

**Advantages**

- Very good performance on experimental results.
- Low dependency on data set dimensionality.

**Disadvantages**

- Pre-processing is required in case of categorical or missing value.
- Difficult interpretation of resulting model.

### 2.1.1.3 Decision Tree Learning

Decision tree classifier decomposes the training data space based on a condition on the attribute value in a hierarchial manner [27]. A decision tree is learned from the training dataset, and that tree is later used to predict the class attribute value for instances in the test dataset. Construction of decision trees is based on heuristics, as different heuristics generate different decision trees from the same dataset. Decision trees classify examples based on their feature values. Each non-leaf node in a decision tree represents a feature, and each branch represents a value that the feature can take. Instances are classified by following a path that starts at the root node and ends at a leaf by following branches based on instance feature values. The value of the leaf determines the class attribute value predicted for the instance.

When selecting features, we prefer features that partition the set of instances into subsets that are more pure. A pure subset has instances that all have the same class attribute value. When reaching pure subsets under a branch, the decision tree construction process no longer partitions the subset, creates a leaf under the branch, and assigns the class attribute value for subset instances as the leaf's predicted class attribute value. A common measure for determining purity of subsets is entropy. Over a subset of training instances, T, with a binary class attribute (values E {+, -}), the entropy of T [28] is defined as:

$$entropy\,(T) = -p^{+} \log p^{+} - p^{-} \log p^{-} \qquad (5)$$

Where $p^{+}$ the proportion of instances with + class is attribute value in T and $p^{-}$ is the proportion of instances with – class attribute value.

#### 2.1.1.4 Nearest Neighbor Approach

The principle of Nearest Neighbor classifier assumes that, the class of a new yet unseen occurrence, for a given set of instances in a training set, is likely to be that of the majority of its closest "neighbour" instances from the training set. As the name suggests, k-nearest neighbor or kNN uses the k nearest instances, called neighbors, to perform classification. The instance being classified is assigned the label (class attribute value) that the majority of its k neighbors are assigned. Thus the **k-Nearest Neighbour algorithm** classifies a new occurrence by checking for its k closest instances in the data set, and making a prediction [29] based on what classes the majority of the k neighbours belong to. The notion of closeness is formally given by a **distance function** between two points in the attribute space, specified a priori as a parameter to the algorithm. Standard **Euclidean distance** is typically used as the distance function between two points in an n-dimensional space, where n is the number of attributes in the data set.

#### 2.1.1.5 Classification with Network Information

Consider a friendship network on social media and a product being marketed to this network. The product seller wants to know who the potential buyers are for this product. Assume the network of the list of individuals who decided to buy or not buy the product is given to us. Our goal is to predict the decision for the undecided individuals. This problem can be formulated as a classification problem based on features gathered from individuals. However, in this case, we have additional friendship information that may be helpful in building more accurate classification models. This is an example of classification with network information.

Assume we are not given any profile information, but only connections and class labels (i.e., the individual bought/will not buy). By using the rows of the adjacency matrix of the friendship network for each node as features and the decision (e.g., buy/not buy) as a class label, we can predict the label for any unlabeled node using its connections. Let $P(y_i = 1 \mid N(v_i))$ denote the probability of node $v_i$ having class attribute value 1 given its neighbors. Individuals' decisions are often highly influenced by their immediate neighbors. Thus, we can approximate $P(y_i = 1)$ using the neighbors of the individual [29] by assuming that

$$P(y_i = 1) \approx P(y_i = 1 \mid N(v_i)) \qquad (6)$$

We can estimate $P(y_i = 1 \mid N(v_i))$ via different approaches. The **weighted-vote relational-neighbor (wvRN)** classifier is one such approach.

**Weighted-Vote Relational-Neighbor Classifier (wvRN)**

The wvRN is a neighbor-based classifier proposed by Mackassy and Provost (Macskassy and Provost) in the year 2003, which estimates class-membership probability of each node as the weighted mean of the class-membership probabilities of its neighbors [30]. It estimates $P(y_i = 1 \mid N(v_i))$ as

$$P\big( y_i = 1 \mid N(v_i) \big) = \frac{1}{\mid N(v_i))\mid} \sum_{v_j \epsilon N(v_i)} P\Big( y_i = 1 \mid N(v_j) \Big) \qquad (7)$$

In other words, the probability of node $v_i$ having class attribute value 1 is the average probability of its neighbors having this class attribute value. Note that only for $v_i's$ that are unlabelled, $P\big( y_i = 1 \mid N(v_i) \big)$ is calculated. $P\big( y_k = 1 \mid N(v_k) \big) = 1$ for node $v_k$, which is labelled 1, and the probability is never calculated. Likewise$P\big( y_k = 0 \mid N(v_k) \big) = 1$, if $v_k$ will not buy the product. Here the probability of the node is affected if the probabilities of its neighbors change. Thus, the order of updating nodes can affect the estimated probabilities. In practice, one follows an order sequence for estimating node probabilities. Starting with an initial probability estimate for all unlabelled nodes and following this order, we estimate probabilities until probabilities no longer change (i.e., converge).

#### 2.1.1.6 Rule-based Approach

In rule based classifiers, a set of rules are used for modelling data space. On the left hand side, is a condition on the feature set expressed in disjunctive normal form whereas, on the right hand side is the class label. The conditions are on the term presence. Term absence is once in a while utilized on the grounds that it is not instructive in inadequate information. There are numbers of criteria in order to generate rules, the training phase uses these criteria to construct all the rules. Support and confidence are the most two common criteria [31]. Support represents absolute number of instances in the training data set that are relevant to the rule. The Confidence refers to the conditional probability that the right hand side of the rule is satisfied if the left-hand side is satisfied. Some combined rule algorithms were proposed in [32]. Both decision trees and decision rules tend to encode rules on the feature space, but the decision tree tends to achieve this goal with a hierarchical approach. Quinlan [27] has studied the decision tree and decision rule problems within a single framework; as a certain path in the decision tree can be considered a rule for classification of the text instance. The main difference between the decision trees and the decision rules is that DT is a strict hierarchical partitioning of the data space, while rule-based classifiers allow for overlaps in the decision space.

### 2.1.2 Unsupervised Learning

The **unsupervised learning** methods overcome the difficulties faced in supervised learning methods, that is, the difficulty to create the labelled training documents required in supervised learning for classification purposes, by making use of the unlabelled documents (that are easier to collect than labelled documents). Out of the many research works presented in this field, in the method given by Ko and Seo [33], the documents are divided into sentences, and based on the keyword lists of each category and sentence similarity measure, the categorization of each sentence is done. An example of unsupervised learning approach is given as:

**Opinion Digger**

Opinion Digger is a Sentence level Sentiment classification approach. The solution Opinion Digger introduced in 2010 by Moghaddam and Ester [34] is an example of a completely unsupervised machine learning method. It uses a set of known aspects on a product and a ratings guideline (correspondence between ratings and adjectives i.e. 5 means “excellent”, 4 means “good", and so on) as input. Based on these components, a set of other aspects and the ratings of each aspect according to the guideline are determined in output. Opinion digger operates in two steps:-

i. First, the set of aspects are extracted. Then pre-processing is done, after that it uses POS to tag each sentence. Here aspects are assumed as nouns and hence detaches the frequent nouns as potential aspects. On matching the sentences with known aspects, it determines opinion patterns- sequence of POS-tags that expressed opinion on an aspect. Opinion patterns are regarded as the frequent patterns that are used with known aspects. Opinion digger considers the noun as an aspect if reviews with a “potential aspect” noun match no less than two different opinion aspects.

ii. In the second phase, the aspects are given rating. An adjective is associated to the opinion by Opinion Digger for each sentence containing an aspect. The guideline in the WordNet synonymy graph are searched for two synonyms. The weighted average of the corresponding rating in the guideline accounts for the estimated rating of the aspect. Here, the minimum path distance between the opinion adjective and the guideline's adjective in the WordNet hierarchy is inversed for estimating the weight.

**Advantages**

A ratings guideline and seed-aspect is present to determine all aspects.

**Disadvantages**

- Its working is dependent on guideline and known aspects.
- Also dependent on WordNet database.

### 2.1.3 Weakly and Semi-Supervised Learning

The idea of **weak** and **semi-supervised learning** is used as a part of numerous applications. In the strategy given by Youlan and Zhou, a weak supervision is given at the level of features rather than instances. By consolidating the prior information obtained from an existing sentiment lexicon (knowledge base) into sentiment classifier model learning, they developed an initial classifier. Prior information are referred as labelled features and are used to inhibit the model’s predictions on unlabelled instances by utilizing the generalized expectation criteria. In their work, they identified the existence of domain-specific polarity words elucidating the possibility that the polarity of a word might not be the same in different domains.

## 2.2 Lexicon-based Approach

An example of Lexicon based approach is given below:

**Sentiment Classification using Lexical Contextual Sentence Structure**

The arrangement exhibited in the article by Khan et al. [35] is a **domain independent rule based method** used for sentiment classification of customer reviews. In this method, based on the pattern structure of the sentence, the contextual information and the sense of each individual sentence are extracted. The steps of Sentiment analysis using Lexical Contextual Sentence Structure are summarized as:

1. Make a Bag of Sentences (BOS) by splitting reviews into sentences.
2. Noise is removed from sentences using spelling correction, special characters and symbols (photonics) are convert to their text expression, each word of the sentence is tagged using POS tagging and the position of each word in the sentence is stored.
3. Create an exhaustive dictionary (feature vector) of the important feature with its position in the sentence.
4. Subjective and objective classification of sentences is done.
5. Check for the polarity of the subjective sentence as positive, negative or neutral by making use of a lexical dictionary as a Knowledge base.
6. Using the sentence structure and contextual feature of each term in the sentence, check and update polarity.

**Advantages**

- It is independent of the domain (subject) of the review.
- Rule-based method.

**Disadvantages**

Dependent on WordNet database.

**2.3 Concept-level Sentiment Analysis by Joining Leaning and Lexicon-based Approaches**

This approach given by Mudinas et al. [36] is **a concept-level Sentiment analysis system** called **pSenti** in which a blend of lexicon based and learning based approaches is used.

This is a document-level sentiment classification method. Here the overall sentiment of a review is measured and a score is given in the form of positive, negative or neutral or 1-5 stars classification. It operates in four parts:-

i. Each word is tagged and stored by the method Part Of Speech (POS) after noise (idioms and emoticons) are being removed from the review.
ii. Second, a list of top 100 aspect groups and top 100 views is generated.
iii. A “sentiment value" is assigned to any sentiment word using the lexicon-based approach and features for the supervised machine learning algorithm are generated.
iv. Finally, a “feature vector" corresponding to each aspect is obtained. This feature vector can either be the sum of the sentiment value for a particular word or the number of occurrences of this word with respect to other descriptive words.

**Advantages**

- The collaboration of lexicon/learning methods, concept level sentiment recognition and estimation and the lesser sensitivity to changes in subject area are the fundamental advantages.
- The results of sentiment analysis are provided by pSenti in a structured and readable way by dividing the overall sentiment into aspects (e.g., product features) and their corresponding views.

**Disadvantages**

A neutral score is often assigned to the reviews with a lot of noise (irrelevant words for the subject of the review) by virtue of the inability of the method to detect any sentiment for those reviews.

**2.4 Regression Analysis**

In regression, unlike classification, class attribute values are real numbers, for instance, stock market value is predicted using regression. The input to the regression method is a dataset where attributes are represented using $x_1$, $x_2$… $x_m$ (also known as regressors) and class attribute is represented using Y (also known as the dependent variable), where the class attribute is a real number. We want to find the relation between Y and the vector X = ($x_1$, $x_2$… xm) [28].Two basic regression techniques are: **linear regression** and **logistic regression**.

**2.4.1 Linear Regression**

In linear regression, we assume that the class attribute Y has a linear relation with the regressors (feature set) X by considering a linear error E (belongs to). In other words

$$Y = XW + E \qquad (8)$$

Where, W represents the vector of regression coefficients. The problem of regression can be solved by estimating W using the training dataset and its labels Y such that fitting error E is minimized. One of the methods of solving the linear regression problem is least squares or maximum-likelihood estimation.

### 2.4.2 Logistic Regression

Logistic regression provides a probabilistic view of regression. For simplicity, let us assume that the class attribute can only take values of 0 and 1. Formally, logistic regression finds probability p such that

$$P(Y = 1|X) = p \qquad (9)$$

Where X is the vector of features and Y is the class attribute. We can use linear regression to approximate p. In other words, we can assume that probability p depends on X; that is,

$$p = \beta X \qquad (10)$$

Where β is a vector of coefficients. Unfortunately, $\beta X$ can take unbounded values because X can take on any value and there are no constraints on how β's are chosen. However, probability p must be in range [0, 1]. Since $\beta X$ is unbounded, we can perform a transformation $g(.)$ on p such that it also becomes unbounded. Then, we can fit $g(p)$ to$\beta X$. One such transformation $g(.)$ for p is

$$g(p) = \ln \frac{p}{1-p'} \qquad (11)$$

Which for any p between [0, 1] generates a value in range [-1, +1]. $g(.)$ is known as the **logit function**. The transformed p can be approximated using a linear function of feature vector X,

$$g(p) = \beta X \qquad (12)$$

Combining Equations (2) and (3) and solving for p, we get

$$p = \frac{e^{\beta X}}{e^{\beta X}+1} = \frac{1}{e^{-\beta X}+1} \qquad (13)$$

This function is known as the logistic function and is plotted in Figure 4. An interesting property of this function is that, for any real value (negative to positive infinity), it will generate values between 0 and 1. Moreover, it acts as a probability function.

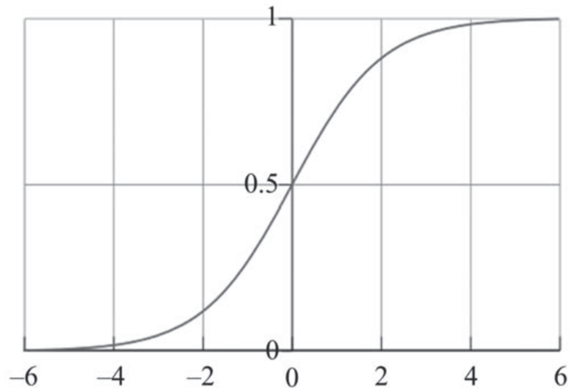


Figure 4. Logistic Function.

We need to find $\beta's$ such that $P(Y|X)$ is maximized. Unlike linear regression models, there is no closed form solution to this problem, and it is usually solved using iterative maximum likelihood methods.

After $\beta's$ are found, similar to the Naive Bayes Classifier (NBC), we compute the probability $P(Y|X)$ using Equation (4). In a situation where the class attribute takes two values, when this probability is larger than 0.5, the class attribute is predicted 1; otherwise, 0 is predicted.

An example of Regression based approach is given as:

**Latent Aspect Rating Analysis (LARA) - a rating regression approach**

This arrangement treats a unique issue called Latent Aspect Rating Analysis with a Model-based method, called **the Latent Rating Regression (LRR)** model [14], created by Wang et al. in 2010. Ratings on different aspects in a review as well as the emphasis of the author on each aspect is estimated here. LRR makes use of a given set of aspects and the overall ratings of the review. This is a Sentence level / reviewers level sentiment classification approach - the latent aspect rating is determined for each review. The task of LARA is [37] to take as input a set of review text documents about an entity with overall ratings and generate as output:

1) Ratings on a set of predefined aspects of the entity, and

2) Relative weights placed by a reviewer on each aspect when writing the review.

So the overall rating is calculated by latent ratings on different aspects determined by the words and not directly by the words used in the review.

**Advantages**

- A new model - The global rating is drawn from latent aspect ratings and not the direct result of the words in the review.
- Latent Aspect Ratings Analysis problem is resolved.

**Disadvantages**

- Performance in simple aspect ratings is poor.
- Prior information of aspects is required.

## 3. Conclusions and Future Scope

From the above classification approaches, we have concluded that a lot of research is present in literature for detecting sentiment from the text. Still, there is a huge scope of improvement of these existing sentiment analysis models. Existing sentiment analysis models can be improved further with more semantic and common-sense knowledge.

For Lexical contextual Sentence structure, the limitation include the dependency on lexicons and the lack of term sense disambiguation. In future, research could be done to improve extraction of the acute sense of sentence and remove noisy text for an efficient semantic orientation. Furthermore, the knowledge base will be improved for the semantic scores of all parts of speech.

In future, research could be done to improve the objectivity/subjectivity detection of a sentence by developing or using a more effective subjectivity detection algorithm as far as the combination of learning based and lexicon based approaches are concerned.

As for probabilistic models, the correspondence between generated clusters and latent variables has not yet been rigorously explained. It is a path that must be explored in order to improve the performance.

For combination of Feature mining and sentiment analysis, it has not yet been applied on reviews written in other languages than English. In future, this must be addressed to improve the performance of sentiment analysis using feature mining, that is, in the long term future we plan to apply the method to deal with documents written in languages other than English.

Regardless of the considerable number of difficulties and potential issues that undermines Sentiment analysis, one can't overlook the value that it adds to the industry. It is bound to become one the major drivers of many business decisions in future because of the reliance of Sentiment analysis results on factors that are so inherently humane. Enhanced precision and consistency in content mining procedures can help overcome some present issues confronted in Sentiment analysis. Looking ahead, what we can see is a true social **democracy** that will be created using Sentiment analysis, where we can harness the wisdom of the crowd rather than a select few "experts"- a democracy where every opinion counts and every sentiment affects decision making.

## References

[1] B. Pang, et al. "Thumbs up? Sentiment Classification Using Machine Learning Techniques," Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP), ACL Press, pp 79-86, July 2002.
[2] Bing Liu, "Sentiment analysis and subjectivity", Handbook of natural language processing, 2:568, 2010.
[3] Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis", Foundations and trends in information retrieval, 2(1-2):1–135, 2008.
[4] J. Wiebe, et al (2004)., "Learning Subjective Language," The Association for Computational Linguistics, vol. 30, no. 3, pp. 277-308.
[5] R. Mukras, J. Carroll (2004), "A comparison of machine learning techniques applied to sentiment classification", pp 200-204.
[6] Mori Rimon (2004), "Sentiment Classification: Linguistic and Non-linguistic Issues", pp 444-446.
[7] Alec Go (2005), "Twitter Sentiment Classification using Distant Supervision", Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005), Vancouver, CA.
[8] Yan Dang, Yulei Zhang, Hsinchun ChenA (July a2010), "Lexicon Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews", Department of Management Information Systems, vol. 25, no. 4, pp. 46-53.
[9] Changli Zhang, Wanli Zuo, Tao Peng,Fengling He (2008), "Sentiment Classification for Chinese Reviews Using Machine Learning Methods Based on String Kernel", Third 2008 International Conference on Convergence and Hybrid Information Technology.
[10] Kudo et al (2001), "An operational system for detecting and tracking opinions in on-line discussion", In SIGIR Workshop on Operational Text Classification, pp 449-454.
[11] Bing Liu. 2012, "Sentiment analysis and opinion mining", Synthesis Lectures on Human Language Technologies, 5(1):1–167.
[12] Kushal Dave, Steve Lawrence, and David M Pennock. 2003," Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", In Proceedings of the 12th international conference on World Wide Web, pages 519–528, ACM.
[13] Soo-Min Kim and Eduard Hovy, 2004," Determining the sentiment of opinions", In Proceedings of the 20th international conference on Computational Linguistics, page 1367, Association for Computational Linguistics.
[14] Hongning Wang, Yue Lu, and Chengxiang Zhai. 2010, "Latent aspect rating analysis on review text data: a rating regression approach", In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 783– 792, ACM.
[15] Yohan Jo and Alice H Oh, 2011, "Aspect and sentiment unification model for online review analysis", In Proceedings of the fourth ACM international conference on Web search and data mining, pages 815– 824, ACM.

[16] Minqing Hu and Bing Liu. 2004, “Mining and summarizing customer reviews”, In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 168–177, ACM.
[17] Lun-Wei Ku, Yu-Ting Liang, and Hsin-Hsi Chen. 2006, “Opinion extraction, summarization and tracking in news and blog corpora”, In AAAI spring symposium: Computational approaches to analyzing weblogs, volume 100107.
[18] Vasileios Hatzivassiloglou and Kathleen R McKeown, 1997, “Predicting the semantic orientation of adjectives”, In Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the European chapter of the association for computational linguistics, pages 174–181, Association for Computational Linguistics.
[19] Jorge Carrillo de Albornoz, Laura Plaza, Pablo Gerv_as, and Alberto Diaz, “A joint model of feature mining and sentiment analysis for product review rating”, In Advances in Information Retrieval, pages 55-66, Springer, 2011.
[20] Vachaspati, P and Wu, C., 2012, “Sentiment Classification using Machine Learning Techniques”.
[21] Hemalatha1, Dr. G. P Saradhi Varma and Dr. A.Govardhan,”Sentiment Analysis tool using Machine learning Algorithms”, International Journal of Emerging Trends and Technology in Computer Science (IJETTCS), Volume 2, Issue 2, March – April 2013.
[22] Samaneh Moghaddam and Martin Ester, “Ilda: interdependent Lda model for learning latent aspects and their ratings from online product reviews”, In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, pages 665-674, ACM, 2011.
[23] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation, “The Journal of machine learning research”, 3:993-1022, 2003.
[24] Cortes C, Vapnik V, “Support-vector networks”, presented at the Machine Learning; 1995.
[25] Vapnik V, “The nature of statistical learning theory”, New York; 1995.
[26] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze, “Introduction to Information Retrieval”, Cambridge University Press, 2008.
[27] Quinlan JR, “Induction of decision trees, Machine Learning 1986; 1:81–106.
[28] Reza Zafarani, Mohammad Ali Abbasi and Huan Liu,”Data Mining essentials," in Social Media Mining, Cambridge: Cambridge University Press, 2014.
[29] P.Kalaivani, Dr. K.L.Shunmuganathan,”Sentiment classification of movie reviews by supervised machine learning approaches”, Indian Journal of Computer Science and Engineering (IJCSE), vol 4, Aug-sep 2013.
[30] Leman Akoglu,” Quantifying Political Polarity Based on Bipartite Opinion Networks”, In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media; 2014.
[31] Liu Bing, Hsu Wynne, Ma Yiming, “Integrating classification and association rule mining”, In: Presented at the ACM KDD conference; 1998.
[32] Medhat W, Hassan A, Korashy H, “Combined algorithm for data mining using association rules”, Ain Shams J Electric Eng 2008; 1(1).
[33] Ko Youngjoong, Seo Jungyun, “Automatic text categorization by unsupervised learning”, In: Proceedings of COLING-00, the 18th international conference on computational linguistics; 2000.
[34] Samaneh Moghaddam and Martin Ester, “Opinion digger: an unsupervised opinion miner from unstructured product reviews”, In Proceedings of the 19th ACM international conference on Information and knowledge management, pages 1825-1828, ACM, 2010.
[35] Aurangzeb Khan, Baharum Baharudin, and Khairullah Khan, “Sentiment classification from online customer reviews using lexical contextual sentence structure”, In Software Engineering and Computer Systems, pages 317-331, Springer, 2011.
[36] Andrius Mudinas, Dell Zhang, and Mark Levene, “Combining lexicon and learning based approaches for concept-level sentiment analysis”, In Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, page 5, ACM, 2012.
[37] Hongning Wang, Yue Lu, and ChengXiang Zhai.” Latent aspect rating analysis without aspect keyword supervision.” In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 618-626, ACM, 2011.