

Support Vector Machine Based Sentiment Analysis Process for Twitter Streams

Rekha Malik¹, Sugandha Hooda², Jyoti Bharadwaj³

¹Department Of Computer Science and Engineering , PDMCE, Bahadurgarh

²H.O.D, Department Of Computer Science and Engineering , PDMCE, Bahadurgarh

³Assist. Professor, Department Of Computer Science and Engineering , PDMCE, Bahadurgarh Haryana,
Bahadurgarh

Abstract: Sentiment Research in twitter is quite tough due to its short length. Attendance of emoticons, slang words and misspellings in tweets compelled to have a preprocessing pace beforehand feature extraction. There are disparate feature extraction methods for accumulating relevant features from text that can be requested to tweets also. But the feature extraction is to be completed in two periods to remove relevant features. In the early period, twitter specific features are extracted. Next these features are removed from the tweets to craft normal text. Later that, once more feature extraction is completed to become extra features. This is the believed utilized in this paper to produce an effectual feature vector for analyzing twitter sentiment. As no average dataset is obtainable for twitter posts of electronic mechanisms, we crafted a dataset by accumulating tweets for a precise era of time. By acting EmotionResearch on a specific area, it is probable to recognize the impact of area data in selecting a feature vector. Disparate classifiers are utilized to do the association to find out their impact in this particular area alongside this particular feature vector. This paper prepossess an SVM established EmotionResearch procedure for twitter streams.

Keywords: Opinion Mining, Opinion Summarization, Sentiment Analysis, Support Vector machines.

I. Sentiment Analysis

Sentiment Research in twitter is quite tough due to its short length. Attendance of emoticons, slang words and misspellings in tweets compelled to have a preprocessing pace beforehand feature extraction. There are disparate feature extraction methods for accumulating relevant features from text that can be requested to tweets also. But the feature extraction is to be completed in two periods to remove relevant features. In the early period, twitter specific features are extracted. Next these features are removed from the tweets to craft normal text. Later that, once more feature extraction is completed to become extra features. This is the believed utilized in this paper to produce an effectual feature vector for analyzing twitter sentiment. As no average dataset is obtainable for twitter posts of electronic mechanisms, we crafted a dataset by accumulating tweets for a precise era of time. By acting emotionResearch on a specific area, it is probable to recognize the impact of area data in selecting a feature vector. Disparate classifiers are utilized to do the association to find out their impact in this particular area alongside this particular feature vector. This paper prepossess an SVM established EmotionResearch procedure for twitter streams. Analyzing tweets emotion will answer this question. You can additionally discover why people contemplate BJP is good, by removing the remove word indicating why people did or did not like BJP. Example. "Development".

Sentiment Research can be utilized to ascertain emotion on a collection of levels. It will score the whole as affirmative or negative, and it will additionally score the emotion of people words or phrases in the document. For example, if someone writes Twitter comment that reads:

"I affection the summertime in New York, but I disgust the winter."

The individual scores should display "love the summer" as affirmative and "hate the winter" as negative. Though, the emotion for the whole comment should be impartial, because the affirmative emotion for the word affection should annul out the negative emotion for the word hate.

Because emotionResearch can trail a particular case, countless firms use it to trail or monitor their produce, services or standing in general. For example, if someone is attacking your brand on communal mass media, emotionResearch will score the post as tremendously negative, and you can craft alerts alongside the post alongside hyper-negative emotions score.

Sentiment Research usually studies the polarity of posts such as affirmative, Negative and Neutral. Calculating emotion accuracy of emotionResearch is extremely vital as undeviatingly results the aftermath of Research.

II. Applications

Sentiment Research is consenting an increasingly producing attention from the usual speech processing area that is chiefly motivated by the wide-spread demand for opinion established requests, such as product reviews, entity pursuing and Research and opinion summarization. Emotion excavating has come to be a functional instrument for the business hobbies of both firms and individual consumers. They desire to sort out opinions concerning produce, services, or brands that are dispersed in online texts such as product study articles or forums [1].

In the following paragraphs we sum up a few important applications of sentiment analysis.

1. Sentiment Research can be utilized for ascertaining critics' opinions concerning a given product by categorizing online product reviews from websites such as Amazon and C|Net, RottenTomatoes.com and IMDb, and can additionally clarify extremely helpful for opinion oriented inquiries in question answering.
2. Tracking the advancing attitudes of the finished area in the direction of a governmental candidate by excavating online forums is additionally a functional application. It can furthermore be utilized to alert client services of dissatisfied clients that utter their frustrations on forums or discussion boards. Pursuing trends of bloggers is additionally becoming a valued Research earth, as it can be utilized for Research in trends or customer preferences.
3. Sentiment Research can additionally be helpful for recommendation arrangements, as those arrangements ought to not counsel something that receives negative feedback, and for the progress of new kinds of find engines.
4. The detection of "flames", overly heated or antagonistic speech, in e-mails or on communal networking websites will additionally benefit from emotion classification. Monitoring newsgroups and forums, whereas fast and automatic detection of blazing is vital, will additionally discern remarkable improvements.
5. Sentiment Research that is becoming hugely vital is that of opinion spam detection. As e-mail and Web spam are quite acquainted, opinion spam is yet new to the finished public. Because of the large development of user-generated content on the Web, it is nowadays a public exercise for people to find and elucidate other's opinions. Opinion spam mentions to human hobbies that endeavor to deliberately mislead readers or automated opinion excavating arrangements by providing undeserving affirmative opinions to a little target objects and/or by providing unjust, malicious or fake negative opinions to supplementary object.

III. Challenges

There are various Challenges which are explained below:

1. One of the biggest setbacks in the earth of Computational Linguistics is ambiguity. There are three spans in that we have to resolve ambiguity [2]:
 - Semantically,
 - Lexically and
 - Syntactically ambiguous text.

This setback can merely be tackled if computational arrangements owning a little form of globe vision or at the extremely least a basic lexicon and decision making skills come to be available. Semantic ambiguity has to do alongside a choice amid each numbers of probable interpretations, and is consequently closely connected to vagueness, e.g. idiomatic expressions that scarcely or not ever have well-defined definitions.

Lexical and syntactic ambiguity setbacks always involve semantic ambiguity as well. Lexical ambiguity deals alongside a choice amid a finite number of recognized and meaningful context-dependent interpretations for example 'bank' and 'bank'. One method of dealing alongside lexical disambiguation is to gaze at supplementary lexical agents oftentimes transpiring in combination alongside or in proximity of the unclear word. Aside from these two spans of ambiguity, syntactic ambiguity is additionally a huge obstacle, and deals alongside grammatical ambiguities; e.g. disambiguating the well-known sentence "Flying planes can be dangerous": is hovering planes hazardous, or are hovering planes dangerous? Context is additionally of critical significance here.

Emotions and opinions can be expressed explicitly and implicitly. Implicitness is a trial for computational arrangements, as even for humans it is not facile to recognize and examine these expressions correctly. This additionally applies for humor, sarcasm, irony, etc.

2. Another setback is inference, the procedure of illustrating conclusions by requesting precise clues to observations or hypotheses. Inference has been a accepted earth of Research, and requests such as expert arrangements and company law engines have followed. The aforementioned globe vision is utilized by an inference arrangement to ascertain ambiguities, and to find and recognize inherent relations. Unlike text categorization that aims at categorizing documents by case emotion association has moderately insufficient classes that generalize across countless areas and users. In fact, opinion association is the most complex after it additionally has to forecast 'neutral'.

3. Lastly, **domain-independence** is one of the biggest setbacks in contraption discovering and classification. A prudently selected feature set can produce extremely elevated accuracies for a precise corpus, but present extremely poorly after requested to one more kind of corpus. Finding innovative and competent ways to vanquish this setback is a valued Research earth.

IV. Problem Formulation

With the quick progress of the Internet& popularity of Twitter, huge number of Twitter posts are consented every single date in a extremely short span of period above extremely critical and vital issues.

These dates, Research on Twitter data focuses on scalable methods applicable to twitter posts to remove functional information from posts and emotionResearch that can be helpful in decision making policies. For example Twitter data Research can be utilized in Power strategy formation for education, for pre-planning of critical faith connected subjects, Company strategies, Advertisement Policies, Election, Arranging etc. Lots and lots can be completed by Research of Twitter data. There is an alternative method, that is focuses by researcher that is storage and computing of mass documents data in a distributed arrangement that makes it extra safeguard and efficient.

Its Brilliant communal implications additionally motivated me work on this case because this does not restrict on one span or one case but can be requested to each timely subject and can furnish vital Research that can be utilized in a functional way. Can be applied to:-

- Government policy formation
- Pre-planning for critical religious issues based on people sentiments
- Advertisement based on competing product twitter posts
- Business policies
- Election policies etc.

V. Proposed Work

The target is to examine all the comments generated as a hot tweet is retweeted. By excavating and analyzing the opinions encompassed in those comments, we can notice a little temporal or spatial outlines, e.g. emotion propagation features, emotion orientation curving points and emotion variation peaks. Setback meaning is firstly demanded.

Definition 1: An opinion in an opinion tweet document is a triple,

(g_i, w_{ij}, s_i) ,

Where g_i is an opinion target, w_{ij} is an opinion word commenting on the opinion target g_i , and s_i is the sentiment orientation of the document about the opinion target g_i .

The meaning is easy yet competent for imitating the opinions generated in microblogs. With the meaning, we nowadays have the task of opinion excavating.

Definition 2: Given an opinion documented, the task of opinion mining is to mine and summarize all the opinion triples (g_i, w_{ij}, s_i) in d .

It quite seems that the task of excavating opinions in corpus D consists of four main sub-tasks:

- (1) Opinion target extraction that extracts entity nouns or noun phrases in D .
- (2) Opinion words extraction that extracts opinion words for every single target.
- (3) Emotion orientation association that decides the emotion tendency of every single document on every single target.
- (4) Opinion summarization that summarizes all the opinion triples.

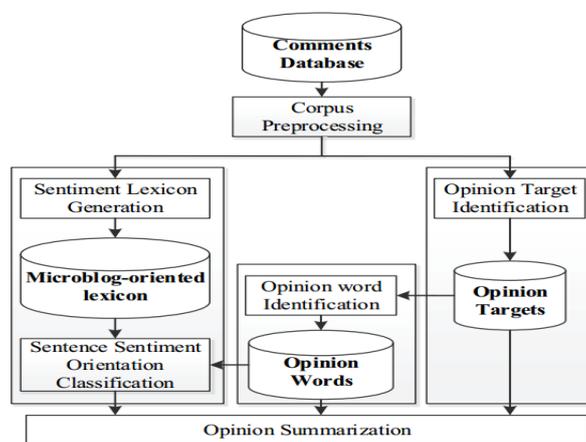


Figure1 Opinion mining framework

Architecture of our opinion excavating framework is shown in Fig. Given a corpus of microblog comments, the framework will early do a little preprocessing such as Chinese segmentation and POS tagging. It next generates a microblog-oriented emotion lexicon employing SO-PMI. At the alike period, the arrangement extracts the opinion targets that users have conversed concerning, and sorts the consequence by the frequencies of targets materializing in corpus. Later that, employing those targets and semantic hints, the arrangement extracts the opinion words commented on every single target. The emotion orientation of every single target in every single document is selected established on the emotion lexicon. In the last pace, a synopsis of all the opinion triples is generated. The key subtasks are debated below.

A. Corpus Preprocessing

A tweet encompasses countless redundant data; therefore a corpus preprocessing is needed. Useless halt words are removed first. NLPiR is utilized, that is a Chinese segmentation instrument, to cut Chinese tweets into solitary words or phrases. The instrument can additionally produce part-of-speech tag for every single word or phrase that decides whether the word or phrase is a noun, adjective, etc.

The preprocessed corpus will be utilized in two methods in the subsequent steps. The early one is for opinion target identification that generally uses nouns. We craft a deal file in that every single record stands for one tweet and encompasses the nouns and noun phrases of this tweet. The subsequent one is for emotion lexicon creation and emotion words extraction that generally uses all the adjectives, adverbs and verbs in corpus.

B. Opinion Target Extraction

The arrangement aims to notice what people are conversing concerning, so find the opinion targets is a critical task. At the alike period, the targets are sorted because we are interesting concerning what targets is the discussion hotspot.

As debated in Serving 2, there are countless methods to remove Opinion Targets. Center on the observation that after people comment on opinion targets, the words they use encounter, we use association excavating to find the hot opinion targets. The deal database is the deal file generated in the corpus preprocessing pace, and an item is a noun or a noun phrases.

Definition 3: Each record is the nouns or noun phrases in a tweet document, all records consist of a tweet transaction database,

$$r = \{\text{noun}_1, \text{noun}_2, \dots, \text{noun}_n\},$$

$$R = \{r_1, r_2, \dots, r_m\}.$$

Apriori algorithm is utilized to produce 1 tokfrequent itemsets that are candidate opinion targets. We delineate an itemset as recurrent if its minimum prop is 5%. Because microblogs posts are normally extremely short, the worth of k is set as 5 to elucidate the calculation.

We nowadays produce and sort opinion targets from those recurrent itemsets. Comments on a hot tweet normally encompass things that materialized in the early tweet. Moreover, comments also converse concerning a little supplementary connected thing, which makes the tweet far extra meaningful. Therefore if a comment encompasses extra nouns or noun phrases, the comment will have higher potential to be a subjective opinion sentence and those nouns or noun phrases could be opinion targets.

Based on that observation, the nouns or noun phrases in k frequent itemsets have higher potential to be opinion targets contrasted alongside those in 1-frequent itemsets. By traversing the itemsets from k-frequent to 1-frequent, noun or noun phrase are locale into opinion targets database one by one. If the noun or noun phrase exists in the database, next skip it. Next the opinion targets are removed afterward finishing traversing 1-frequent itemsets. The worth of k is adjustable according to the real situation.

C. Opinion Words Extraction

We are interested in how people contemplate of a target, that is, what words they use to comment a target. Opinion words that express subjective opinions are crucial. Words displaying opinions generally are adjectives, adverbs and verbs.

Based on opinion targets generated in the last pace, opinion words are removed, consequently merely those tweets that encompass those opinion targets are seized into account. For every single tweets, if it encompasses an opinion target, we remove the adjectives/adverbs/verbs adjacent the target. We say, if an adjective is adjacent to the target, that is, if the distance amid them is no extra than a threshold, the adjective modifies the targets. The distance is compute by the number of words amid them in a sentence. Convoluted syntactic or semantic constructions are not believed here for efficiency, but it is yet an enhancement work in our upcoming work.

D. Sentiment Orientation Classification

After removing opinion targets and opinion words, the emotion orientation of a document in the direction of every single target are decided. The emotion orientation of a word/sentence/document indicates the association that it deviates from the emotion neutrality.

Sentiment words, that are the frank and pondered reflections of emotion features, are words that have seeming emotion tendency, encompassing adjectives, nouns, adverbs, etc. A emotion lexicon is a set of emotion words, that can be tear into affirmative emotion lexicon and negative emotion lexicon. There are emotion lexicons for English, e.g. SentiWordNet. Though, they are not adequate for microblogs because expressions in microblogs are newfangled. That is why a microblog-oriented emotion lexicon is crafted.

VI. Tweet Sentiment Analysis Steps

Performing EmotionResearch on Twitter is trickier than acting it for colossal reviews. This is because the tweets are extremely short (only concerning 140 characters) and normally encompass slangs, emoticons, hash tags and supplementary twitter specific jargon. For the progress intention twitter provides streaming API that permits the builder an admission to 1% of tweets tweeted at that period centers on the particular keyword. The object concerning that we desire to present emotionResearch is presented to the twitter API's that does more excavating and provides the tweets connected to merely that object. Twitter data is usually unstructured i.e. use of abbreviations is extremely high. Additionally it allows the use of emoticons that are manage indicators of the author's think on the subject. Tweet memos additionally encompass of a timestamp and the user name. This timestamp is functional for estimating the upcoming trend request of our project. User locale if obtainable can additionally aid to gauge the trends in disparate geographical spans.

The colossal number of feature powers us to use methods for selecting the top features to use for training the classifier, Word Frequency, Inverse Document Frequency (TF-IDF) is a numerical statistic method to filter the features by weighting and scoring every single of the unigrams and N-grams employing the frequency of words in the text. There are continuing algorithm obtainable established on lexicon established way and contraption discovering way Additionally Published Directory are obtainable like SentiWordNet ,SentiWord to find polarity of tweets. These are undeviatingly utilized to find polarity of posts

Tweets Pre-Processing

Before tweets emotionResearch pre-processing is the one of the viral method to become effectual aftermath in the pursuing delineated pre-processing method

Vectorization:

The data has been modified to vector form by employing the TF-IDF purpose and find out # the exceptional words from all the tweets has been provided. It provides a solitary text file of encompassing all the data demand to need for the tweet classification. Vector space ideal is the most extensively utilized method in tweet representation. This ideal uses feature entries and their weights to express the document information. To embody document object vector space ideal have been introduced. Vector $d = (w_1, w_2, w_3, \dots, w_m)$ stands for the feature entry and its heaviness in document d , the number of all feature entries. w_i ($i= 1, \dots, m$) is the heaviness of the entry t_i in document d . The document vector set is the outline or data object of the document clustering.

Part-of speech Tagging (POS): The procedure of part-of-speech tagging permits to automatically tag every single word of text in words of that portion of speech it belongs to: noun, pronoun, adverb, adjective, verb, interjection, intensifier, etc. The aim is to remove outlines in text established on Research of frequency allocations of this part-of-speech.

Stemming and Lemmatization:

Stemming is a procedure of substituting words alongside their stems, or roots. The dimensionality of the BOW is decreased after root-related words, such as "read", "reader" and "reading" is mapped into one word "read". Though one ought to be careful when applying stemming, as it could rise bias. For example, the biased result of stemming appears after merging different words "experiment" and "experience" into one word "exper".

Stops-Word removal: Stops words are words that hold a relating purpose in the sentence, such as prepositions, articles, etc. There is no definite catalog of halt words, but a mechanism, are employing a little of the most public, short purpose words, such as "the", "is", "at", "which" and "on". These words can be removed from the text beforehand association as they have a elevated frequency of occurrence in the text, but do not alter the final emotion of the sentence.

Negations Handling: Negation mentions to the procedure of conversion of the emotion of the text from affirmative to negative or from negative to affirmative by employing distinct words: "no", "not", "don't" etc.

These words are shouted negations. The simplest way to grasp negation is to revert the polarity of all words that are discovered amid the negation and the early punctuation marks pursuing it.

But-clauses: The phrases like “but”, “with the exclusion of”, “except that”, “except for” usually change the polarity of the portion of the sentence pursuing them. In order to grasp these clauses the opinion orientation of the text beforehand and afterward these phrase ought to be set opposite to every single other.

Tokenization into N-grams: Tokenization is a procedure of crafting a bag-of-words from the text. The incoming thread gets broken into including words and supplementary elements. The public separator for recognizing individual words is whitespace. Tokenization of social-media data is considerably extra tough than tokenization of the finished text as it encompasses countless emoticons, URL links, abbreviations that cannot be facily separated as finished entities. It is a finished exercise to join conveying words into phrases or n-grams that can be unigrams, bigrams, trigrams, etc.

VII. Feature Generation & Feature Selection

Features are text qualities that are functional for seizing outlines in data. Features are words in the finished set of twitter tweets. All of the words in form of unigram. Bi-gram, N-gram could not be useful in tweet classification. The most accepted features utilized in contraption discovering classifications. The attendance or the frequency of n-grams removed across the pre-processing step. In the presence-based representation for every single instance a binary vector is crafted in that “1” way the attendance of a particular n-gram and “0” indicates its absence. In the frequency-based representation the number of occurrences of a particular n-gram is utilized instead of a binary indication of presence. In cases whereas text length varies greatly, it could be vital to use word frequency (TF) and inverse word frequency (IDF) measures. Though, in short memos like tweets words are unlikely to recap inside one instance, making the binary compute of attendance as informative as the counts. Separately from the n-grams, supplementary features can be crafted to enhance the finished quality of text classification. The most public features that are utilized for

1. Feature selection

Since the main features of a text classifier are N-grams. The dimensionality of the feature space grows proportionally to the size of the dataset. There can be 10,000 of words (N-grams) in feature vector. This melodramatically development of the feature space makes it in most cases computationally impossible to compute all the features of a sample. Countless features are redundant or irrelevant and do not considerably enhance the results. Feature selection is the procedure of recognizing a subset of features that have the highest predictive power. This pace is critical for the association procedure, as elimination of irrelevant and redundant features permits to reduce the size of feature space rising the speed of the algorithm, circumventing above fitting as well as giving to the enhanced quality of classification. Results

Sentiment Research is additionally extremely vital as can aid in lots of decision making as clarified earlier. People post their sights on assorted subjects possessing affirmative, negative and impartial sentiments. Posts can be categorized on the centers of their polarity. This gives a main think concerning each subjects and can aid in strategy making.

2. SVM Kernel Classification

Now that we have the believed of a margin we can devise the maximum margin classifier. We will early delineate the hard margin SVM, applicable to a linearly separable dataset, and next adjust it to grasp non-separable data. The maximum margin classifier is the discriminant purpose that maximizes the geometric margin $1/||w||$ which is equivalent to minimizing $||w||^2$. This leads to the following constrained optimization problem:

$$\begin{aligned} \underset{w,b}{\text{minimize}} \quad & \frac{1}{2} ||w||^2 \\ \text{Subject to:} \quad & y_i(w^t x_i + b) \geq 1 \quad i = 1, \dots, n. \end{aligned}$$

The constraints in this formulation safeguard that the maximum margin classifier classifies every single example accurately, that is probable as we consented that the data is linearly separable. In exercise, data is frequently not linearly separable; and even if it is, a larger margin can be attained by permitting the classifier to misclassify a little points. To permit errors we substitute the inequality constraints in Eqn. (10) alongside $y_i(w^t x_i + b) \geq 1 - \xi_i \quad i = 1, \dots, n,$

where $\xi_i \geq 0$ are slack variables that allow an example to be in the margin ($0 \leq \xi_i \leq 1$, also called a margin error) or to be misclassified ($\xi_i > 1$). Since an example is misclassified if the value of its slack variable is greater than 1, $\sum_i \xi_i$ is a bound on the number of misclassified examples. Our objective of maximizing the margin, i.e. minimizing $\frac{1}{2} ||w||^2$ will be augmented with a term $C \sum_i \xi_i$ to penalize misclassification and margin errors. The optimization problem now becomes:

$$\underset{w,b}{\text{minimize}} \quad \frac{1}{2} ||w||^2 + C \sum_{i=1}^n \xi_i$$

subjected to: $y_i(w^t x_i + b) \geq 1 - \xi_i, \xi_i \geq 0.$

The constant $C > 0$ sets the relative importance of maximizing the margin and minimizing the amount of slack. This formulation is called the soft-margin SVM, and was introduced by Cortes and Vapnik. Using the method of Lagrange multipliers, we can obtain the dual formulation which is expressed in terms of variables a_i ,

$$\underset{w,b}{\text{minimize}} \quad C \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j a_i a_j x_i^T x_j$$

subjected to: $\sum_{i=1}^n y_i a_i = 0, 0 \leq a_i \leq C$

The dual formulation leads to an expansion of the weight vector in terms of the input examples:

$$w = \sum_{i=1}^n y_i a_i x_i.$$

The examples x_i for which $a_i > 0$ are those points those are on the margin or within the margin when a soft-margin SVM is used. These are the so-called support vectors. The expansion in terms of the support vectors is often sparse, and the level of sparsity (fraction of the data serving as support vectors) is an upper bound on the error rate of the classifier.

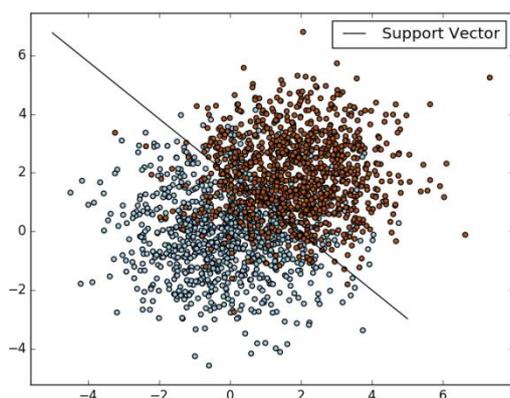


Fig: 2 Decision Boundary for Positive (red) and Negative (blue) tweets using Support Vector Machines

VIII. Experiment Evaluation

An examination is projected to assess our emotion orientation association method. Conduct a emotion association employing two methods: lexicon-based method and SVM Kernel established method. The aftermath is shown in TABLE below. The aftermath indicates that in the context of microblogs, SVM Kernel established method achieves far higher precision and recall that displays the affectivity of our arrangement.

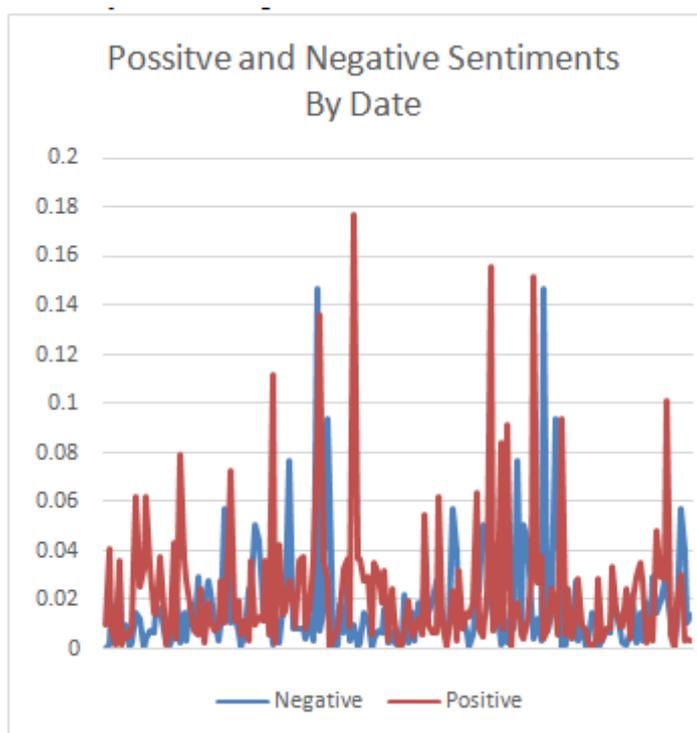
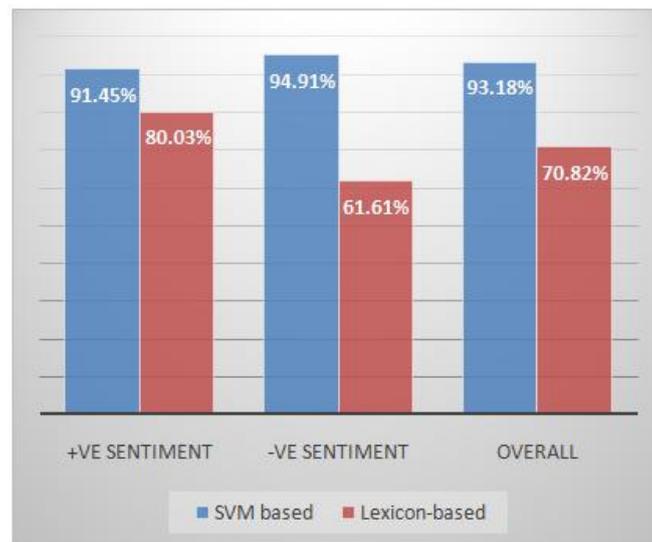


Fig: 3 Positive and Negative Sentiment by Date

Table 1 Results of sentiment orientation classification

Method	+vesentiment	-vesentiment	Overall
SVM based	91.45%	94.91%	93.18%
Lexicon-based	80.03%	61.61%	70.82%

**Fig 4:** Comparative Analysis of SVM based Twitter Sentiment Analyzer

IX. Conclusion and Future Scope

Sentiment Research is speech processing task that uses a computational way to recognize opinionated content and categorize it as affirmative or negative. The unstructured textual data on the Web frequently carries expression of opinions of users. EmotionResearch attempts to recognize the expressions of opinion and mood of writers. A easy emotionResearch algorithm endeavors to categorize a document as 'positive' or 'negative', established on the opinion expressed in it.

This work introduces an way to emotionResearch that uses prop vector mechanisms (SVMs) to hold jointly varied origins of potentially pertinent data, encompassing countless favorability measures for phrases and adjectives and, whereas obtainable, vision of the case of the text. Models employing the features gave are more joined alongside unigram models that have been shown to be competent in the past and lemmatized editions of the unigram models. The work gave in this thesis can be pursued more in countless directions. One of the tasks to be addressed is to discover the relation amid emotion groups and intensity. A little steps might be seized to address the distinct needs of the kind of casual speech utilized in online communication. This should aid enhance performance. One more association for upcoming work is to ponder the emotion intensity association setback as that of ordinal association – that is, association alongside arranged groups.

References

- [1]. WojciechGryc, and KaroMoilanen. "Leveraging textual sentiment analysis with social network modelling." *From Text to Political Positions: Text Analysis Across Disciplines* 55 (2014): 47.
- [2]. Hassan Saif, Yulan He, Miriam Fernandez, and Harith Alani. "Semantic patterns for sentiment analysis of Twitter." In *The Semantic Web- ISWC 2014*, pp. 324-340. Springer International Publishing, 2014.
- [3]. Li Dong, Furu Wei, Ming Zhou, and Ke Xu. "Adaptive multi-compositionality for recursive neural models with applications to sentiment analysis." In *Twenty-Eighth AAAI Conference on Artificial Intelligence*. 2014.
- [4]. Grégoire, Mesnil, Tomas Mikolov, Marc'AurelioRanzato, and YoshuaBengio. "Ensemble of generative and discriminative techniques for sentiment analysis of movie reviews." *arXiv preprint arXiv:1412.5335* (2014).
- [5]. Robert,West, Hristo S. Paskov, Jure Leskovec, and Christopher Potts. "Exploiting social network structure for person-to-person sentiment analysis." *arXiv preprint arXiv:1409.2450* (2014).
- [6]. Li, Fangtao, Sheng Wang, Shenghua Liu, and Ming Zhang. "Suit: A supervised user-item based topic model for sentiment analysis." In *Twenty-Eighth AAAI Conference on Artificial Intelligence*. 2014.
- [7]. Zhu, Linhong, Aram Galstyan, James Cheng, and Kristina Lerman. "Tripartite graph clustering for dynamic sentiment analysis on social media." In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pp. 1531-1542. ACM, 2014.
- [8]. Khan, Aamera ZH, Mohammad Atique, and V. M. Thakare. "Combining lexicon-based and learning-based methods for Twitter sentiment analysis." *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCE)* (2015): 89.
- [9]. Shefrin, Hersh. "Investors' Judgments, Asset Pricing Factors and Sentiment." *European Financial Management* 21, no. 2 (2015): 205-227.
- [10]. Pak, Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." In *LREc*, vol. 10, pp. 1320-1326. 2010.