International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

VISUALIZATION OF CUSTOMER OPINION USING SENTIMENTAL ANALYSIS

K. Charan Kumar Reddy, M. Chandra Gangadhar Reddy, M. Bharath, Ritesh Kumar Jaiswal

Madanapalle Institute Of Technology & Science, JNTU-A, Angallu, Madanapalle,517325.
Email ID: (18699A0420@mits.ac.in), (18699A0430@mits.ac.in), (18699A0413@mits.ac.in), (drriteshkumarj@mits.ac.in)

ABSTRACT

Social media has redefined the nature of how companies strategize their businessprocesses. It contains a massive volume of unstructured data (e.g. tweets, comments, user post, and reviews) that can be used for business intelligence suchas customer profiling and content analysis. Twitter, which is a social networking online service, is mainly used as a marketing and promotion tool by most companies. Specifically, twitter data contains not only user information, but also texts that contain subjective information (such as user sentiments) towards a particular issue. From a business perspective, the wealth of tweets is enough forcompanies to gather sufficient feedback about their products and services from their customers without having to spend for costly customer surveys and interviews. Data miners, on the other hand, face a difficult problem in assessing and extracting information from unstructured data. To address this problem, we use a combination of natural language processing and machine learning techniques in our project. This project used a popular food brand from Swiggy to judge a stream of Twitter client remarks. For the analysis, several measures in knowledge classification and clustering were applied. The Twitter corpus is gathered using a Twitter API. This project aims to debate the technical and business aspects of text mining analysis of Twitter data, as well as provide recommendations for future development potential in this burgeoning sector.

1. INTRODUCTION

Other people's opinions have a big influence on how we make decisions on a daily basis. These choices range from purchasing smart phones to making investments to selecting a school all of which have an impact on many areas of everyday lives. Before the internet, people would seek evaluations on products and services from friends, families or customer reviews. However, with the age of the internet, it is much easier to gather a variety of viewpoints from various folks all throughout the world people hunt for information on review sites, e-commerce sites, and other internet resources, blogs and social media to seek opinions on how to identify a specific product or service in the market place.

Sentiment Analysis and Opinion Mining April 22, 2012 Bing Liu.Padmaja, S., & Fatima, S. S. (2013). Opinion Mining and Sentiment Analysis–An Assessment of Peoples' Belief: A Survey. International Journal.

There are several sorts of sentimental classifications, each of which focus on how to convert unstructured material into structured opinions and address the field's present issues.

2. SENTIMENT CLASSIFICATION LEVELS

Sentimental analysis can occur at different levels:

- Document level
- Sentence level
- Aspect/Feature level

2.1 SYSTEM DESIGN:

- Review Dataset
- Pre-Processing
- Tokenizer and Stop words removal
- Transformation
- Classification

- Evaluation

3. SYSTEM IMPLEMENTATION

The purpose of system implementation is to make the new system available to the prepared group of users, as well as to keep the system supported and maintained within the performing company. Transitioning the system and maintenance mode of operation, with ownership of the new system changing from the project team to the performing organization.

3.1 ALGORITHMS

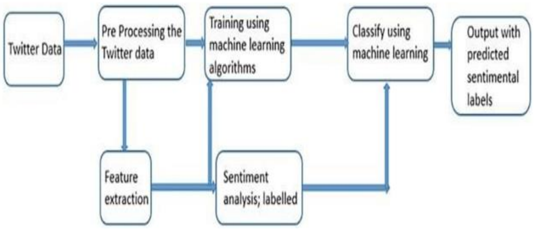
In our application we have used four algorithms

- Naïve Bayes
- Random Forest
- Logistic Regression
- Linear SVM

3.2 SYSTEM ARCHITECTURE

The definition and modelling of an architecture dedicated to the activities of analysis of big data, as the ones produced by social networks as twitter, is currently still at an initial stage of its development and consolidation. Unlike traditional data warehouse or business intelligence systems, whose architecture is designed for structured data, or so called "raw data", i.e., without a particular structure. It should also be pointed out that such systems should be able to allow processing and analysis of data not only in batch mode, but also in a real-time fashion.

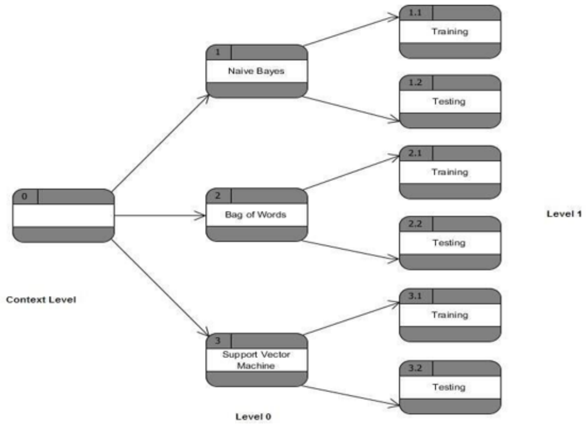
Nowadays a huge amount of data, daily produced by social networks, can be processed and analysed for different purposes. These data are provide with several features, among which: dimension, peculiarities, source, reliability.



3.3 DATA FLOW DIAGRAM

A data flow diagram is a graphical representation of data flow through an information system, complete with process elements. A DFD is frequently used as a preliminary in the development of a system overview that can then be expanded upon.

- **Process**: A process is a collection of stages that takes data as input and outputs it
- **External entity**: Objects that interact with system operations but are not included in the model.
- **Data Store**: A process's input and output data are saved in files or data storage.
- **Data Flow**: The data flows from process to process.



3.4 DATA SET

We work with the 568,454 reviews in the food reviews dataset. The dataset is a single CSV file that contains the product and reviewer ids, the reviewers scores ranging from 1 to 5, the data for each review, a brief synopsis for each review, and the text of the reviews. As labels and raw inputs, we extract the columns of the scores and the review texts.

Train Data:

The experience that the algorithm learns is based on the results in the training set. Each observation in supervised learning problems consists of multiple observed input variables and an observed outcome variable.

Test Data:

The test set is a set of data that is used to evaluate the model's performance using a metric. The test set must not contain any observations from the training set.

Score Distribution:

Here in score distribution the scores from 1 to 5 are visualized by the Bar graph. We obtain

- 1 as very negative
- 2 as negative
- 3 as neutral
- 4 as positive

The distribution of reviews over each score is represented in the below graph.

We plot this graph based on dataset:

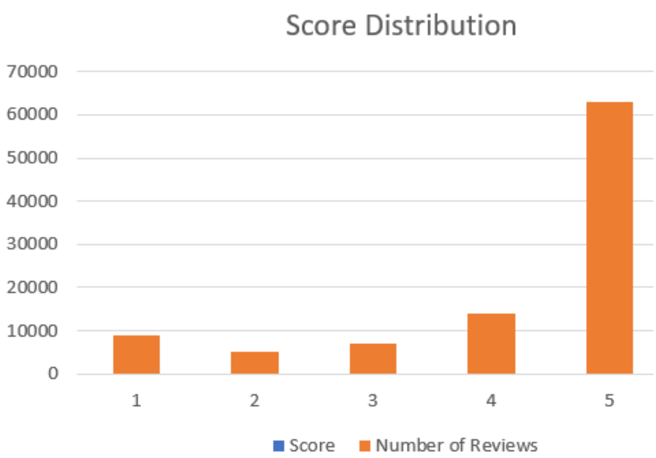


Fig 1. Score Distribution

Positive and Negative Bar Graph

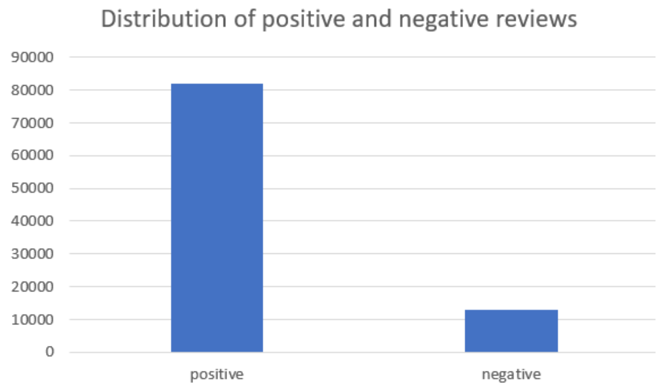


Fig2. Positive and Negative Bar Graph

Confusion Matrix:

A confusion matrix is a table that is frequently used to summarize a classification model's performance on a set of test data for which true values are known. It enables the visualization of an algorithm's performance.

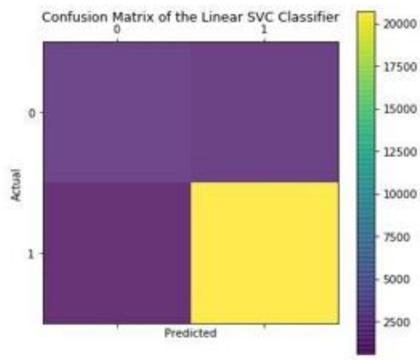


Fig3. Confusion Matrix

4. RESULTS

We'll start with our outputs for the objective and subjective classifications, as well as positive and negative classifications. These findings serve as the first phase in our classification process. We only use the shortlisted features for both of these results. This means that for the objective or subjective classification we have 5 features and for positive or negative classification we have 3 features. For both of these results we use different classification algorithms, because these are the algorithms, we are employing in our actual classification approach at the first step.

We make a condition while reporting the results of classification that only subjective labelled tweets are used to calculate these results. In the final categorization approach, however, any such condition is eliminated, and objectivity and classifications.

4.1 POSITIVE TWEET

Fig 4. Positive Tweet

When a customer gave a good review, it was analyzed as appositive tweet by visualizing the tweet and obtaining a pie-chart by analyzing 20 tweets. It was determined that 90 percent people believed it was positive and 10 percent it was negative.

4.2 NEGATIVE TWEET

Fig 5. Negative Tweet

When a customer gave a negative review, it was analyzed as a negative tweet by visualizing the tweet and analyzing 56 tweets. It was determined that 25 percent of people thought it was a nice review and 146 percent felt it was a bad review.

5. CONCLUSION

Interest in sentiment classification has grown as a result of the rise of social media. Individuals and organizations must be able to recognize and identify sentiment from text quickly and accurately. Most reliable approaches are predicted to minimize misclassifications in the creation of prediction models to classify reviews. In this project, the results of various methods are empirically evaluated on datasets of different size for use in sentiment mining. Among the methods used, linear support vector machine (LSVM) is highlyrobust in nature which is studied through various quality parameters. For practically all of the prediction approaches, the compound combination of unigram, bigram, and trigram works exceptionally well. The results show that utilizing SVMs for class prediction can be influenced by data balance in real-time applications, even if SVMs can respond well to some degree of data balance.

5.1 FUTURE ENHANCEMENT

In future, the effect of various other feature reduction techniques like latent allocation can be investigated. Further experiments should be conducted in the future to evaluate the impact of various domain and region specific parameters. Extending sentiment mining to different domains could result in some surprising new findings. In the future, other combinations of n-grams and feature weighting that provide a higher level of accuracy than this could be examined. This project's effort is limited to categorizing sentiment into two classes (binary classification), namely, a positive class and a negative class. A multiclass of sentiment classification, such as positive, negative, or neutral, and in a short time frame, may be considered in the future. The goal of this study is to find features that appear in the reviews as nouns or noun phrases. The detection of implicit features will have to wait till later. Because ensemble learning methods take a long time to compute, parallel computing techniques should be investigated to address this issue.

The inability to interpret the findings of ensemble learning methods is a key drawback, and the knowledge obtained by ensembles is difficult for humans to comprehend. As a result, enhancing ensemble interpretability is a prominent research direction.

Future sentimental mining systems will necessitate a larger and more comprehensive knowledge base of common and commonsense data. This will enable a better understanding of natural opinions as well as a more efficient connection between multimodal and machine processable data. Combining scientific theories of emotion with the actual engineering goals of analyzing sentiments in natural language text will result in more bio-inspired approaches to the design of intelligent sentiment systems capable of handling semantic knowledge, making analogies, learning new affective knowledge, and detecting, perceiving, and feeling emoji emotions.

REFERENCES

[1] N Cristianini, J Shawe-Taylor, "An Introduction to Logistic Regression and Other Kernel-based Learning Methods".2004.

[2] Hiroshi Shimodaira, "Text classifying using Naïve Bayes Document models".Learning and Data Note7 (Pg.no 1-9),2014.

[3] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson publications. (KDD)2021.

[4] Dan Jurafsky "Text Classification and Naïve Bayes", The Task of Text Classification.2007.

[5] Akkaya, C., J. Wiebe, and R. Mihalcea. Subjectivity word sense disambiguation. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing.2009.

[6] Dave, S. Lawrence, and D. Pennock. Mining the food reviews: opinion extraction and semantic classification of product reviews.2012.

[7] Davidov, D., O. Tsur and A. Rappoport. Enhanced sentiment learning using twitter emoji simleys.2010.

[8] Ding, X., B.Liu, and L.Zhang. Entity discovery and assignment for opinion mining applications. In proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2009).