Sentiment Analysis on Restaurant Reviews : A Multimodal Approach

Charugatra Tidke, Sreekanth Kankara, Manasa Alla

December 10, 2018

Abstract

In the last decade, the area of Sentiment Analysis captured the attention of all the researchers and developed at a vast extent. Various methods from the areas of Data Mining, Natural Language Processing, Statistics, Machine Learning, Mathematics contributed towards Sentiment Analysis. In machine learning approach, different classifiers are trained on the basis of labelled training data. In this project, a multi-modal approach is presented to improve the accuracy and confidence of Sentiment Analysis. This modal implements Naive Bayes classifier and its variants, Support Vector Machines and Recurrent Neural Networks as the base of the majority theorem.

Keywords: Sentiment Analysis, Bayes Rule, Naive Bayes, Support Vector Machines, Recurrent Neural Networks.

1 Introduction

Recommendations and reviews are in use even before the birth of modern online era. Similarly now a days everything works on reviews, from a product to buy to where we should plan our next vacation. Reviews works as the best adviser to recommend, as it is experience of a person. Mostly people like to visit restaurants based on the online ratings and the reviews of that restaurant. Ratings are easy to understand in less time than reading a long explanation of the restaurant review. But sometimes ratings provide the false information. Many reviews can be found which says really bad about the restaurant but the ratings are 4 out of 5. This misleading information arose so many questions like :

- Is the ratings given for a restaurant really mean what they say?
- Is the ratings given by websites actually say about how good/bad the restaurant is?
- Is there a way that instead of reading long reviews, can we just get the polarity of it?
- How much patience does a person need to have to go through all the reviews and find out what exactly the restaurant is good for?

Thus to answer all these questions and making the understanding of review quick this project presents a multi-modal based approach for Sentiment Analysis of Restaurant Reviews. This modal makes the uses of different classifiers such as Naive Bayes, Support Vector Machines and Recurrent Neural Networks.

2 Review of Related Works

There are various research done on different types of sentiment analysis approach. Theresa et al.,[1] proposes a module of sentiment analysis which is used to recognize Contextual polarity. It uses machine learning and a variety of features to identify the polarity of given text. Kolchyna et al.,[2] implemented Lexicon based sentiment analysis with the Machine Learning approach and a hybrid model combining both approches to perform Twitter Sentiment Analysis. The results of their analysis showed that the Naive Bayes Classifier and SVM classifier performs better than the Lexicon based Sentiment Analysis. The analysis also confirms that combining both the methods leads to more accurate results. Inspired with this approach this paper presents a multi-modal approach combining results from multiple classifiers.

Kouloumpis et al.,[3] used supervised learning approach based on lexicons to perform sentiment analysis on Twitter hashtags. SentiWordNet is mostly used resource in opinion mining area. Kreutzer and Witte[4] detailed information on SentiWordNet and its use and application in Opinion Mining. Pang and Lee[5] provided a very useful and detailed survey on "Opinion Mining and Sentiment Analysis, which focuses on the aspects like its implementation and challenges. The survey also provides a list of resources

and a flow of the implementation of different methodologies.

All the approched follow similar kind of preprocessing approaches for e.g., tokenization, lemmatization, removal of stop words, Parts of Speech Tagging. After the preprocessing step each approach follow different algorithm based on method of implementation

3 Solution Methodology

Various algorithms have been implemented in this project which follow different methodologies for training the machine and then performing the sentiment analysis based on that. Though the algorithm follows different method for training analysis the preprocessing steps are same. Following are preprocessing steps followed by our Multimodal Approach:

- Reading data from the file - This file contains the prelabelled reviews as 0 - negative and 1 as positive.
- Tokenization - The tokenization step separate the words in the sentence as tokens.
- Lemmatization/Stemming - This stem converts each word into its root or sometimes simply cut the suffix from the given input word.
- Stopword Removal - This is a very important stem as it removes all the unnecessary words and provide with the most meaningful words of the review.

3.1 Naive Bayes

Naive Bayes Approach uses a future set for training the classifier. The feature set is nothing but the dictionary of the unique word contained in the the data base with the True and False tag according to the review. A part of this feature set is used to train the data set and the remaining part is used to test the dataset. For the building of feature set we selected only the adjectives from the reviews as adjectives contains 90% of the feelings of the sentence.

Original Naive Bayes Classifier : Naive Bayes Classifier is based on the Bayes Rule for the classification of the events and objects:[6]

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

where,
P (A | B) is the probability of 'A' being true given that 'B' is true,
P (B | A) is the probability of 'B' being true given that 'A' is true,
P (A) is the probability of 'A' being true,
P (B) is the probability of 'B' being true.
It makes a naive assumption that each word is independent of other word in the context.

Bernoulli Naive Bayes Classifier : It is a type of Naive Bayes classifier and also works on the probability theory. It considers only two situation, presence or absence of the attribute, i.e, it considers whether the word is present or not in the review.

Multinomial Naive Bayes Classifier : It is a categorical distribution based on the probability theory, mostly used for the discrete counts. It considers the occurrance of the feature i.e how often the word occurs in the context.

Complement Naive Bayes Classifier : The Complement Naive Bayes classifiers instead of checking the probability of feature in the particular class, it checks what the probability of feature not belonging to the other class.

3.2 Support Vector Machines

Algorithm

1. Read the data from the excel
2. Taking Reviews into X array and target in Y array
3. Normalization of the reviews
4. Removing stop words from the reviews
5. Excluding punctuation and tokenization of data

Support Vector Machines contain a set of supervised learning methods used for classification, regression and outliers detection. SVM classifies the data using a hyper plane and support vectors as shown in the below figure. It classifies the data with a median vector.

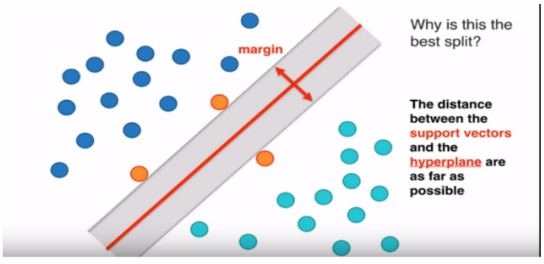

Figure 1: SVM classifies the data using a hyper plane

It uses Mathematically Lagrange Multiplier technique to maximise the margin. Here, we have used

classification with two methods: i) Stochastic Gradient Descent ii) Support Vector Classification Continuing to the mentioned algorithm at the top of the document:

- Vectorization of the tokenized data
- Dividing the data for training and testing
- Passing the data to the below classifiers for training the machine

3.2.1 Stochastic Gradient Descent (SGD)

Stochastic Gradient Descent is a classification of Support vector Machine which is used for discriminative learning and an efficient approach of linear classifiers using convex loss functions. Like most of the vector classifiers, SGD needs to be fit into two arrays, X with size [samples, features] as the training samples and array Y [samples] contains the target values for the samples. This makes up the training dataset and helps the machine to train with the vectorized data. Once the machine is trained, it can be used to predict target values of the new data using predict function. **Process of Implementation** A plain stochastic gradient descent learning is used by the class SGD Classifier, which supports various loss functions. However, we have used loss="log" which is a "logistic regression". This enables a method predict_prob, which results in a vectorized form of probability estimates P(y/x) per each sample x. sgd_clf = SGDClassifier(loss="log", class_weight="balanced") sgd_clf.fit(X_train, y_train) Predicts the data using the below syntax, y_pred_sgd = sgd_clf.predict(X_test)

```
Accuracy: 0.9174565560821485
F1 score: 0.9172483676432528
              precision    recall  f1-score   support

           0     0.9097    0.9165    0.9131      1198
           1     0.9245    0.9183    0.9214      1334

   micro avg     0.9175    0.9175    0.9175      2532
   macro avg     0.9171    0.9174    0.9172      2532
weighted avg     0.9175    0.9175    0.9175      2532
```

Figure 2: Result of Stochastic Gradient Descent Classifier

3.2.2 Support Vector Classification

Support Vector Classification is one of the Support vector machine classifiers similar to SGD which takes vectorised form data as input and trains the machine using the data using different mathematical formula. Like SGD, it also takes two arrays as input where X contain the samples and features and Y contain the target value of the samples. The machine has been trained using the sample data as input. **Process of Implementation**

$svmModel = svm.SVC(decision\_function\_shape =' ovr', C = 100, probability = True, gamma =' auto')$

$svmModel.fit(X\_train, y\_train)$ Predicts the data as below, $svmModel.predict(X\_test)$

Results: Support Vector Classification

```
Accuracy: 0.9052132701421801
F1 score: 0.9051322384907583
```

Figure 3: Result of Support Vector Classification

3.3 Neural Networks

The Neural network is a machine learning algorithm, which takes trained data and gains the knowledge. This knowledge obtained is used to predict the polarity of the sentence. The structure of artificial neural network is like neural network of the brain. This network consisted of nodes (like neurons) in multi layers organized in a hierarchical manner. This system contains a single input and output layer, with one or more intermediate hidden layers. The information is passed from one node to another node, over the layers using weight associated with each one. The output is calculated by applying activation function.

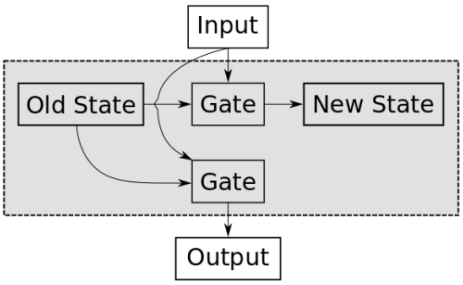

Figure 4: Working of Neural Network[7]

Raw Text: The data-set(review and their labels) is loaded and divided among the training and testing data.

Tokenizer: The text string connot be directly used by the neural network. So the text should somehow be converted.The first step of conversion is tokenizer, where each word of the data set before being the input of the neural network is converted into an integer

The output of data may result in arbitrary length. In order to maintain the same length, all the integer token are padded, by considering the mean value of the length. As the maximum length conversion may led to a lot of memory wastage.
embedding: Embedding layer is the integrated part of neural network, which converts integers into real valued vectors
This vectors are defined according to the meaning of the word. So one can recognize the word similarity either by comparing the vectors or by calculating the euclidean distance.

Recurrent Neural Networks(RNN): One of the simplest member of the family of neural network. First the parent is determined and then all its children are computed. The figure 7 depicts the formation of a RNN struture. In the tree , p_1 has its two children's vectors since both are words, the polarity of the child is transfered to the oarent node and the total polarity is computed, thereby producing the overall polarity of the statement

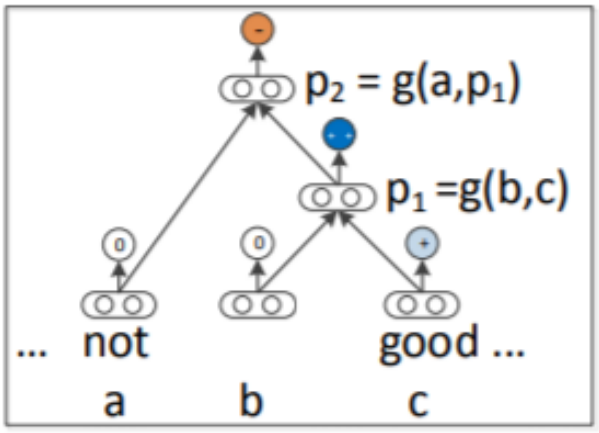

Figure 5: Recursive neural network model approach for sentiment:The parent vectors are computed in a bottom up fashion with a compositionality function and these node vectors are further used as features for a classifier at the node.[8]

Sigmoid:A non-linear logistic funxtion F(x), which gives an output in the range of 0 to 1. The number closer to zero is considered to have a negative polarity and the text that gives the result near one have positive polarity.

4 Experiment : Multimodal Approach

The ensemble technique in sentimental classification is used to predict the polarity of the review with maximum accuracy

4.1 Majority Theorem

Robustness, accuracy and generalization is enhanced by combining redundant and complementary classifiers. The aim of this is to predict the polarity of review with maximum accuracy. Naive Bayes, SVM and Neural networks are considered to predict the classification score because of their heterogeneous techniques in terms of strenghts and the philosophies. The results from these techniques are taken as base input for majority theorem and the maximum polarity is given as output. The benefits of this approach has shown improved accuracy in restaurant reviews classification.

5 Result Analysis

5.1 Biased Data Analysis

This approach when trained with biased data of 800 positive reviews and 200 negative reviews, the results were also biased towards positive side. When tested against biased positive data, this approach achieved 99% accuracy for the positive results, which looks far different from negative reviews which is inappropriate. Then the machine has been trained with 500 positive and 500 negative reviews and out results look promising with 60% accuracy. However, addition of more unbiased training data is made and the machine is trained with 12000+ reviews of which 8000 for training and 4000 for testing and our results gave 90-91% of accuracy for SVM classifiers which is very promising. 70-75% accuracy for Naive Bayes and 96% accuracy for RNN, which combinely gave 82% accuracy for the multimodal approach.

5.2 Accuracy calculation of Multimodal Approach

In the multi-modal approach, data/review is passed as input to all the classifiers which include Naive Based, SVM and Neural Network classifiers. Result of each classifier is taken as a array of data consisting of 1's and 0's. Now, calculate the percentage of 1's in the total data and result that as positive if the percent is more than 50%. This mean that more

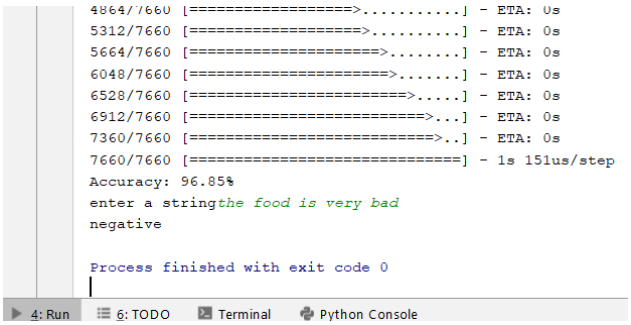

Figure 6: Output of Neural Network Approach

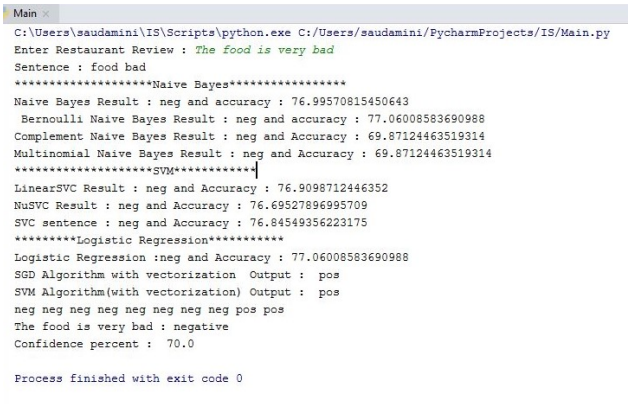

Figure 7: Output of Multi-modal Approach

classifiers classified the review as positive review and vice-versa. The multi-modal approach produced 82% of accuracy which is more promising when compared to the output of each classifier individually.

6 Improvisations to be made in our Approach

The proposed model filters the stop words and considers only adjectives, more parts of speeches can be added with the improvement. Negative words are not handled in this approach(only RNN handles it) which may sometimes change the meaning of the actual sentence. This is the most important feature that needs to be added. Also, the implementation of lexicon based approach i.e based on the meaning of the word can be added. To increase the accuracy of the algorithm, more data is required. The current accuracy of the multi modal algorithm is 82 %(for 200 reviews).

7 Future Scope

What could be done with the sentiment analysis we have done on review? The information about a restuarant could be collected along with the variety of food items available in it. The item names are considered as a key word. Analyse the review and find the sentiment of the review. Check for the item name in the review along with sentiment of the review. If the sentiment of the review with the item name in it is positive, rate that item with 1. If the sentiment of the review with the item name in it is negative, rate that item with 0. Similar work will be done on all the reviews. Collect all the 1's and 0's for a particulat item and find the percentage of 1's and rate the item accordingly in the restaurant.

Similarly, the sentiment analysis we implemented can be done can be done on any reviews like reviews on the product of Amazon, eBay. It Can also be implemented on reviews of locations in websites like TripAdvisor,goibibo etc.

8 Conclusion

Our project implements individual modules of the Naive Bayes and its three variants, Support vector machines with SVC and SGD Classifier and the Recurrent Neural Network successfully. The Naive Bayes and its variants shows the accuracy between 70-75%. The SVC and SGD gives the accuracy of 90-91%. While the RNN model with the accuracy of 96%. Moreover, we combined all these approaches in a majority based multimodal approach with the accuracy of 80-82%. By conducting more research on classifiers and taking appropriate classifiers for calculating the results in multi-modal approach the accuracy can be improved upto 93%.

References

[1] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), Vancouver.*, pp. 347–354, October 2005.

[2] O. Kolchyna, T. T. P. Souza, P. C. Treleaven, and T. Aste, "Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination," *Available online http://arxiv.org/abs/1507.00955*, 2015.

[3] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG!," *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[4] J. Kreutzer and N. Witte, "Opinion Mining Using SentiWordNet," 2015.

[5] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval Vol. 2, No 1-2 (2008)*, 2008.

[6] S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach,"

[7] "www.github.com."

[8] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank."