Restaurant Review System

J COMPONENT PROJECT REPORT

Submitted by

Gaurav Kumar Singh (16BCE0242)

Pushan Srivastava (16BCB0097)

in partial fulfillment for the award of the degree of

B. Tech

in

Computer Science and Engineering

Vellore-632014, Tamil Nadu, India

School of Computer Science and Engineering

October, 2019

Abstract

During the last few years, reviews have become crucial to the success of a restaurant, as every restaurant owner is aware of the fact that good reviews can boost popularity and profitability, whereas terrible reviews even have the potential of closing businesses down. In our project we aim to make a restaurant review system with the help of which customers can view restaurant ratings based on the reviews given by other customers. Customers can also view top restaurants according t a particular cuisine and submit their own review or feedback via voice or text. The reviews given by the customers are classified into positive, neutral and negative reviews using our novel approach in text classification. Text classification is the process of categorizing text into organized groups. By using Natural Language Processing (NLP), text classifiers can automatically analyze text and then assign a set of pre-defined tags or categories based on its content. The literature review done in this regard consists of different techniques and development in the field of text classification and other relevant fields regarding our project. It includes new algorithms which are more flexible and efficient than the traditional approaches. It also includes customization in existing approaches to make them more efficient as well as comparisons between various techniques which can for tell us which technique to use. Text classification is becoming an increasingly important part of businesses as it allows to easily get insights from data and automate business processes.

Keywords: Text Classification, Natural Language Processing, Reviews.

Introduction

The Internet today plays a big part in everybody's life compared to 4 or 5 years ago and many factors are responsible for it, for example, an increasing amount of people are now having reliable connectivity, an increasing amount of information on the internet, etc. People for finding information or knowledge about anything search on the internet before asking their family or friends, the people now believe more on the internet than their friends or family. Food is one of the things humans cannot live without and we often go outside to a restaurant to eat. Survey reveals that around 91 percent of people regularly or occasionally read online reviews, and 84 percent trust online reviews as much as a personal recommendation. And they make that decision quickly: 68 percent form an opinion after reading between one and six online reviews. According to another survey done in the USA, it was revealed that about 31.5% of people have never written a review in 2016, 53.3% less than 6 in 2016 and 15.2% more than 6 in 2016 and the reason given by the customer was the review takes lots of their valuable time.

Over the past few years, reviews were written only way to write the review was in text form. It is known that verbal communication is the most appropriate modem of passing on and conceiving the correct information, avoiding misquotations. To fulfill the gap over a long distance, verbal communication can take place easily on phone calls. A path-breaking innovation took place in the SMS technology using speech recognition technology, where voice messages are being converted to text messages. Quite a few applications used to assist the disabled make use of Text to Speech, Speech to Text, and translation. They can also be used for other applications, taking an example: Siri an intelligent automated assistant implemented on an electronic device, to facilitate user interaction with a device, and to help the user more effectively engage with local and/or remote services makes use of Nuance Communications voice recognition and text-to-speech (TTS) technology. Speech recognition is a technology that converts the audio to text.

Text classification problems have been widely studied and addressed in many real applications. Especially with recent breakthroughs in Natural Language Processing (NLP) and text mining, many researchers are now interested in developing applications that leverage text classification methods. Most text classification and document categorization systems can be deconstructed into the following four phases: Feature extraction, dimension reductions, classifier selection, and evaluations. Here we are developing a system that can capture the audio and rate the restaurant according to the audio given by the people. The people can then find the top restaurant according to cuisine and also find the rating of a particular restaurant. The rating can be given by converting the audio to text and then performing text classification.

Background Study

A classification scheme is the product of arranging things into kinds of things or into groups of classes. Our project relates to classifiers of two types, text classifiers and voice-based classifiers. Classifiers classify data into categories based on training data and any particular classifier-based algorithm among numerous classifier algorithms. One example of a text-based classification framework which uses the point of view of Support Vector Machine (SVM) along with the Euclidian distance function for the training period [1]. Another approach is using SVM along with membership function [2], which helps classify multi-label documents. Another approach to classify text-based documents revolves around keyword extraction [3] which uses a robust algorithm for the training phase based on all possible mixture of keywords. Another approach, which uses a semi-supervised practice, is using the web as a corpus for self-training text classification [4]. It extracts data from the web and uses it as training data, hence labelled data is not a prerequisite for this approach. Another approach is a class-dependent projection-based technique for text classification [5]. This technique is done so by projecting onto a lot of individual subspaces, the examples having a place with various document classes are isolated to such an extent that they are effectively to be classified. This is accomplished using supervised feature weighting algorithm to become familiar with the improved subspaces for all the document classes. Another approach is using and improved k-nearest-neighbors algorithm based on eager learning for text classification [6] to overcome to lazy learning of standard k-nearest-neighbor algorithm. Another approach is to use Rocchio algorithm, or an improved Rocchio algorithm to classify text documents [7]. The improved Rocchio algorithm [7] is meant for Chinese text classification done based on Rocchio algorithm. There is another approach of using an improved k-nearest-neighbors algorithm for text classification [8] which fabricates the classification model by constrained one pass clustering algorithm and k-nearest-neighbors text classification.

From the literature review we did, we find that the most common methods used to implement text classification and analysis are Support Vector Machine, K-Nearest Neighbors and some other specific text classification techniques based on NLP, Pattern Recognition or Neural Networks, out of which Support Vector Machine being the most used one. Support Vector Machine also had a lot of different ways in which it was implemented like SVM using Euclidian Distance, SVM using Membership Function and Multiple Binary class SVM. K-Nearest Neighbor was used mostly for small scale projects and analysis as it was not computationally efficient enough to be able to scale up and deliver fast results on larger datasets. For this limitation a more efficient KNN algorithm has also been developed which builds the classification model by combining constrained one pass clustering algorithm and KNN text categorization. Other than these there were also some other algorithms used for text classification like the Rocchio algorithm which was used for the purpose of classifying Chinese language, it can also be used for regular English text classification as well. Another way was by using the world wide web as a corpus for self-training and then categorizing

text based on that. Naïve Bayes was also used for text classification and produced good accuracy results and was even able to outclass SVM under the condition when external enriching is done through an external knowledge base. Natural Language Processing methods made the use of the nltk library in Python for the purpose of identifying and classifying text.

Methodology

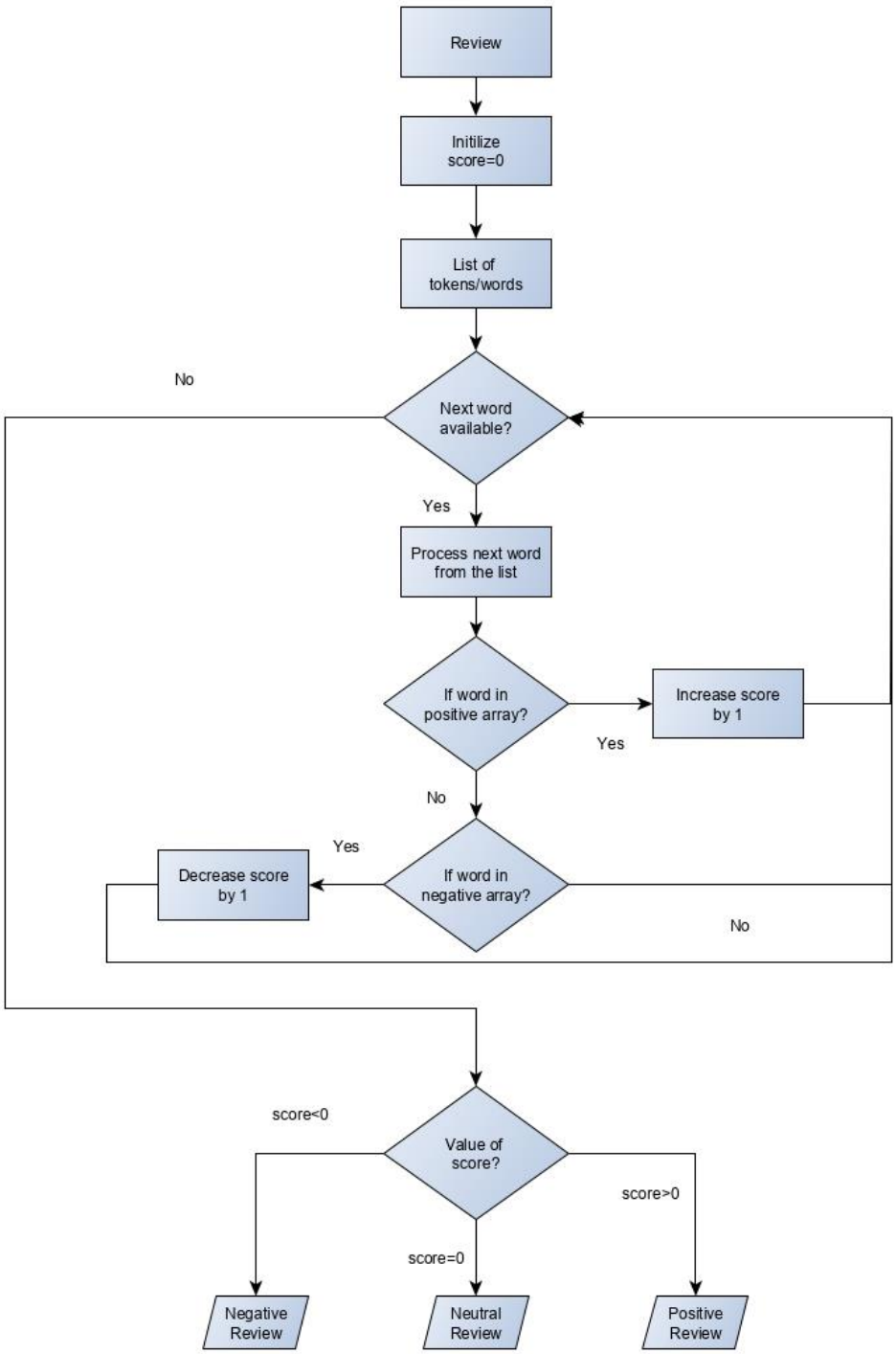
For our project we drew idea from the way all of these above-mentioned methods implement themselves to classify text and made an efficient and simpler algorithm for classifying and analyzing the restaurant review text. We focused more on the natural language aspects of our algorithm rather than diving deep into machine learning and prediction methods of classification. This helped us achieve faster response times compared to the training and predicting procedure of machine learning algorithms.

Hence for the purpose of classifying the text as positive, negative or neutral, we defined 2 arrays containing tokens which are positive and negative respectively. As our project focusses on restaurant review classification it was easier to narrow down the tokens that each array contains based on the terms that would commonly occur in a restaurant review. The review text is tokenized and then then tokens are lemmatized using the nltk library in python to get the synonyms of the tokens which can help in more efficient search of the sentiment behind the review. All of these tokens are then checked for presence in both the arrays, if a token is present in the positive array then a positive score is given else if it is present in the negative array a negative score is given. At the end all the scores are added and if the final score is positive then the review is stated as a positive review, if the score is in negative then the review is stated as a negative review else if it is equal to 0 then the review is stated as a neutral review. This is a simple NLP based algorithm used for the classification of the restaurant reviews.

In speech to text, voice is analyzed as parts of the speech and then using a specific algorithm/model it is converted it into text. This requires a very versatile model which has been trained on a big dataset of voices containing many different types of voices and accents to provide better accuracy. Hence to implement speech to text, many papers used different APIs which converted the voice into text. Among all the APIs that are there the Google Speech Recognition API had the best results as it has been trained on a very vast dataset and hence was able to deliver good accuracies. It is also available for free unlike some other APIs developed by other companies. Hence in the project also we use the Google Speech Recognition API for the purpose of converting the restaurant review spoken by the user into text which is then stored for further processing.

The following diagram describes the classification approach which is implemented in our project. The positive array and negative array in the diagram below are:

- **positive_words:** ['perfect', 'great', 'good', 'tasty', 'friendly', 'spectacular', 'awesome', 'delicious', 'yummy', 'best', 'soothing', 'juicy']
- **negative_words:** ['bad', 'tasteless', 'sad', 'mild', 'foul']



Results and Discussion

On executing the python program, the first screen encountered is the home page. The Home Page can be seen in Figure 1 below. As clearly visible it offers 3 main options to the user which are:

- View Ratings of a restaurant
- View top restaurants according to a specified cuisine
- Submit feedback or review for a particular restaurant

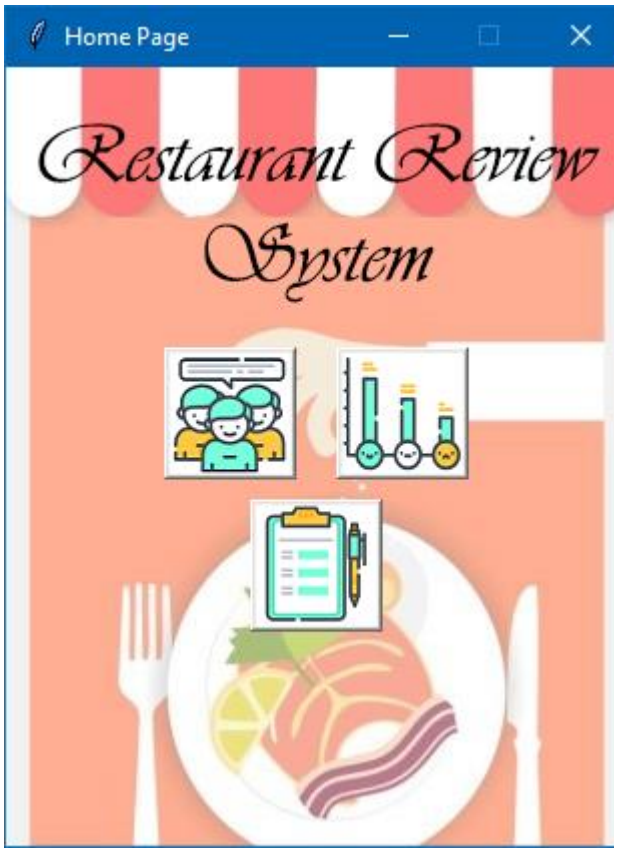


Figure 1: Home Page

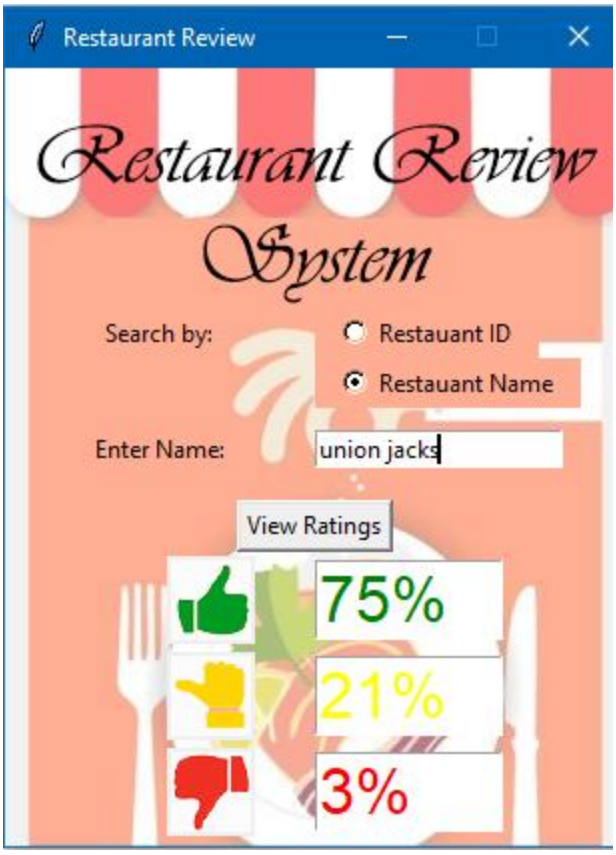


Figure 2: View Ratings Page

The first button takes the user to another screen of the page which View Ratings Page seen above in Figure 2. This page allows the user to view ratings of restaurant. For this purpose, user can search the restaurant with its name or distinct id by choosing the appropriate radio button of his choosing. Then the user enters the id or name based on the option he chose earlier. After entering the value, the user is supposed to click View Ratings button which will give the approx. percentage of the positive reviews, neutral reviews and

negative reviews beside green, yellow and red thumb pic with appropriate meaning respectively. This allows the user to decide whether the restaurant is good enough to eat or not.

Figure 3: Cuisine Selector Page

Figure 4: Submit a Review Page

On clicking the second button on the Home Page, the user is directed to the Cuisine Selector Page which is shown in Figure 3. This page allows the user to view top restaurants serving a type of cuisine which is specified by the user. For this purpose, the user chooses the cusine he is interested in from the combobox provided beside 'Choose Your Cuisine' label. After choosing a cuisine, the user is supposed to click the View Top Restaurants which will give a list of top 10 restaurants in decreasing order of positive reviews for the cuisine specified.

On clicking the third button on the Home Page, the user is directed to the Submit a review page which is shown in Figure 4. This page allows the user to submit his or her own feedback or review about a restaurant and its services. Since it is sensitive situation, the user has to enter restaurant id for submiting his/her review. The user specifies the review title and then proceed to the write his detailed review in the text area provided below. Here the user is also given the option to give his review via speech input. This is achieved by clicking the green microphone button below the textarea. After clicking the microphone button the user is supposed to speak and after the timeout, the speech is converted into text and written into the text area. The user can

then submit his/her review by clicking ‘Submit review Button’ and this review will be stored into the dataset. The screen shot of the dataset is provided below to show that the review has been appended to the dataset and will be now used to in calculating rating of the particular restaurant.

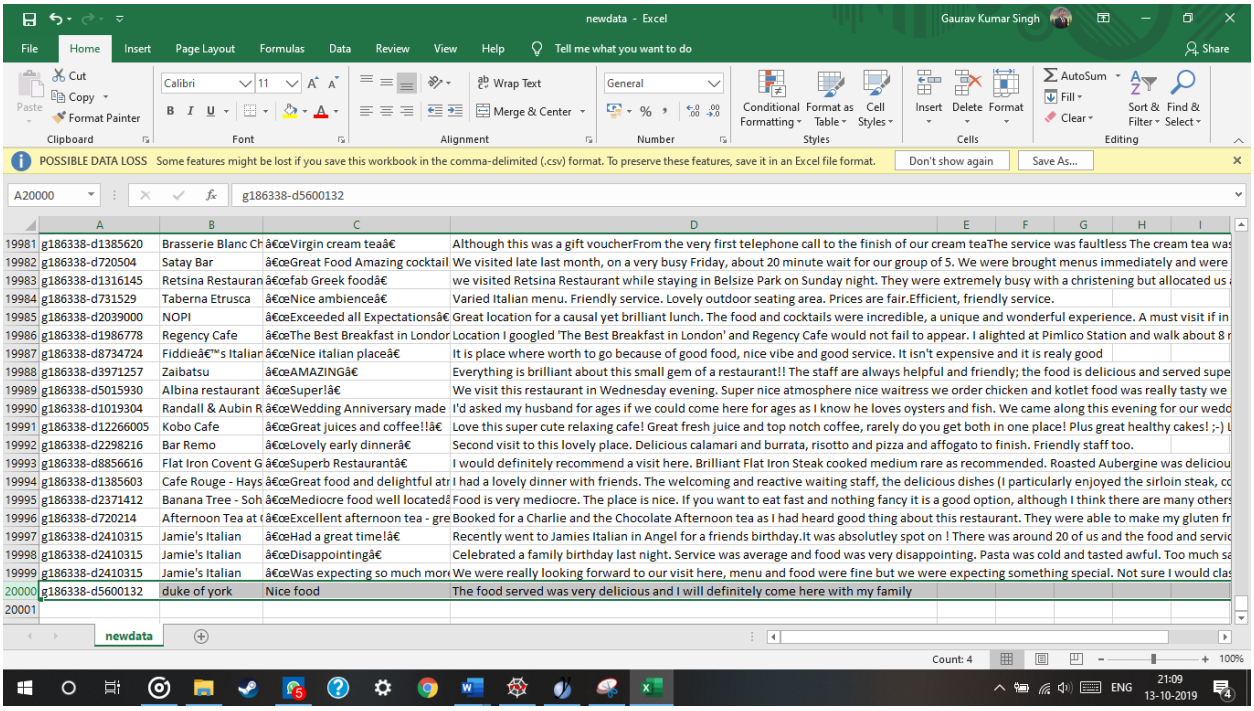


Figure 5: The review appended to the dataset after submitting the review from Figure 4

Conclusion

This project of “Restaurant Review System” made use of the natural language techniques of implementing sentiment analysis. It used a positive and negative array and checked for presence of each work in both of them to calculate the score. The project’s implementation was done completely in python using several libraries like nltk etc. The GUI is made using Tkinter and Google speech recognition used for the purpose of speech to text.

The literature review done showed us the various approaches that are present for the purpose of classifying and analyzing text. These approaches ranged from Machine Learning algorithms to simple Natural Language Processing techniques. The research showed that NLP techniques tend to be faster than ML algorithms in certain scenarios, but won’t be very efficient for complex ones.

Overall to conclude it was a successful project which was implemented well and provided us with the opportunity to learn about the applications of natural language processing.

References

[1] L. Lee, C. Wan, R. Rajkumar and D. Isa, "An enhanced Support Vector Machine classification framework by using Euclidean distance function for text document categorization", Applied Intelligence, vol. 37, no. 1, pp. 80-99, 2011.

[2] T. Wang and H. Chiang, "Solving multi-label text categorization problem using support vector machine approach with membership function", Neurocomputing, vol. 74, no. 17, pp. 3682-3689, 2011.

[3] Jiyuan An and Yi-Ping Phoebe Chen, "Keyword extraction for text categorization", Proceedings of the 2005 International Conference on Active Media Technology, 2005. (AMT 2005).

[4] R. Guzmán-Cabrera, M. Montes-y-Gómez, P. Rosso and L. Villaseñor-Pineda, "Using the Web as corpus for self-training text categorization", Information Retrieval, vol. 12, no. 3, pp. 400-415, 2008.

[5] J. Zhang, L. Chen and G. Guo, "Projected-prototype based classifier for text categorization", Knowledge-Based Systems, vol. 49, pp. 179-189, 2013.

[6] T. Dong, W. Cheng and W. Shang, "The Research of kNN Text Categorization Algorithm Based on Eager Learning", 2012 International Conference on Industrial Control and Electronics Engineering, 2012.

[7] G. Gao and S. Guan, "Text categorization based on improved Rocchio algorithm", 2012 International Conference on Systems and Informatics (ICSAI2012), 2012.

[8] S. Jiang, G. Pang, M. Wu and L. Kuang, "An improved K-nearest-neighbor algorithm for text categorization", Expert Systems with Applications, vol. 39, no. 1, pp. 1503-1509, 2012.

[9] M. Kumar and M. Gopal, "A comparison study on multiple binary-class SVM methods for unilabel text categorization", Pattern Recognition Letters, vol. 31, no. 11, pp. 1437-1444, 2010.

[10] S. Hassan, M. Rafi and M. Shaikh, "Comparing SVM and Naïve Bayes classifiers for text categorization with Wikitology as knowledge enrichment", 2011 IEEE 14th International Multitopic Conference, 2011.