EPICURE SENTIMENT ANALYSIS

R. ARUNA FLARENCE
Assistant Professor, Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and
Technology, Bachupally, Nizampet, Hyderabad-500070
aruna1202@gmail.com

K. ANUSHA
Assistant Professor, Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and
Technology, Bachupally, Nizampet, Hyderabad-500070
anusha.kollu87@gmail.com

Abstract: - This paper proposes a system for extract and summarize sentiments given by various customers on restaurant services. The reviews are collected manually and internet. The Summarization can help marketers evaluate the success and new find item launch, determine which food items are popular and identify the demographics like or dislikes. This system helps in improving the performance of restaurants with their respective cuisines through their websites. This system classifies the review data into positive, negative and neutral from the preclassified data set.

Key-Words: - Sentiment, customer reviews, summarization, classification

1 Introduction

There are many popular food websites providing comments and reviews given by different customers for food courts. Users with the same interests forming the community to help each other in sharing, searching, advertising, and decision making. In addition, members in food communities can comment on food and services provided by different food courts and restaurants and exchange their experiences [1]. Some comments agree that dishes taste good while some comments disagree and give the information to improve the recipe of dishes and services. Therefore, these user comments about food items from other persons are valuable resources to help members make a decision and choose the food item from that particular restaurant.

Although there is a star rating for food items on popular food websites, the rating may not be reliable because members of the community can vote the food items by giving scores without the practical preference consistency. Moreover, the preference rating summarizing from all food items comments will be more trustworthy than those of star rating [2]. Therefore, if there is the system which can automatically analyze information from all user comments about food items, the summary of score rating and the classification of comment groups are the valuable information. The benefit of an analysis system is also shown on studying consumer behaviour using sentiment analysis [3].

What is an Opinion? "What people think?"
- Which Restaurant should I go to?
- Which cuisine do I need to order?
- Which Service do I need to use?

It is a personal belief or judgment that is not founded on proof or certainty. Sentiment analysis is the process of determining whether a piece of comment is strong, *moderate or neutral* .It is also known as opinion mining deriving the opinion or attitude of a speaker [4].

Who is an epicure?
A person who takes particular pleasure in fine food and drink. Opinions or comments from other people are core factors of the consuming manner and actions because most people often seek out the opinions of others before they make the decision to choose the right things that they want. In addition, consumers or users always post reviews of services or give comments about food items which express their opinions and exchange personal experiences about them on the internet, such as reviews, blogs, and forum discussion in the online communities [5]. Furthermore, the industry always want to find consumer opinions about their quality and services.

The human system will have difficulty identifying relevant comments and accurately summarizing the information and opinions contained in them. Most of the cases human analysis of text information has limitations because users pay attention to comments that are defined with their own favourites. Additionally, users can input a sentiment target as a query (e.g. topics, subjects or products), and search for positive or negative sentiments towards the target. Therefore, it is also widely accepted that extracting sentiments from text is a hard semantic problem even for human beings [6].

Moreover, Sentiment Analysis of Food items analysis is still domain specific because the polarity of some terms depends on the context in which they

are used. For example, the word small in the mobile devices is the positive feature, while this word is the negative polarity in the agricultural products, such as fruits. There is the relationship between the context of text and the sentiments of text, thus the subject dependent sentiment analysis is more informative and more useful than the subject independent analysis [7].

2. Method

In the olden days analysis is performed using the traditional approaches and they are developed using coding techniques for collecting the data from the food based websites by using codes. For those they are going to download the libraries that are provided by food websites from this they are creeping the data that want mainly. After getting the raw data they will filter by using old techniques like Word Net.

Word net is a lexical database organized by meanings and it is developed at Princeton University and also they will find out the positive, negative words from the list of collected words. All these words should be poised by us to filter out for sentiment analysis. These words can be termed as lexicon sets by which they will achieve soppiness study, after performing all these things they want to store in database. Creating data base will have limitations .After eliminating the restrictions data can be retrieved effectively, so to improve this we are creating our own proposed system.

By knowing the downsides, here we are going to dazed them by solving this issue by proposing a system .This research proposed the automated sentiment analysis of food reviews or comments' using text analytics. The aim of this analysis is to classify' food related' comments into three groups that are impartial, confident and adverse groups by detecting positive and negative words in the foodstuff domain. Opinion are classified into two main categories they are 1) the classification based on supervised learning using the machine learning; 2) The classification based on unsupervised learning with the semantic orientation approach. Sentiment classification with a supervised learning uses the training data to learn the classification model for determining the testing data into three classes: neutral, positive or negative [8].

Any existing supervised learning methods can be applied to sentiment classification, such as decision tree classifier, naive Bayesian classification, and support vector machines (SVMs).But in our approach we have used naive Bayesian classification. Finally, based on the dictionary assignment of score, and referring to SentiWordNet, the proposed system interprets whether the review is positive, negative or neutral. So to improve our analysis, we have used SentiWordNet.

Sentiwordnet is a vocabulary resource for opinion mining. Sentiwordnet allocates to each synset of word net three sentiment scores: positive, negative and neutral. The main objectives of the sentiment analysis with the semantic oriented method are to measure and to classify the subjectivity and opinion in text [9].

The proposed sentiment analysis of user comments about food comments in this research is the improved methods for obtaining the higher performance of analysis capturing evaluative factors and potency or strength towards subject topics, or ideas. In addition, the aggregation of sentiment for each entity and certain lexicons with sentiment words are very informative and efficient. Therefore, the proposed sentiment analysis of food comments using the semantic orientation approach is the intensive analysis of words and their meaning about foods. In addition, the summary information of the sentiments or opinions of the software implementing this sentiment analysis is adequate to satisfy the members of the food community [10].

The analysis processes are composed of pre-processing, detecting polarity words, calculating polarity scores of sentences and comments. This technique analyses recipe comment messages by words' information from the polarity lexicon to detect polarity words for the sentiment classification.

Sentiment analysis on food reviews data is based on two important parts viz Data Extraction, pre-processing of extracted data and classification. The process we are using for this analysis is as follows and diagrammatically shown in Figure 1:

- **Collection of reviews: We** collect reviews from different cuisine based sites**.**
- **Pre-processing Data**:We remove all the noisy, inconsistent and incomplete data
- **Feature Extraction:**We collect different types of parts of speech from the data.
- **Feature Selection:**We select the adjectives from the extracted data**.**
- **Detecting the polarity: We** calculate the polarity of the comments**.**
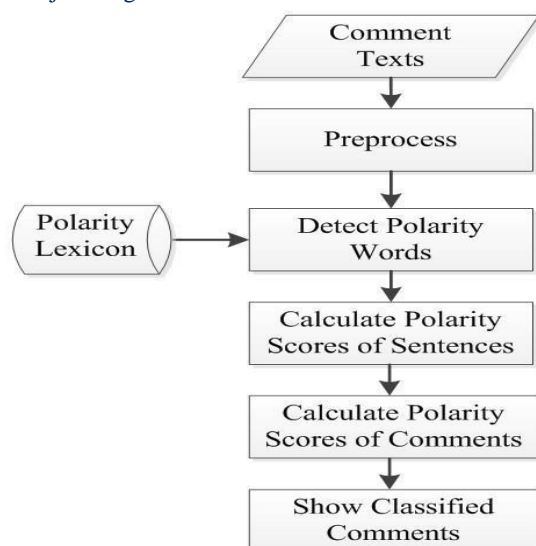
Fig.1: Steps in Sentiment Analysis

Finally, all words in all sentences of the user comment are separated into individual words in the final step of pre-processing, using space between two words and some punctuations, for example comma(,) and underscore(_).

After the pre-processing, the word sequences of all sentences in the user comment are collected and some words are handled by syntax to provide knowledge for the next process. For example, some words which are usually used in abbreviated forms in text messages and the meaning is (not) are labelled as words presenting the opposite meaning of sentiments. Some words with their common abbreviated forms are shown in table.

Table 1
THE WORDS IN ABBREVIATED FORMS

don't(do not)	didn't(did not)
isn't(is not)	aren't(are not)
wasn't(was not)	shouldn't(should not)
wouldn't(would not)	couldn't(could not)

Delicious! I used fresh skinless, boneless chicken breasts and olive oil instead of melted butter. Chicken was moist and tasty! Thanks for the great recipe! The output of the pre-processing process for the comment is described as follows [11]: All letters are transformed into lowercase letters and this comment consists of four sentences which are divided by the exclamation point (!) and the full stop (.).

Delicious, I used fresh skinless, boneless chicken breasts and olive oil instead of melted butter, chicken was moist and tasty, thanks for the great recipe. Then, all individual words are separated by the space and the comma (,).The word sequences of

all sentences are ordered. Another recipe's comment input (Comment 2): I was very excited to try this recipe, but I was so disappointed at the outcome. It didn't taste as good as all the reviews made it out to be. The output of the pre-processing process for the comment is explained as follows: All letters are converted into letters in lowercase and this comment consists of two sentences which are divided by a full stop (.).I was very excited to try this recipe, but I was so disappointed at the outcome. It didn't taste as good as all the reviews made it out to be. Then, all individual words are separated by the space and the comma (,). The word with opposite meanings (didn't) in the abbreviated form is labelled.

2.1 Detecting Polarity Words

To detect polarity words, this research creates the new polarity lexicon for the good domain based on the **SentiWordNet**.SentiWordNet performs the following operations internally:
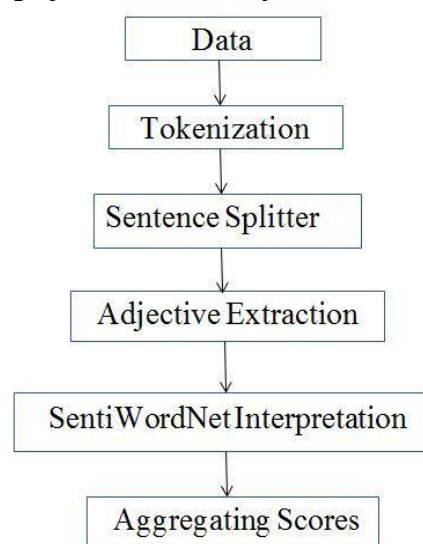

Fig.2. Detecting Polarity

Many words from many comments about food items from different food based websites are collected to filter subjectivity. For identifying the sentiment for all the words given in the comment we use Sentiwordnet.

What SentiWord does is that it takes a word and also the POS of that word has in a given sentence. Using the combination of POS and the word itself SentiWord contributes it a numeric score between -1 and 1 where lower value refers to negative sentiment and higher value refers to positive sentiment. As a observation or a criticism consists of a few words we can take the SentiWord score for each of those words and then

IJECRT- International Journal of Engineering Computational Research and Technology, Volume 2, Issue 1, December 2017
ISSN (Online): 2456-9852
www.ijecrt.org

sum them up to get a numeric mark for each observation.

Table 2
SENTIMENT SCORE RANGE RATING

Sentiment score range	Assigned sentiment
Mark<=-0.5	Worst
-0.5<Mark<=0	Bad
Mark=0	Neutral
0<Mark<=0.5	Good
Mark>=0.5	Excellent

The subjectivity words in the created polarity lexicon are assigned the polarity to be positive or negative using PosScore and NegScore of words from the SentiWordNet

In addition, some subjectivity words are inserted into the polarity lexicon, while some existing words are reassigned the positivity or the negativity manually after considering many comment messages about food recipes. However, polarity words in the lexicon are reviewed by the expert to identify polarity scores (positive or negative). Therefore, our created polarity lexicon is suitable for text analytics in the food domain because of focusing on words from this domain.

Table 3
THE WORDS WITH POSITIVE SCORES

amazing	awesome	beautiful
best	better	delicious
nice	perfect	tasty

Table 4
THE WORDS WITH NEGATIVE SCORES.

awful	bad	disappoint
hate	mess	problem
salty	terrible	worse

However, there is no process for word stemming in the proposed sentiment analysis. Stemming is to reduce words to their base forms or stems. For example, 'agree' is the stem or the base form of the words 'agrees', 'agreed', and 'agreeable'. Therefore, the generated polarity lexicon contains words in all different forms as shown in table

Table 5
THE WORDS WITH DIFFERENT FORMS

best	better
flavour	flavourful
worry	worried

Furthermore, words with the opposite meaning when interpreting with other words, i.e. not and never, are marked as words representing the reverse meaning of sentiments. In conclusion, the individual words of all sentences in recipes' comments are compared to polarity words in our polarity lexicon. The words found in the polarity lexicon are detected and are labelled with the polarity. The sequence of words in the sentence is also used to interpret the meaning of sentiments. After this process, the subjectivity words or polarity words in the sentence are detected and tagged the polarity scores.

Delicious I used fresh skinless, boneless chicken breasts and olive oil instead of melted butter chicken was moist and tasty thanks for the great recipe. All these polarity words (delicious, moist, tasty, and great) also have sentiment scores more than zero. Delicious (+) I used fresh skinless, boneless chicken breasts and olive oil instead of melted butter chicken was moist (+) and tasty (+) thanks for the great (+) recipe.

Comment 2 contains two detected polarity words that are disappointed and good. I was very excited to try this recipe, but I was so disappointed at the outcome. It didn't taste as good as all the reviews made it out to be. The first sentiment word disappointed has the negative polarity score (less than zero), while the second word good has the sentiment score more than zero and also the negative verb didn't is marked as follows.

I was very excited to try this recipe, but I was so disappointed (-) at the outcome. It didn't taste as good (+) as all the reviews made it out to be.

2.2 Calculating Polarity Scores

The calculating polarity score process is composed of two steps that are calculating polarity scores of thecomment.In calculating polarity scores of the comment, the summation of all polarity word scores in each sentence is calculated. Then, the polarity scores of the sentence are defined by the result of the summation. Unfortunately, some words, presenting opposite meaning or representing reverse meaning when are interpreted with other words expressing sentiments, occur in the comment.

Therefore, the polarity word scores of these sentiment words may change into opposite values that are positive to negative (more than zero changed to less than zero) or negative to positive (less than zero changed to more than zero). The previous situations depend on the sequence of words.To calculate polarity scores of the comment,

IJECRT- International Journal of Engineering Computational Research and Technology, Volume 2, Issue 1, December 2017
ISSN (Online): 2456-9852
www.ijecrt.org

the summation of polarity scores of all sentences in the comment are calculated. If the polarity scores of the comment are more than zero, these comments are classified to positive comments. On the other hand, comments are classified into negative groups, when the summation of sentences' polarity scores less than zero. If the summation of the scores is equal zero, comments are identified as neutral comments [12].

According to the comment examples in all individual sentences contain at most one polarity word, so the polarity scores of each sentence equal the polarity score of the word found in the sentence. Consequently, the polarity scores of the first, the third and the fourth sentence of Comment 1 are more than zero, while the second sentence has zero polarity score [13]. The details are displayed as follows. Delicious (+) + I used fresh skinless, boneless chicken breasts and olive oil instead of melted butter 0 chicken was moist (+) and tasty (+) +thanks for the great (+) recipe.

3 Results

The data set contains 900 reviews posted by various customers .Reviews has been manually classified in positive and negative. In data set there are 300 positive, 300 negative and 300 neutral reviews. The following table shows the prediction of reviews.

Table 6
POLARITY OF REVIEWS

	Actual Reviews	Predicted Reviews	Others
Positive	300	255	45
Negative	300	168	132
Neutral	300	132	168

System Classification performance was evaluated using precision, recall and accuracy [14]. The accuracy rate of neutral, positive and negative classification is calculated by

% Neutral Accuracy =correct number of neutral comments ×100/ number of actual neutral comments.

% Positive Accuracy =correct number of positive comments ×100 /number of actual positive comments.

%Negative Accuracy=correct number of negative comments ×100/ number of actual negative comment.

%Neutral Precision = correct neutral comments ×100/number of predicted neutral comments.

%Positive Precision=correct positive comments ×100/number of predicted positive comments.

% Negative Precision = the number of correct negative comments ×100/ the number of predicted negative comments.

The following figures 3, 4 & 5 shows the predicted reviews, accuracy and precision.
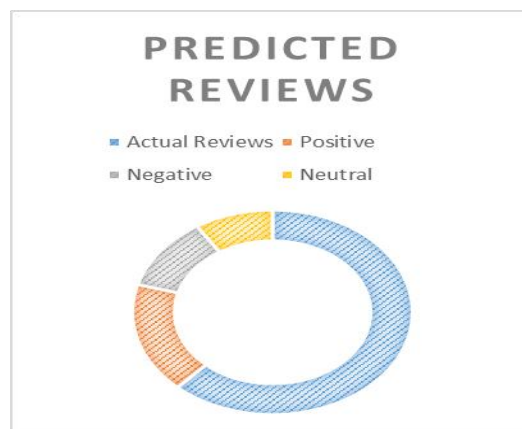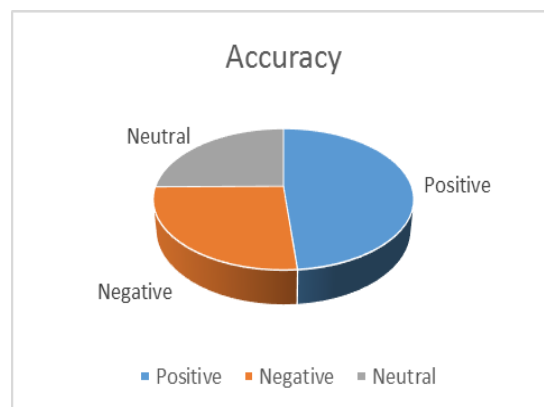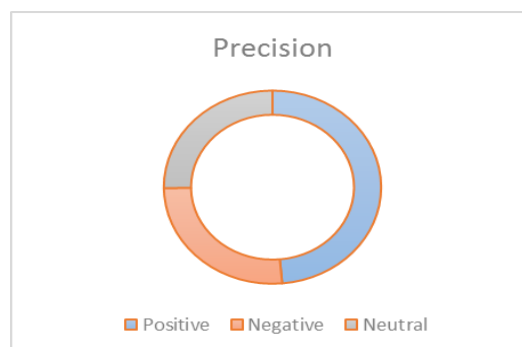

Fig.3. Predicted Reviews

Fig.4. Accuracy

Fig.5. Precision

4 Conclusion and Future Scope

According to the results showed that these analysis techniques are very beneficial for users, consumers and product providers in different domains. Therefore, the comment analysis of food domain using the sentiment analysis can generate the new knowledge and summarize the valuable information

about food reviews for users. It provides the ability to act on the customer suggestions. It identifies the Strengths, Weaknesses. It provides an insightful and accurate customer perceptions and feedback. This helps the restaurants in improvising their service and innovating new ideas for customer satisfaction. Also, new customers are attracted towards them. Thereby, a communication is established between the customers and the restaurants and business effectively increases.

In the future work, the personal profiles of people who comment the recipes, e.g. nationality and age, will be collected to analyse comments by the groups of people visiting restaurants. Rewards will be given to the best reviews which have helped more number of customers and the one whose comments have been accepted by the public.

References:

[1] Potgieter, Marius, Johan W. de Jager, and Neels H. van Heerden. An innovative marketing information system: A management tool for South African tour operators. *Procedia-Social and Behavioral Sciences* 2013; 99: 733-741.

[2] Raut VB, Londhe DD. Survey on opinion mining and summarization of user reviews on web. *International Journal of Computer Science and Information Technologies*. 2014; 5(2): 1026-30.

[3] Bansal P, Ahmad T. Methods and Techniques of Intrusion Detection: A Review. *In International Conference on Smart Trends for Information Technology and Computer Communications, Springer, Singapore* 2016; 518-529.

[4] Xiang Z, Schwartz Z, Gerdes JH, Uysal M. What can big data and text analytics tell us about hotel guest experience and satisfaction? . *International Journal of Hospitality Management* 2015; 44: 120-130.

[5] Zhang W, Wang J. Integrating Topic and Latent Factors for Scalable Personalized Review-based Rating Prediction. *IEEE Transactions on Knowledge and Data Engineering* 2016; 28(11): 3013-3027.

[6] Tsytsarau M, Palpanas T. Managing Diverse Sentiments at Large Scale. *IEEE Transactions on Knowledge and Data Engineering* 2016; 28(11): 3028-3040.

[7] Philander K, Zhong Y. Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management* 2016; 55(2016): 16-24.

[8] Hu HW, Chen YL, Hsu PT. A Novel Approach to Rate and Summarize Online Reviews According to User-Specified Aspects. *Journal of Electronic Commerce Research* 2016; 17(2):132.

[9] García-Pablos A, Cuadros M, Linaza MT. OpeNER: Open Tools to Perform Natural Language Processing on Accommodation Reviews. *InInformation and Communication Technologies in Tourism Springer, Cham.* 2015; 125-137.

[10] García-Pablos A, Cuadros M, Linaza MT. Automatic analysis of textual hotel reviews. *Information Technology and Tourism* 2016; 16(1): 45-69.

[11] A Study of Recommendation System Combining Social Information. *Master's Thesis 2015.*

[12] Dingding Wang, Shenghou Zhu, Tao Li. SumView: A Web-based engine for summarizing product reviews and customer opinions. *Expert Systems with Applications 2013; 40: 27–33.*

[13] Dim Nyaung DE, Thein TL. Feature-Based Summarizing and Ranking from Customer Reviews. *International Journal of Computer, Electrical, Automation, Control and Information Engineering* 2015; 9(3).