International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 22 (2017) pp. 12637-12644
© Research India Publications. http://www.ripublication.com

Semantic Analysis of Reviews Provided by Mobile Web Services Using Rule Based and Supervised Machine Learning Techniques

Roshan Fernandes
Department of Computer Science and Engineering,
Nitte Mahalinga Adyanthaya Memorial Institute of Technology (NMAM),
Nitte 574110, Udupi District, Karnataka, India.

Orcid id: 0000-0002-7625-3296

Dr. Rio D'Souza G. L.
Department of Computer Science and Engineering, St. Joseph Engineering College,
Vamanjoor 575028, Mangalore, Karnataka, India.

Abstract:

Most of the applications such as MakeMyTrip, Oyo and many others, attract customers by giving good offers for various location based mobile services. These services typically include hotels, flights, cabs, holiday plans and many more. The main objective of this research work is to automatically analyze the feedbacks given by the customers into positive, negative and neutral categories. The work also identifies the subject which is positive or negative and gives a summarized review in case of multiple sentences are present in the feedback. In this paper, we have considered the various sources of data; namely from MakeMyTrip and Twitter. The method to analyze the data include collecting the data from the mobile/web application sources, filtering the unwanted data, Parts of Speech tagging, preprocessing and finally analyzing and summarizing the reviews using rule based and supervised machine learning techniques.

Results show that when the reviews follow a particular pattern, rule based analysis technique gives better performance. But when the review sentences follow a random pattern, Naïve Bayesian classifier gives better performance as compared to rule based and Support Vector Machine techniques.

This paper collects the data from MakeMyTrip and Twitter and explores the rule based and supervised machine learning techniques to analyze and summarize the reviews. The proposed techniques gives reasonably good accuracy.

Keywords: POS Tagger, Semantic Analysis, Bayesian Classifier, Support Vector Machines, Parser, JSOUP Library

INTRODUCTION

Today, the role of social media applications is very crucial in every field; whether in business or any other area. The success of a business is determined through customer satisfaction. At present, almost all services are digitized and people prefer to use their smartphones or web for searching the best service in the area of his/her interest. To determine the quality of the service, customer feedback has become crucial. The feedback given by customers or critics are in the form of text that expresses their feelings about the service. The text feedback provided by the user or a critic is in unprocessed or unstructured format and it is in the language they use.

A framework for finding semantic Web Services by making use of natural language processing technique is defined. The proposed framework matches user query with a set of sematic Web services by using keywords. Three techniques, called part-of-speech tagging, lemmatization, and word sense disambiguation have been used. The article concluded that the three proposed methods can effectively perform matching and approximate matching[1]. The natural language processing technique (NLP) for opinion mining is discussed and the authors identified different NLP techniques required for text processing and later investigated different approaches for opinion mining. The authors also reviewed deep learning approaches for opinion mining and the challenges and open problems related to opinion mining are discussed[2]. Imran Sarwar Bajwa and Muhammad Abbas Choudhary[3] used natural language processing based automated system to understand speech language text. Their system analyses the text and extracts relative meaning from it. Their proposed system provides meaning with an accuracy of 90%. This system is limited to only Active-voice sentences and sentences in Passive-voice are yet to be considered. Mohammad AL-Smadi et al.[4] proposed a method for identifying paraphrase and analysis of Arabic news tweets. Several phases of text processing were carried out to solve the weaknesses and limitations of the present technologies in solving these tasks for the Arabic language. Two algorithms, namely Maximum Entropy and Support Vector Regression classifiers are used in the process. Their results shows that the proposed method performs better than baseline results. Ning Kang et al.[5], proposed a technique to perform dictionary-based concept normalization using Natural Language Processing system. A comparison of two biomedical concepts normalization systems namely MetaMap and Peregrine has been carried out with and without the use of rule based NLP module. Yang Zoua et al.[6] proposed a method to improve the efficiency and performance of risk case retrieval, by using two Natural Language Processing techniques called Vector Space Model and Semantic Query Expansion. The results showed that their system is capable of retrieving similar cases automatically for top 10 similar risk case. The two models used for NLP by Michael Tanana et al.[7] are discrete sentence features (DSF model) and more complex recursive neural networks (RNN model). The results showed that DSF model performed better than RNN model. The articles[8, 9, 11, 13] discuss the various approaches for opinion mining including neural network and fuzzy logic and their performance metrics. The articles[10, 12] discuss the general approaches for applying data semantics to the location based services.

The current work focuses on the semantic analysis of the feedback obtained for location based services. In the current paper, the review data is collected from location service provider mobile/web applications such as MakeMyTrip and Twitter. The analysis of the feedback is done using rule based semantic analysis and also by using machine learning techniques. The main idea behind this implementation work is to automate the analysis of feedback review given by various customers for the location based services used. Most of the times the customers have to manually read the reviews and then decide which service to choose depending on the positivity of the previous reviews. This paper considers the multiple sentences and finally concludes with the service name and its final opinion in terms of rating.

There are many challenges in automatically analyzing the feedback. They include the language, the way of framing the sentences, identifying the nature of service and many more. This implementation work focuses on two techniques to semantically analyze the content of a sentence. The first approach is rule based and the second approach is using machine learning techniques. In the rule based approach, a set of predefined grammar rules are designed and the sentence is parsed on these rules. Based on the rules, the opinion or feedback of the given sentence is analyzed. In the machine learning technique, the predefined rules are not considered but the prediction of the destination class of opinion of the reviews is determined using probability based approach. The supervised machine learning algorithms considered in this paper include Bayesian Classifier and Support Vector Machines. These two techniques are called supervised machine learning techniques because they need to be trained before a random variable (random sentence in our case) is classified using these techniques. The rule based and machine learning approaches are compared to consider how accurately the predictions are done.

The rest of the paper is organized as follows. Section 2 gives the detailed implementation methodology, section 3 discusses the results and discussion and section 4 highlights the conclusion and future scope.

METHODOLOGY

The semantic analysis of the location based services' feedback is carried out using two techniques. The first technique is Rule Based Analysis and the second approach is using Supervised Machine Learning techniques. Section 2.1 discusses the implementation of the semantic analysis using Rule Based technique and section 2.2 discusses the implementation details of Semantic Analysis using Supervised Machine Learning technique.

Rule based approach

The Rule Based approach uses predefined grammar rules for the sentences. The random sentence is parsed against these rules and then the semantic analysis is applied on this sentence. The various steps followed in the Rule Based approach are shown in Figure 1.

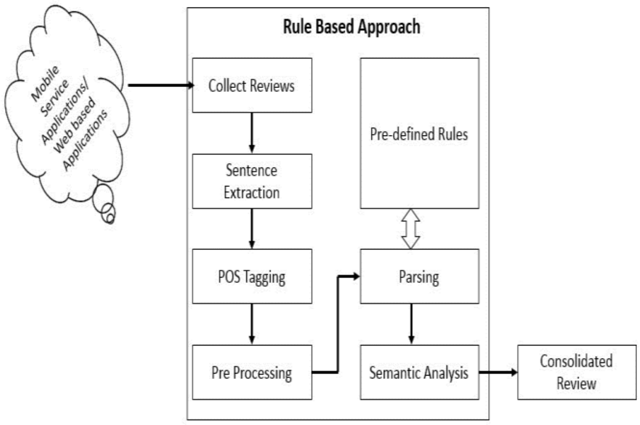

Figure 1: Rule Based Approach

Step 1 (Collect Reviews): In the first step, the reviews were collected from mobile service applications or from web based applications. Most of the reviews had multiple sentences. All the reviews were collected and stored for further processing. In this implementation work, the reviews were collected from the MakeMyTrip service application and Twitter social media. The reviews were collected by extracting the HTML contents from the web application. This extraction was implemented using JSOUP library in Java. The unwanted content was removed from the extracted content and only the reviews content was stored for a specific location based service as in hotel service. The Twitter data was collected by first registering to the Twitter application and then specifying the topic of interest. The Twitter data is normally in the JSON format. This format was analyzed and only the review content was stored and rest of the unwanted content was kept aside. At the end of Step 1, the review sentences by each customer is available and was stored in the local machine for further processing.

Step 2 (Sentence Extraction): In the second step, each review was processed. The reviews had multiple sentences. Each sentence was extracted by considering sentence delimiters. The delimiters considered in this implementation work were '**.**' and '**;**'. Once each sentence was extracted, it was passed to step 3 for the Parts of Speech tagging process. Step 3 through step 6 were repeated for each and every review sentence.

Step 3 (POS Tagging): This process tags the words of an English sentence according to Parts of Speech (POS). POS tagging categorises linguistically by breaking the sentence grammatically and each word in the sentence are tagged with extension whether it is noun, adjective, verb, preposition and so on. The POS tagger used in this implementation is Stanford POS tagger. The tags used in the Stanford POS tagger tool is shown in Figure 2.

1. CC	Coordinating conjunction	25. TO	*to*
2. CD	Cardinal number	26. UH	Interjection
3. DT	Determiner	27. VB	Verb, base form
4. EX	Existential *there*	28. VBD	Verb, past tense
5. FW	Foreign word	29. VBG	Verb, gerund/present participle
6. IN	Preposition/subordinating conjunction	30. VBN	Verb, past participle
7. JJ	Adjective	31. VBP	Verb, non-3rd ps. sing. present
8. JJR	Adjective, comparative	32. VBZ	Verb, 3rd ps. sing. present
9. JJS	Adjective, superlative	33. WDT	*wh*-determiner
10. LS	List item marker	34. WP	*wh*-pronoun
11. MD	Modal	35. WP$	Possessive *wh*-pronoun
12. NN	Noun, singular or mass	36. WRB	*wh*-adverb
13. NNS	Noun, plural	37. #	Pound sign
14. NNP	Proper noun, singular	38. $	Dollar sign
15. NNPS	Proper noun, plural	39. .	Sentence-final punctuation
16. PDT	Predeterminer	40. ,	Comma
17. POS	Possessive ending	41. :	Colon, semi-colon
18. PRP	Personal pronoun	42. (	Left bracket character
19. PP$	Possessive pronoun	43.)	Right bracket character
20. RB	Adverb	44. "	Straight double quote
21. RBR	Adverb, comparative	45. '	Left open single quote
22. RBS	Adverb, superlative	46. "	Left open double quote
23. RP	Particle	47. '	Right close single quote
24. SYM	Symbol (mathematical or scientific)	48. "	Right close double quote

Figure 2: Tags used in Stanford POS Tagger

The word '*good*' in a typical review sentence will be tagged as '*good_JJ*' or '*good_JJR*' or '*good_JJS*' meaning adjective. Similarly, the word 'not' in a typical sentence will be tagged as '*not_RB*' or '*not_RBR*' or 'not_*RBS*' meaning adverb. Similarly, the verb will be tagged with _VB and its alternatives. In this implementation work, the Parts of Speech words considered for analysis are noun, verb, adverb and adjective. This is because most of the review sentences have these Parts of Speech words.

Step 4 (Preprocessing): Once the Parts of Speech tagging process was completed, the tagged sentence was tokenized. The conditions checked for tokenizing a sentence included white space(s), punctuation marks and delimiters. The tokens having the POS tags as Articles (the, a, an), Prepositions (at, for, except, in, of, on, to, up, with), Conjunctions (and, or, but), Pronouns (I, me, she, her, he, him, we, it, you, they, them) were removed and were not considered in the further stages of analysis. The white spaces, punctuations were also removed. In this implementation, only the tokens with tags of noun, verb, adverb and adjectives were retained for the next stages of analysis process.

Step 5 (Parsing): The tokens retained after applying step 4 were parsed with the pre-defined rules. The pre-defined rules are shown in Figure 3. According to the rules, sentences having a noun or verb followed by zero or more number of nouns or verbs followed by an adverb and adjective were considered. Here, the noun may represent the service name or person name or food name. The adverb may represent additional features added to the adjective. The rule may also be a noun followed by an adverb and followed by zero more noun or verbs. Here too, noun specifies the service name. One more rule may be zero or more number of nouns/verbs followed by an adjective. The typical two reviews are given below:

Review 1: This hotel is very good. Breakfast is also excellent. The staffs are cooperative. Location of this restaurant is very good. Dinner is also not bad.

Review 2: The service in this hotel is really bad. Food is worst. Rooms are not clean. Staffs are not cooperative. Location is very good.

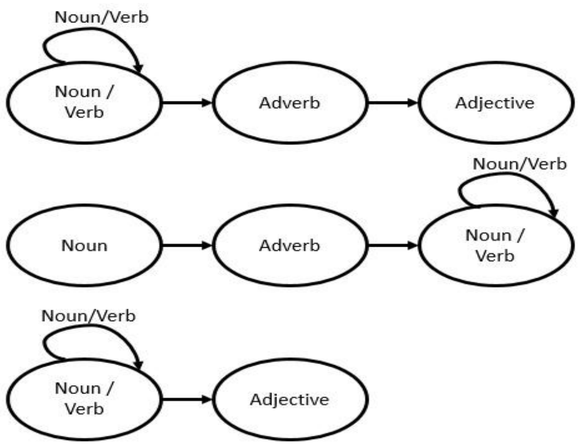

Figure 3: Pre-defined rules for semantic analysis

The above mentioned reviews need to be parsed by the rules mentioned in Figure 3. The sentence, '*Dinner is also not bad*' must be carefully handled since bad signifies negativity but not bad signifies positivity. These situations were considered in the implementation of semantic analysis algorithm. The algorithm for matching the tokens against the predefined rules is given in Algorithm 1.

Algorithm **RuleBasedParsing** (*Rules R, Tokens T*)

Input: Set of predefined rules *R*, Set of Tokens *T*

Output: *Service_Name, Weight, Impression*

Service_Name gives the name of the service

Weight represents the weightage of the adjectives

Impression gives the measure whether the adjective is positively used or negatively used.

```
Get the next token T.token;
If T.token is Noun or Verb then
        Service_Name←Token value of Noun or Verb;
        While the next token T.token is Noun or Verb do
                Service_Name←Service_Name +”,” +Token value of Noun or Verb;
        End While
        If T.token is Adverb then
                If T.token is ‘very’ or ‘always’ or ‘really’ or then
                        Impression←’good’;
                Else if T.token is ‘usually’ or ‘often’ then
                        Impression←’average’;
                Else if T.token is ‘not’ then
                        Impression←’bad’;
                End if
        Else if T.token is Adjective then
                Get the weightage of token, Weight, using SentiWordNet tool;
        Else
                Print “Error… Cannot recognize the token”;
                Return;
        End if
        Get next token T.token;
        If T.token is Adjective then
                Get the weightage of token, Weight, using SentiWordNet Tool;
        Else if T.token is Noun or Verb then
                Service_Name←Service_Name +”,” + noun or Verb;
                While the next token T.token is Noun or Verb do
                        Service_Name←Service_Name +”,” + Token value of Noun or Verb;
                End While
Else
                Print “Error… Cannot recognize the token”;
                Return;
        End if
Else
Print “Error… Cannot recognize the token”;
        Return;
End if
```

Algorithm 1. Rule Based Parsing

Once the values *Service_Name*, *Weight* and *Impression* are extracted from the above algorithm 1, the final semantic analysis is done. This is explained in Step 6.

Step 6 (Semantic Analysis): In this step, the final semantic analysis of the review sentences is performed. The input for the semantic analysis algorithm is *Weight*, *Service_Name,* and *Impression.* These values are obtained from the previous step after applying the Rule Based Parsing algorithm. The *Weight* values are designed in such a way that for negative adjectives, negative values are assigned, for neutral adjectives, 0 is assigned and for positive adjectives, positive values are assigned. If the *Impression* is ‘good’ then a constant value of 5.0 is added to the *Weight*, if the *Impression* is ‘average’ then a constant value of 3.0 is added to the *Weight* and if the *Impression* is ‘bad’, no constant value is added to the *Weight*. If the *Impression* is ‘bad’ and the *Weight* is negative, then the *Weight* is changed to positive *Weight* because it signifies a positive response. The algorithm for the rule based semantic analysis is given in algorithm 2. The step 3 through step 6 is repeated for each sentence present in the review paragraph. The results obtained for a set of reviews using Rule based approached is discussed in section 3.

Algorithm **RuleBasedSemanticAnalysis** (*Weight*, *Service_Name*, *Impression*)
Input: *Weight, Service_Name* and *Impression*
Output: Consolidated review *Cons_Review*

```
If the Impression is ‘good’ and Weight > 0 then
        Weight ← Weight+5.0;
Else if Impression is ‘good’ and Weight < 0 then
        Weight ← Weight – 5.0;
Else if Impression is ‘bad’ and Weight > 0 then
        Weight ← Weight -5.0;
Else if Impression is ‘bad’ and Weight < 0 then
        Weight ← Weight +5.0;
Else if Impression is ‘average’ and Weight > 0 then
        Weight ← Weight +3.0;
Else if Impression is ‘average’ and Weight < 0 then
        Weight ← Weight – 3.0;
End if

If Weight < -3 then
        Cons_Review ← Service_Name + “Worst”;
Else if Weight is between – 3 and 0 then
        Cons_Review ← Service_Name + “Not good”;
Else if Weight is between 0 and 3 then
        Cons_Review ← Service_Name + “Good”;
Else if Weight is > 3 then
        Cons_Review ← Service_Name + “Best”;
End if
```

Algorithm 2. Rule Based Semantic Analysis

The next session discusses Supervised Machine Learning approach for the semantic analysis.

Supervised Machine Learning approach

The machine learning approach used in this work is supervised. The two models used for classification are Bayesian Classifier and the Support Vector Machine (SVM). In the supervised machine learning approach, a set of predefined reviews are manually analyzed and are used to train the classification model. The process of analyzing the reviews using supervised machine learning techniques is shown in Figure 4. Initially, a set of reviews with multiple sentences are stored offline for training the classifier models. The classifier models used in this implementation work are Bayesian and Support Vector Machines (SVM). Once the models are successfully trained, the random reviews are collected from web applications or mobile applications and classified according to the predefined classes. The steps include collecting random reviews, sentence extraction, Parts of Speech (POS) tagging, Preprocessing, Semantic analysis and finally generating consolidated reviews. The steps; collecting the random reviews, sentence extraction, Parts of Speech (POS) tagging and preprocessing are similar to the steps explained in section 2.1 (step 1 through step 4). The remaining steps are explained in this section.

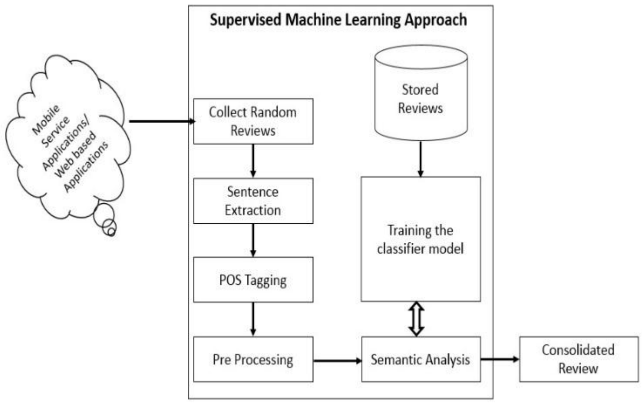

Figure 4: Supervised machine learning approach

Stored reviews: A set of 1000 reviews with multiple sentences were downloaded offline. These review sentences were used to train the classifier models. The stored review sentences consist of maximum 8 words. The words considered in these review sentences include noun, verb, adjective and adverb. The typical adjectives considered include *appearance adjectives* (*beautiful, clean, neat, elegant, and magnificent*), *personality adjectives* (*kind, nice, lazy, scary)*, *sound adjectives* (*noisy, quiet, loud)*, *size adjectives* (*big, great, small, short, less, more*), *time adjectives* (*early, late, modern, old, quick, slow*), *taste/touch adjectives* (*cold, cool, damaged, dirty, dusty, hot, warm, hot, wet, broken*) and *general adjectives* (*good, excellent, bad, worst, horrible, ok*). The adverbs considered in this work include *very, always, really, usually, often,* and *not*. The sequence of these words was analyzed for the above mentioned 1000 review sentences and accordingly, a training table was prepared. This is explained in the training the classifier model section. Once the training model is ready, the semantic analysis was done using the Bayesian and Support Vector Machines (SVM).

Training the classifier model: As discussed in the previous section, the maximum number of words considered in this implementation work after the preprocessing step is 8. The sequence of appearance of words in the training set of reviews is shown in Table 1.

Table 1: Training set for the classifier model

Word 1	**Word 2**	**Word 3**	**Word 4**	**Word 5**	**Word 6**	**Word 7**	**Word 8**	**Class**
Noun	Noun / NULL	Noun / NULL	Very	Very	Clean	Beautiful	NULL	C1
Noun	Noun / NULL	Noun / NULL	Really	Very	Nice	Kind	NULL	C1
Noun	Noun / NULL	Noun / NULL	Usually	Good	NULL	NULL	NULL	C2
Verb	Verb / NULL	Verb / NULL	Not	Good	NULL	NULL	NULL	C3
Noun	Verb / NULL	Noun / NULL	Always	Very	Lazy	Scary	NULL	C4
Noun	Noun / NULL	Noun / NULL	Dirty	NULL	NULL	NULL	NULL	C4
Verb	Verb / NULL	Noun / NULL	Usually	Dusty	Damaged	NULL	NULL	C5
Noun	Noun / NULL	Noun / NULL	Not	Bad	NULL	NULL	NULL	C6
Noun	Verb / NULL	Verb / NULL	Really	Very	Good	Excellent	Noun	C1
⋮								

The total number of attributes considered are 8 which are named *word1, word2…word8*. The last column of this Table 1 represents the predefined classes. These classes are predefined as a set of 6 classes *C1, C2, C3, C4, C5* and *C6*. A maximum of three consecutive nouns/verbs, a maximum of two consecutive adverbs and a maximum of two consecutive adjectives are considered. This is done by analyzing the offline stored reviews. The NULL value in each attribute represents that the values are not present or the values can be optional. The Table 1 shows few entries in the training set for the classifier model. In the implementation work, the table is constructed for all the 1000 review sentences. Once the table is ready, it is applied to classify the reviews to one of the predefined classes, namely *C1* through *C6*. This process is discussed in the next section.

Semantic analysis: Once the training data table is ready, the next process is to apply the Naïve Bayesian and Support Vector Machine classifier on this data. The Naïve Bayesian algorithm classifies the tweets based on their score. Naïve Bayes follows the Bayes theorem which is given in the equation (1).

$$P(Y/X) = \frac{P(X/Y) * P(Y)}{P(X)} \qquad (1)$$

Where,

P (Y|X) = Posterior Probability

P (X|Y) = Class Conditional Probability

P(Y) = Prior Probability

P(X) = Total Probability

In order to classify a random review sentence after the preprocessing stage, the calculation of P(X) is not necessary. This is because the classifier calculates the maximum probability among various classes and then classifies the review to one of the classes *C1…C6* depending upon the maximum probability. For example, consider a random review sentence with Word1 = Noun, Word2=Noun, Word3='*Really*' and Word4='*Excellent*'. The Naïve Bayesian classifier calculates the posterior probability for each classes C1…C6. The class name with the maximum probability is chosen as the destination class label. The values of X and Y of equation (1) are given below which demonstrates how we can calculate the posterior probability for the class label C1.

Y = (Class=C1)

X = (Word1=Noun AND Word2=Noun AND Word3='Really' AND Word4='Excellent')

The probabilities P(X/Y) and P(Y) is computed from the given training set and P(Y/X) is computed. This step is repeated for calculating the posterior probability for the class labels C2…C6. The class label with the maximum probability is given as the destination class label. The training and calculating the posterior probabilities is carried out using the R tool.

Support Vector Machine (SVM) is a machine learning algorithm which also classifies instances based on their score. Each review sentence is considered as a data point and has score attributes. Here, the data values are linearly separable. In this work, Support Vector Machine classification algorithm has been implemented using R tool. The Table 1 is given as the input for the SVM in R tool and the target classes are obtained.

The *Service_Name* is the combination of the noun and/or verb obtained in the review sentence. Once the target class for the random review sentence is obtained using the above mentioned techniques, the following algorithm (Algorithm 3) is applied on the review sentence. The input for this algorithm is the predicted class name *Class_Name* and *Service_Name*. The output is the consolidated review *Cons_Review*.

Algorithm **MachineLearningBasedSemanticAnalysis** (*Class_Name, Service_Name*)

Input: *Class_Name* obtained from either Naïve Bayesian or SVM classifiers;

Service_Name;

Output: Consolidated review *Cons_Review;*

If *Class_Name* = C1 then

Cons_Review ← Service_Name + "Best";

Else if *Class_Name* = C2 then

Cons_Review ← Service_Name + "Average";

Else if *Class_Name* = C3 then

Cons_Review ← Service_Name + "Bad";

Else if *Class_Name* = C4 then

Cons_Review ← Service_Name + "Worst";

Else if *Class_Name* = C5 then

Cons_Review ← Service_Name + "Bad";

Else if *Class_Name* = C6 then

Cons_Review ← Service_Name + "Average";

End if

Algorithm 3. Machine Learning Based Semantic Analysis

These steps are repeated for each sentence in the review. The results obtained are discussed in the next section.

RESULTS AND DISCUSSION

A set of 1000 reviews with multiple sentences was collected for carrying out the experiment. The reviews were collected from web and mobile applications. The reviews are concentrated on hotels and restaurants and hence, most of the set of adjectives for these categories are considered. The rule based semantic analysis has the advantage of not using the training set. The rules are framed by manually analyzing the patterns of the sentences. The parsing phase becomes simpler by matching each token with the predefined rules. The parsing will become complicated when the number of rules increases.

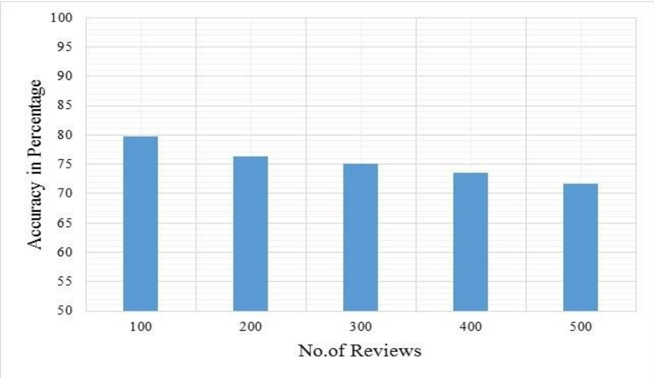

Figure 5: Accuracy obtained for Rule Based Semantic Analysis

The *SentiWordNet* tool was used to compute the weights of the adjectives. This is an overhead in this technique. The results obtained for random 500 review sentences for the rule based semantic analysis is shown in Figure 5.

The Naïve Bayesian classifier is applied on the same random 500 reviews and the results are obtained. The accuracy obtained is shown in Figure 6.

Figure 6: Accuracy obtained for Naïve Bayesian classifier

The Support Vector Machine classifier is applied to the 500 reviews. The accuracy obtained is shown in Figure 7.

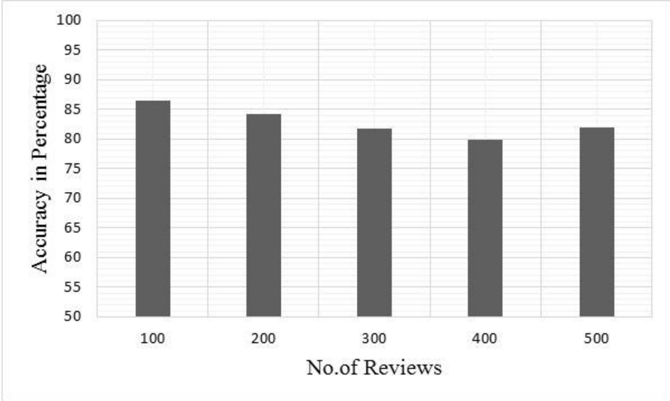

Figure 7: Accuracy obtained for SVM classifier

Figure 8 gives the comparison between the various techniques used for the semantic analysis.

For the given random 500 reviews, our experiment gives the best accuracy for the Naïve Bayesian Classifier approach as compared to SVM and Rule Based approach. The SVM gives better accuracy as compared to the Rule Based approach. The results may vary depending on the reviews taken.

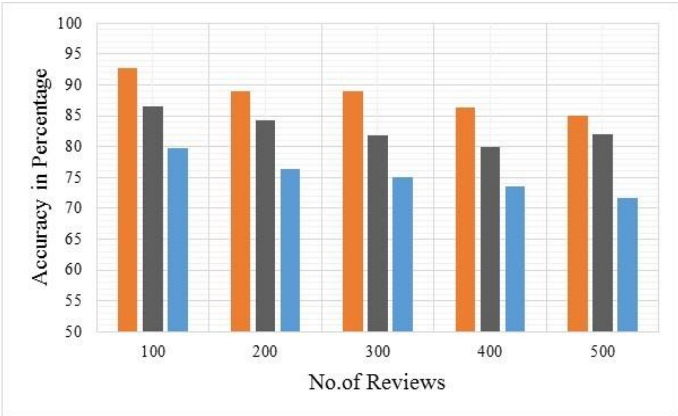

Figure 8: Comparison of Rule Based, Naïve Bayesian and SVM approaches

CONCLUSION AND FUTURE ENHANCEMENT:

In recent times, business depends on the feedback given by the customers and most of the web based or mobile based applications give more importance to the feedback given by the customers. Automatic analysis of these reviews help the organizations to improve the service quality. In this implementation work, we proposed a solution for analyzing the reviews automatically by applying rule based and supervised machine learning approaches. Rule based approach is better when the pattern of the sentences is known in advance. When the pattern of the sentences changes randomly, rule based approach will fail to give better accuracy. It is also seen that as the rule becomes more and more complex, the parsing becomes more complex too. The supervised machine learning approach is better off compared to rule based approach for review sentences with random patterns. But these techniques need more training samples to get better accuracy. It is also noticed that when the number of maximum words in a sentence increases, the training phase becomes more complex. When the amount of review data increases, then some sophisticated techniques must be used to manage this huge data. Improving the accuracy of the semantic analysis is always an open challenge.

REFERENCES

[1] Sangers, Jordy, et al. **Semantic Web service discovery using natural language processing techniques. Expert Systems with Applications**. 2013, 40.11: 4660-4671.

[2] Sun, Shiliang, Chen Luo, and Junyu Chen. **A review of natural language processing techniques for opinion mining systems. Information Fusion**. 2017, 36: 10-25.

[3] Bajwa, Imran Sarwar, and Muhammad Abbas Choudhary. **A rule based system for speech language context understanding. Journal of Donghua University (English Edition)**. 2006, Vol 23.06.

[4] Mohammad, AL-Smadi, et al**. Paraphrase identification and semantic text similarity analysis in Arabic news tweets using lexical, syntactic, and semantic features. Information Processing & Management**. 2017, 53.3: 640-652.

[5] Kang, Ning, et al. **Using rule-based natural language processing to improve disease normalization in biomedical text. Journal of the American Medical Informatics Association**. 2013, 20.5: 876-881.

[6] Zou, Yang, Arto Kiviniemi, and Stephen W. Jones. **Retrieving similar cases for construction project risk management using Natural Language Processing techniques. Automation in Construction**. 2017.

[7] Tanana, Michael, et al. **A comparison of natural language processing methods for automated coding of motivational interviewing. Journal of substance abuse treatment**. 2016, 65: 43-50.

[8] Neeraj Kanger and Gourav Bathla. **Recognizing Emotion in Text using Neural Network and Fuzzy Logic. Indian Journal of Science and Technology**. March 2017, Vol 10(12).

[9] Osamah Ali Mohammed Ghaleb and Anna Saro Vijendran. **Survey and Analysis of Recent Sentiment Analysis Schemes Relating to Social Media. Indian Journal of Science and Technology**. November 2016, Vol 9(41).

[10] Tryfona N. and Pfoser D. **Data Semantics in Location-Based Services. Journal on Data Semantics III. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg**. 2005, Vol 3534.

[11] Irina Pak and Phoey Lee Teh. **Machine Learning Classifiers: Evaluation of the Performance in Online Reviews, Indian Journal of Science and Technology**. December 2016, Vol 9(45).

[12] Liangcun Jiang, Peng Yue, Xia Guo. **Semantic location-based services. Geoscience and Remote Sensing Symposium (IGARSS), IEEE International**. 2016.

[13] Sapna Zol and Preeti Mulay. **Analyzing Sentiments for Generating Opinions (ASGO)-A New Approach. Indian Journal of Science and Technology**. February 2015, Vol 8(S4).