Sentimental Analysis and Its Applications - A Review

Akriti Kaura, Er. Gitika Sharma
Student Assistant Professor
CSE Department CSE Department
Chandigarh University Chandigarh University
Punjab Punjab
Email: akriti317@gmail.com, gitikasharma41@gmail.com

Abstract : Sentimental Analysis over the years has received a lot of recognition as it has shown a tremendous growth. Nowadays Sentimental analysis can be applied to any field as to bring out the emotions attached to it and also we can be able to know what the other person wants to convey. In our work we will be applying the sentimental analysis on the dataset of 15 hotels from the city and apply the new technique of Statistical Analysis. The Statistical Analysis Technique has taken the attributes like Food Quality, Ambience, area of Location etc into the account for calculating Document Index. The Reviews are taken from a trusted website. It is observed that people before making any decision do visit the reviews before making out the further decisions so e-WOM is the most effective to way to convey the views .Nowadays people rely a lot on electronic form of reviews because it is considered as the most trusted way that people can convey the views.

_____*********_____

1. Introduction

Sentiment Analysis [1] is way to analyzing the emotions attached for any area of interest Sentiments are considered to be precious as people are attached to there views and thinking on a great scale. Sentimental analysis is an application of NLP(Natural Language Processing). It all started with the emotions[2][3] that people used to have and all humans tend to have i.e happy, sad, depressed Satisfied etc. Firstly the work was done to recognize these kind of emotions in there work that was called Emotion Detection. Slowly the work began to extend and the work of emotion detection extended to face recognition with emotions detection. Now after some years of extensive research Sentimental analysis has reached the highest position among researchers for there work. The work was extended firstly to the field of Internet and then it extended its grip on the other areas of research.

1.1 History

Sentimental Analysis is an extensive application of Natural Language Processing(NLP). NLP grew as an activity for the behaviour modelling task of three renowned scientists as Frank Pucelik, John Grinder & Richard Bandler who were studying the subjets to know the activity of the behaviour carried out for analytics modelling.It all started in the late year 1970 when NLP[3] had not gained much attention by other research scholars and these 3 scientists started to find out the relationship of modelling and data analytics.Slowly and slowly the researchers came to know about the earlier work if possible had been done in past years or not but unfortunately they didnot find anything.The most appropriate way they found out was to firstly see the computer applications of this area and find out how they are compatible with that time computers i.e the data available or not.

The researchers[4] were unable to track the data so there curiosity grew and with much more effort they tried there level best to give a new direction to the field of NLP.So this field is old and due to new technologies growing day by day so they had to make the system adaptive to the new technology.In the mid 1970's the researchers they took a small data set to try there technique for studying the concept of sentimental analysis so as to observe that what kind of behavioural activity will the results show and also they were startled to find that the results would be self sufficient to prove there theory or not[5].So the very next day they decided to test the set for the given problem and found that there theory was correct and was solvable.

This gave the scientists Pucelik et al,.a great confidence that analysis is possible to do and now they had to perform the Sentimental analysis for the bigger data set so that the theory can be tested on a bigger level.But this time they found it little bit hard because of the reason that the technology began to change and the system was not showing great results.So the major challenge here was first to understand that what technology is demanding and then to perform there tests[6].By now the year 1990's had been arrived and social media also started to develop.Social Media here means that social networking sites started developing that gave people to express there feelings without showing the real identity of the people.Now the Bandler et al,. Got an instinct to test there theory on social networking sites. They used to record the data on tapes and drives so that they can store there data for future use also.Slowly they got success

122

and now the field of sentimental analysis has been expanded to a great level and much work of research has been done on it and and much work is still going in process where on a large scale the experiments are been working so that in future they can be implemented for the betterment of the society.

1.2 Types Of Sentimental Analysis[7]

Sentence Level[7]: The very first level where sentiments are analyzed in a single sentence.

Document Level[7]: The second Level where a whole document is analyzed. There is not much difference between sentence and document level as document is also composed of sentences only.

Aspect Level[7]: The last level where sentiments are analyzed regaurding a particular topic . Example. In our work we are taking into account the views of hotels so this an aspect level sentimental analysis.

1.3Applications of Sentimental Analysis

The application for this area are much but in other words we can say that they are endless.But main work can be seen in these areas particularly

a) social media monitoring
b) VOC to track customer reviews
c) survey responses
d) Sentiment analysis is in demand because of its efficiency. Thousands of text documents can be processed for sentiment in seconds, compared to the hours it would take a team of people to manually complete.

2. Literature Survey

The extensive survey [7] reveals that people oftenly write the reviews online for any industry but hotel industry is one such industry that needs to be analysed.

 The approaches which earlier are were followed earlier are clubbed under 3 categories namely Machine Learning, Lexicon based Approch and Hybrid Technique. Below is the table showing all the 3 approaches with there advantages and limitations

Table 1 Approaches For Sentimental Analysis

Sentiment Classification Approaches	Features/Techniques	Advantages and Limitations
Machine Learning - Bayesian Networks Naive Bayes Classification Maximum Entropy Neural Networks Support Vector Machine	Term presence and frequency Part of speech information Negations Opinion words and phrases	ADVANTAGES the ability to adapt and create trained models for specific purposes and contexts LIMITATIONS the low applicability to new data because it is necessary the availability of lablled data that could be costly or even prohibitive
Lexicon based - Dictionary based approach Novel Machine Learning Approach Corpus based approach Ensemble Approaches	Manual construction, Corpus-based Dictionary based	ADVANTAGES wider term coverage LIMITATIONS finite number of words in the lexicons and the assignation of a fixed sentiment orientation and score to words
Hybrid Machine learning Lexicon based	Sentiment lexicon constructed using public resources for initial sentiment detection Sentiment words as features in machine learning method	ADVANTAGES lexicon/learning symbiosis, the detection and measurement of sentiment at the concept level and the lesser sensitivity to changes in topic domain LIMITATIONS noisy reviews

 The above table shows the various techniques that have been used in the past years. The extensive survey discusses about the how various scientists have gone in different directions in order to provide the best results in the field of sentimental analysis that is also an application of data mining field and also a sub field for the research purpose. We have seen that people before taking any decision also consider WOM(Word of Mouth) i.e the good/bad talks that have been said by the people regaurding any industry so that people may try it out or not. Various researchers have also been finding out ways in order to correctly classify the views that people give regaurding any place or item. The sentimental analysis is one such field that is not restricted to one field only.The Sentimental Anlaysis have been even

used for the crime prediction in order to judge that whether the person would commit the crime or not.

The Researcher Sagadevan et al,[7].discusses about the importance of writing text for communication among humans. The inventions of online networking shows the rapidly growth of the size of the textual data and digital data available and also inproptionately questions the approaches being followed to mine the information . He suggested the mass audience to study the hidden talent of personality by applying automatic personality detection approaches. As such, this study develops the Three Factor Personality (PEN) Model as a personality framework to guide our understanding and revealing the role of words in depicting

the characteristics of a user. This first stage study revealed how the sentiment analysis of public towards particular words can assist us in deciding the personality of social media users by decoding their status messages. Another researcher Ajitha et al,.[8] Coined a new term as Emotion Analysis or Opinion Anlysis which discussed about techniques for analyzing and identifying the sentiment expressed in the blog post. The sentence in the blog post can, be categorized as objective and subjective sentence. The Subjective Sentence plays a major role in expressing the opinion and the objective sentence deals with facts. So we have to consider only the subjective sentence for predicting the emotion. Valence and Arousal plays a major role in predicting the opinion. The impact of Valence and arousal on the affective states predict the emotion and feelings in the blog post. Valence can be defined as the angle of experience that refers to satisfaction and dissatisfaction and arousal describes the level of energy used for expressing the feelings for a particular task . If a term is having high valance and high arousal it has positive opinion. Similarly if a term has low valance and low arousal it has negative opinion.

Generally emotion can be defined as the expression of affective states of the person and we will map the affective state of the sentence with the ordinal states based on the rating of the valence and arousal.

Great Researchers like Bader et al.,[9][10]did a fantastic job by giving another dimension to the field of multi lingual texts for sentimental analysis where we integrate sentiment analysis and a multilingual vector-space approach to information retrieval to address the problem of multilingual sentiment analysis.

Another Researcher Dengel et al.,[11][12] discussed about the increasing popularity of social networking, blogging and micro-blogging websites, where every day a huge amount of information subjective text statements are fetched online. The information collected from these texts, can be used for specific surveys for a social perspective.

From the above discussions it is observed that sentimental analysis has grips in the software industry from past many years and a lot of work has been done in this fields and in the sub fields like for micro blogging sites facebook twitter and of course the multilingual texts but for the hotel industry much work is needed to be done and the processing is ongoing. The work in the hotel industry is also going using various approaches/techniques like Hybrid Supervised and Unsupervised have been used for sentimental analysis .So from this we have been proposed a new Technique that will help to calculate the polarity of Hotel Reviews so that people can make decisions wisely.

3. Proposal

The main problem with the hotel industry[13][14][15] lies in the fact that people whenever visit to any new hotel or any restaurant and if they do not find any of the attributes like the ambience food quality or the location right about the restaurant or hotel they will write about it online so that there views could be reached to the common people. It has been found[16][17][18] that 65-70%of the people always visit the web site of the hotel or the restaurant so as to check that are they making the right decision or not and also there hard earned money is not wasted in a wrong direction. So our main motive becomes to accurately find out the result so that people may not feel cheated. Secondly, people sometimes does not express there feelings openly and hence they write neither negatively nor positively about the hotels. So in that case people are not able to figure as to what they should do because views are neither going on the positive side nor on the negative side. So a new approach has been proposed in order to give the correct views about the hotel. The hotel Industry is one such industry in the world that is ever growing and will prosper always to cater the people all around the world but all it needs is the support of people and without that we cannot judge any hotel or restaurant and if the something went wrong then negative views will come in the name of hotels that is considered not good for the goodwill of hotels. The exact polarity of the online views till now so our approach to find out polarity is an effort to produce good results so that people can think better before making any decisions.

The exact way as to how to calculate the polarity of the reviews is Diagram showing the correct way
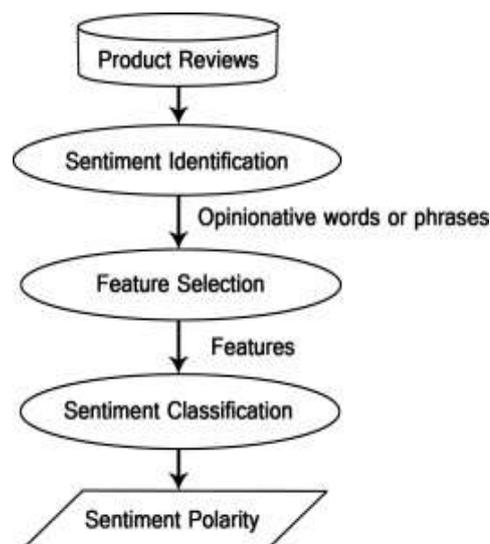

Fig 1 Showing the way for calculating the polarity

The technique which which has been proposed is below showing as how the polarity of hotel reviews will be calculated
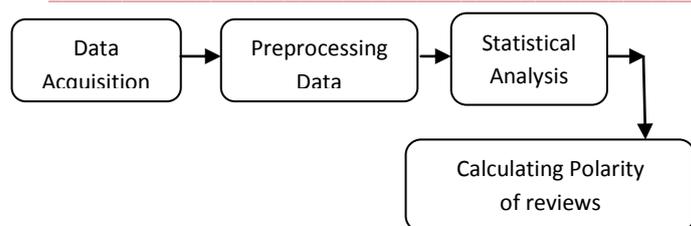
Fig 2 The New Approach For Calculating Polarity of Hotel Reviews

4. Conclusion

The main motive for all the researchers always remains the same to give better results for the work that is in demand of the society. So hotel industry is one such industry that is always in demand for researchers to experiment new ways to find out the sentiments attached to it. From past few years the 3 approches were followed in order to find out the sentiments from it.The techniques were the Hybrid learning Supervised learning and Unsupervised Learning. These 3 approaches were mainly applied in order to calculate the polarity of views of various fields like micro blogging site, twitter, facebook and atlast the multilingual texts but now the the another field that researchers have found is the hotel industry where a lot have still to be found. From the Literature Survey we have concluded to a new rather a good approach for finding polarity of online reviews of hotels/restaurants so that it may help the society in future that will help to judge the sentiments of the reviews of people correctly and also will benefit the society that is help to make the decisions carefully

References:

[1]   Xu Chen, Hui Gao, Yan Fu "Situation Analysis and Prediction of Web Public Sentiment" 2008 International Symposium on Information Science and Engieering,2008.

[2]   Kerstin Denecke "Using SentiWordNet for Multilingual Sentiment Analysis"2008.

[3]   Brett W. Bader, W. Philip Kegelmeyer, and Peter A. Chew "Multilingual Sentiment Analysis Using Latent Semantic Indexing and Machine Learning", 11th IEEE International Conference on Data Mining Workshops, 2011.

[4]   Lin Dai Hechun Chen "Improving Sentiment Classification Using Feature Highlighting and Feature Bagging", 11th IEEE International Conference on Data Mining Workshops,2011.

[5]   Bin Lu Myle Ott, Claire Cardie and Benjamin K. Tsou " Multi-aspect Sentiment Analysis with Topic Models", 11th IEEE International Conference on Data Mining Workshops,2011

[6]   Zied Kechaou, Mohamed Ben Ammar "Improving e-learning with sentiment analysis of users' opinions", IEEE Global Engineering Education Conference

(EDUCON) – "Learning Environments and Ecosystems in Engineering Education", 2011

[7]   Dominique Ziegelmayer ,Rainer Schrader "Sentiment polarity classification using statistical data compression models", IEEE 12th International Conference on Data Mining Workshops,2012

[8]   Lumin Zhang, Yan Jia, Bin Zhou,Yi Han "Microblogging Sentiment Analysis Using Emotional Vector", Second International Conference on Cloud and Green Computing,2012

[9]   Seyed-Ali Bahrainian, Andreas Dengel "Sentiment Analysis using Sentiment Features", IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT), 2013

[10]   Wenjing Duan, Qing Cao, Yang Yu "Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality", 46th Hawaii International Conference on System Sciences,2013

[11]   Simon Fong, Yan Zhuang, Jinyan Li "Sentiment Anlaysis of Online News using MALLET", International Symposium on Computational and Business Intelligence,2013

[12]   Niloy Gupta, Abhinav K.R. "Fuzzy Sentiment Analysis on Microblogs for Movie Revenue Prediction",2013

[13]   Chung-chi Huang "Interest Analysis using Social Interaction Content with Sentiments",2013

[14]   Shoushan Li, Zhongqing Wang, Sophia Yat Mei Lee,Chu-Ren Huang "Sentiment Classification with Polarity Shifting Detection", International Conference on Asian Language Processing,2013

[15]   Bingwei Liu, Erik Blasch, Yu Chen, Dan Shen , Genshe Chen "Scalable Sentiment Classification for Big Data Analysis Using Na¨ıve Bayes Classifier", IEEE International Conference on Big Data,2013

[16]   Prashast Kumar, Arjit Sachdeva, Dhruv Mahajan "An approach towards feature specific opinion mining and sentimental analysis across e-commerce websites",2014

[17]   Yilin Wang, Baoxin Li "Sentiment Analysis for Social Media Images", IEEE 15th International Conference on Data Mining Workshops,2015

[18]   Pavitra.R, PCD.Kalaivaani "Weakly Supervised Sentiment Analysis Using Joint Sentiment Topic Detection With Bigrams", IEEE SPONSORED SECOND INTERNATIONAL CONFERENCE ON ELECTRONICS AND COMMUNICATION SYSTEMS(ICECS '2015),2015