



## Opinion Mining for monitoring social media communications for Brand Promotion

Mini U.

Department. of Computer Science,  
Cochin University of Science & Technology  
Cochin, Kerala, India

Poulose Jacob. K.

Department. of Computer Science,  
Cochin University of Science & Technology  
Cochin, Kerala, India

**Abstract:** The recent years research has shown a great interest in text mining leading to opinion mining and analysis of the social media postings. The influence of social media in opinion formation of public, is widely used in brand promotion. The casual public posts are classified. The sentiments are measured. A feedback loop is created. This paper intends to identify the sentiments behind the casual social media transcripts and convert the general viewers to potential visitors to confirmed client. This is a case study of tourism industry. In this study NER algorithm for opinion mining technique is applied to an eco tourist resort property domain. Here we are trying to discover consumer preferences about eco tourism, speciality features of the resort property and restaurant, from Facebook posts as reviews. The model created helps in determining the sentiment orientation of opinions and the feedback to improve the services offered which in turn increases the traffic.

**Keywords:** Opinion Mining, Sentiment Analysis, Social Network, Text mining, Brand promotion

### I. INTRODUCTION

Social media democratised communication medium ignoring all the existing norms of hierarchies, structures, gender, caste bridging other categories of social divide. The explosion of social media sites promoted marketing. The worldwide trend of shifting the marketing from the traditional channels to social media has resulted in building a new kind of customer relationships. In any domain, there is an increasing trend to check online to get the opinion before taking a final buying decision. A product or service promoted by word of mouth, had a strong viral effect. Social media which started as a casual medium of communication has emerged as electronic word of mouth today targeting social, functional and emotional factors turning them into favourable factors. Social networking platforms such as Twitter, Facebook, YouTube and LinkedIn helps the connectivity promoting the reach as part of their broader customer engagement programs. Social media marketing is an effective way to engage audiences. This mode of marketing helped in faster information diffusion and expanded the reach of the brand by driving more traffic into the sites.

#### *Facebook – Social Media Platform*

Facebook has become the de-facto medium of social mobilization and medium of learning. Facebook is a social media platform where people started to share photos, communicate with friends and stay in touch with people. As the Facebook User base grew, more and more companies started choosing Facebook as a media for brand awareness and presence through Facebook Fan Pages. The advantage of being in direct contact with their fan base allows two way communication with their customer base. This helps in promoting the brand and also to fix up and tailor according to customer feedback. Facebook platform can be treated as one of the best representations of the world's population. The evenly spread demographics across age, gender, geographic location, education and income level is a best fit. The interactions are explored across the available dimensions like geography, gender, hashtags. These interactions are further classified reflecting the topics and categories of the shared

giving a more structured understanding of the stories. All the provided insights can be anonymised and aggregated.

### II. LITERATURE REVIEW

#### *A. Related Work*

Sentiment analysis has attracted a lot of attention in recent years because of its applications. The main focus areas are lexicon construction, feature extraction, and then determine the polarity which can be taken as a feedback for improvements in many cases. The emergence of crowdsourcing created new opportunities in data collection and annotation methods.

The term opinion mining was introduced by Dave et al. [2]. The paper explains opinion mining tool combining all the qualitative featuring of product attribute in such a way that search results highlight the attributes. Opinion mining is defined by Bing Liu [7] as the study analysing people's opinions, emotions and sentiments based on evaluations of entities and their attributes like products, services, events etc. Hu and Liu [6] reviews machine learning methods to analyse the user reviews. challenges due to the noise in the text and complexities of natural language processing are also considered. Sentiment analysis can be performed at document level, sentence level and aspect level. Pang et al. [9] uses overall sentiment to classify documents. The well-known and much-cited paper of Hu and Lie (2004) [5] represent component (which are product, person, event etc.) and associated set of attributes as aspects. The majority of the algorithms sentiment analysis use machine learning classifier

The entire document is classified as positive or negative in document level sentiment analysis as described by Pang et al, [9] and Turney, [11]. Pang et al. [8] investigate the effectiveness of sentiment classification of documents by machine learning techniques. They compared the effectiveness of applying Naive Bayes, Maximum Entropy and Support Vector Machines and found that SVM is the best. The aspect level sentiment analysis aims at identifying the target of the opinion. The basis of this approach is that

every opinion has a target and an opinion without a target is of Numerous works are done in aspect extraction. Hu & Liu used part of speech (POS) tagger to identify nouns and noun phrases as candidate aspects, and chose frequent ones as aspects [6][7][8]. Popescu & Etzioni[16] used point wise mutual information as an improvement.

The opinion mining analysing the reviews is done to improve user experience and also for the organisation to improve based in feed back. But another side to this is that purposful writing reviews to promote/demote a certain product can influence users' decision making. This is called opinion spamming. This is studied by Choo [1] where the sentiments of the user communications are aggregated to identify strongly connected communities from which spammers are identified.

Apart from studies on extraction of sentiments from tweets, many studies are conducted on explicit reviews. Usage of NLP techniques to categorize Amazon reviews according to their sentiment ng[is explained by Fang[3] The method leverages on lexical resources for sentiment analysis available in English[12]. The synonyms and antonyms in wordNet are used by the set of opinion words tagged. Hoogervorst *et al* [4] employs a discourse parser implementing Rhetorical Structure Theory (RST). In this case, context of each aspect is determined from the parser and expressed sentiment is computed with respect to the weightage of the discourse relations between word. Determine the polarity of comments whether it is positive, negative or neutral by extracting features and components of the object is Opinion mining [10].

### III. SOCIAL MEDIA, SENTIMENT ANALYSIS AND OPINION MINING

Social media promoted establishing digital relationship among online communities. It emerged as a powerful tool more than an entertainment media, with business intelligence to communicate with existing as well as potential customers. Social media platforms changed the way consumers communicate and permitted consumer reviews to reach a large audience. It offered a way to map relevant conversations in the influential order. This platform promoted users to share and engage with other users or friends. This increased brand exposure. This viral and collaborative character of social media percolated into the community promoting brand to build trust and loyalty among users. Business receive feedback. Social media permits instant availability of customers' voices and opinions. Also it became critical for the business to understand how the content is diffused, shared and consumed. This feedback information includes geo-location data and helps in analysing customer-preference. This can refine the segmentation and reach the target audience. The contextual text analytics helps in understanding market-trending information. This gives a better understanding of market dynamics which was earlier not available to the industry. Business houses are trying to unleash the potential and quality of the insights that sentiment analysis can deliver. Sentiment analysis is slowly creeping into the marketing mix of industries. Sentiment analysis helps in understanding the customer opinion about brand, products, and services. In this paper we look into how sentiment analysis help in making better business decisions. Social media with compelling content, with proper analysis can monitor the sentiments of

limited use, (Hu and Liu, [6]). the brand. Most of the business houses are integrating marketing communications with social media to reach out to all sections of consumers. Companies are ensuring that social media is a part of the marketing eco system and online exposures are directly tied to the brand image.

Social media marketing starts with creating content page. More and more readers are attracted towards that page. If the contents are attractive and feel a worth, encourages readers to share it other social networks. These reposts have more value among the consumers as they come from a trusted source, which is a favourable publicity for the brand. The virality character of the social media helps in broadcasting. Social media campaigns are targeted at consumer products. These conversation data volumes are huge. Dynamic and responsive social media presence will be able to enhance customer interaction and influence key decisions using feedback mechanisms turning negative opinions into favourable decisions. Sentiment analysis or opinion mining is the process of dissecting the emotional tone behind the words used and understanding the attitudes, opinions and emotions of the expression. Sentiment analysis finds its place in social media monitoring. The wide ranging applications of sentiment analysis are very powerful. Real time monitoring of social data helps in extracting insights and gauge consumer attitudes and take effective corrective action if needed.

#### A. Case Study

In this paper, authors are studying the the opinons expressed by the customers about an eco nature tourist property situated in the state of Kerala in India. In our case study we consider only online promotion through social media Facebook. NER algorithm for opinion mining technique is applied to a eco tourist resort property domain. Here we are trying to discover consumer preferences about eco tourism, speciality features of the property and restaurant, using opinions available on the Facebook as reviews. Results are extracted from product reviews available on Facebook containing valuable information about customer preferences. The model created helps in determining the sentiment orientation of opinions and the feedback to improve the services offered which inturm increases the traffic.

#### Developing a Response Review Strategy

A customer review is a user generated content about the property. Probable customers read this review. This can be influential in the process of deciding to use the services of the property. The large amount of information available and the difficulty due to the information overload makes the user difficult to read all and decide. Hence the automatic extraction of the data and presenting the relevant information is very important from the consumer angle. Many aggregators are making use of this functions. There are many methods to automatically collect and extract the reviews and summarise the opinion.

Here we are presenting a system that extracts all the opinions from the social media summarises the review and present an overall perspective. The feedback helps in improving the services of the property.

- Data acquisition – The property constantly monitor the social media conversations. People use Facebook platform to express their opinions.

- Mining the contents of social media conversation- Customers perceptions are expressed in the social media through their reviews.
- Applying Sentimental Analysis tools- The opinion expressed can be classified in the first step as positive, negative and neutral based on the content polarity of the sentiments expressed.
- Feedback – Business can be improved by an effective social media strategy which helps in sales by increasing customer loyalty.
- New CRM strategies are devised to establish and betterment of the customer relations
- Impact of Social media findings and events on business

There are multiple data sources available. Here for this study, we are considering only social media communications through Facebook.

In this case study, from the fan's key behavioral pattern, we are trying to analyse

- <sup>1</sup> What people talked about their experience in the property?
- <sup>2</sup> What do they like/dislike about the property?
- <sup>3</sup> What are they talking about you?
- <sup>4</sup> How many people are talking about you?
- <sup>5</sup> How to capitalise the influence?

#### Systems Architecture

Opinion extraction from the reviews posted can be done using different ways. Various approaches based on NLP, lexical analysis, machine learning are discussed in literature.

#### Stages involved

- Acquisition – This module consist of acquiring the content, curation and analysis. The contents for extraction and analysis are collected using a crawler program. Each sentence is split into words through tokenisation. Polarity of each work
- Rule Engine - Rules are built. It is a continuous process. Machine learning algorithms work on it. The different keywords are to be trained to find the themes and to pull out the sentiment from the word stream.
- Connecting with CRM Customer feedback from social media, contains many useful business information. It helps in understanding more than what customers are talking about, how they feel. changes in sentiment scores is a place to think into making product improvements, or changing marketing strategies
- Sentimental Analysis and Visualisation- Sentiment analysis is critical as it helps in understanding the customer like and dislike about the brand.
- Feedback Sentiment analysis is a continuous process. The review of customer feedback helps to be more proactive in the changing dynamics in the market place.

#### How the analysis is done?

Ontology is created. This is fed to rule engine. Weightage table is created. From the extracted data, the first step is keyword tagging. Keywords are matched against the taxonomy (based on weightage allotted). Keywords that do not match are listed. Taxonomy is evolving.

The Acquisition module takes care of data cleaning. The data is unstructured and may be of different formats. There may be overlapping data also. Data preprocessing is also done in this module. After curation, analysis is done.

#### Language Identification

This is the first step involved in the process. Language identifier receives plain text and identifies the language of the input text. Here in our case study only "English" language is used.

#### Tokenising

The tokeniser accepts the plain text as input and a language parameter. The input text is split into paragraphs, sentences and tokens and generates a token with these information. Extracting higher level information from the document is easier through tokens. Each token is a representative of a type or a group of words. These tokens are converted to a standard form and the process is called stemming or lemmatization

#### POS Tagger

Part of Speech Tagger identifies each word as a noun, a verb, etc. This provides Part of Speech Perceptron and Maximum Entropy Models for English. Apache OpenNLP Machine Learning API is used for training. Lemmatisation is done based on dictionary-based lemmatisation, identifying lemma (dictionary entry) for the word.

#### Named Entity Recognition

and Classification identifies names of persons, places etc and classifies them in a semantic class (PERSON, LOCATION, etc.). Apache OpenNLP API is used to get Perceptron and Maximum Entropy models trained.

#### Parsing

Parser provides syntactic tree representation of a sentence. The mainstream of dependency parsing is conducted on lexical elements and relations are built between single words. The labeled corpus is made as single phrase, and is identified as noun phrase (NP) or verb phrase (VP). Each candidate NP or VP is checked by the model with a threshold value.

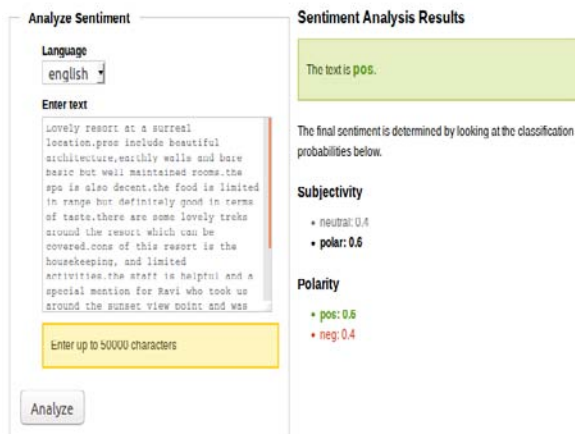
#### Data Analysis:

Sentiment Analysis is done on the opinion expressed by customers as Facebook post. The opinion expressed may be direct or indirect. The data is cleaned, pre processed and undergo stages of analysis. Classification is based on machine learning. The text is considered as positive or negative based on attitude of expression. The tagged data is compared against the large corpus/ lexicon using machine learning.

SentiWordNet a freely down loadable lexical resource for opinion mining is used here. It cross checks every word in the database and assign scores with positive, negative and neutral. In the starting phase, synsets are labeled manually. Since this is an evolving corpus, it is expanded through Machine learning algorithms. Here we have used Naive Bayes classifier.

## SENTIMENT ANALYSIS WITH PYTHON NLTK TEXT CLASSIFICATION

The sentiment analysis tool is written in Python and is using NLTK text classification process. The entered post is classified as positive sentiment, negative sentiment, or neutral. Here we are following hierarchical classification. Neutral sentiment is determined first. If not neutral, polarity is determined.



**Fig 1 – Sample Screen shot of the Analyser**

The words are classified using classifiers trained from the data sets created in the corpus using nltk-trainer. The accuracy of the results depend on identifying text which is similar to the original training data. If the result is not as expected, it may be because of unrecognized words. Machine learning improves the accuracy over a period of time and increased words.

### Challenges:

- Only posts in english language is considered. Reviewers may be more confident in expressing in their native language.
- Authenticity of extracted data is important as there is a tendency to give biased reviews, at least in some cases
- Contextual classification – A word could be considered positive in one occasion and also as negative in a different occasion.
- Linguistic nuances like sarcasm expressed and also ironical expressions are difficult to identify.

## IV. RESULT ANALYSIS AND CONCLUSION

The opinion mining by sentiment analysis of Facebook post considered here is the guest reviews that will provide insights about the feel of customer satisfaction and dissatisfaction which is otherwise not explicitly known by just examining the numerical scores of emotions. Very many industries especially in hospitality and healthcare ask the customer rating in stars, a kind of numerical rating in addition to providing

comments. Though numerical rating is easy way, it is essential that managers use text analytics to understand the underlying customer sentiments. Such an approach will provide better insights for quality and process improvement.

### Further Enhancements

In the current setup the review considered is from Facebook posts only. This can be extended to other social media posts including tweets, blog posts, forums, aggregators etc. In this case of opinion mining, the program focuses solely on the final sentiment in the review, but with multiple emotions in the comment, this is not very efficient.

## V. ACKNOWLEDGMENT

Data has been shared by a commercial agency with an agreement that all the information provided will be used for academic usage only.

## VI. REFERENCES

- [1] Choo, E., Yu, T., & Chi, M. (2015,). Detecting opinion spammer groups through community discovery and sentiment analysis. In IFIP Annual Conference on Data and Applications Security and Privacy (pp. 170-187). Springer International Publishing.
- [2] Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web (pp. 519-528). ACM.
- [3] Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data*, 2(1), 5.
- [4] Hoogervorst, R., Essink, E., Jansen, W., van den Helder, M., Schouten, K., Frasinca, F., & Taboada, M. (2016, ). Aspect-Based Sentiment Analysis on the Web Using Rhetorical Structure Theory. In International Conference on Web Engineering (pp. 317-334). Springer International Publishing.
- [5] Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177). ACM
- [6] Liu, B. (2007). Web data mining: exploring hyperlinks, contents, and usage data. Springer Science & Business Media.
- [7] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- [8] Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.
- [9] Pang, B., Lee, L., & Vaithyanathan, S. (2002, ). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 79-86). Association for Computational Linguistics.
- [10] Popescu, A. M., & Etzioni, O. (2007). Extracting product features and opinions from reviews. In Natural language processing and text mining (pp. 9-28). Springer London.
- [11] Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting on association for computational linguistics (pp. 417-424). Association for Computational Linguistics.
- [12] K. Denecke, (2008). Using SentiWordNet for multilingual sentiment analysis, In Proceedings of IEEE 24th International Conference on Data Engineering Workshop, Cancun, pp. 507-512.