

Sentiment Analysis of Tourists Opinions of Amusement, Historical and Pilgrimage Places: A Machine Learning Approach

Venu Dave^{#1}, DhvaniShah^{#2}, DikshiSuthar^{#3}, Bhagirath Prajapati^{#4}, Priyanka Puvar^{#5}
Research Scholar, Dept. of Computing Engineering, A. D. Patel Institute of Technology, Anand, Gujarat, India^{#1,2,3}
Associate Professor, Dept. of CE, A. D. Patel Institute of Technology, Anand, Gujarat, India^{#4,5}

Abstract- In today's social media trend everyone is using Internet for every aspect of their lives whether it is shopping, interfacing with each other or getting information. The motive of this study is to cover one another, but important aspect that is tourism. As we know tourism plays a major role in any country's economy so it is very important to throw light on this aspect. *Customer experience is consequential in any business, so to improve the customer experience, reviews of different places, hotels and restaurants are taken and after analyzing them an overall review of that place is generated so customer could decide which place to visit and which not to, so this saves both time and money and customer experience is also positive. The fake and redundant reviews need to be eliminated for better accuracy of the results.*

Keywords- Sentiment Analysis, Opinion Mining, Natural Language Processing, R Language.

I. INTRODUCTION

Opinion mining- which is also known as sentimental analysis is the process of determining the emotional intonation behind a concatenation of words, used to gain an understanding of the attitudes, opinions and emotions expressed within an online mention.

As we know in recent years the data collected and stored in networking sites has become an important topic to research topic for opinion mining application since networking sites took attention of many users all around the world. At that point computational intelligence comes into the picture. As we know that today's world deal with large amount of data and to organize and analyze them is difficult, but at the same time very important.

Objective of this project is to process the reviews of tourist places, collected from a website based on sentiment analysis. And segregate those reviews based on emotions behind the text such as positive, negative and neutral. Also provide reliability of system through generating good accuracy (precision and recall) of each place.

II. PREVIOUS WORK

Existing researches in the document based opinion mining are mentioned below. The most prominent work was done by richa sharma [1]. "Poor" and "Excellent" seed words are used by him to calculate the semantic orientation, point wise mutual information method is used to calculate the semantic orientation. The sentiment orientation of a document was calculated as the average semantic orientation of all such phrases. 66% accuracy was achieved for the movie review domain. Taboada et al. [2] used lexicon-based method to perform sentiment classification. For classification positive and negative words dictionaries are used and semantic orientation calculator (SO-CAL) is built that incorporate intensifiers and negation words. This approach has been shown to have 59.6% to 76.4% accuracy on 1900 documents of the movie review dataset.

III. PROPOSED WORK

The purpose of this study is to obtain sentiment behind reviews written by tourists. It analyze dataset of reviews and classify them into three categories – Positive, Negative and Neutral. It also gives accuracy of system to show how reliable it is. Through this, visitors will be able to identify whether place is good or not without reading bunch of reviews.

IV. METHODS and ALGORITHMS

A. Sentiment Analysis

Sentiment analysis is the conjecture study of people's opinions, attitudes and emotions to an entity. The entity can represent individuals, events or topics. These topics are most likely to be covered by reviews. Sentiment Analysis is to identify the sentiment expressed in a text then analyzes it.

There are three main classification levels in Sentiment Analysis: document-level, sentence-level, and aspect-level.

1) **Document Level:** Document-level SA aims to classify an opinion document as expressing a positive or negative opinion or sentiment. It considers the whole document a basic information unit (talking about one topic).

2) **Sentence Level:** Sentence-level SA aims to classify sentiment expressed in each sentence. The first step is to identify whether the sentence is subjective or objective. If the sentence is subjective, Sentence-level SA will determine whether the sentence expresses positive or negative opinions.

3) **Aspect Level:** Aspect-level SA aims to classify the sentiment with respect to the specific aspects of entities. The first step is to identify the entities and their aspects. The opinion holders can give different opinions for different aspects of the same entity like this sentence “*The voice quality of this phone is not good, but the battery life is long*”. This survey tackles the first two kinds of Sentiment Analysis. [3]

B. R Language and R Studio

R is a programming language which provides environment for statistical computing and graphical representation for data analysis. R Studio is an integrated development environment for R.

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. [4] Among other things it has:

- An effective data handling and storage facility,
- A suite of operators for calculations on arrays, in particular matrices,
- A large, coherent, integrated collection of intermediate tools for data analysis,
- Graphical facilities for data analysis and display either directly at the computer or on hardcopy, and
- A well-developed, simple and effective programming language (called ‘S’) which includes conditionals, loops, user defined recursive functions and input and output facilities. (Indeed most of the system supplied functions are themselves written in the S language.)

R is very much a vehicle for newly developing methods of interactive data analysis. It has developed rapidly, and has been extended by a large collection of packages. However, most programs written in R, are essentially ephemeral, written for a single piece of data analysis.

R is a modern, functional programming language that allows for rapid development of ideas, together with object-oriented features for rigorous software development. The rich set of inbuilt functions makes it ideal for high-volume analysis or statistical simulations, and the packaging system means that code provided by others can easily be shared. Finally, it generates high-quality graphical output so that all stages of a study, from modelling/analysis to publication, can be undertaken within R. [5]

C. Naïve Bayes classifier

The Naive Bayes classifier is a simple probabilistic classifier which is based on Bayes theorem with strong and naive independence assumptions. It is one of the most basic text classification techniques with various applications in email spam detection, document categorization, language detection, sentiment detection and automatic medical diagnosis. It is one of the most basic text classification techniques used in various applications. It is highly scalable.

Naive Bayes is that it only needs a small amount of training data to estimate the parameters necessary for classification. It is fast to train data and fast to classify. It is not sensitive towards irrelevant features. It handles real and discrete data and it handles data stemming also.

FIG.1 FLOW DIAGRAM OF SYSTEM



V. DATASET

Three categories of places is being analyzed. i.e. Amusement park, Pilgrimage place and Holly place we have taken reviews of one place from each three categories as a sample.

TABLE I. EXTRACTED REVIEWS OF ESSEL WORLD

Sr.no.	Essel world theme park Reviews
1	Just enjoy all the rides. The place may be crowded but you should enjoy. You should walk fast in order to reach the places you wish so it is advisable that avoid going here with old people.
2	This is a rides based theme park having lots of fun rides. Most of the rides are good for all including kids. My personal favourites is the Top Spin ride - please try this only if u can take the thrill. The roller coaster ride here is however extremely pathetic.
3	Amazing place of Mumbai ride lover, one of the must visit place of Mumbai .
4	Worst amusement park in Mumbai
5	Ridiculous place with high price

TABLE II. EXTRACTED REVIEWS OF MOHEN-JO-DARO

Sr. No.	Mohen-jo -daro Reviews
1	I want to go here.. I love this place.. this is place where Hindu religion Start.. birth place of Hinduism.. India name is come from here.. hindh name come from indus valley civilization
2	amazing place even in present times the ingenuity of the

	residents of this place inspires us...
3	A mysterious place asking you to explore every inch of it. It was a fascinating evening in Moenjo Daro
4	It's great place to experience our heritage and understand great that humans are lived in better culture before western people came this side and even before so called different Gods taken birth on this earth. Means humans first Gods next a must visit place for everybody
5	Love this place very awesome for photography

TABLE III. EXTRACTED REVIEWS OF SOMNATH MAHADEV TEMPLE

Sr. No.	Somnath Mahadev Temple Reviews
1	This is very good temple. This is attached to sea, so you can actually enjoy that too. No mobile phone, shorts, camera allowed. But they do provide lockers if you are carrying those.
2	This place resides at sea shore and gives a mental peace. I don't have any word to explain, attraction but just amazing. I attraction of this place is light and sound show at evening time. Must VISIT place.....
3	Visited Somnath temple on Jan 26, 2017. Its a world class monument located at unique spot. A major tourist attraction on the west coast of India bus sadly poorly managed.
4	The temple is spacious, quiet and clean. The energy there is amazing. If you are in India for a spiritual quest, this is certainly a place worth visiting. There's just temple in this town, so no need to stay overnight.
5	Wonderful temple. Visit to be cherished for lifetime. Clean place, organised with no pushing in lines. Even at busiest time of year end, the management wasn't perfect.

VI. RESULTS

TABLE IV. POLARITY OF ESSELWORLD

Sr.no.	Essel world theme park Reviews	Emotion	Class
1	Just enjoy all the rides. The place may be crowded but you should enjoy. You should walk fast in order to reach the places you wish so it is advisable that avoid going here with old people.	Joy	Neutral
2	Very boring	Unknown	Negative
3	Amazing place of Mumbai ride lover, one of the must visit place of Mumbai.	Surprise	Positive
4	Worst amusement park in Mumbai	Unknown	Neutral
5	Ridiculous place with high price	Anger	Negative

TABLE V. POLARITY OF MOHEN-JO-DARO

Sr. No.	Mohen-jo -daro Reviews	Emotion	Class
1	I want to go here.. I love this place.. this is place where Hindu religion Start.. birth place of Hinduism.. India name is come from here.. hindh name come from indus valley civilization	Joy	Positive
2	amazing place even in present times the ingenuity of the residents of this place inspires us...	Surprise	Positive
3	A mysterious place asking you to explore every inch of it. It was a fascinating evening in Moenjo Daro	Joy	Neutral

4	It's great place to experience our heritage and understand great that humans are lived in better culture before western people came this side and even before so called different Gods taken birth on this earth. Means humans first Gods next a must visit place for everybody	Joy	Positive
5	Love this place very awesome for photography	Joy	Positive

Table VI. Polarity of somnath mahadev temple

Sr. No.	Somnath Mahadev Temple Reviews	Emotion	Class
1	This is very good temple. This is attached to sea, so you can actually enjoy that too. No mobile phone, shorts, camera allowed. But they do provide lockers if you are carrying those.	Joy	Positive
2	This place resides at sea shore and gives a mental peace. I don't have any word to explain, attraction but just amazing. I attraction of this place is light and sound show at evening time. Must VISIT place.....	Surprise	Positive
3	Visited Somnath temple on Jan 26, 2017. Its a world class monument located at unique spot. A major tourist attraction on the west coast of India bus sadly poorly managed.	Sadness	Neutral
4	The temple is spacious, quiet and clean. The energy there is amazing. If you are in India for a spiritual quest, this is certainly a place worth visiting. There's just temple in this town, so no need to stay overnight.	Surprise	Positive
5	Wonderful temple. Visit to be cherished for lifetime. Clean place, organised with no pushing in lines. Even at busiest time of year end, the management wasn't perfect.	Surprise	Positive

VII. ANALYSIS

The strategy steps can be enlisted as follows:

1. **Review extraction:** The first step is to extract reviews of tourist places. We will extract review from different websites including trip-advisor and goibibo.com. We will take around 100 reviews of three different places like holy, amusement and historical place. Then we will store in .csv format. (Table [I], [II],[III])

2. **Install and load required packages in R tool:** In this step we will install **tm** package for text mining, **SnowballC** for text stemming, **RColorBrewer** for color palettes, **ggplot** for creating plot and **sentiment** package for giving a label (positive, negative, neutral here) and for other sentiment analysis.

3. **Load the .csv file:** The .csv file of reviews will be loaded into R using read.csv () function. (Table [IV], [V], [VI])

4. **Text pre-processing:** We will clean up the corpus by eliminating numbers, punctuation, white space, and by converting to lower case. In addition, we discard common stop words such as “his”, “our”, “hadn’t”, “couldn’t”, etc. We use the tm_map () function from the ‘tm’ package to this end. (Table [VII])

TABLE VII. SAMPLE OF TEXT-PROCESSING ON REVIEWS

Review	Tokenization	Stopwords removal	Stemming
Just enjoy all the rides. Don't miss any of them. The place may be rowded but you should enjoy. You should walk fast in order to reach the places you wish so it is advisable.	Just enjoy all the rides Dont miss any of them The place may be crowded but you should enjoy You should walk fast in order to reach the places you wish so it is advisable	enjoy all rides Dont miss any place may be crowded should enjoy walk fast reach places advisable	enjoy all ride Dont miss any place crowd enjoy walk fast reach place advise

5. **Transformation/Build a term-document matrix:** The process of representation figures calculated from textual data. Binary representation that is commonly used and only count a presence or absence of a word in the document. How many times a word appears in the document is also used as a weighting scheme of textual data. Here, we will create truth table.

6. **Develop a training dataset:** A dataset will be created in .csv format. This dataset will contain the text and the corresponding classes. We will use 3 labels for classifying the reviews namely positive, negative and neutral. As we are using the libraries of R tool therefore our dataset will be of .csv format. Here, we will train 70% of our data and test on rest of dataset.

7. **Classify the input using some machine algorithm:** In this stage we will classify the reviews stored in the .csv by using some machine learning algorithm. We will use Naïve Bayes for classifying. This classification will be done on the basis of predefined classification of reviews done in the .csv file. The classifier will compare the incoming review with the .csv file and then classify it in one of the category of labels.

8. **Generate the Summary, accuracy and plot:** We will develop a formula to calculate the score for a particular application and also accuracy and summary of reviews will be given by confusion matrix. Also we will generate graph for each place. The prediction accuracy of classification model is given by the proportion of the total number of correct predictions. (Fig. [II],[III] and Table [VII], [IX])

TABLE VII. POLARITY CLASS TABLE

POS	NEG	POS/NEG	BEST_FIT
3.0000	2.0000	1.4999	Neutral
3.0000	0.5000	5.9999	Positive
2.0000	1e-06	2000001	Positive
2.0000	1e-06	2000001	Positive
1e-06	1.0000	9.999991e-07	Negative

FIG.II CLASSIFICATION BASED ON POLARITY

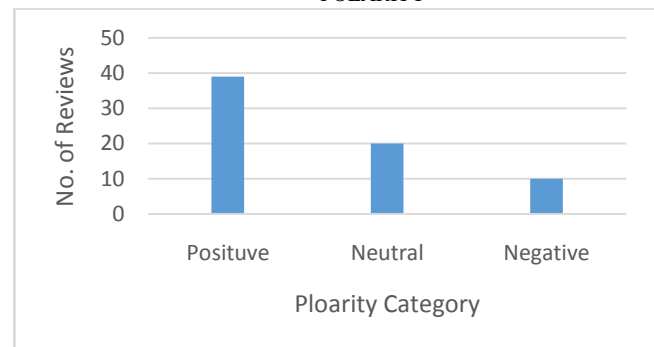


FIG. III CLASSIFICATION BASED ON EMOTION

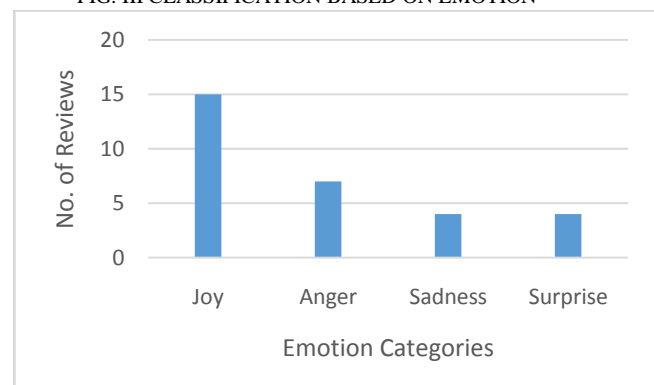


TABLE IX. ACCURACY OF ALL THREE PLACES

Sr. No.	Place name	Accuracy
1	Esselworld	55.25%
2	Mohenjo-Daro	83.95%
3	Somnath temple	67.52%

VIII. CONCLUSION

Opinions on different places are available on websites but it takes lots of time to read and analyze them therefore for that purpose sentiment analysis is needed. Result obtained from sentiment analysis decision can be taken for different places without wasting time to read reviews. In this study tested the model using Naïve Bayes Algorithm on the data of tourist place reviews which are positive or negative or neutral. Data review and 4 emotion related to the sentiment of tourist, namely joy, surprise, anger and unknown. The resulting models were tested to obtain the value of accuracy, precision, recall and our algorithm to obtain the test data by using naïve Bayes obtained accuracy is 83.00% in Mohenjo Daro's data. While other place Obtained 50% accuracy rate. It can be concluded testing Mohenjo-Daro review data using naïve Bayes is better than other two places. Thus the results of testing the model above it can be concluded that the Mohenjo-Daro provide solutions to problems of classification more accurate place reviews.

ACKNOWLEDGEMENT

We are using this opportunity to express our gratitude to our guide prof. Bhagirath Prajapati and prof. Priyanka Puvar for their aspiring guidance, indispensably constructive criticism and friendly advice throughout the work. We are fervently grateful to them for sharing truthful and irradiating views on issues of our work.

REFERENCES

- [1] movies.ndtv.com
- [2] Web opinion mining: how to extract opinions from blogs? By *ali harb, michel planti'e, g'erard dray, mathieu roche, fran,cois trousset, pascal poncelet*.
- [3] Sentiment analysis algorithms and applications: a survey by *walaa medhat,ahmed hassan,hoda korashy*
- [4] Notes on r: a programming environment for data analysis and graphics by *w. N. Venables, d. M. Smith and the r core team*.
- [5] A quick guide to teaching r programming to computational biology students by *stephen j. Eglén*.
- [6] Opinion mining of movie reviews at document level by *richa sharma, sweta nigam and rekha jain*
- [7] Tourview: sentiment based analysis on tourist domain by *deepanshi sharma, achal kulshreshtha, priyanka paygude*
- [8] Opinion mining in tourism: an emotional tagging and affective representation technique for analyzing customer response by *massimo morellato and veronica scuotto*
- [9] Applying supervised opinion mining techniques on online user reviews by *ion smeureanu, cristian bucur*
- [10] A survey on feature based opinion mining for tourism industry by *patel bhoomi, patel pankita and tanvi verm*