International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Sentiment analysis using Singular Value Decomposition

Veena Dubey* **and Dharmendra Lal Gupta**

Dept of CSED, K.N.I.T, Sultanpur (U.P.) India

Accepted 10 Aug 2016, Available online 15 Aug 2016, **Vol.6, No.4 (Aug 2016)**

Abstract

Customer Opinions play a very important role in our daily life. Opinions of other individuals are also very important component to take a decision. Now-a-days many of web users post their opinions for many products through blogs, review sites and social networking sites. Business organizations and corporate organizations are always eager to find consumer's or individual's view regarding their products, support and services. In e-commerce, online shopping and online tourism, it is very crucial to analyze the good amount of social data present on the Web automatically therefore, it is very important to create methods that automatically classify them. Sentiment Classification is another method for opinion mining which may be defined as mining and analyzing of reviews, views, emotions and opinions taken either from text, big data and speech by means of various methods. In the paper we have focused on sentiment analysis on various reviews of hotel Taj and perform singular value decomposition regression to get the estimated weights of the attributes of the hotel, which may be utilized for further rating improvement.

Keywords: *Sentiment classification, customer reviews, opinion mining, singular value decomposition.*

1. Introduction

Sentiment analysis (Ahmed Abbasi *et al*,2008) is a procedure for following the views of the clients around a specific item or subject. Sentiment analysis, which is likewise called opinion mining (Andrea Esuli,2008) ,includes in building a framework to gather and look at opinions about the item made in blog (Andrea Esuli *et al*,2005) entries, remarks, audits or tweets. The web has all of a sudden changed the way how individuals express their perspectives and opinions (Richa Sharma *et al*,2013).

They can now post audits on different sites, take an interest in discourse on different discussions, compose a site depicting their experience, redesign their status on social sites like Linkedin, Facebook, Google+ and etc. This information on audit sites, exchange gatherings, websites, and interpersonal organizations might be combined and called as customer produced substance (Lu´ıs Sarmento *et al*,2009; John Krumm *et al*,2008). Each of these customers created substance has their special property. In this paper, we focus on reviews given by the customers for hotels (a case of Mumbai Taj and Delhi Taj).

The user gives a subjective feedback to a service like hotel service, web service, schools, hospitals and also an overall rating. From the overall rating and the subjective feedback, it is hard to rate the sub-components of the services. The proposed work associates an objective rating to each sub-component of services. The estimated weights of the attributes are computed by using singular value decomposition (B Valarmathi *et al*,2011) regression.

This paper is organized into the following parts. Part II describes the data used in our work, part III describes problem definition of proposed work, Part IV describes experiment to get estimated weights as results and part V conclusion.

data set

For data set, 400 reviews of Hotel Taj, Mumbai has been collected during the period of October 2015 to January 2016 (https:// www. tripadvisor.in/ Hotel_Reviewg304554d302179ReviewsThe_Taj_Mahal _PalaceMumbai_Bombay_Maharashtra.html) and 185 reviews of Hotel Taj, Delhi has been collected from February 2016 to March 2016 (https:// www.tripadvisor.in/ Hotel_Review-g304551-d302182 Reviewsor230Taj_Mahal_HotelNew_Delhi_National_Capital_Territory_of_Delhi.html#REVIEWS).

We can take Dataset from many websites like Amazon, myntra, hotel reviews and other online shopping sites for performing Opinion Mining process which includes analyzing the reviews that are posted by the customers or users on various products. Bo Pang and Lillian Lee (Movie review data) have given collections of movie-review documents labeled with respect to their overall sentiment polarity (positive or negative).

*Corresponding author **Veena Dubey** is a M.Tech Scholar and **Dharmendra Lal Gupta** is working as Associate Professor

Problem definition of the proposed work

The SVD regression model has been used to estimate the weight of the each attributes.

$Ax = B$

Problem Definition: Estimate the weight x using SVD regression .Where A is the attribute value (matrix) and B is the actual rating (array).

Architecture of the proposed model

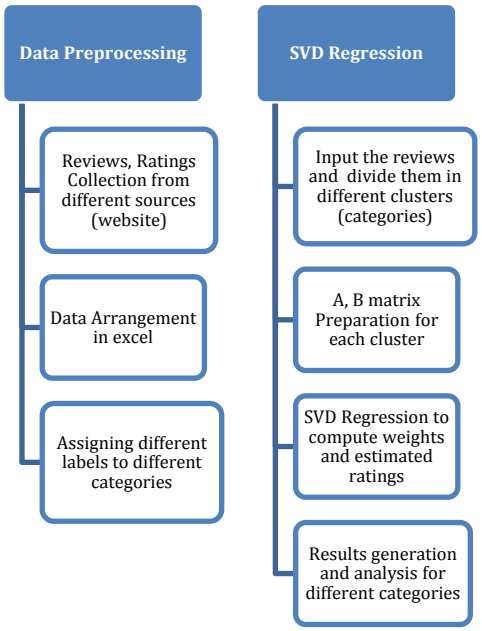

Figure 1 Architecture of the proposed model

Figure 1 represents the sequence of proposed approach work. First reviews are collected from the site and data are arranged in excel sheet such a way that each row denotes one review with first column being reviewer name, second column refers to nationality, family, overall rating given by the customer and rating against each attributes as value, location , sleep quality ,room, cleanliness and services like wise. Also we have classified the customer into four categories (Indian, Foreigner, Group and Solo). Then this sheet has been applied as input to prepare A and B metrics. Through programming SVD regression has been used to compute the weights of the attributes and estimated ratings.

Implementation and results

In this section, we discuss about proposed approach for performing opinion mining and what techniques we are going to use to perform Opinion Mining and Sentiment Analysis for getting useful information from online customer reviews. Our methodology includes the following steps:

1) Gather the subjective feedback and overall rating from the reviews of different user, consumer.
2) Collect the reviewer/user/consumer personal information for example nationality, businessman, family, friends and solo.
3) Reviewer clustering: By selecting different categories such as solo, group, couple etc. have been arranged.
4) Extract the service sub-component: according to data set there are six type of services such as value, location, sleep quality ,room, cleanliness and services. The rating of each services has been collected.
5) Identify the regression model between overall rating and service sub-component.
6) Apply the model to estimate the objective rating to service sub-component

Result analysis for Hotel Taj Mumbai

There are various methods for regression .Here we are going to use SVD. Estimated weight is computed by SVD (singular value decomposition) regression. From figure 2 analysis shows that Service attribute of the hotel has highest weight (significance) and Sleep quality has less weight for all reviews.

Figure 3 shows that the estimated rating follows the actual rating, which proves that the weights computed from SVD regression is good fit.

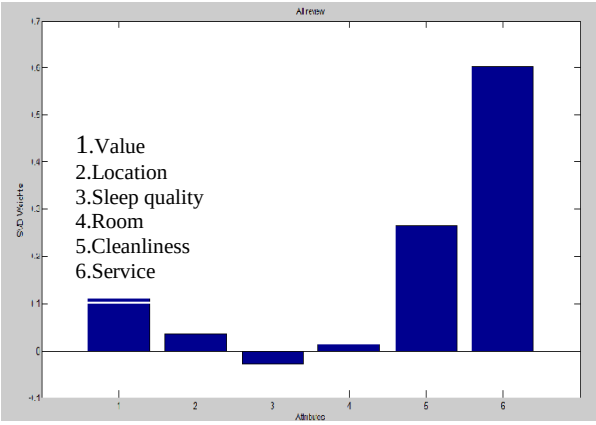

Figure 2 SVD weights of Mumbai Taj

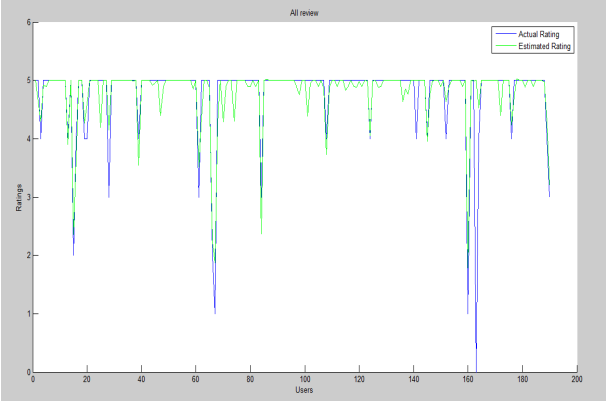

Figure 3 Actual rating and estimated rating of Mumbai Taj

Categorical results for Hotel Taj Mumbai

We have divided the reviews into four categories as Foreigner, Indian, solo and group. Comparison has been established between Foreigner v/s Indian and Solo v/s Group.

Foreigner v/s Indian

From the figure 4 and figure 5 , most of the NRI have given high ratings to service and cleanliness attributes. In overall Service attribute has the highest weight for both Foreigner and Indian. Room attributes also seems to play a role in overall rating for Indians

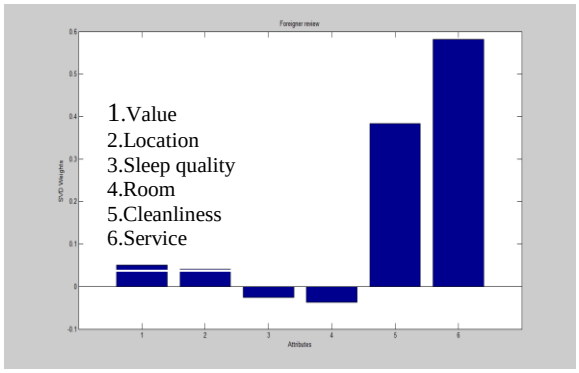

Figure 4 SVD weights of foreigner reviews

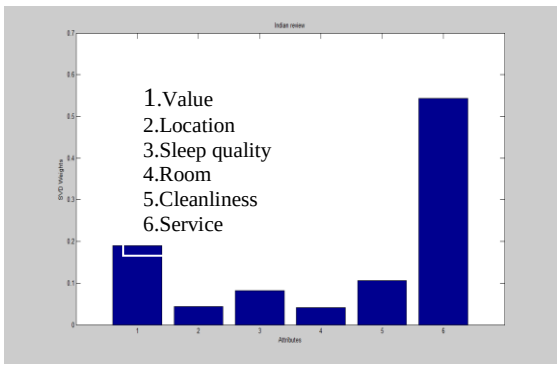

Figure 5 SVD weights of Indian reviews

Solo v/s Group

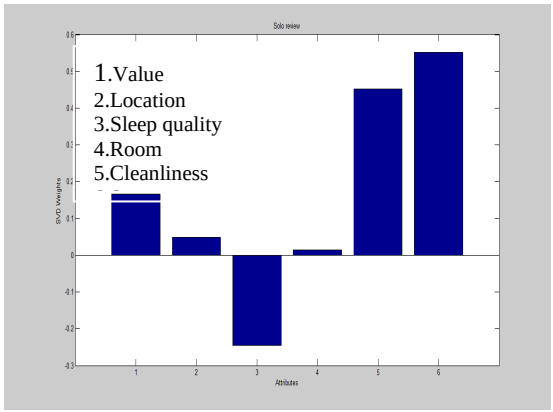

Figure 6 SVD weights for Solo reviews

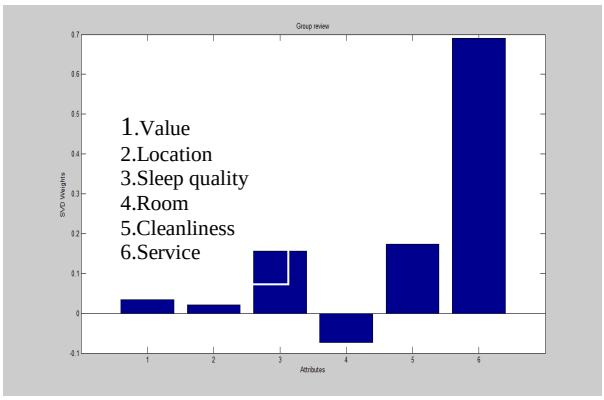

Figure 7 SVD weights for group reviews

From the Figure 6 and Figure7, Service attribute has the highest weight (significance) for both Solo and Group. For Solo, Sleep quality attributes seems to have less significance in overall rating.

Result analysis for Hotel Taj Delhi

The reviews which has been collected from the sites, for labeling of data we have used zero value for business or single, one value for family , two value for friends and three value for couples . In this section analysis has been done for Delhi Taj Hotel as above mentioned for Taj Hotel Mumbai.

Figure 8 depicts the estimated weight of service sub- component of Taj Delhi Hotel for all reviews. In it Service attribute has the highest weight and Location has less significant for Delhi visitors.

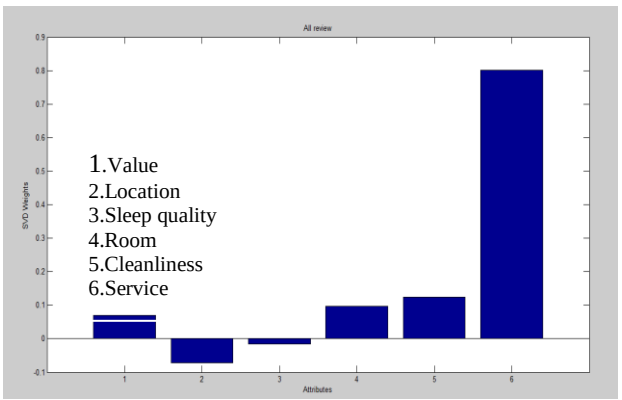

Figure 8 SVD weights of all review of Taj Delhi

Categorical results for Hotel Taj Delhi

Foreigner v/s Indian

From the figure 9 and figure 10 shows that the cleanliness attribute is most important for foreigners whereas for Indians Rooms, service and Location attributes have significance.

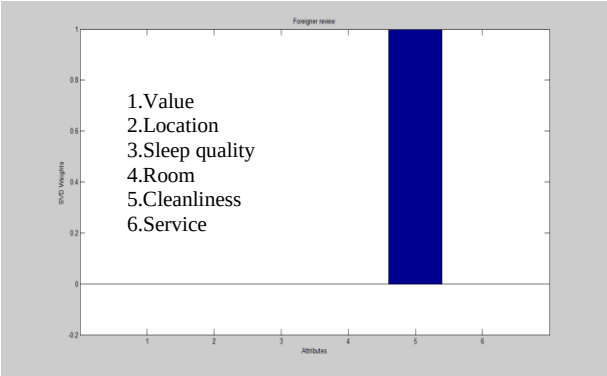

Figure 9 SVD weights of Foreigner reviews of Taj Delhi

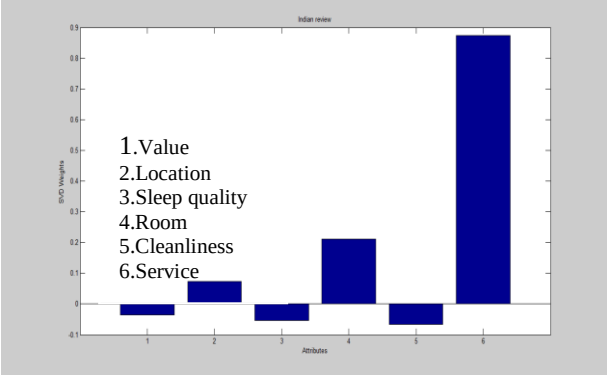

Figure 10 SVD weights of Indian reviews of Taj Delhi

Solo v/s Group

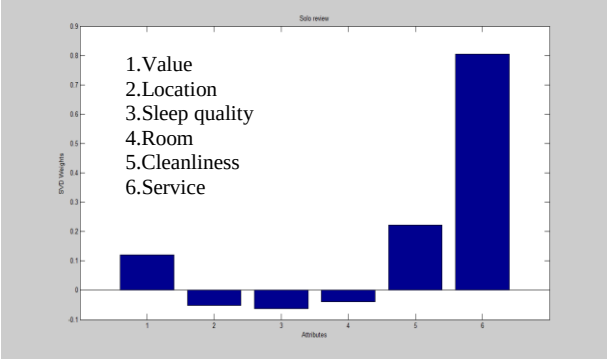

Figure 11 SVD weights of Solo reviews of Taj Delhi

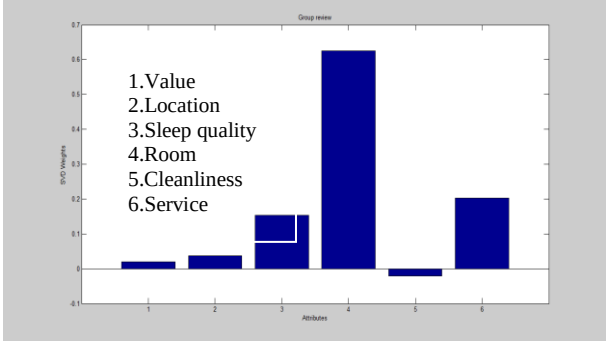

Figure 12 SVD weights of Group reviews of Taj Delhi

Figure 11 and figure 12 the analysis shows that for solo service has highest significance but for group rooms attribute has highest significance.

Accuracy calculation

Accuracy has been calculated as following:
Accuracy=100-Percentage Error.
Percentage Error=100 * mod (average Predicted Output - average Actual Output)/average Actual Output
For the accuracy calculation we have taken sample of 50 and 185 dataset of Taj hotel of Mumbai and Delhi.
From Table 1 we can conclude that accuracy can improve on larger data set. For 50 data set it is 99.85% and 99.54% where as on 185 data set it is 99.93 % and 99.91%.

Table 1 Accuracy result analysis

Sample	Taj hotel Mumbai		Accuracy (%)	Taj hotel Delhi		Accuracy (%)
	Average actual rating	Average estimated rating		Average actual rating	Average estimated rating	
50	4.68	4.687	99.85	4.429	4.449	99.54
185	4.75	4.753	99.93	4.644	4.648	99.91

Conclusion

For both the hotels, Service attribute has highest significance in overall rating. For Taj hotel Delhi, location and sleepquality has less significance provided by the customers which reveals that the improvement is required in this field. . In the paper we have shown how SVD Regression can be used in Opinion Mining process. Finally the work is concluded as following:

1) It is seen that for both Hotels, Delhi Taj and Mumbai Taj , estimated rating follows the actual rating for all the categories.
2) The weights computed can be generalized for any other reviews; it's not just limited to hotel review.
3) It is also observed that the service remains the most significance attributes for all the categories, which is quite obvious. Other minor attributes appears like Value, Cleanliness etc. for different categories.
4) The attributes weights can be utilized to further improvement of the ratings. The attributes with less values also indicate that this is still a scope of improvement, while the attribute like Service need to be maintained in the higher side for high overall rating.
5) The different attributes have different significance for different categories (NRI, Indian, Solo, Group). The service providers can set priorities among these attributes based on frequency of each category. For example, if foreigners are more

frequent customers for hotels, then the attributes which has less value for foreigners can be worked upon.

For the further study the proposed concept may apply on data from different applications like from flip cart reviews, Amazon product reviews, other social media sites etc.

References

Ahmed Abbasi, Hsinchun Chen and Arab Salem(2008), Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums, *ACM Transactions on Information Systems,*volume 26,No.3,article12.

Andrea Esuli (2008), Automatic generation of lexical resources for opinion mining: models, algorithms and applications, *ACM SIGIR Forum*, volume 42, pp 105–106

Andrea Esuli and Fabrizio Sebastiani (2005), Determining the semantic orientation of terms through gloss analysis, *14th ACM International Conference on Information and Knowledge Management*, CIKM-05, pp 617–624, Bremen,DE

Richa Sharma, Shweta Nigam, and Rekha Jain(2013), Supervised opinion mining techniques: A survey, *International Journal of Information and Computation Technology,*volume 3,pp 737-742

Lu´ıs Sarmento, Paula Carvalho, M´ario J Silva, and Eug´enio de Oliveira(2009), Automatic creation of a reference corpus for political opinion mining in user-generated content, *ACM 1st international CIKM workshop* on Topic-sentiment analysis for mass opinion, pp 29–36.

John Krumm, Nigel Davies, and Chandra Narayanaswami(2008), User-generated content. *IEEE Pervasive Computing*, 7(4) , pp10–11

B Valarmathi et al. (2011), Opinion Mining Classification Using Key Word Summarization Based on Singular Value Decomposition, *International Journal on Computer Science and Engineering (IJCSE)*, volume 3,pp 212-215.

Movie Review Data, http:// www.cs.cornell.edu/ People/pabo/movie-review-data/