



7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India

## Aspect based Sentiment Oriented Summarization of Hotel Reviews

Nadeem Akhtar<sup>a\*</sup>, Nashez Zubair<sup>a</sup>, Abhishek Kumar<sup>a</sup>, Tameem Ahmad<sup>a</sup>

<sup>a</sup>Department of Computer Engineering, Zakir Husain College of Engineering & Technology, AMU, Aligarh 202002, India

---

### Abstract

Hotel booking websites use online ratings and customer feedback to help the customer's decision making process but reviews provide a better insight about the hotel but most travellers don't have the time or patience to read all reviews. This study analyzes the hotel reviews and gives information that ratings might overlook. The reviews and metadata are crawled from website and classified into predefined classes as per some of the common aspects. Then Topic modelling technique (LDA) is applied to identify hidden information and aspects, followed by sentiment analysis on classified sentences and summarization. Finally we discuss results and future work, ultimately building towards Hotel Recommender System.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 7th International Conference on Advances in Computing & Communications.

*Keywords:* Review; sentiment analysis; topic modeling; classification; lda; MALLET; nltk;opinion mining; wordnet; sentiwordnet

---

### 1. Introduction

The Internet, in more ways than one has been a boon to many people around the world. The users tend to move towards convenient methods to find what they want and with the advent of the Internet, there has been an influx of web surfers searching for a good place to stay during their trip. The Hotel Industry, therefore is in a completely new phase now, and relies on the web for advertisement and publicity, for which there are tremendous websites available these days. Some of the prominent ones in India are TripAdvisor, GoIbibo, Yatra, etc. Each of these provide preview of what the user should expect from a certain hotel when he/she visits it or whether or not it would suit the budget.

---

\* Corresponding author. Tel.: +91-9450658150.

E-mail address: [nadeemalakhtar@gmail.com](mailto:nadeemalakhtar@gmail.com)

Mainly, what these sites rank the hotels according to are the ‘Ratings’ that some of the previous travelers provide from the feedback. Even though some of the websites might provide with ratings based on some aspect, but still research shows us that for proper customer satisfaction, interaction with service providers and other customers is necessary[1]. Also the reliability of these ratings is questionable [2]. For this, the reviews of users mentioned on the website are necessary and do provide a good insight into what the other customers experienced while visiting this particular hotel. Customers are more sensitive to personalized information that they find in the reviews and more often than not use it as a basis for decision making. Although the traditional review set is helpful but it fails when a customer’s interests fall outside of the information provided by the reviews[3].

The complications in this kind of study are plenty. There are a number of factors that make the data analysis difficult [4]. For example, the requirements of one user may vastly differ from the others; one may be on to a business trip, another on a vacation. Also the writing style of each reviewer is different from the other. Different aspects of the reviewer also affect his decision making, like their living standards, native place, lifestyle, budget, etc. Another important factor is the size of review set. Most of the customers are likely to read the first fifty or the most up voted reviews which may be missing out on some of the features described by the others.

This study analyses the data from hotel reviews and applies various different natural language processing techniques to reveal some important information that is apparently not visible to the viewer. Some predefined aspects were used to classify customer reviews into various categories and then ran over to topic modeling techniques which revealed some hidden topics over the raw classified data.

The study target for this was the website TripAdvisor ([www.TripAdvisor.com](http://www.TripAdvisor.com)) and since the dataset wasn’t freely available so a custom scrapper was built in Python to crawl customer reviews and their metadata. Discussion of a little bit of background is followed by delving into the brief concepts of text analytics. Then some related works are discussed that have been done in this field, building towards the proposed approach of handling the problem. A discussion of the experiments conducted and the received results follows, concluding towards the future scope of improvement.

## 2. Background

### 2.1. Modeling Customer Experience

Guest Experience and satisfaction is the backbone of the Hotel industry. It is a complex human experience within a hospitality setting. Customer satisfaction can be more commonly defined as interaction between his expectation and post purchase evaluation.

Modeling guest experience is a complex and challenging task because the demands and expectations of all users are not alike. Moreover depends on writing style of user. Although consumer surveys can be used to gather the data and it is very efficient but more often the customers are uninterested in writing about their experience so this method suffers from poor response rates[5].

### 2.2. Text Analytics

Due to unpredictable size of review sets and customer generated content, various text analysis techniques like sentiment analysis, opinion mining, topic modeling, classification, etc. play an important role. Classification is an important step and is a supervised method and is used to classify the text into various classes. In our study, we use it to classify the sentences of the text into predefined aspects that we already know are present in hotel reviews. Topic Modeling is a type of machine learning technique that can be used as a statistical model to discover the abstract topics in a collection of documents. It is a valuable tool to identify hidden semantic structures in the textual data and performs this function over our dataset[6].

Sentiment Analysis can be used easily for extracting opinions from the data about a certain aspect[7]. It is particularly useful for unstructured human authored documents and is a very important factor in business intelligence. It has become the central part of Information Retrieval process. The approaches towards short text summarization have improved sentiment analysis techniques [8].

Download English Version:

<https://daneshyari.com/en/article/6902341>

Download Persian Version:

<https://daneshyari.com/article/6902341>

[Daneshyari.com](https://daneshyari.com)