Text mining approach on on-line guest reviews: A case study from hospitality industry

N. I. Hewage[*] and Shantha Jayalal

Department of Industrial Management, Faculty of Science,
University of Kelaniya, Sri Lanka
*Email: ishakahewage@gmail.com

Social media and consumer-generated context on the Internet have become an integral part of the modern society with millions of users. With the increase of its popularity, tourism industry has shifted towards electronic transactions. Tourists now tend to use online social media reviews and ratings posted by hotel guests to make decisions before booking a hotel, which is an impossible task for a single user due to its high volume of reviews. Hence, to make better decisions, ranking of hotels for a specific region will be beneficial for the tourists who are willing to travel in that region and for the management of the hotel as well.

However, while a handful of studies have employed on hotel guest satisfaction and experience by analyzing online hotel guest reviews collected from online travel agencies, there is a significant research problem with ranking hotels by analyzing hotel guest reviews in aspect level consideration. Expedia.com, Agoda.com and Booking.com are some of the leading online travel agencies that have millions of users. Online reviews used for this study are collected from 18 hotels that belong to all these three online travel agencies, and from that dataset, 6 hotels are selected as testing dataset. The dataset contains reviews from year 2010 to 2016.

In this study, we propose a ranking mechanism, that ranks hotels by using the overall rating values, sentiment scores and the reviewed year. For computing sentiment scores, each review is split into sentences and they are categorized in to six attributes as Location, Service quality, Cleanliness, Comfort of rooms, Value for money and other. Thereafter, the sentiment analysis is done by considering the weight of the positive and negative words. In this research, we present a novel ranking algorithm to rank hotels, considering the reviewed year and computing the ranking score by getting the variance of the polarity rate and variance of rate of overall rating from initial year to the last. The results were taken by considering specific time period and without considering specific time period. Therefore, when using all the reviews without considering a time period, the rankings deviate from the Booking.com and TripAdvisor.com rankings. When using the reviews within 3 years of time, the ranking results are almost equal to the TripAdvisor.com rankings. When the time period is reduced for 18 months, the accuracy is 50% with Booking.com rankings and 33.3% with TripAdvisor.com rankings. This could be due to the fact that the above mentioned online travel agencies use online reviews more than two years, and therefore it perhaps causes for the deviation of rankings. Since sample dataset is used in this study, the accuracy can be increased by using a large dataset. Since this ranking mechanism considered variance to clarify the performance of hotels, and it is not only depending on the number of positive reviews or star ratings of the hotel, this is beneficial for the hotels which are not much popular, but having good standards.

Keywords: Hotel attributes, On-line hotel reviews, Social media, Text analytics