****

9B19E019

UNDERSTANDING TEXT MINING AND SENTIMENT ANALYSIS IN HOTEL BOOKING

Sakariya Ahmed wrote this exercise under the supervision of Rasha Kashef solely to provide material for class discussion. The authors do not intend to illustrate either effective or ineffective handling of a managerial situation. The authors may have disguised certain names and other identifying information to protect confidentiality.

This publication may not be transmitted, photocopied, digitized, or otherwise reproduced in any form or by any means without the permission of the copyright holder. Reproduction of this material is not covered under authorization by any reproduction rights organization. To order copies or request permission to reproduce materials, contact Ivey Publishing, Ivey Business School, Western University, London, Ontario, Canada, N6G 0N1; (t) 519.661.3208; (e) cases@ivey.ca; www.iveycases.com. Our goal is to publish materials of the highest quality; submit any errata to publishcases@ivey.ca. i1v2e5y5pubs

Copyright © 2019, Ivey Business School Foundation Version: 2019-09-06

In February 2016, a management science professor checked a search engine to book a hotel in New York City for an upcoming business trip. She found a hotel that cost US$130 per night, an appealing price, with an overall customer rating of 4.2 (out of 5) and a free breakfast. Convinced this was her best option, she decided to go ahead and book three nights. However, her experience was not as pleasant as she expected. She was confused at the hotel’s seemingly high ratings and wondered if overall customer ratings were enough evidence to decide on the best hotel to book.

Soon enough, the professor had to plan another trip to New York City. This time she was curious about looking beyond the customer ratings, as they had led her awry previously. She noted that many customers left reviews along with their scores (on the 1–5 scale), and with her expertise in text analysis, she knew that insight could be derived from this information.

HOTEL BOOKING

Finding the right hotel when deciding on a vacation was a crucial decision for many travellers. The abundance of data on the Internet made vacation planning difficult. Hotel booking websites were commonly used to search for hotels, and reviews were useful in making accommodation decisions. Most of the online reviews were available with numerical ratings (usually scaled from 1–5, where 5 meant strongly recommended). It would be more interesting to consider what previous customers had thought about a hotel and, if possible, amalgamate their opinions to make an informative decision about a given hotel. The sheer number of reviews available on the Internet was overwhelming, and hotel reviews were considered unstructured text data. Using text mining and sentiment analysis would help to structure that data, find a typical pattern in the data, easily fit the data into any machine-learning technique, and efficiently provide recommendations.

TEXT MINING

Text mining was a useful method of solving problems through pattern recognition and analysis, among other things. Text mining had become more prevalent, as the production of data had exploded exponentially, with 2.5 quintillion bytes of data being created daily! Data had become the core of almost all large business decisions, and thus the management and analysis of data were becoming crucial to organizational success. Data that we typically interacted with (spreadsheets, forms, and tables) was known as structured data, and this data could be managed with conventional database methods. Unstructured data, like textual data, came in diverse formats, which increased the analytical rigour required. Consequently, unstructured data had become increasingly difficult to analyze, as it made up 90 per cent of the data we created, and most tools dealt only in structured data that was stored efficiently and was accessible. Through various methods, text mining could derive insights and relationships from textual data.

SENTIMENT ANALYSIS

Sentiment analysis was a linguistic analysis method utilized to identify the opinion being conveyed, whether explicit or implicit, in textual materials. Sentiment analysis allowed organizations to ascertain overall public opinion, observe customer satisfaction, monitor brand reputation, and assist in customer market research. Social media platforms such as Facebook and Twitter provided brands with the ability to monitor reactions to any changes. Additionally, marketplaces like Amazon had a bevy of customers who consistently left reviews, creating data that could be valuable in product marketing decisions. In this same vein, hotel companies could derive sentiment from reviews, which would allow for better direction of resources and more informed strategic decision-making. To build a machine-learning model to conduct sentiment analysis, sentences or character strings needed to be represented in a vector space model. A vector was an object that possessed both a direction and a magnitude. Vector-based models represented similarities between objects as distances or angles. The vector space model consisted of the following steps: (1) document indexing, which allowed for extraction of only relevant terms and weighting of the indexed terms; and (2) term frequency–inverse document frequency (TF-IDF) weighting, which was one of the most common methods used by large companies like Google for search ranking.

Document Indexing

With text analysis, parsing the data for valuable content was crucial, as many words conveyed little to no meaning (e.g., “is,” “the,” “there”). The simplest indexing methods depended on a list of stop words—the high-frequency words that did not add value and could hinder analysis. Documents and queries were represented as vectors:

(2)

Dimensions represented individual terms, so if a term was included in the document, then the vector value would be non-zero. These values, or weights, were then computed using the numerous methods created to quantify these weights.

Weighting of the Indexed Terms: TF-IDF

TF-IDF was a ranking technique used for information retrieval, text analytics, and determining the importance of a word in a document. The TF-IDF weight consisted first of term frequency (TF), which computed the number of times a word appeared in a document divided by the total number of words in the document. This number increased along with the frequency of the word in the document.

(3)

The second term, inverse document frequency (IDF), was the logarithm of the total number of documents divided by the number of documents in which the specific term appeared. Unlike TF, this number was inversely correlated with a high IDF, implying rarity of a word in a collection of documents.

(4)

Combining both these equations allowed for the computation of the TF-IDF:

(5)

= weight of term *i* in document *j*   
 = frequency of term *i* in document *j*  
 = total number of documents   
 = number of documents in which *i* occurs at least once

Using the TF-IDF score, additional analysis could be conducted to extend the vector space model and calculate similarity. However, the TF-IDF score itself allowed for a final weight matrix, which provided insight into the corpus of documents so as to recognize trends and common sentiment.

For example, a document, *d*, had 100 words, and the word “hotel” appeared three times. The TF for the word “hotel” was TF (“hotel” *d*) = 3. If “hotel” appeared in 1,000 out of 10 million documents, the inverse document frequency of the term “hotel” was IDF (“hotel”) = log (10,000,000/1,000) = 4. The TF-IDF was then calculated using the equation 3 × 4 = 12 (product of TF and IDF).

ANALYZING CONSUMER REVIEWS

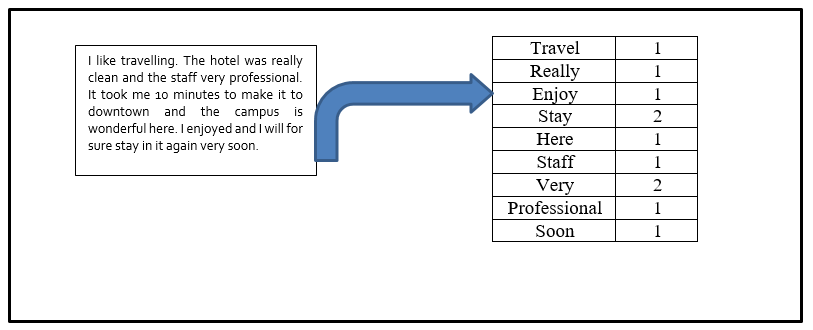
The management science professor needed to find out if the quantification of the text reviews would provide any additional insight. The professor collected 100 customer text reviews, with most reviews coming from English-speaking guests. The reviews were, on average, about 70 to 100 words in length (see Exhibit 1). Additionally, through breaking up the text, the text reviews could also be represented as lists of words (see Exhibit 2). Analyzing the text reviews indicated whether the customer ratings aligned with the textual data from the reviews. Additionally, specific trends could be present but not easily conveyed with a simple rating. With this data, the professor could better ascertain the hotel’s standing and choose a hotel in New York City with confidence.

Exhibt 1: Sample Customer reviews

|  |  |
| --- | --- |
| **Reviews** | **Text** |
| **Review 1** | I'm always travelling and have a lot of experience with different hotels. I enjoyed my time at the hotel as it met expectations. It was very clean and the hotel staff were friendly and helpful. Travelling from the hotel was very easy as it was located in the centre of the city in a beautiful landscape. Looking forward to my next stay! |
| **Review 2** | The room was very tidy and the stay was spacious and comfortable. The utilities were pretty standard however and the A/C unit was too loud. Nevertheless, none of these concerns were problematic and I had an excellent time. There is free-Wi-Fi and the hotel restaurant's food was below average. |
| **Review 3** | Was looking forward to this stay as I was traveling with a bunch of my friends for a big yearly festival. Having been here before, I only expected excellence from the hotel. My experience was once again, very pleasant and the staff were as welcoming as usual. The shower pressure wasn't great but I enjoyed the spacious area. There was less food options around than expected so make sure you do research ahead of your stay! |
| **Review 4** | I love this place! A frequent visitor, it is among my top hotels to stay at, especially when I'm in the area. As I've gotten to know the staff I can only say nice things as they always go above and beyond to help you with anything, it doesn’t have to be hotel-related. I got help with lost toothbrushes, directions, forgotten clothes, food recommendations, and coffee have always been stellar. |
| **Review 5** | My sister got really sick and was hospitalized in the city so I had to rush to Texas in order to go see her. Taking the night flight I was exhausted and was met with the friendly staff at the hotel. She provided early check-in along with a complimentary breakfast. Additionally the rooms are very large and clean, as expected. If I'm every back in the area, you can bet I'll be staying here. |
| **Review 6** | I am often in Houston and I love staying at this specific Clifton hotel. They provide flowers, call you by your name, send a letter of thanks from the hotel manager, and promptly respond to any requests. Although it may be expensive, it is well worth the money. They took care of me like royalty and I'm always itching to come back. Waiting for my next trip anxiously! |
| **Review 7** | During the NCAA semifinals I decided to come from neighbouring Dallas and book for me and my best friend. I think we weren't as diligent as we should've been with research and the location of the hotel was bad. Especially if you want to spend a lot of time in the downtown core where there are strips of high quality restaurants and pubs. The manager was great however and the complimentary continental breakfast was amazing. This hotel was good. You could do better but you could also do a lot worse. |
| **Review 8** | Coming to the Barry Hotel for the 5th time and second time this month, I really enjoy the atmosphere and the very traditional style. The staff at the hotel were so kind and very helpful in getting us situated. The location was beautiful and overlooked Rockets stadium! A great view. There was always a big NBA store nearby, and being a huge basketball fan it was a great time. I was able to get items delivered to the hotels and the staff were wonderful about it. |

Source: Created by the author.

Exhibit 2: Breaking up text in Sample Customer Reviews



Source: Created by the author.