University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Summer 7-15-2018

Multi-Dimensional Explorations into Visitors' Experience Sharing through TripAdvisor Using Social Media Analytics: An Investigation on Jasper National Park

Farshid Mirzaalian
University of Alberta, mirzaali@ualberta.ca

Elizabeth Halpenny
University of Alberta, elizabeth.halpenny@ualberta.ca

Pariya Pourabedin
pariya.porabedin@gmail.com

Follow this and additional works at: https://scholarworks.umass.edu/ttracanada_2018_conference

Part of the Marketing Commons, and the Tourism and Travel Commons

Multi-Dimensional Explorations into Visitors' Experience Sharing through TripAdvisor Using

Social Media Analytics: An Investigation on Jasper National Park

Farshid Mirzaalian[1], Elizabeth Halpenny[2], Pariya Pourabedin[3]

[1] **Farshid Mirzaalian**

PhD Student
Faculty of Kinesiology, Sport, and Recreation
University of Alberta
3-156 University Hall
Edmonton, Alberta, CANADA, T6G 2H9
Tel: 780-807-2884, Email: farshid.mirzaalian@ualberta.ca

[2] **Elizabeth Halpenny, PhD**

Associate Professor, Faculty of Kinesiology, Sport, and Recreation
2-130G University Hall, Van Vliet Complex, University of Alberta
Edmonton, Alberta, CANADA, T6G 2H9
Tel: 780-492-5702, Email: elizabeth.halpenny@ualberta.ca

[3] **Pariya Pourabedin**

B.S. in Computer Science
Edmonton, Alberta, CANADA
Email: pariya.porabedin@gmail.com

Abstract

User-generated content have been studied a lot in hospitality and tourism (H&T), however, existing studies tend to consider short samples which affects generalizability features of social media analytics research. The current study tries to examine the whole existing data on nature and park attractions of a specific destination on TripAdvisor (TA), the so-called big data. Big data analytics is a rising research paradigm that uses various data sources and analytical tools such as natural language processing and Web textual mining tools to make inferences and predictions about reality. By targeting Jasper Nation Park, around 13 K English online reviews about the natural attractions and park areas of this outdoor destination was collected, and some analytical methods such as semantic and sentiment analysis were applied on the existing corpus. Using a Latent Dirichlet Allocation topic modeling method, five major topics of visitors' experience sharing were identified as: Place, Trekking, Value, Landmarks, and Watershed. Also, results of the sentiment analysis show a high percentage of positive opinion among the reviewers (68%). This study will be of interest to the both academic and industrial realms, and helps to have a better understanding of visitor's experience sharing behaviors in online travel communities.

1. Introduction

Social media analytics has recently received considerable attention in organizational level, and companies try to analyze collected data for business purposes by merging different methods such as computational linguistics, machine learning, and statistical. The goal of social media analytics is to track trendy topics and to find popular sentiments which in turn can be useful in identifying consumers' opinions and beliefs about products and services (Fan & Gordon, 2014; Lazer et al., 2009). Existing literature on user-generated content (UGC) tend to focus on review samples which could be defined as representative of the whole existing data. This phenomenon has limited the generalizability and contribution of social media research. Considering this in mind, the current study tries to examine the whole existing data on one of the online review platforms in travel and tourism context, the so-called big data. This article tries to address following questions; what are the main topics and themes of tourists' experience sharing in the collected reviews about natural attractions and park areas of Jasper National Park (JNP)? And, what is the sentiment analysis of online reviews about parks and natural attractions of this tourism destination? Through using multiple analytical tools and various techniques such as Latent Dirichlet Allocation (LDA) topic modeling and sentiment analysis, this study seeks to answer these questions.

2. Literature review

Travel and tourism industry seems to be a potential field of social media analytics. It has been vital for tourism organizations to recognize tourist trends so that they can deliver unique services. Social media can also provide valuable sentiment (valence of an opinion) and semantic information, which is helpful in predictive analytics. Online review platforms and UGC in H&T can be categorized into community-based websites like TripAdvisor (TA), and transaction-based online travel agencies such as Booking.com (Gligorijevic, 2016; Xiang, Du, Ma, & Fan, 2017). TripAdvisor, one of the largest travel community websites in the world, has a unique feature of "Top Things to Do" for each specific tourism destination, in where travelers can limit their searching results based on different types of attractions such as "Nature and Parks", "Outdoor Activities", etc. This destination-based attribute has made TA as an appealing avenue for hospitality sectors and tourism destinations. In a study on 5K TA reviews of 843 hotels in Texas, USA, relationships between sentiment, rating, volume and variation of reviews and hotel performance was examined, and results revealed that overall and specific ratings, cleanliness, variation and volume of reviews, and the number of management responses are significantly associated with hotel performance (Xie, Zhang, & Zhang, 2014). Another research on 373 TA reviews related to Costa Rica Eco lodges found influential factors on ecotourists' satisfaction using exploratory content analysis and linear regression (Lu & Stepchenkova, 2012). Pearce and Wu (2015) also used an exploratory content analysis on 350 TA reviews about an entertaining performance at an attraction site in China, and reported that international tourists were generally positive toward the entertainment while sharing their experiences in TA.

3. Method

All the existing traveler's reviews extracted from TA about top ten natural attractions and park areas of the Jasper National Park were examined in this research. JNP is the largest national park in the Canadian Rockies and part of UNESCO's Canadian Rocky Mountain Parks World Heritage Site (Parks Canada, 2017). Top ten natural attractions and parks areas listed by TA is as follows; Athabasca Falls (3310 English reviews), Maligne Canyon (2930 ER), Columbia Ice Field (3251 ER), Maligne Lake (833 ER), Pyramid and Patricia Lakes (1149 ER), Mt. Edith Cavell (450 ER), Athabasca Glaciers (508 ER), Sunwapta Falls (215 ER), Mount Edith Cavell

Trail (137 ER), and Sunwapta Falls and Canyon (255 ER). Regarding the semantic analysis, frequent words and latent dimensions (topics) of the corpus have been presented. The valence of reviewer's opinion (positive vs. negative), sentiment analysis, has been also explored. For data extraction, client-side software named Octoparse was used (Octoparse, 2017).

4. Data analysis

After pre-processing the corpus by techniques such as tokenization, stop words removal through using Python's "nltk.tokenize" package, Natural Language Toolkit (NLTK), basic explorations like word frequencies were applied on the cleaned corpus (Xiang, Du, Ma, & Fan, 2017). Other lightweight text analytics methods such as word frequency lists, frequency distribution plots, and key word in context (KWIC) were applied through using Voyant Tools, an open source web-based text reading and analysis application (Voyant Tools, 2017). LDA topic model method, an unsupervised statistical model, was used for discovering main topics and themes of tourists' experience sharing in online reviews (Blei, Ng, & Jordan, 2003). Two pre-defined available lexicons of positive and negative words were applied on the corpus, and the valence of reviewer's opinion (positive vs. negative) was explored.

5. Results

Five most frequent words in the corpus were: glacier (6539); walk (4634); falls (4243); lake (3984); and ice (3577). Five main topics were identified as: Place, Trekking, Value, Landmarks, and Watershed from the bag of content tokens by means of LDA topic modeling method. That said, this study demonstrates that visitors of the top ten natural attractions and parks areas of the JNP tend to discuss about scenery aspects of their visit while sharing their experiences in TA. In fact, travelers to the JNP used terms such as "Glacier", "Falls", "Lake", "Ice", or "Canyon" more than other words to share their experiences with others in their reviews. A sentiment analysis based on two existing lexicons of positive and negative words was done on the aggregate online corpus, and the overall emotional valence of the whole review text was identified, suggesting that 68% of the whole reviews seem to be positive, compared to only 7% negative and 25% neutral, which verifies that visitors' overall perception of the top ten natural attractions and parks areas of the JNP is highly pleasant. Among the frequent positive words are terms like: "Beautiful", "Great", "Worth", and "Amazing". On the other hand, top negative words are terms such as: "Cold", "Busy", "Crowded", Steep", and "Deep".

6. Discussion and implications

This paper is one of the first studies which shed light for evaluating and exploring large scales of online reviews in outdoor tourism context. This study also sheds light for developing theoretical frameworks related to tourism experience through considering online reviews. In terms of practical and managerial implications for travel businesses and tourism organizations, this research helps managers to discover how visitors experienced their destination and shared their understandings with others in online platforms, what kind of topics they tend to discuss more, and finally which aspects of the destination are more perceived as either positive or negative. One of the biggest limitations for the current study is that the extracted data is only about park and natural attractions of a specific destination (JNP), and other aspects such as sights and landmarks, transportation, or outdoor activities were not included in the initial sample. Therefore, results must be interpreted carefully, and should not be generalized to other outdoor tourism destinations. Future research can include more aspects of a specific outdoor destination to develop the targeted sample. Associations between sentiment results and visitors overall rating is also one of the potential avenues for future studies.

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Fan, W., & Gordon, M. D. (2014). The power of social media analytics. Communications of the ACM, 57(6), 74-81.

Gligorijevic, B. (2016). Review platforms in destinations and hospitality. In Open Tourism (pp. 215-228). Springer Berlin Heidelberg.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... & Jebara, T. (2009). Life in the network: the coming age of computational social science. Science (New York, NY), 323(5915), 721.

Lu, W., & Stepchenkova, S. (2012). Ecotourism experiences reported online: Classification of satisfaction attributes. Tourism Management, 33(3), 702-712.

Octoparse (2017). About Octoparse. Retrieved from https://www.octoparse.com

Parks Canada (2017). About Jasper National Park. Retrieved from https://www.pc.gc.ca/en/pn-np/ab/jasper/info

Pearce, P. L., & Wu, M. Y. (2015). Entertaining international tourists: An empirical study of an iconic site in China. Journal of Hospitality & Tourism Research, 1096348015598202.

Sinclair, S. & G. Rockwell. (2017). Cirrus. Voyant Tools. Retrieved December 10, 2017, from http://voyant-tools.org

Scott, S. V., & Orlikowski, W. J. (2012). Reconfiguring relations of accountability: Materialization of social media in the travel sector. Accounting, organizations and society, 37(1), 26-40.

TripAdvisor (2017). About TripAdvisor. Retrieved from https://tripadvisor.com

Voyant Tools (2017). About Voyant-Tools. Retrieved from http://voyant-tools.org

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. Tourism Management, 58, 51-65.

Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. International Journal of Hospitality Management, 43, 1-12.