

Summer 7-9-2021

Text mining approach to explore dimensions of national parks visitors' experience and satisfaction using online customer reviews

Alireza Zolfaghari
University of Guelph, azolfagh@uoguelph.ca

HS Chris Choi
University of Guelph, hwchoi@uoguelph.ca

Follow this and additional works at: https://scholarworks.umass.edu/ttracanada_2021_conference



Part of the [Tourism and Travel Commons](#)

Zolfaghari, Alireza and Choi, HS Chris, "Text mining approach to explore dimensions of national parks visitors' experience and satisfaction using online customer reviews" (2021). *TTRA Canada 2021 Conference*. 16.

Retrieved from https://scholarworks.umass.edu/ttracanada_2021_conference/16

This Article is brought to you for free and open access by the TTRA Canada at ScholarWorks@UMass Amherst. It has been accepted for inclusion in TTRA Canada 2021 Conference by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Text mining approach to explore dimensions of national parks visitors' experience and satisfaction using online customer reviews

Introduction:

Natural parks are gaining global popularity with an estimated eight thousand million visits per year (Balmford et al., 2015). However, studies on evaluating these visitors' experiences, satisfaction, and motivations are limited to traditional methods, such as direct observations, spatial analysis using global positioning system (GPS) trackers, interviews, surveys, and focus groups. Although these methods provide useful insights, they are spatially and temporally limited and require considerable resources to conduct (da Mota & Pickering, 2020).

As an alternative to these conventional methods, user-generated content (UGC) provides available, easily accessible, and consumers' reliable recent experiences with services in multiple forms, including online ratings (number of stars) and online reviews (personal opinions in a text format) (Flanagin and Metzger, 2013). This study aims to investigate visitors' reviews of selected Canadian national parks to explore the dimensions of the visitors' quality of experience and the drivers of satisfaction based on online ratings.

Literature Review:

Existing literature differentiates between performance quality and satisfaction. Performance quality is conceptualized as the attributes of a service that a tourism supplier controls, while satisfaction refers to a tourist's emotional state after exposure to the service opportunity (Baker & Crompton, 2000). Studies that investigated factors affecting visit quality mainly discuss spatial and temporal factors, including the environment (Hausmann et al., 2018), infrastructures at the destination (Hamstead et al., 2018), viewsheds for the landscapes (Dunkel, 2015), and social-demographic characteristics of the visitors (Heikinheimo et al., 2017).

On the other hand, most visitor satisfaction studies discuss cultural ecosystem services. These services are defined as the nonmaterial benefits people obtain from ecosystems through spiritual enrichment, cognitive development, reflection, recreation, and aesthetic experiences. Studies in this category discuss aesthetic values of landscapes (Depellegrin et al., 2012), the attractiveness of wildlife (Willemen et al., 2015), and the sense of place (Ardoin, 2006).

Method:

Based on the popularity rate for travelers, a list of 10 national parks of Canada was obtained, and the reviews for each park were collected from TripAdvisor.com. The national parks studied in this research were Gros Morne, Thousand Islands, Waterton Lakes, Gaspesie, Cape Breton Highlands, Mont-Tremblant, Jasper, Yoho, Banff, and Pacific Rim. For each review, the review date, topic, body text, and the reviewer's rating for the park were extracted.

In this study, Latent Dirichlet Allocation (LDA), a generative topic modeling approach, was adopted to analyze the reviews. To analyze the data, raw reviews were preprocessed, and the model was trained with a list of preprocessed words from the reviews. LDA assumes that each document (i.e., the reviews) can be represented as a probabilistic distribution over the latent topics, and the topics form a Dirichlet prior distribution (Lucini et al., 2020). Based on the prior distribution and the pruning settings,

the model calculates a probability distribution for words in each topic. Semantic coherence was used to evaluate the performance of the LDA method.

Along with the generated themes, LDA produces a probability distribution of the topics for each review by showing the extent to which, each review is related to each topic. Logistic regression was performed using the generated scores to distinguish the topics that differentiated the satisfied visitors (4- and 5-star ratings) from the dissatisfied visitors (1- and 2-star ratings).

Findings:

In this study, 39,286 reviews were preprocessed and analyzed. Over an iterative procedure, the 16-topic model generated the highest semantic coherence score (0.5917), so it was chosen as the optimal model. Hereafter, based on the most frequent words within each topic and the top 10 reviews that best represented the topic, each topic was named. Topics contained visitors' experience with parks amenities (e.g., parking lots, information centers, gift shops, restaurants), weather and seasonality, feelings about the scenic views, crowding, and quality of trails. Appendix.1 shows the table of topics and frequently used words within each topic.

To understand which topics are more important in classifying the visitors, a structured dataset was created and divided into training and testing with the probability distributions for each review as input and ratings as output. A logistic regression classifier was trained by the training dataset and tested using the test dataset with five-fold cross-validation. The mean accuracy of the testing set was 91%.

According to the classification results, schedule and weather-related hassles (topic 3); food, tickets, and shopping experiences (topic 12); and visitor information center and exhibitions (topic 11) were among the most prevalent topics that distinguished dissatisfied visitors (1- and 2-star ratings) from others. On the other hand, positive vibes expressiveness (topic 15), photography opportunities (topic 4), and daytime and weather experience (topic 10) were the top determinants of satisfaction (4- and 5-star) for visitors. For those who were neither satisfied nor dissatisfied, parking and access routes were the most important determinant of their satisfaction class. Appendix.2 demonstrates the importance of each topic in classifying the reviewers based on their satisfaction level.

Conclusion:

This study examined the user-generated content of top Canadian national parks using a text mining method called LDA. The analysis yielded various topics ranging from visitors' pleasant feelings about trails, mountain views, and water activities to their unpleasant experiences regarding food, crowds, lineups, and parking lots.

Furthermore, this study classified the visitors based on the topics they discussed in their reviews. Results of the classification model highlighted the most important drivers of satisfaction and dissatisfaction between the visitors. This study discovered hidden underlying factors in examining national park visitors' experience quality and satisfaction and can be further examined in future empirical studies.

References:

- Ardoin, N. M. (2006). Toward an interdisciplinary understanding of place: Lessons for environmental education. *Canadian Journal of Environmental Education (CJEE)*, 11(1), 112-126.
- Baker, D. A., & Crompton, J. L. (2000). Quality, satisfaction and behavioral intentions. *Annals of tourism research*, 27(3), 785-804.

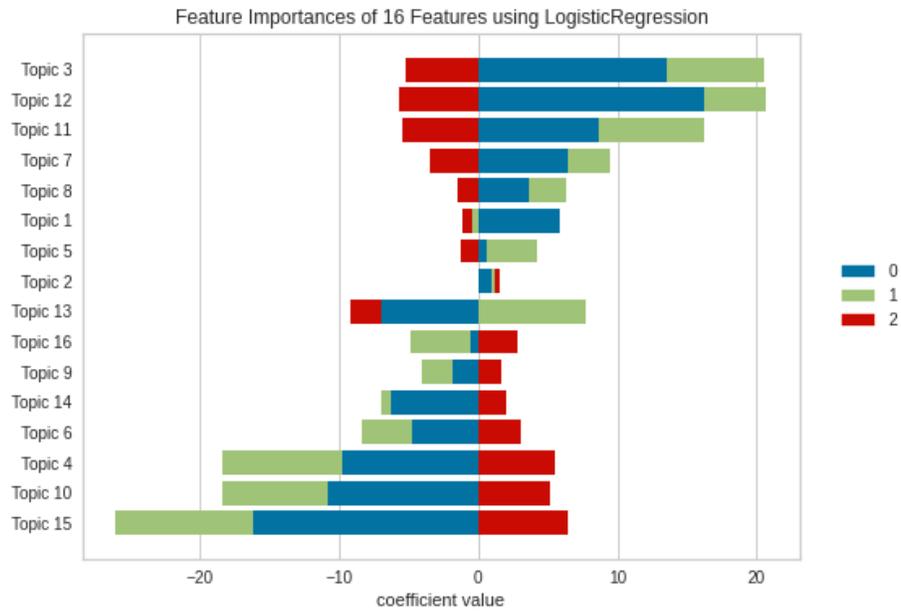
- Balmford, A., Green, J. M., Anderson, M., Beresford, J., Huang, C., Naidoo, R., ... & Manica, A. (2015). Walk on the wild side: estimating the global magnitude of visits to protected areas. *PLoS biology*, *13*(2), e1002074.
- da Mota, V. T., & Pickering, C. (2020). Using social media to assess nature-based tourism: Current research and future trends. *Journal of Outdoor Recreation and Tourism*, *30*, 100295.
- Depellegrin, D., Blažauskas, N., & Vigl, L. E. (2012, May). Aesthetic value characterization of landscapes in coastal zones. In *2012 IEEE/OES Baltic International Symposium (BALTIC)* (pp. 1-6). IEEE.
- Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. *Landscape and urban planning*, *142*, 173-186.
- Hamstead, Z. A., Fisher, D., Ilieva, R. T., Wood, S. A., McPhearson, T., & Kremer, P. (2018). Geolocated social media as a rapid indicator of park visitation and equitable park access. *Computers, Environment and Urban Systems*, *72*, 38-50.
- Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E. (2018). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conservation Letters*, *11*(1), e12343.
- Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. *ISPRS International Journal of Geo-Information*, *6*(3), 85.
- Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of pragmatics*, *59*, 210-220.
- Sarukhán, J., Whyte, A., Hassan, R., Scholes, R., Ash, N., Carpenter, S. T., ... & Raudsepp-Hearne, C. (2005). Millenium ecosystem assessment: Ecosystems and human well-being.
- Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews
- Willemen, L., Cottam, A. J., Drakou, E. G., & Burgess, N. D. (2015). Using social media to measure the contribution of red list species to the nature-based tourism potential of African protected areas. *PloS one*, *10*(6), e0129785.

Appendices

Appendix 1

TOPIC NO.	TOPIC NAME	FREQUENT TERMS
1	Group tour experience	Tour, ice, bus, experience, guide, part, informative, group, excellent, snow, special, icefield, minute, huge, driver, wonderful, young, local, big
2	Driving and wildlife watching	Drive, road, wildlife, bear, car, parkway, scenic, highway, route, black, main, animal, big, close, side, stop, chance, slow, wild, elk
3	Schedule and weather-related hassles	Time, hour, back, minute, people, cold, long, bit, half, day, couple, warm, thing, wind, car, windy, wife, return, dog, bad
4	Photography opportunities	Place, beautiful, photo, visit, lovely, great, spot, nature, beauty, area, opportunity, plenty, stunning, enjoy, love, location, wonderful, natural, camera, photograph
5	Water-based activities	Lake, boat, beautiful, island, cruise, picnic, area, peaceful, lovely, spirit, small, gorgeous, canoe, quiet, lakes, shore, large, end, kayak, rent
6	Mountain views	Top, mountain, ride, weather, station, boardwalk, peak, summit, spectacular, walkway, high, views, degree, deck, bottom, height, observation, smooth, pricey, minute
7	Winter family activities	Kid, village, activity, family, snow, ski, resort, good, year, day, hotel, hill, week, fun, mountain, lift, skiing, weekend, side, night
8	Crowdedness, seasonality, and visit admissions	Time, people, early, busy, summer, morning, winter, year, crowd, friendly, staff, season, afternoon, late, helpful, evening, visited, frozen, high, June
9	Hiking quality and challenges	Hike, easy, back, lot, steep, path, foot, part, end, short, low, pretty, high, upper, difficult, level, shoe, hard, step, effort
10	Daytime and weather experience	Day, good, amazing, beautiful, clear, spectacular, mountain, incredible, stunning, perfect, sunny, miss, sun, sight, blue, rain, lucky, sky, air
11	Visitor information center and exhibitions	Area, interesting, site, small, spring, history, hot, free, information, visitor, visit, center, center, attraction, board, pool, pass, sign, town, display
12	Food, tickets, and shopping experience	Restaurant, shop, ticket, food, price, expensive, gift, line, lunch, money, cafe, good, staff, excellent, coffee, facility, free, dinner, service, person
13	Parking and access routes	Nice, lot, walk, parking, easy, short, town, pretty, area, hotel, access, close, downtown, distance, nearby, spot, quick, pleasant, main, street
14	Water-based sceneries	Water, river, bridge, waterfall, falls, rock, fall, path, walk, impressive, accessible, easily, power, point, deep, close, stop, viewpoint, sight, powerful
15	Positive vibes expressiveness	Great, good, amazing, experience, awesome, fun, highly, fantastic, bit, wonderful, spectacular, recommend, family, glad, husband, time, clean, friendly, quick, fire
16	Description of trails	Trail, beach, hiking, side, beautiful, long, forest, tree, point, loop, Cabot, head, skyline, mile, gorgeous, landscape, moose, ocean, west, coast

Appendix 2:



Note: Class "0" : Dissatisfied (1-&2-star); Class "1" :Neither dissatisfied nor satisfied (3-star); Class "2": Satisfied (4-&5-star)