



Research & Policy Brief Series

Analyzing Online Reviews: New Tools for Evaluating Visitor Experiences

By Hari Prasad Udayapuram¹ and Srinagesh Gavirneni, Cornell University

What is the Issue?

The New York State (NYS) park system consists of 214 parks and historic sites, over 2,000 miles of trails, 67 beaches, and 8,355 campsites. It attracts approximately 60 million visitors every year. The State Office of Parks, Recreation, and Historic Preservation is responsible for operating and maintaining the state park system, and one of its strategic priorities is to “Increase, Deepen, and Improve the Visitor Experience”. Visitor feedback is integral to achieving this objective, but traditional feedback methods – public meetings, web-based surveys and comment cards – are often tedious, expensive, and limited by low participation. Public online review platforms such as TripAdvisor offer a large volume of visitor feedback that could vastly improve how NYS park managers and other community leaders concerned with tourism or business development currently understand and improve visitor experiences.

The Challenges of Utilizing Public Online Reviews

The NYS park system could develop a deeper understanding of diverse public opinions about its parks by harnessing public online reviews. However, the data on these sites lacks structure, is voluminous, and is not easily amenable to manual analysis. In order to tap into this rich source of visitor information, facility managers need new ways to get feedback online and new software tools for analysis.

A research group² from Cornell’s Johnson Graduate School of Management and the Water Resources Institute is designing, developing, and implementing software systems to tap online review content for the benefit of state agencies and the general public. Among the many online social platforms that host visitor reviews, Yelp, TripAdvisor, and Google are the most popular, and thus were used to develop a pilot decision support system for NYS park managers.

Data Collection and Analysis

Given the large volume of text associated with publicly available online visitor comments, a software system was developed to “scrape”, or copy and download, text from Yelp, Google and TripAdvisor sites. Once the reviews are gathered and their text content is cleaned and organized, a method called “sentiment analysis” is used to separate the positive reviews from the negative ones. Themes from these two sets of reviews are then extracted in a process called “topic modeling”.

Sentiment Analysis: Each phrase in a review’s text is classified as conveying either a positive or negative sentiment. These sentiment

measures are aggregated to arrive at a numerical “sentiment rating” for the review as a whole. The software uses an established dataset of words classified according to their sentiment value. For example, phrases such as “dazzling, excellent, mesmerizing, fantastic, beautiful” are categorized as positive, whereas “terrible, awful, boring” are categorized as negative.

Topic Modeling: Once the reviews are segregated by sentiment, they are analyzed to draw out useful themes, or topics. The researchers developed a software system to apply a technique known as “probabilistic topic modeling” to the reviews. Topic modeling is a statistical model used to uncover themes in large volumes of text, where the computer model assigns each word in the review to one of a specified number of topics that arise from the text itself. For example, a topic could contain the words “mountain, hike, views, trails, walk, tower” and refer to the hiking trails in the park. Another topic could contain the words “pool, zoo, picnic, tables, crowded, pretty” and refer to the picnic facilities. The software is designed to identify the most likely topics, the distribution of words in each topic, and the weight for each topic in a given review. The advantage of probabilistic topic modeling is that these topics can be automatically distilled and then used to determine managerial actions targeted towards improving the visitor experience.

An Illustrative Example

To demonstrate the utility of these tools and analytic approaches, an analysis was performed on the reviews collected for Bear Mountain State Park. Located on the west bank of the Hudson River, Bear Mountain attracts more than two million visitors per year and is visited for a wide variety of activities such as fishing, swimming, hiking, and boating. As of June 2014, there were 70 reviews on Yelp, 191 on TripAdvisor, and 34 on Google for this park. The reviews were downloaded into a database that could be searched and sorted in a cursory manual review. Automated analyses in the areas of sentiment rating and topic modeling were then applied to the more than 25,000 words of text found in the data.



Kayaking at Bear Mountain State Park.

Photo credit: Matthew Pugliese, www.flickr.com/matpugliese

¹ An independent consultant on the Johnson School research team.

² This research project was funded by the Hudson River Estuary Program, a program of the NYS Department of Environmental Conservation. The original report is available at www.wri.cals.cornell.edu.

Figure 1: Distribution of sentiment ratings among available reviews for Bear Mountain State Park

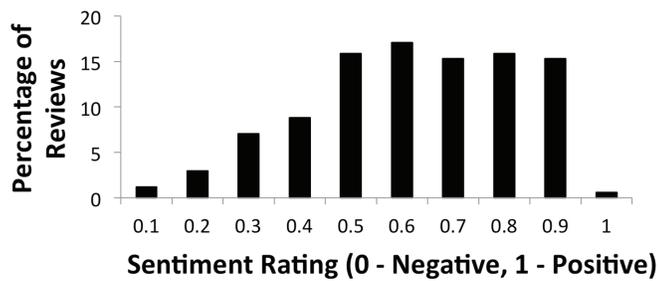


Figure 1 illustrates the distribution of sentiment among the reviews. About 45% of the reviews have a sentiment rating of 0.7 or higher, mostly positive. As an example, a review that received a sentiment rating of 0.9 contained phrases such as “beautiful attraction,” “absolutely gorgeous,” and “fun to watch.” A review that received a sentiment rating of 0.4 contained phrases such as “incredibly crowded and dirty,” “trash all over,” and “bathrooms were disgusting.”

The negative and positive reviews were then separated and probabilistic topic modeling was performed on both sets of reviews. From the negative reviews, the following major topics (with quotes from related reviews, edited for grammar) were distilled:

Crowded

...The park grounds were incredibly crowded and dirty...
 ...Arriving at the pool... we had to make a line because it was crowded going in...

Parking

...Worn out, run down, poor access/parking, filled with drunk young punks... parking was \$8! I have been to a bunch of parks and this is the first one I've had to pay.

Trash

...There was trash all over the place...
 ...There were tons of plastic cups & trash (p.s. trash bin literally 3ft away) on the floor & even in the pond...

From the positive reviews, the following topics were distilled:

Trails

...Great outdoors for all to enjoy all seasons. Trails, zoo and lake walk to be enjoyed...
 ...There are lots more trails and good bird watching at Iona Island...

Animals

...Took my granddaughter and was thrilled to see all the natural beauty... with its natural rock, flora and animals...
 ...We were able to enjoy the zoo (which is small, but cute) with plenty of animals...

Hike

...It was a nice hike (LOTS of steps, our butts felt it later!) with a beautiful view at the top...
 ...Lake, carousel, hikes, play fields, zoo, WOW!...

This type of analysis helps a facility manager to not only identify the overall sentiment about their park and facilities, but to recognize which features should be appropriately leveraged in marketing and future investments. Likewise, the analysis identifies areas that need to be addressed to improve visitor experience. Throughout this process, the topics, themes, and even language used by visitors in reviews can also be used by facility managers to target and improve marketing and communications efforts.

Ideas to Improve the Review Analysis

The research team continues to update their analysis tools, making them available and applicable to groups such as land trusts. Depending on the needs of state or local agencies and other managers of public spaces, the tools could be enhanced in several ways:

Automatic Downloading of Reviews: Online review platforms continue to increase in popularity and new reviews are submitted on a regular basis. Manual downloading is time consuming, especially for managers who want “real time” reports. A next step could be to develop a tool that automatically downloads and organizes online reviews into a database.

Topic Modeling on Negative/Positive Review Segments: The current system extracts themes from whole reviews that have been labeled as positive or negative. But reviews are rarely, if ever, completely positive or negative. Each review typically contains segments that are positive alongside segments that are negative. In order to get a more accurate collection of themes, future versions of the software could perform topic modeling on collections of review segments as opposed to whole reviews.

Topic Modeling Incorporating Expert Feedback: A topic is simply a collection of words. When the topics are chosen by computer software, some of the words in the topic may not fit according to the needs of park managers. In such cases, the managers can identify words that should be dropped from a topic and the model can be re-run. Such a recursive approach will lead to a more accurate extraction of themes and an improved analysis.

Verify Reviews: Reviews from third party online platforms, are unsolicited and often cannot be verified. With the advancement and proliferation of technologies like mobile phones and microchip wristbands, the use of devices that track key personal information are increasingly common. These devices carry important information which visitors could voluntarily share with the facility management to create verified or more detailed reviews.

Conclusion

Online reviews are a valuable resource for NYS communities, organizations, and facilities of all types with the right software to analyze them. The analyses used in the Bear Mountain State Park example could be adopted by other community and economic development leaders working on tourism or business development activities in order to better understand and connect with their clients and users. In partnership with state, regional, and local community leaders, the researchers developing these tools are interested in refining and testing their software systems further in order to better serve managers of public and community spaces.

