Technology Review on Sentiment Analysis in Tourism Industry

## Tourism Industry

Tourism is one of the billion-dollar industries that is actively involved with building experience to satisfy a wide range of personal needs. Modern technology is the key factor in the rise of user-generated content and drives the revolution of tourism industry turning into a heavily digitally supported and “omnipresent travel service network”. Millions of people read, post and share their travel experiences and travel-related opinions online which has great impacts on decisions making process of potential travelers and the reputation and success of any tourism business.

Sentiment analysis basically refers to the use of computational linguistics and natural language processing to analyze text and identify its subjective information (Alaei, Becken, & Stantic, 2017) (when broad interpretations are applied, “sentiment analysis” and “opinion mining” denote the same field of study, we use these terms interchangeably in this paper). Opinion mining or sentiment analysis is based on the idea of unlocking the hidden value of opinions to achieve deeper understanding of customers’ need and more informed and actional business insights. In the past, tourism businesses can only gain reviews through survey responses (often costly and potentially biased), but today through the power of modern technology, they can access, process and interpret large, authentic, and free-of-charge volumes of customer reviews in text format. The purpose of this paper is to review different approaches and applications of sentiment analysis in the context of tourism industry and the newly developed aspect-based approach.

# Key Methods

Over the past decades, sentiment analysis, has gained increasing attention in both industry and academia, but only recently it has entered the domain of tourism research (Alaei, Becken, & Stantic, 2017). In order to identify and extract opinions with text, sentiment analysis also extract attributes of the expression, for example (Zhai & Massun, 2016):

* Opinion holder: Whose opinion is this?
* Opinion target: What is this opinion about?
* Opinion content: What exactly is the opinion?
* Opinion context: Under what situation (e.g., time, location) was the opinion expressed?
* Opinion sentiment: What does the opinion tell us about the opinion holder’s feeling (e.g., positive vs. negative)?

Since there’re plenty of papers reviewed the sentiment analysis theories and classification algorithms, we will briefly touch this part.

Supervised machine learning approaches:

* Support Vector Machine (SVM)

A SVM is a classifier which uses annotated data for training to construct an optimal separating hyperplane/line in a multi-dimensional space which can be used to categorize new samples data into different groups. It is one of the key machine learning methods widely used for sentiment analysis.

* Naïve Bayes

A Naïve Bayes classifier is a family of probabilistic algorithms, which uses Bayes' theorem in the classifier's decision rule, with an independent assumption between features. Despite the great advances of machine learning techniques, Naïve Bayes has proven to not only simple but also fast, and reliable with natural language processing (NLP) problems.

* Decision trees

Decision tree is a non-parametric learning method that predicts the value of a target variable by learning decision rules. Tree models where the target variable can take a discreate set of values are called classification trees and leaves represent class labels and breaches represent connections of features that lead to those labels. One of the advantages of decision trees is to learn inherent rules available in the dataset that are not available to the user. And Random Forest Classiﬁer can correct decision trees’ shortcoming of overfitting to the training set by constructing a multitude of decision tress at training time.

* K-nearest neighbor (KNN) algorithm, Neural Networks and Similarity learning are all used in existing research for sentiment analysis.

Unsupervised machine learning approach:

* Cluster analysis

Clustering is the task of grouping a set of data in a way that items in a cluster are more similar to each other compared to those in other clusters. It is a general data mining technique very useful for revealing overview of the data. Clustering techniques has been applied to pattern recognition, information retrieval, data compression, and computer graphics and began to gain popularity in sentiment analysis too.

Dictionary-based approach:

* VADER

Valence Aware Dictionary for Sentiment Reasoning is a parsimonious rule-based model that requires no training data, but is constructed from a generalizable, valence-based, human-curated gold standard sentiment. Compared with other sentiment analysis lexicons, VADER performed well in the social media context and readily generalized to multiple domains (Hutto & Gilbert, 2014).

Through this brief review of a number of main classification methods employed on tourism data, we can see that researchers have been exploring and demonstrating the utility of sentiment analysis in tourism domain. Sentiment analysis has been proved to be a great supplement to tourism research methods and to improve theoretical and practical understanding of tourism industry. However, with opportunities also come challenges: some of these challenges stem from the sheer rate and volume of user generated content, some combine with text data Itself which is rich in complex and chaotic tourism experience.

## Aspect-based sentiment analysis

There are three different levels of sentiment analysis: Document-based, Sentence-based and Aspect-based. Hence Aspect-based sentiment analysis is the trending topic and this review will specifically focus on its most recent research (from 2011 till now). A traditional sentiment classification task tends to treat an entire entity as a whole, but a great challenge and major concern of Aspect-based sentiment analysis is how to estimate sentiment score for different aspects of an entity, how positive or negative the opinions are on average for each aspect and how do they contribute to the overall sentiment score.

In Hu and Liu’s KDD-2004 paper, they first proposed the Feature-Based Opinion Mining model, which was then called Aspect-Based Opinion Mining. Since then, they worked on mining two kinds of opinions (regular opinions and comparative opinions) and fake review and opinion spam detection (Liu, 2012).

Shi and Li proposed a supervised machine learning approach using unigram feature with two types of information (frequency and TF-IDF) to realize polarity classification of documents. (Shi & Li, 2011)

Brob provided some distant supervision techniques to reduce the amount of required human supervision, especially for two main subtasks of aspect-based sentiment analysis: identifying relevant product aspects and determining and classifying expressions of sentiment. He experimented these techniques with dictionary-based and supervised machine learning approaches. The task of terminology extraction at the word level and multi-label text categorization were both completed. His work presented detailed studies of sentiment lexicon acquisition and sentiment polarity classiﬁcation and a more automated process of examining semantic relationships and meaning in reviews can beneﬁt from this proposed method (Brob, 2013).

Another study extended Bing Liu’s (Liu, 2012) aspect-based opinion mining technique to apply it to the tourism domain. To address special issues that aspect-based sentiment analysis facing in tourism industry, a set of rules to determine the sentence orientation was developed. Final results showed that this extension performed better than the original model for tourism domain by improving Accuracy and Recall for the tasks of subjective and sentiment classification. Also, the approach is largely effective in finding the sentiment orientation of opinions (Marrese-Taylora, Velasqueza, Bravo-Marquezb, & Matsuoc, 2013).

A sentiment topic recognition model was introduced to distinguish objective from subjective propositions and determining sources and topics of different opinions expressed in textual data sets. The model was based on Correlated Topics Models (CTM) with Variational Expectation-Maximization (VEM) algorithm. The researchers validated the effectiveness and efficiency of this model using airline data from Twitter. They also examined the reputation of three major airlines by computing their Airline Quality Rating (AQR) based on the output from their approach (Adeborna & Siau, 2014).

Joseph and his peers implemented an Aspect-based opinion miner for restaurant reviews, which automatically found important aspects and opinions of a restaurant and ranked all restaurants by location and aspect-based reviews. Their approach used SentiWordNet, two-word phrases and linguistic rules together for opinion orientation detection, with automatic acquisition of aspects (T & Joseph, 2014).

Aurchana et al. (2014) found three main uses of data mining techniques in the tourism industry are: (1) forecasting expenditures of tourists, (2) Analyzing profiles of tourists, and (3) Forecasting number of tourist arrivals. And their paper made the conclusion that Implicit aspects play a significant role in determining sentiments from customer reviews so extracting both implicit aspects and explicit aspects would improve the accuracy of sentiment analysis results (PAurchana.P, PIyyappan.R, & PPeriyasamy.P, 2014).

Online applications:

As an interesting expansion of our technology review, we would like to see how sentiment analysis applied to some newly developed applications(those applications are not limited to tourism domain).

Sentiment Analysis on Social Media:

Twitter Advanced Search

<https://twitter.com/search-advanced?lang=en>

Twitter Advanced Search provide users a feature to tailor search results based on a range of dates, people, hashtags, and specific phrases so that business can better understand current and previous trends and capture a snapshot of what people were talking about.

NCSU Tweet Visualizer

<https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/>

This free tool takes single or multiple keywords as input and pulls tweets containing the keyword(s) form Twitter and visualize in the Sentiment in different ways. It offers a choice of visualizations: topics showing clusters, heatmap showing tweets counts, tag cloud showing common words indicating emotions, timeline showing posted time, map showing posted location and so on.

Social Mention

<http://www.socialmention.com>

Social Mention is a real-time social media search and analysis platform that monitors 100+ social and aggregates user-generated-content from across universe into a single stream of information. It enables user to easily track and measure what happened on social media landscape. It sends out reports with sentiment score, top keywords, top users, top hashtags, and source.

Social Searcher

<https://www.social-searcher.com/social-buzz/>

Social Searcher is another social media search engine. Free users can save their search results and set up email alerts. Advanced filter search is available where users can choose filter by post types or by sources. And posts are distributed by time range, sentiment, links, users, types and keywords. It recognizes multi-languages including English, German, French, Italian, Portuguese, Russian, Dutch, Spanish. Social Searcher just launched a Social Searcher Mobile App for social and web mentions search.

General sentiment analysis application:

MeaningCloud

<https://www.meaningcloud.com>

MeaningCloud implements a detailed multilingual sentiment analysis API. It can determine local and global polarity, detect aspect-based sentiment and irony, and distinguish opinions from facts. It also allows users to use custom dictionaries in sentiment classification to detect the polarity of entities and concepts they define themselves, which makes this tool applicable to any kind of scenario.

Sentiment Analyzer

<https://www.danielsoper.com/sentimentanalysis/default.aspx>

Sentiment Analyzer is another free tool that could conduct a sentiment analysis on English text. The tool is designed to provide a general-purpose sentiment score ranging from -100 to 100 and is not oriented toward any specific domain.

SentiStrength

<http://sentistrength.wlv.ac.uk/>

SentiStrength is designed to estimate the strength of positive and negative sentiment in short texts, even for informal language. It is alleged to have human-level accuracy for short social web texts in English, except political texts. It can provide a positive sentiment strength ranges from 1 (not positive) to 5 (extremely positive) and a negative sentiment strength from -1 (not negative) to -5 (extremely negative). The term weights can be optimized and other domains can be easily adjusted by adding new relevant words and sentiment strengths to the term list.

# References

Adeborna, E., & Siau, K. (2014). An Approach to Sentiment Analysis – The Case of Airline Quality Rating. *Pacific Asia Conference on Information Systems (PACIS) .*

Alaei, A., Becken, S., & Stantic, B. (2017). Sentiment analysis in tourism: Capitalising on Big Data. *Journal of Travel Research*.

Brob, J. (2013). *Aspect-oriented sentiment analysis of customer reviews using distant supervision techniques.* PhD Thesis, Department of Mathematics and Computer Science, University of Berlin.

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Association for the Advancement of Artificial Intelligence*.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining.* Synthesis Lectures on Human Language Technologies.

Marrese-Taylora, E., Velasqueza, J. D., Bravo-Marquezb, F., & Matsuoc, Y. (2013). Identifying Customer Preferences about Tourism Products using an Aspect-Based Opinion Mining Approach. *17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013*, 182 - 191.

PAurchana.P, PIyyappan.R, & PPeriyasamy.P. (2014). Sentiment Analysis in Tourism. *IJISET - International Journal of Innovative Science, Engineering & Technology*, Vol. 1 Issue 9.

Shi, H.-X., & Li, X.-J. (2011). A sentiment analysis model for hotel reviews based on supervised learning. *2011 International Conference on Machine Learning and Cybernetics.*

T, C. C., & Joseph, S. (2014). Aspect based Opinion Mining from Restaurant Reviews. *International Journal of Computer Applications*.

Zhai, C., & Massun, S. (2016). *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining.* New York: Association for Computing Machinery and Morgan & Claypool.

Team member:

* + Xin Qu: [xinq2@illinois.edu](mailto:xinq2@illinois.edu);
  + Biruo Zhao: [biruoz2@illinois.edu](mailto:biruoz2@illinois.edu)