**Tourism attraction recommendations for US major cities**

Bhagyashree Kulkarni

Sunanda Unni

Vidya Patil

**Abstract**

With increasing use of social media, it is noted quite frequently that tourists are researching about the top places to visit in a city. In addition, people love to share their thoughts on the social media about their sentiments over a location. Our goal for this project is to extract tweets from twitter, identify the top locations by doing sentiment analysis, and then predicting a score. Hence, major goal of this project is to study different algorithms for doing sentiment analysis, identifying, and extracting locations to predict the best possible tourist spots within a city.

**INTRODUCTION**

Recommendation engine and branding has been a very common business objective of Twitter sentiment analysis. We have seen a lot of papers where micro blogs (Oku, Hattori, & Kawagoe , 2015), Twitter (Shimada, Inoue, Maeda, & Endo, December 2011), yahoo searches have been extensively mined for analyzing the top tourist spots and tourist behaviors. A lot of travel sites (https://www.stratosjets.com, jatralog- Choudhury, 2016) have also invested in doing an analysis on the same to improve branding and to drive the tourism industry.

The recommendations would help the users to plan his/her trip of the target city effectively.

The characteristics of the target city would help the user in his/her decision making whether to visit the target city or not at the first place based on his inclination or liking. Indirectly this would also help local business to allocate resources in an optimized manner in the major cities for their operations. For example, transportation services like Uber, Lyft can focus on driving their operations primarily near recommended areas within the target cities.

One thing is missing in these researches is mapping the most recommended places within a city with the objective to assist the users in decision making to visit a place, to plan their trip effectively within the target city and to drive local businesses.

**LITERATURE REVIEW**

As a tourist in a city the main question that arises is to find out which places to visit. People used to rely on recommendations of family and friends, but with the popularity of online social media, people are resorting to the online information. The article by Ghani, Alowibdi, Jalal, Mokbel, & Mohamed (2014) discusses about a program called “vacation finder” which is about finding the top locations based on people’s tweets before and after the vacation. The article also discusses about travelling during specific holidays. It’s quite possible that some users might not necessarily tweet before and after leading to lot of missed information. To fill that gap, we will take into consideration user’s current location as well as any tweets for the popular destination irrespective of the location or the time.

Having access to social media has given access to so much information that aids in predicting a pattern for the travelers. The article by Abbasi, Rashidi, Maghrebi, & Waller (2015) not only talks about geolocation tweets, it also discusses about additional attributes such as departure time, traffic, location route, activity duration which helps in predicting certain patterns of the travelers. The article is very informative, but the scope of the project increases requiring more data and further analysis.

Online media is great place to get the information, but it comes with its own challenges. When users tweet messages, they do not necessarily tweet the location. Or they might have never been at the physical location. Hence, analyzing the text data getting the location could pose a challenge. As per the article by Oku, Hattori, & Kawagoe (2015), to overcome this challenge, the article also looked for pictures taken at the tourist spot. This is to ensure that the extracted data is valid, and the system can make a better prediction. The picture at the tourist spot would confirm the location. The article also refers to extracting the latitude and the longitude which provides information for a specific region which helps in getting the exact location of the spot. The article makes some good points about the user tweeting after the vacation, or maybe the tweet before the vacation. It gives some good pointers about excluding such information. The article does not take into consideration the text messages. So, in our case we will consider the text messages.

Pictures are certainly a great way to find out more about a tourist spot and its popularity. We reviewed an article by García-Palomares, Gutiérrez, & Mínguez (2015) which provided great information about extracting images from various websites such as Flickr, Instagram, Panoramio to get the geolocation of the spot with increased accuracy. The article also suggests that people could pretty much create maps referring to these images which is very impressive. The results were based on spatial popularity and the spots were dispersed in a wide region indicating some outliers as well. The pictures taken by local residents were heavily located within the cities whereas the pictures by tourists indicated spots outside the cities. We felt gathering so much images and doing this kind of analysis would be out of scope for us. This article mainly focused on image basis and missed the text messages. In addition to get the tweets with pictures, we will fill the gap of analyzing the tweets as well.

Considering the challenges of extracting information based on images, we plan to focus on text messages that would still provide great information. As per Maghrebi, Abbasi, Rashidi, & Waller (2015), text mining is easier and cheaper to implement utilizing Sentiment analysis based on the context of the text messages. The article discussed the overall advantages of text mining and sentiment analysis which we plan to implement. However, it did not focus specifically on top tourist spots. The article suggested analyzing the data with only the text messages but did not focus on mapping the distances. We will be covering that.

In the research article by Shimada, Inoue, Maeda, & Endo (December 2011), emphasis is on extracting the data from web and analyzing the tweets to find the negative/positive opinion of the tourists to get the feedback on the locations. The article takes the approach of doing sentiment analysis working with unsupervised learning by extracting the seed words from twitter. This article is aligned with our interests of creating a recommendation system for tourists to target the top spots. We found this article to be the most relevant to what we want to achieve. However, the article does not discuss about mapping the most recommended locations and giving some characteristics about the city. We plan to fill that gap.

**DATASET DESCRIPTION**

We will be using Tweepy Geo-Search API to extract last 7 days tweets for our target cities. We will filter the tweets for the current day and incrementally add it to our existing historical data in a rolling window fashion. There might be possibility that users disable or turn off Geolocation service while tweeting. To make sure we don’t miss tweets and enrich our data, we will augment that API call with the predefined hashtag patterns and user accounts related to target city to fetch additional tweets. For first time attraction recommendations in a city, we would use Third party API’s to get historical tweets data about target cities of ~2-3 months.

To build the predefined hashtag patterns and user accounts we will make use of location entity recognition or leverage Google places APIs. The user accounts would be some official tourism twitter accounts for cities which publishes vacation ideas, travel tips, & local happenings. These twitter accounts tweets will be fetched. The above module will help us to generate the characteristics of a target city.

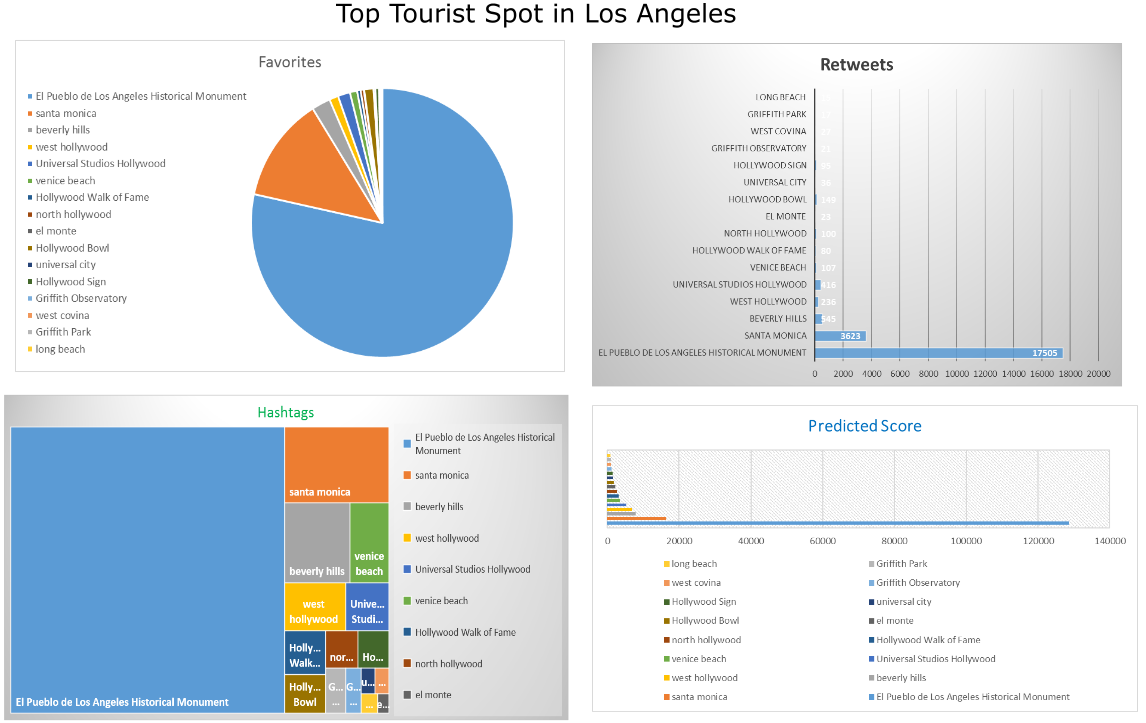
**RESEARCH DESIGN AND METHODS**

In the current scope we are only considering Los Angeles city from USA. The goal of the project is to provide top N attractions as recommendations for a target city based on the analysis of twitter user’s sentiments about the places. The system would also highlight the top N characteristics of the target city. For example, if a city has lot of beaches and amusement parks- “beaches, parks” would be displayed as some of the characteristic of the city.

We would require fetching location entities from the tweets and identify whether the sentiment is positive, negative or neutral. We would map this sentiment into ratings for each place. This would be an important feature as it would directly refer to user’s overall sentiments over a place. We will do a count for those tweets which mentions about place/location in a positive sentiment context. We will also consider count of number of retweets, favorite counts, hashtag used for that place. With the help of these features we would list down top N recommended places for target cities by twitter users.

Predefined official tourism twitter handles of the target city will be used to build characteristics of the target city. The method we will use is to extract the high frequency keywords from the tweets and map the keywords with the attraction type to get a characteristic of the place.

**VISUALIZATIONS**



**STATISTICAL ANALYSIS/RESULTS**

**TEAM MEMBERS – CONTRIBUTION**

We identified below areas of work and each section was broken down into 3 parts and handled by each one of them. The best approach was implemented after reviewing the results. So overall each of us contributed to each section in our own way.

Research

Coding

Visualization

Research Paper

**CONCLUSIONS**

The significance of our proposal is to develop a system which will analyze user’s tweets sentiment over various places and provide accurate ‘top N’ attraction recommendations of the city. Alongside recommendations we will provide characteristic of target cities like beach friendly or having lots of monuments or landmark structures. We will come up with attraction types for cities which will help user to choose its target city of their liking.

Study of such huge enormous twitter data is helpful to understand tweet sentiments around tourist attractions in major US cities. The big challenge lies in analyzing the tweet sentiments. The tweet itself is limited to 140 characters which makes it difficult to analyze the context and sentiment of the tweet. The tweet might contain informal language, local language words, misspellings, slangs which needs to be understood and handled properly.

In our design we are considering how users are thinking about their visit to some place, are they happy, are they willing to visit the same place again, or dissatisfied altogether. The tweet date and time does not necessarily guarantee the accuracy of the data, as people might tweet about visited place at later point in some other location. So, to tackle this we are also considering official tourism twitter accounts which will provide good level of accuracy about the most commonly mentioned places. The analysis from this regarding local events/activities, local museums or parks or other attractions, restaurants/café around the city would help us build characteristics of the city.

The direct advantage for the end user would be to plan his trip effectively. Besides, the study has great potential to leverage data to build n-day itineraries (1-day itinerary, 2-day or 3-day and so on) for the end user taking into consideration distances between various spots, recommended location ratings, approximate time required to spend at a recommendation. Also, the local businesses would benefit from this system.

**REFERENCES**

Abbasi, A., Rashidi, T. H., Maghrebi, M., & Waller, S. T. (2015). Utilising Location Based Social Media in Travel Survey Methods. *Proceedings of the 8th ACM SIGSPATIAL International Workshop on Location-Based Social Networks - LBSN15*. doi:10.1145/2830657.2830660

Choudhury, N. (2016, December 12). Sentiment Analysis of Twitter Data for a Tourism Recommender System in Bangladesh. Retrieved February 03, 2018, from https://aaltodoc.aalto.fi/handle/123456789/23967

García-Palomares, J. C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography,* *63*, 408-417. doi:10.1016/j.apgeog.2015.08.002

Ghani, Sohaib & Alowibdi, Jalal & F. Mokbel, Mohamed. (2014). VacationFinder: A Tool for Collecting, Analyzing, and Visualizing Geotagged Twitter Data to Find Top Vacation Spots.. . 10.1145/2755492.2755495.

K. Oku, K. Ueno and F. Hattori, "Mapping Geotagged Tweets to Tourist Spots for Recommender Systems," *2014 IIAI 3rd International Conference on Advanced Applied Informatics*, Kitakyushu, 2014, pp. 789-794. doi: 10.1109/IIAI-AAI.2014.159

Kenta Oku, Fumio Hattori, Kyoji Kawagoe, Tweet-mapping Method for Tourist Spots Based on Now-tweets and Spot-photos, Procedia Computer Science, Volume 60, 2015, Pages 1318-1327, ISSN 1877-0509,<https://doi.org/10.1016/j.procs.2015.08.202>.

Maghrebi, M., Abbasi, A., Rashidi, T. H., & Waller, S. T. (2015). Complementing Travel Diary Surveys with Twitter Data: Application of Text Mining Techniques on Activity Location, Type and Time. *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. doi:10.1109/itsc.2015.43

Shimada, K., Inoue, S., Maeda, H., & Endo, T. (2011, December). Analyzing tourism information on twitter for a local city. In *Software and Network Engineering (SSNE), 2011 First ACIS International Symposium on* (pp. 61-66). IEEE.