Mustafa Telab

CUNY SPS Capstone

Professor Dr. Bailo

9/23/22

#getmewhere

[working title]

The goal of this project is to build a tool that builds a custom vacation itinerary based on a simple personality test. The bedrock data for this tool will user review/posts related to hotels, restaurants, vistas, museums, and more.

Relevant Literature:

[*Approach for Social Media Content-Based Analysis for Vacation Resorts*](https://hrcak.srce.hr/file/329140)

*JOURNAL OF COMMUNICATIONS SOFTWARE AND SYSTEMS, VOL. 15, NO. 3, SEPTEMBER 2019*

This is a targeted and detailed article that directly aligns with the text mining and sentiment analysis of this project. The authors outline their process for producing a sentiment analysis of tweets regarding vacation resorts during a particular holiday period over the course of three years. It begins with an interesting consideration of the APIs that Twitter offers. They also touch on a few available text processing software and why they chose to use RapidMiner. Finally, the paper walks through the chosen machine learning techniques and their implementation.

[*VacationFinder: A Tool for Collecting, Analyzing, and Visualizing Geotagged Twitter Data to Find Top Vacation Spots*](https://www-users.cse.umn.edu/~mokbel/papers/lbsn14.pdf)

[*Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks*](https://dl.acm.org/doi/proceedings/10.1145/2755492)*November 2014*

“Vacation Finder” serves as a helpful reference for my initiative. Although the authors offer less details in the way of code and implementation, they touch on one of the key challenges in my project, geography. Rather than searching the text for locations or venues, geotags are used to label the associated sentiment.

*Predicting Personality with Social Media*

[*Extended Abstracts on Human Factors in Computing Systems*](https://dl.acm.org/doi/proceedings/10.1145/1979742)*May 2011*

The authors tackle the challenge of deducing a personality profile from a social media profile. Their approach sparks interest because they are able to pull information from data outside of the strict text content such as number of friends or decision to include certain personal data in their bio.

Problem Statement:

Vacations are a significant expense for the average household. This makes trip planning a stressful and length process. It often starts with extensive research into locations, activities, restaurants, site seeing, and beyond. The problem with this approach is that someone can do extensive research on what other people enjoyed but how can that be translated to what the individual wants.

Solution:

The solution begins with a personality test, and a short list of questions related to the potential trip. The application should isolate a location and build an itinerary that optimizes utility with a collection of appointments of various categories according to the users profile.

Process:

1. Text mine social media platforms - Gather data related to vacation elements attached to a particular location.
2. Sentiment Analysis – Establish logistic scoring format towards the vacation element.
3. Network Analysis – Build personality profile to associate with the data points.
4. Recommender System
   1. Classifier to place in personality bucket
   2. Regression analysis to rank top vacation elements for user.
   3. Optimization algorithm to build itinerary off of results.

Considerations:

* Establishing rules for itinerary configuration(I.e. cant go on three hikes in one day)
* Establish reasonable scope
* Factor in seasonality into model
* Social media platform selection