Mustafa Telab

CUNY SPS Capstone

Professor Dr. Bailo

10/16/22

#getmewhere

Abstract:

The travel industry present distinct challenges to the consumer. The cost is high, and there is no avenue for refund based on any lack of enjoyment. Travel is rare, so the consumer is often unable to draw from past experience to make an informed decision. And lastly, it is difficult to identify the attributed of a particular location that would contribute the most towards satisfaction. This paper surrounds the idea of creating a passive recommendation for consumers based on their social media habits and that of their peers.

Key words:

Travel, twitter, recommender, NLP, mining

Literature review:

Naturally, there is a robust academic body of knowledge surrounding the travel recommender systems given the economic and social significance of the industry.

An important concept that repeats across the literature is the “cold start problem.” The idea surrounds the questions of how you include elements in a recommender system that do not have extensive data or any data at all. In the context of this paper that may include new resorts, restaurants and even the users themselves. The article “The long tail of recommender systems and how to leverage it” describes their approach for handling the sparse items in a system by categorizing them as such[[1]](#endnote-1). Our approach will pursue expanding on this idea, as it seems less “popular” travel items will likely be suited for certain personality affinities.

In regard to collecting data, this paper will draw insights from the authors of “Recommendations in location-based social networks: a survey[[2]](#endnote-2).” As the title suggest, the paper surrounds the topic of drawing geographically linked data from social media. The authors provide a list of social media platforms and methods for locational analysis outside of geo tagging such as, travel paths, visit sequence, duration, and trajectories.

Ulrike Gretzel and contributors describe their methodology towards developing a recomender system based off a personality testing[[3]](#endnote-3). The foundation of their data was a generic personality test mailed out to participants. The participants also were asked to describe their travel/vacation habits and preferences. Analysis was done on the results to classify personality traits into 12 different travel profiles(City Slicker,Family Guy, Avid Athelete, etc). This paper is similar in the area of personality testing. However, our aim is to remove assumptions of prototypyipcal travel habits and instead let the avaiable sentiment data drive the clustering through unsupervised learning. Drawing from Burke’s “Hybrid” approach [[4]](#endnote-4), we will attempt to combine the sentiment results with demographic and relationship data.

Not to be overlooked is the challenge of arranging recommendations in a schedule that taking into account time, distance and diversity. This is a combinatorial challenge that is addressed in "The city trip planner: an expert system for tourists[[5]](#endnote-5). The authors forgo more computational intesive algothyms to settle for the “Greedy Randomised Adaptive Search Procedure” (GRASP). GRASP iterates through combinations of activities within different time slots untill a certain utility measure hits a given threshold. Our approach will instead be modeled off reserch on trip planning featured in the journal Tourism Analysis[[6]](#endnote-6). The authors found that the typical approach to trip planning contain three components; Pre-planned core decisions, pre-planned flexible decisions, and improptu. With that in mind, our model will establish core recommendations for each day; then insert secondary activities into avaible time slots. To avoid overloading the schedule, we can draw techniques described by Li[[7]](#endnote-7) to infer typical duration of the activity.

For sake of simplicity, we will forgo providing any type of ranked recommendations in the results. This decision also stems from observations that users lose interest in lower ranked items faster than a linear proportion[[8]](#endnote-8).

It is well to assume that the user is unlikely to adopt the trip recommendation verbatem. Manual edits to the itinerary by the user can serve as valuable information to future models as discussed in “Recommending improved configurations for complex objects with an application in travel planning[[9]](#endnote-9).” However, this layer of data collection is out of scope for this paper.

Methodology: discuss the key aspects of your model/problem. If you work on real data, explain how you verified your model. If you work on theoretical aspects of data science advancement, explain your method with details (you can put the theory and proof in the appendix if you see fit).

The first challenge of this initiative is developing the dataset. What the study requires is a collection of geo located entries that are attached to a particular individuals social media account. In addition, a body of social media textual history is necessary to build a model adjacent to the location data. In recent years, twitter and other social media companies have increased restrictions on the data they provide. This poses a challenge for this study in that Twitter’s base API does is not optimal for the amount of data needed to draw concrete results. Therefore, we will continue with a smaller sample that will not provide adequate geo-density, but will hopefully build a scaffolding of the approach to be tackled at a later date.

In reviewing the sample of raw tweets, it confirms that geo-located posts are rather rare <1%.

* Results and/or experimentation: describe what you did (validation, experimental design, study of the theories, etc.), and what you found out (statistical analysis, comparison of various approaches, system performance improvement, optimization, new algorithm development, etc.).
* Summary and future works: conclude your findings, suggest areas for future work.
* Appendices (put your detailed code here.)

1. Yoon-Joo Park and Alexander Tuzhilin. 2008. The long tail of recommender systems and how to leverage it. In Proceedings of the 2008 ACM conference on Recommender systems (RecSys '08). Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/1454008.1454012 [↑](#endnote-ref-1)
2. J. Bao, Y. Zheng, D. Wilkie, M. Mokbel, "Recommendations in location-based social networks: a survey," GeoInformatica , vol. 19, no. 3, pp. 525-565, 2015. [↑](#endnote-ref-2)
3. Gretzel, Ulrike & Mitsche, Nicole & Hwang, Yeong-Hyeon & Fesenmaier, D.R.. (2006). Travel personality testing for destination recommendation systems. 10.1079/9780851990231.0121. [↑](#endnote-ref-3)
4. Burke, R. Hybrid Recommender Systems: Survey and Experiments. User Model User-Adap Inter 12, 331–370 (2002). <https://doi.org/10.1023/A:1021240730564> [↑](#endnote-ref-4)
5. P. Vansteenwegen, W. Souffriau, G. Vanden Berghe, D. Van Oudheusden, "The city trip planner: an expert system for tourists," Expert Systems with Applications , vol. 38, no. 6, pp. 6540-6546, 2011. [↑](#endnote-ref-5)
6. Fesenmaier, D.R., & JiannMin, J. (2000). Assessing structure in the pleasure trip planning process. Tourism Analysis, 5, 13-27. [↑](#endnote-ref-6)
7. Li, CT., Chen, HY., Chen, RH. et al. On route planning by inferring visiting time, modeling user preferences, and mining representative trip patterns. Knowl Inf Syst 56, 581–611 (2018) [↑](#endnote-ref-7)
8. Renjith, S., Sreekumar, A., & Jathavedan, M. (2020). An extensive study on the evolution of context-aware personalized travel recommender systems. Information Processing & Management, 57(1), 102078. [↑](#endnote-ref-8)
9. A. Savir, R. I. Brafman, G. Shani, "Recommending improved configurations for complex objects with an application in travel planning," in Proceedings of the 7th ACM Conference on Recommender Systems (RecSys '13), pp. 391-394, Hong Kong, October 2013. [↑](#endnote-ref-9)