CSE6242 Data & Visual Analytics

Progress Report

Team 30

All-in-One Travel Recommendation

Huizi Shao

Jayden Sun

Jiayuan Bi

Shangru Yi

Yuanxun Shao

Yue Li

Georgia Institute of Technology

11-10-2019

Table of Contents

1. Introduction

Currently, users have to search for different information in separate apps, including Airbnb, Yelp and Google Maps, and manually collect information on their own. This introduce negative user experience since the user needs to switch among Apps and overwhelmed by redundant information. Furthermore, the lack of integration might end up with low-quality information. Our proposed all-in-one travel recommendation App aims to reduce the tedious process in one shot and also enhance the recommendation quality. We will combine various traveling recommendation tasks into a single application with map visualizations and all-in-one recommendations based on user preference and historic reviews.

2. Survey

2.1 User Interface Design

1. User Study

We can conduct our user study based on categories at purpose, users, tasks, setup, the procedure to get the general information for the usability of our developed website [17]. Paper [6] shows that trust-based predicted rating can provide better recommendations for the users.

2. UI Design

The creation of a high-quality user interface implies the principle "the user is above all" [1]. Several UI design principles including familiar context, consistency, and strong visual hierarchy are introduced within this article. We can follow this guideline for our UI and interactive design.

2.2 Recommendation System

1. General Recommendation Systems

Generalized recommend System does not meet diverse demands. We can cluster the information and provide more insight for personalized recommendations [5]. Moreover, by combining the user's preference for dining, traveling, and living, we can offer a mixed recommendation for multiple purposes.

2. Recommendation Diversity

Diversity is important for recommendation systems to influence the behaviors of users. Perceived categorical diversity has a significant coloration for the usefulness of systems [13].Thus we can follow the guidelines to design our website.

2.3 ML in Recommendation System

1. Challenges in Recommendation Systems

Due to limitations associated with correlation-based models, such as sparsity, scalability, and synonymy [18], machine learning has been widely applied to reduce dimensions in recommendation systems.

2. Machine Learning Algorithm for Recommendation Systems

For efficient and high-quality personalized recommendations, we need to discover aggregated user-profiles and cluster user activities. Based on the traditional K-means algorithm [16, 7], a new clustering method combining DBSCAN and Chameleon further boosts the performance on web pages [10].

2.4 Document Database

1. Survey on NoSQL Database

NoSQL database, characterized by the efficient big data storage and high scalability, is increasingly popular in internet development. Especially, document databases with key-value using JSON, like MongoDB, are able to support complex data types and high-speed mass data access in real-life big data projects [9].

2. Extensive Document Query and Clustering

Internet information is presented in text format so that document query and clustering are particularly useful for handling search engine results. Current automatic keyphrase extraction algorithms involve heuristic candidate phrase extraction and phrase selection [11].

3. Proposed method

3.1 Data

3.1.1 Airbnb Datasets

There are two datasets being used from the website provided by Airbnb. The first is called listing, which includes some general information. The second dataset is called the reviews, which includes some review details. SQLite and Open-Refine are being used.

3.1.2 Yelp Datasets

The datasets are collected from Yelp dataset challenge which consist of both business information and their reviews. For this project, the business is filtered with location within Nevada and only restaurants. Similarity, the reviews are filtered using the business id from the filtered business dataset. Databricks and cPython scripts are being used.

3.1.3 Database Design

With large dataset and flexibility of NoSQL database, the database is designed using the normalized data models. The database consists of four tables, two tables for airbnb and yelp business information and two tables for their corresponding reviews. Each table is identified using business id for fast lookup.

3.2 Recommendation ranking

We would like to recommend some merchants to users. One possible way is to recommend merchants which are similar to that users have lived before. We implement item-based collaborative filtering method in our system. Collaborative filtering makes use of rating data from user only.

For collaborative filtering method, given a partially-observed noisy matrix M, we would like to approximately complete it and estimate unrated items. Item-based collaborative filtering is that find preferred items by similar items to this one.

$$\hat{r_{u,i}} = \frac{1}{\sum_{v \in S_i} sim(i,v)} \sum_{v \in S_i} sim(i,v) r_{u,v}$$

Where $\hat{r_{u,i}}$ is a rating on item i by user u;

$sim(i,v)$ is similarity function between item i and item u;

S_i is a set of similar items to item i.

A more straightforward way to illustrate that:

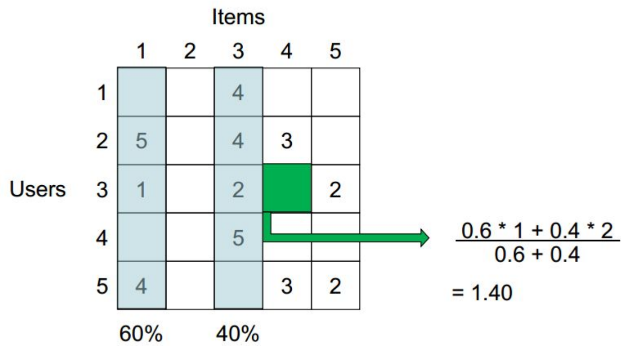


Figure 3.2 illustration for item-based collaborative filtering

Item similarity function:
We already cleaned data which contains each merchants in Los Vegas, including features like location, room type, price, availability, etc.

We define the similarity function in the forms of

$$sim\,(i,v) = e^{-(d(i,v))^2}$$

Where $d(i,v)$ is the distance of features, i.e., if x_i is the feature vector of the first item, and x_v is the feature vector of the second item, then $d\,(i,v) = \,\|\, x_i - x_v \,\|$. We manually assign numbers for each non-numerical features, which is easier for distance calculations.

3.3 Map

3.3.1 Interface Design

The interface consists of three main components: 1. preference interface, 2. map visualization and 3. information dashboard. The preference interface provides support and visualization for the preference search conducted by users. Users can provide keywords and features for retrieving proper results and visualize surrounding traveling locations.

3.3.2 Map Visualization

Our project focuses on the exploration of Las Vegas, one of the most famous travel cities in the United States. The function builds on the foundation of Leaflet.js and Turf.js. Leaflet is one of the leading open-source libraries for interactive maps. Turf focuses on geo and feature analysis. The map provides direct spatial information about the region where the project provides service on. By getting digital map data from Mapbox service, the topographic information is shown, including road network information and range of the region.

By showing attraction pictures, users can easily pick the places they are interested in and the hotels and restaurants nearby will show up on the map. Thereby, users can have a view of the nearby restaurants and hotels within a given distance range. The nearby restaurants and hotels can easily be selected by one single click and unselected by another click on the mark. As user selecting locations in the map, a distance-based geo range for selected locations is shown. With selections, a union geo-range can be retrieved. This distance range functions aim to provide restaurants and hotels in a feasible range. With the range, users can just consider the restaurants and hotels in a feasible range. Moreover, we also provide this function when users select restaurants and hotels so that users can further consider nearby hotels and restaurants. By default, the searching range of attractions is 5 km for local attractions and 3 km for Airbnb hosts and restaurants. In a word, the map intends to provide infrastructure and visualization for our proceeding recommendation.

3.3.3 Preference List

Users can use the preference list to select the restaurants and hotels that they may be interested in. With the information provided by the users, such as price, rating and some keywords, our recommendation system can pick the corresponding restaurants and hotels in the region. The extracted features and characters using NLP techniques and clustering in our recommendation algorithm will be used for location search (preference list is still under development and more detailed information about preference list is unavailable right now).

3.3.4 Information Dashboard

In some circumstances, users only cares about ratings. Since these users may prefer to get a list of restaurants, recreation, and hotels in order of rating so that they can just pick the highest rating places. By clicking one of the three tabs, users can get the ordered lists of restaurants, attractions and hotels ranking.

4. Conclusion

4.1 Datasets

4.1.1 Airbnb

There are 2487 sets of information included in the listing dataset and 6752 sets for the review dataset.

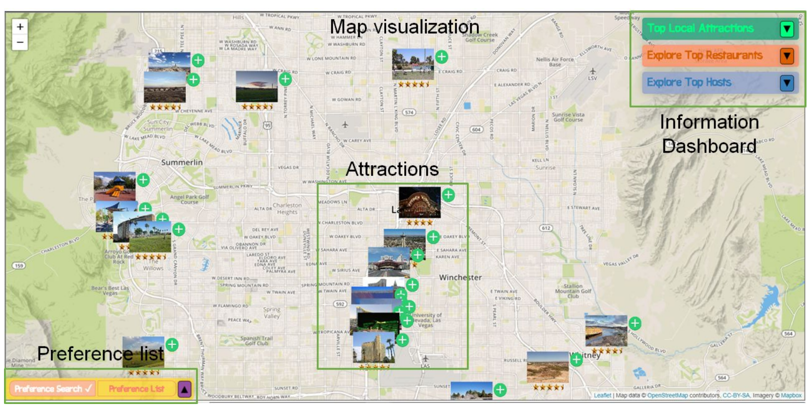
4.1.2 Yelp

The dataset includes 7398 restaurants in Nevada with 1.4 million different reviews of different restaurants. Most of the restaurants are around the Los Vegas area.

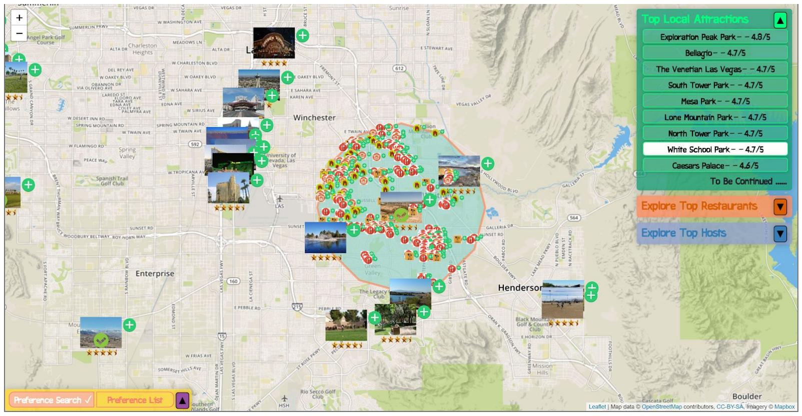
4.2 Map development

Three major functions of map interface are provided, including map visualization, preference list and information dashboard. The map intends to provide infrastructure and visualization for our recommendation.

Welcome page

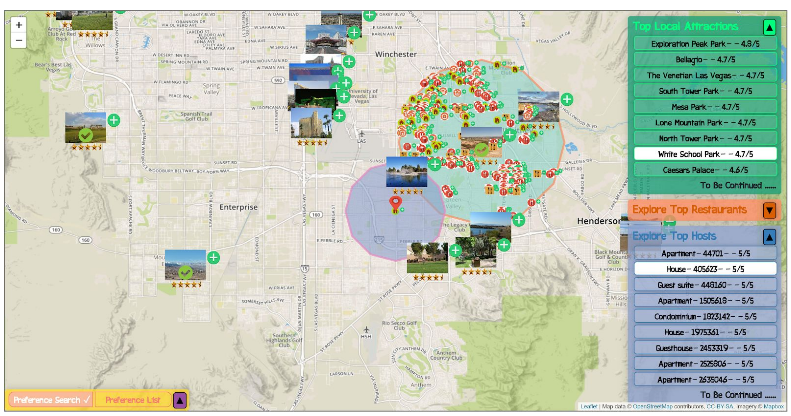


Map functions

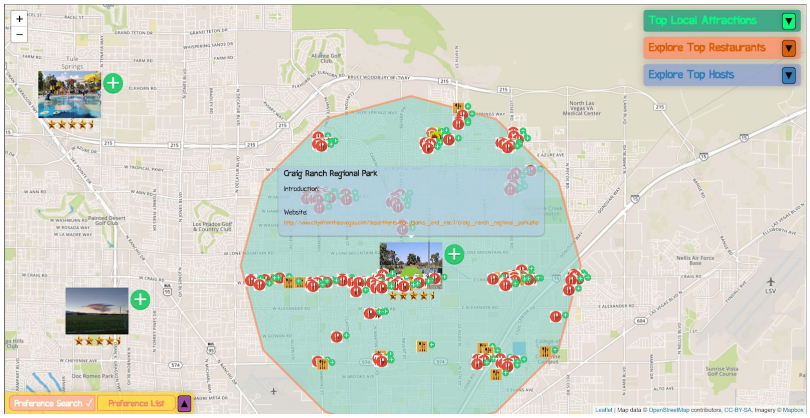


Top attraction

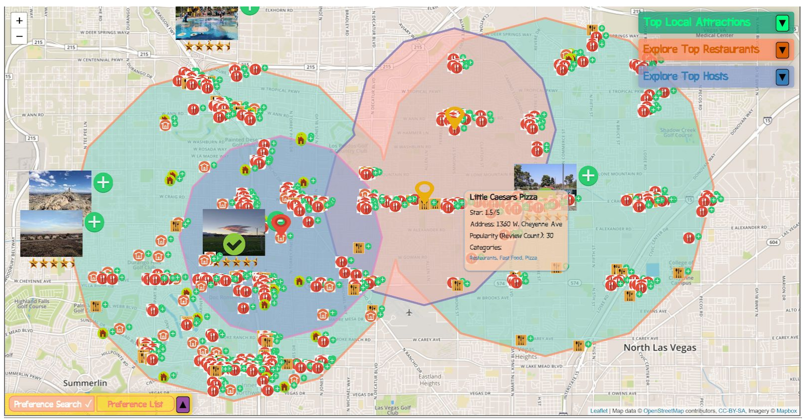
Top Restaurant



Top host



Local attraction selected zoom in



Multiple selection and their related position

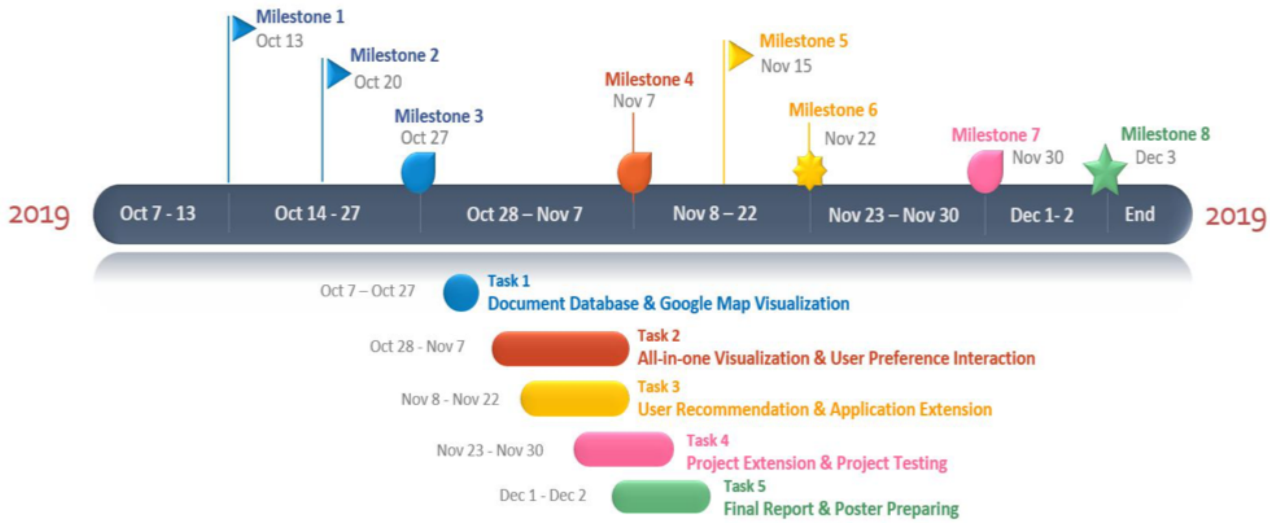
Preference search (not implemented yet)

5. Plan of activities

5.1 Algorithms (Same)

We implement item-based collaborative filtering to provide recommendation and then conduct users' classification by NLP and clustering.

5.2 Scheduled Implementation Agenda (Same)



5.3 Labor Division

Old：

	Front End Visual	Document DB	Recommendation
Huizi Shao	😊		
Jayden Sun	😊		
Jiayuan Bi			😊
Shangru Yi		😊	
Yuanxun Shao		😊	
Yue Li			😊

New :

	Data	Map	Recommendation
Huizi Shao	😊		
Jayden Sun	😊		
Jiayuan Bi		😊	
Shangru Yi		😊	
Yuanxun Shao			😊
Yue Li			😊

5.4 Tools

6. Innovation

6.1 Visualization & User Interaction

The map visualization uses tourist attraction to filter both airbnb and restaurant choices, which allow tourists to quickly locate the place they want to stay and dining nearby. In addition, the application will recommend user similar airbnb or restaurants based on user's previous history.

6.2 Recommendation System

We recommend new merchants that are similar to the merchants where users lived or visited before, which means provide personalized services for each customer. This time we implement item-based collaborative filtering to provide recommendation and then conduct users' classification by NLP and clustering.

References

[1] Top ui design principles to keep in mind. https://uxplanet.org/top-ui-design-principles-to-keep-in-mind-bfb3ad8790c6. Accessed: 2010-09-30.

[2] Elie Aljalbout, Vladimir Golkov, Yawar Siddiqui, and Daniel Cremers. Clustering with deep learning: Taxonomy and new methods. CoRR, abs/1801.07648, 2018.

[3] Liangliang Cao, Jiebo Luo, Andrew Gallagher, Xin Jin, Jiawei Han, and Thomas S Huang. A world-wide tourism recommendation system based on geo-tagged web photos. In 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, pages 2274–2277. IEEE, 2010.

[4] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, Computer Vision – ECCV 2018, Cham, 2018. Springer International Publishing.

[5] Mingming Cheng and Xin Jin. What do airbnb users care about? an analysis of online review comments. International Journal of Hospitality Management, 76:58–70, 2019.

[6] J Golbeck and J Hendler. Filmtrust: movie recommendations using trust in web-based social networks. In CCNC 2006. 2006 3rd IEEE Consumer Communications and Networking Conference, 2006, volume 1, pages 282–286. IEEE, 2006.

[7] Abdelghani Guerbas, Omar Addam, Omar Zaarour, Mohamad Nagi, Ahmad Elhajj, Mick Ridley, and Reda Alhajj. Effective web log mining and online navigational pattern prediction. knowledge-based systems, 49:50–62, 2013.

[8] Khaled M Hammouda and Mohamed S Kamel. Efficient phrase-based document index-ing for web document clustering. IEEE Transactions on knowledge and data engineering, 16(10):1279–1296, 2004.

[9] Jing Han, E Haihong, Guan Le, and Jian Du. Survey on nosql database. In 2011 6th international conference on pervasive computing and applications, pages 363–366. IEEE, 2011.

[10] Sun Hao, Shen Zhaoxiang, and Zhang Bingbing. A user clustering algorithm on web usage mining. In 2017 First International Conference on Electronics Instrumentation & Informa-tion Systems (EIIS), pages 1–4. IEEE, 2017.

[11] Kazi Saidul Hasan and Vincent Ng. Automatic keyphrase extraction: A survey of the state of the art. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1262–1273, 2014.

[12] Liang Hu, Guohang Song, Zhenzhen Xie, and Kuo Zhao. Personalized recommendation algorithm based on preference features. Tsinghua Science and Technology, 19(3):293–299, 2014.

[13] Rong Hu and Pearl Pu. Helping users perceive recommendation diversity. In DiveRS@ RecSys, pages 43–50, 2011.

[14] A. Kappeler, R. D. Morris, A. R. Kamat, N. Rasiwasia, and G. Aggarval. Combining deep learning and unsupervised clustering to improve scene recognition performance. In 2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP), pages 1–6, Oct 2015.

[15] Michael Luca. Reviews, reputation, and revenue: The case of yelp. com. Com (March 15, 2016). Harvard Business School NOM Unit Working Paper, (12-016), 2016.

[16] Bamshad Mobasher, Honghua Dai, Tao Luo, and Miki Nakagawa. Discovery and evaluation of aggregate usage profiles for web personalization. Data mining and knowledge discovery, 6(1):61–82, 2002.

[17] Catia Pesquita, Valentina Ivanova, Steffen Lohmann, and Patrick Lambrix. A framework to conduct and report on empirical user studies in semantic web contexts. volume 11313, pages 567–583, 2018.

[18] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Application of dimen-sionality reduction in recommender system-a case study. Technical report, Minnesota Univ Minneapolis Dept of Computer Science, 2000.

[19] Yang Tang and Paul D. McNicholas. Clustering airbnb reviews. 2017.

[20] Zied Zaier, Robert Godin, and Luc Faucher. Evaluating recommender systems. In 2008 In-ternational Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution, pages 211–217. IEEE, 2008.