CSE6242 Data & Visual Analytics

Project Proposal

Team 30

All-in-One Travel Recommendation

Huizi Shao
Jayden Sun
Jiayuan Bi
Shangru Yi
Yuanxun Shao
Yue Li

Georgia Institute of Technology

10-07-2019

Contents

1 Problem Definition

Realizing the inconvenience of requiring multiple recommendations Apps for completing a single travelling plan, we intend to leverage the functionality to the proposed all-in-one travel recommendation App. By utilizing *Yelp* and *Airbnb* dataset, we will combine various travelling recommendation tasks into a single application with map visualizations, end-to-end automatic all-in-one recommendation based on user preference and historic reviews.

2 Background Description

Currently, a single application can't provide users comprehensive travel recommendations, i.e. combined plans for accommodation, restaurants and city attractions. Users have to search different information in separate Apps, including Airbnb, Yelp and Google Map, and manually collect information by their own. User experience is a bit of negative while shifting among Apps and overwhelmed by redundant information. Furthermore, the lack of integration might end up with low-quality recommendations. Our proposed all-in-one travel recommendation App aims to reduce the tedious process in one shot and also enhance the recommendation quality.

3 Helimeier Questions

1. *What are you trying to do?*
   See Problem Definition.

2. *How is it done today; what are the limits of current practice?*
   Nowadays, there are separate Apps providing different categories of recommendations for travelers: Airbnb for hotels, Yelp for dining, Google for resorts, and so on. Thus users inconveniently need more than one Apps for the a complete traveling plan.

3. *What's new in your approach? Why will it be successful?*
   We propose an all-in-one recommendation App including information for resorts, hotels and dining, which can make travelers' decision making easier and more enjoyable.

4. *Who cares?*
   Travelers, foodies, local explorers.

5. *If you're successful, what difference and impact will it make, and how do you measure them ?*
   We can provide convenient and personalized hotels, foods, resorts recommendations for travelers in one single web. We will measure the success by user studies, which compares the experience of using our single web versus multiple webs for travel plannings.

6. *What are the risks and payoffs?*
   It is difficult to provide high-quality personalized recommendations due to the vague correlations in different datasets, and it is also challenging for machine learning algorithms

handle multi-purpose tasks. The success of the project is to fully achieve "once for all recommendations" features and enhance user experience.

7. *How much will it cost?*
   Free Azure credit was given.

8. *How long will it take?*
   The simple features can be completed within this semester. However, more complex personalized recommendations will take more time to investigate.

9. *What are the midterm and final "exams" to check for success? How will progress be measured?*
   The midterm milestone is to find appropriate datasets, create an interactive web. The final goal is to successful analyze the data and achieve all proposed features. We will conduct user studies to measure the satisfaction of our web.

4 Literature Survey

4.1 User Interface Design

1. User Study
   We can conduct our user study based on categories at purpose, users, tasks, setup, 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. For our website, we can ask several questions to get to know users (traveling style, food preference, etc.).

2. UI Design Principles
   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.

4.2 Recommendation System

1. Airbnb, Yelp, Netflix and Tourism Recommendation Systems
   Generalized recommend System does not meet diverse demands. Based on massive user data and geo-tagged images, Airbnb, Yelp, Netflix and tourism websites have implement personalized ranker, top-N ranker, trending, continue watching, similarity and clustering methods to give specific suggestions for subgroups to improve recommendation quality [19, 15, 20, 3]. We can cluster the information and provide more insight for personalized recommendations [5]. Moreover, by combining 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 behaviors of users. Per-

ceived categorical diversity has a significant coloration for the usefulness of systems [13]. Thus we can follow the guidelines to design our website.

4.3 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 recommendation, we need to discover aggregated user profile and cluster user activities.Based on traditional K-means algorithm [16, 7], a new clustering method combining *DBSCAN* and *Chameleon* further boost the performance on web pages [10]. In addition, combining user preferences and item features, a hybrid collaborative filtering algorithm can provide better recommendation [12]. Moreover, DNN is good at learning representations from rich data, when finding hand-designed low-level features are very challenging. Using refined features from DNN produces better cluster classifiers [14]. A systematic taxonomy of clustering methods and end-to-end framework using DNN for clustering further boost the performance [2, 4]. Those algorithms can be used to provide the all-in-one recommendations for dining, living, and tourist destination. We will try out and find out the best scheme for our problem.

4.4 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 key phrase extraction algorithms involve heuristic candidate phrase extraction and phrase selection [11]. Features with distance measure provides extraction insights. Phrased-Based document index for more accurate clustering, proposed in [8], presents efficient incremental phrase index construction and compact similarity-based clustering. We could utilize those database techniques into our project.

5 Plan of Activities

5.1 Algorithms

- Recommendation problem
- User clustering (user similarity based)
- Document query with phrase indexing
- Key phrases based Document clustering

5.2 Scheduled Implementation Agenda

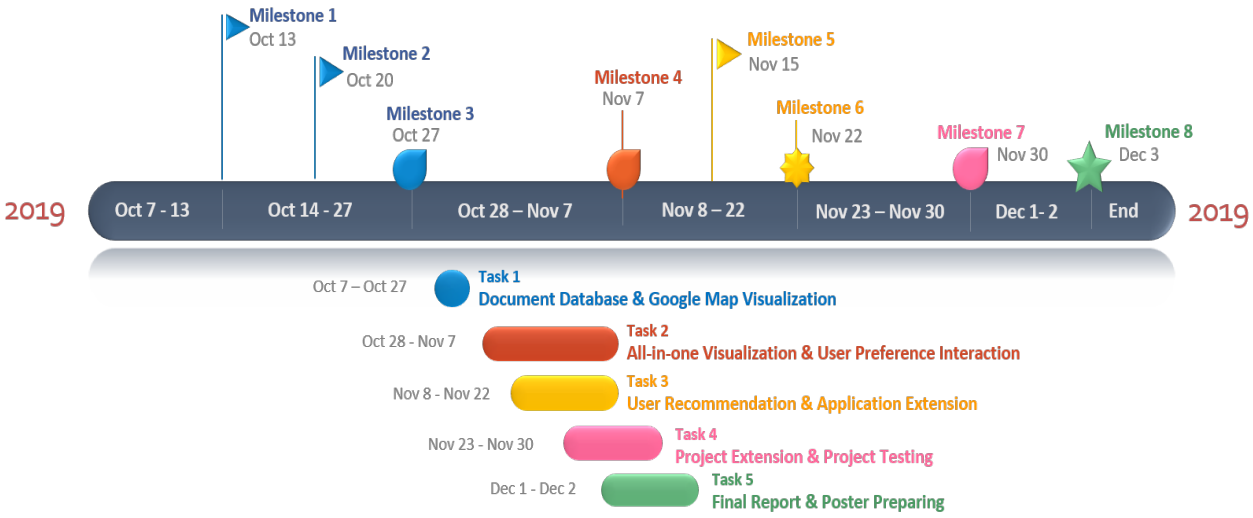

Figure 1: Scheduled Agenda

5.3 Project Data

- Yelp dataset
  Restaurant data (dining conditions, position, pricing, reviews, rating and reservation)
- Airbnb dataset
  Accommodation data (housing conditions, positions, pricing, reviews and booking)
- Google Map API

5.4 Labor Division

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

Figure 2: Labor Division

5.5 Implementation Tools

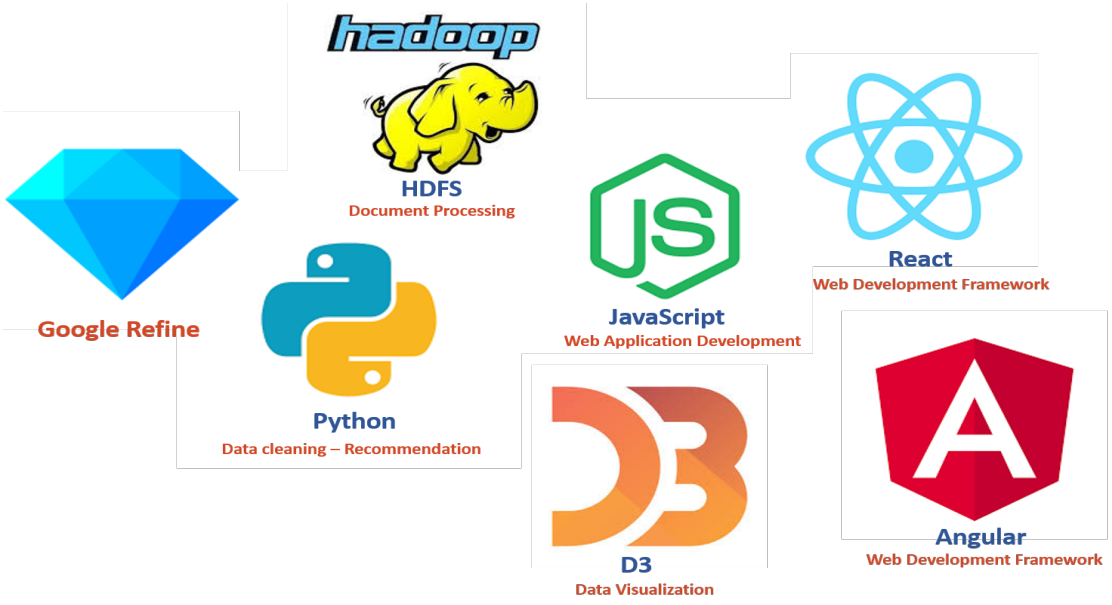

Figure 3: Implementation Tools

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. Aworldwide tourism recommendation system based on geotaggedweb 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 indexing 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 & Information 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 dimensionality 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 International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution*, pages 211–217. IEEE, 2008.