Recommendation Algorithms for User First Booking on Airbnb

Zhao Wang
Northeastern University
Seattle, WA
wang.zhao2@husky.neu.edu

Zerui Ma
Northeastern University
Seattle, WA
zeruima1989@gmail.com

Heng Xu
Northeastern University
Seattle, WA
xu.he@husky.neu.edu

ABSTRACT

Airbnb has become more and more popular under the trend of sharing economy. In order to attract new users to place their first booking on Airbnb and offer a more personalized experience, we aim to build a recommendation system by using collaborative filtering approach, to predict which country these new users will make their first booking. Also the hosting data for various countries will be analyzed to provide better insight around booking area.

Keywords

Learning Algorithms; Recommendation Systems; Collaborative Filtering; Cluster Models

1. INTRODUCTION

Data mining is the process of discovering interesting patterns from massive amounts of data. As a knowledge discovery process, it typically involves data cleaning, data integration, data selection, data transformation, pattern discovery, pattern evaluation, and knowledge presentation [1]. The science of learning plays a key role in the fields of data mining, statistics and artificial intelligence, intersecting with areas of engineering and other disciplines[2]. In a typical scenario, a quantitative or categorical outcome measurement is predicted based on a set of features.

Recommendation algorithms are a kind of learning algorithms which are widely used on e-commerce web sites. they use input about a customer's interests to generate a list of recommended items. Most recommendation algorithms are designed for finding similar customers, where they aggregate items from the similar customers, eliminates items the user has already purchased or rated, and recommends the remaining items to the user. The popular versions of these algorithms are collaborative filtering and cluster models. In the collaborative filtering, the similarity of customers is measured by the cosine of the angle between the two vectors which represent users' interests[3]. Using this algorithm to generate recommendations is computationally expensive, but it can be released by dimensionally reduction techniques[4]. To find the similar customers to the user, cluster models divide the customer base into many segments and treat the task as a classification problem[5]. Some algorithms classify users into multiple segments and describe the strength of each relationship[6]. Besides grouping the user to the similar customers, other algorithms such as search-based methods and item-to-item collaborative filtering focus on finding similar items[7]. Search- or content-based methods treat the recommendations problem as a search for related items[8].

User experience is now a critical factor to keep users and attract new users among web applications. That is the reason Airbnb wants to provide personalized and unique experience for its new users, thus Airbnb need an effective recommendation system to recommend a country for first-time booking.

However, the main challenge here is that Airbnb doesn't have the travel history or other type of the traveling data of new users, the only data available here is basic feature such as age, gender, session log etc., basically like a white paper to a recommendation system. While a typical recommendation system might make recommendation based on a few strongly related features, the system designed here need to focus on correctly classify similar users first, then trying to make recommendation with some relatively strong features. And that is why we choose collaborative filtering as our first-step approach.

2. DATASET DESCRIPTION

2.1 New user data set

The training data consists of 213451 rows and 16 columns, while test user set has 62096 rows and 15 columns. The additional column is a 'country_destination column', and this is where our predictions will be focused on.

About 60% of the users did not making a making (NDF:no destination found). US is the most popular destination among who booked a travel destination.

Quite a few age values are given in the year format. For now we ignore the null values and inconsistent values, and the corrected data will be presented in the next update.

Compared to basic sign-up and Facebook sign up users, those who signed up via Google have a much lower tendency to book a trip to U.S.

The differences between genders on destination country choices are not significant. It might be hard to generate any useful conclusion based on gender.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

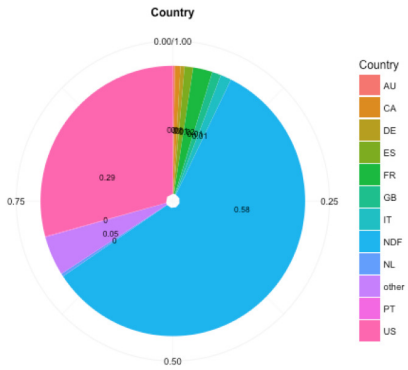

Figure 1: Destination Country Distribution.

Table 1: User Age Statistics

Minimum	1.00
1st Quartile	28.00
Median	34.00
Mean	37.41
3rd Quartile	43.00
Maximum	150.00

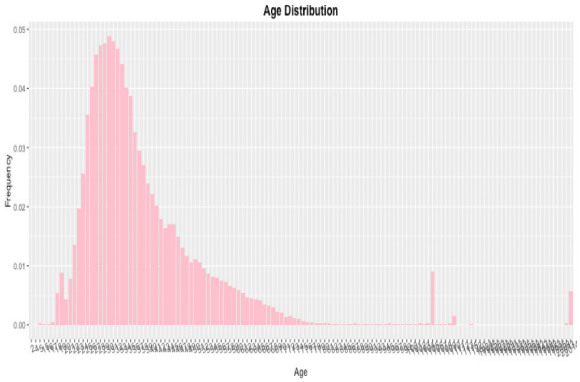

Figure 2: Age Distribution for Users on Airbnb

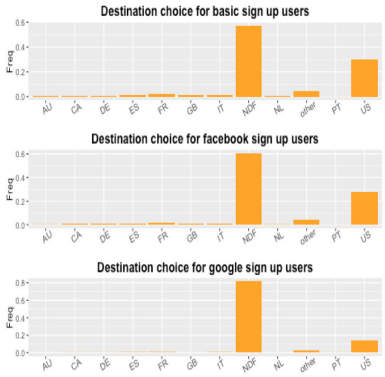

Figure 3: Destination Choice Sign Up Method

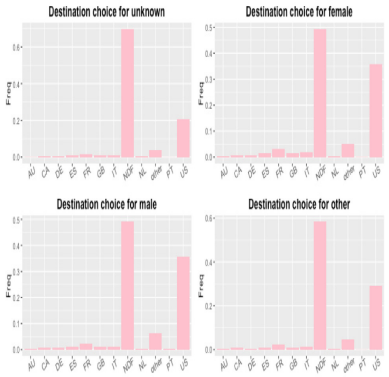

Figure 4: Destination Choice based on Gender

2.2 Listing data set

The other data set is about detailed information of host listings among 16 countries available on Airbnb, although it covers almost all the countries in new user data set, but PT(Portugal) listing data is missing here.

This listing data set contains most of the public information available on Airbnb, and lots of features such as listing price, listing rating score, host registration information and neighbourhood etc., may help recommendation system identify better listings among target country for new users. However, due to the limitation of new user data, the recommendation on this part is most likely to be based on features weighted by common sense: highest rated, best available price etc..

Since listing price is showed in dollar, so the average listing price distribution is influenced by currency exchange rate in some extent, but besides that fact, the average listing price definitely reflects the popularity of these destination countries. BM(Bermuda) is a typical high-end vacation hot spot, but it's surprised to see UY(Uruguay) has such a high average price, maybe because the retrived UY listings is very limited. Other than these two countries, AU and US are the most expensive countries to go on Airbnb. It's interesting to see AU has higher average listing price over US, a possible explaination for this could be the number of listings in US is much greater than AU's, thus competetion over the host has lowered the average listing price in US.

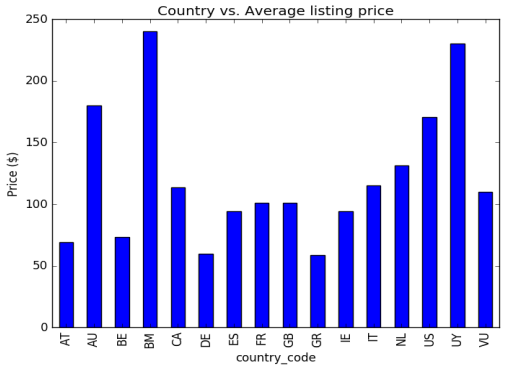

Figure 5: Average price distribution among countries

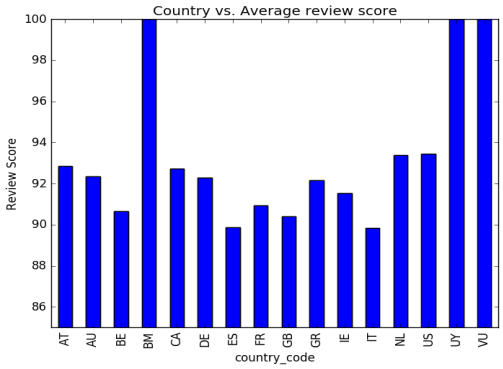

Figure 6: Average rating score distribution among countries

When comes to average review scores, most of countries get over score of 90. BM, UY and VU receives average score of 100, but since the number of listing of these 3 countries in this data set is very limited, it probably has large bias here. Besides these 3 countries, US and NL is probably best reviewed place to go on Airbnb, IT and ES is the worst reviewed, but still has average score of over 89.

3. PROPOSED ALGORITHMIC APPROACH

Using collaborative filtering to aggregate the similar users based on their first booking on Airbnb. The similarity between users is measured by cosine of the angle between the two vectors which indicate users? properties. Dimensionality reduction should be applied to reduce computational complexity and speed up the algorithm.

4. REFERENCE

[1]. Han, J., Kamber, M., and Pei, J. , "Data Mining: Concepts and Techniques", 3rd Edition, 2011.

[2]. Hastie T., Tibshirani R., and Friedman, J. , "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", 2nd Edition, 2009.

[3]. B.M. Sarwarm et al., "Analysis of Recommendation Algorithms for E-Commerce", ACM Conf. Electronic Commerce, ACM Press, 2000, pp.158-167.

[4]. K. Goldberg et al., "Eigentaste: A Constant Time Collaborative Filtering Algorithm", Information Retrieval J., vol. 4, no. 2, July 2001, pp. 133-151.

[5]. P.S. Bradley, U.M. Fayyad, and C. Reina, "Scaling Clustering Algorithms to Large Databases", Knowledge Discovery and Data Mining, Kluwer Academic, 1998, pp. 9-15.

[6]. L. Ungar and D. Foster, "Clustering Methods for Collaborative Filtering", Proc. Workshop on Recommendation Systems, AAAI Press, 1998.

[7]. G. Linden, B. Smith, and J. York, "Amazon.com recommendations: itemto-item collaborative filtering", Internet Computing 7:1, 2003, pp. 76?80.

[8]. Leskovec, J., Rajaraman, A., and Ullman, "Mining of Massive Datasets", 2nd Edition, 2004, pp. 307-341.