Recommend Merchants Based on reviews in Google Map Data

CSE158 Assignment 2

Jinglan Yang

ABSTRACT

In this paper, we investigate the properties of the google map dataset located in Alaska. The relationship between the ratings received by merchants and the feedback given by users is explored. Further, we hope to use these relationships to build recommender systems. This paper discusses three recommendation approaches: first, simply using the average score of a merchant to predict whether a user will like the shop, i.e., the higher the score of the merchant, the more likely the user is about to like the shop; second, recommending a merchant using the number of reviews the merchant has received; and, third, mining the semantics of the review statements and recommending a merchant by comparing the similarity of the semantics in the reviews of different merchants. Finally, we find that a simple average score model can be used to recommend merchants for new users well, while for users with a lot of review history, the use of relevant data mining or natural language processing models can recommend merchants more accurately.

1. INTRODUCTION

I used existing Google map dataset generated by Tianyang Zhang and Jiacheng Li (Li, etc., 2022; Yan, etc., 2023). This dataset contains Google map information up to September 2021 in the United State. In fact, this dataset is divided into multiple smaller datasets based on each state in the U.S. I chose the data from Alaska because of the relatively modest number of merchants and reviews that this region has. The datasets for more populous states such as California, Florida, etc. are too large for data analysis. (Data in these states need more data preprocessing and runtime for models may be significantly longer.)

This dataset has been assembled by researchers for use in demonstrating that certain models can be used to solve specific problems. This dataset has been used to pre-train the UCTopic model to distinguish whether the contexts of two phrases share the same semantics. The UCTopic model is a novel non-universal contrastive learning framework for context-aware phrase representation and topic mining (Li, etc.). Additionally, this dataset is also used to experimentally demonstrate the feasibility of a personalized multimodal framework to solve the personalized showcases problem (Yan, etc.).

Natural language processing is used to classify whether a review is positive or negative, and then combine the user's preferred merchant attributes with the merchant's positivity to recommend a merchant for the user (Gomathi, etc., 2019). Multiple linear regression can be used to determine the weights of the various aspects of the rating criteria to calculate the average score of the merchandise (Jhalani, etc., 2016).

Using an inverse ratio of a word's frequency in a given document to the percentage of documents the word appears in; Term Frequency Inverse Document Frequency (TF-IDF) determines values for every word in a document. High TF-IDF words indicate a close connection to the document they occur in, indicating that the user may find the document interesting if the word were to show up in a query. (Ramos, 2003).

2. THE BODY OF THE PAPER

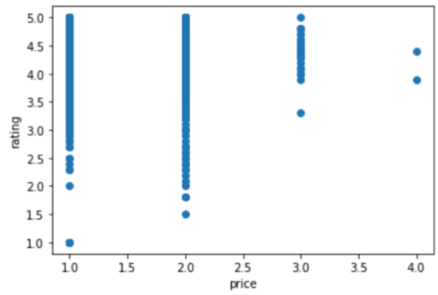
2.1. exploratory analysis of dataset

I explore and discuss the google map dataset in Alaska (Zhang etc., 2021) in this paper. The Alaska business dataset contains two sub-datasets: review data and metadata. The review data contains 1051246 review information. For every information, it includes user id, username, time, user's rating for a specific business, user's review text for this business, pictures of the review, business response to the review including unix time and text of the response, and ID of the business. The metadata contains name of the business, address of the business, ID, description, latitude, longitude, category, average rating, numbers of reviews, price, open hours, MISC information, current status, relative businesses recommended by Google, and URL.

This dataset features 12774 merchants, 278696 users who have made reviews, and a total of 1051246 reviews.

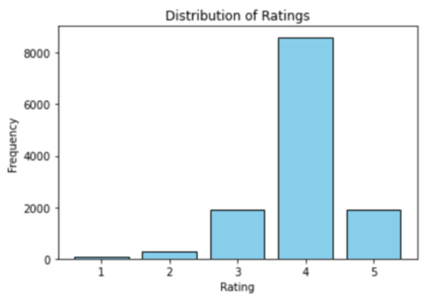
The most popular restaurant (define popularity by number of reviews a restaurant received) is Moose's Tooth Pub & Pizzeria who has 3,677 reviews. Besides, the businesses included in this dataset have at least one review.

People tend to give higher ratings to expensive merchants, but there are high and low ratings for cheap or relatively cheap merchants. As you can see from the image below, the most expensive merchants ($$$$) are rated in a very small and high range: 3.5 to 4.5; the more expensive merchants ($$$) are rated in a smaller but still high range: 3 to 5; the cheaper merchants ($$) are rated in a larger range: 1.5 to 5; and the cheapest merchants ($) have the largest range of ratings: 1 to 5. Taken together. people are inclined to give higher ratings, regardless of whether the merchant is expensive or not.



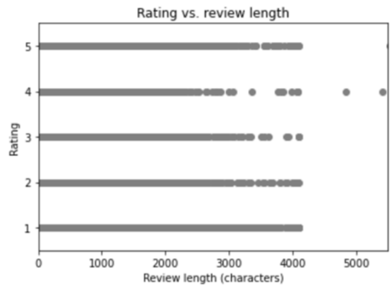
In the picture above, 1,2,3, and 4 of price indicates $, $$, $$$, $$$$. More $ signs mean higher prices

Based on the average ratings received by each merchant, I drew the following bar graph. As you can see, most of the merchants have ratings between 3 and 4, with 4 being the common occurrence. It can be concluded that people tend to give higher scores, but do not give full scores (5).

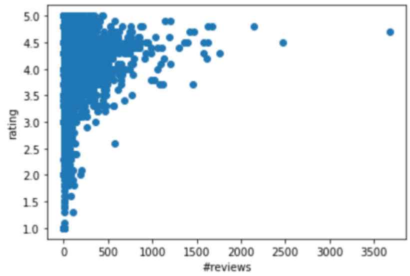


Most merchants tend to cluster together. As you can see from the image below, most merchants are clustered in an area centered on (-150, 60), radiating away. A small portion is around (175, 50).

Overall, the number of words in a review does not determine the rating. There are long and short comments for each rating scale. In particular, some users write longer reviews for businesses that they give a rating of 4.



Merchants that get more reviews tend to get higher ratings. As you can see from the image below, merchants with a lower number of reviews received ratings ranging from high to low. As the number of reviews increases, the range of ratings narrows and tends to be higher.



2.2. identify a predictive task

I try to recommend business to google map users. This recommender system can be broken down into the following categories:

1. The user does not have a specific merchant preference, e.g., does not specify that he/she wants to go to a Chinese restaurant or a Japanese restaurant or others. Or, this is a new user who has no previous records in Google map application.

2. The user wants the algorithm to recommend businesses based on the user's own preferences, e.g., if the user prefers Thai restaurants, the algorithm recommends other Thai restaurants that the user might like.

To solve the first problem, I propose to recommend merchants to users based on merchants' popularity. Here, the popularity can be expressed in terms of the average score of the merchant or the number of reviews received.

To solve the second problem, I propose to recommend merchants based on similarity by extracting the keywords in the user's previous reviews and comparing the sub-user's keywords with the keywords of other users' reviews for other merchants.

I use the accuracy to assess how good the model is. Split the dataset into a training set and a test set, train the model with the training set and then calculate the accuracy on the test set.

2.2.1. predict based on average rating

We assume that the popularity of a merchant can be expressed in terms of the average score it receives, i.e., the higher the merchant's average score, the more popular the shop is. Predict given a (user, business) pair whether the user would give a positive rating (>=4) or negative rating (<4). If the model is classified correctly, then we consider the model to be valid.

Starting with the simplest model, predict the rating that users will give to this merchant based on the average rating (in my case, we need to convert it into positive/negative) that the merchant has ever received.

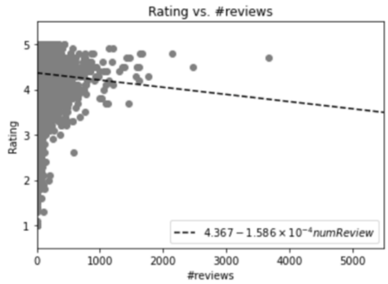
Based on the two datasets I have: review data and meta data, since the meta data has the average score of each merchant in it, I can just utilize that value and then use the whole review data as a test set. Finally, I calculated the Accuracy for this model is 0.7871928594667451.

2.2.2. predict based on number of reviews

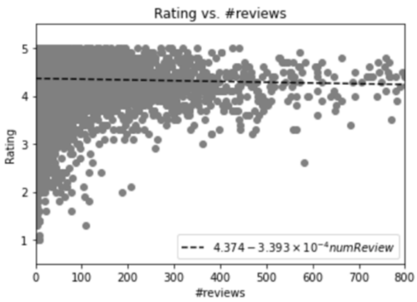
Now, let us assume that the popularity of a merchant can be expressed in terms of the number of reviews it gets, i.e., the more reviews a merchant gets, the more popular it is.

Based on previous observations of the data and the image I drew of the number of reviews vs. the rating, I believe that the number of reviews a merchant receives predicts how users will rate the shop.

I perform linear regression between rating and number of reviews and get the following result. But the line does not fit data well.



I tried to narrow down the dataset to only include restaurants with less than 800 reviews and came up with the following picture.



The straight line fitted by linear regression does not represent the characteristics of the dataset well. When I used the entire dataset, the straight line fitted by linear regression showed a decreasing trend, indicating that as the number of reviews increased, the ratings received by the merchants decreased instead, which is clearly not true. When I try to shrink the dataset to include only merchants with less than 800 reviews, the linear regression fits a straight line that tends to be horizontal, indicating that no matter what the number of reviews the merchant has, the rating he receives will be around 4.4, i.e., the rating is independent of the number of reviews.

2.2.3 predict based on TF_IDF

Compare the similarity between the high-frequency words (the phrases that appear most frequently in the reviews) in the specified user's reviews for a particular merchant and the high-frequency words in the reviews owned by the merchants in the entire dataset and recommend a merchant for this specified user.

For each entry (user, businesses, review texts) in the dataset, sample a negative entry by randomly choosing a business that user hasn't visited and review text from other people.

We assume that the similarity between merchants that the user has visited will be higher than the similarity between merchants that the user has not visited and those that have visited.

For every user, we randomly selected a business that the user had visited as a reference standard. We extract frequencies for terms in this specific review. Find the highest tf-idf words in this specific review. Order the other reviews in the corpus with the cosine similarity between tf-idf vectors in descending order. If this unvisited merchant is located in the second half of the sorted merchant list, then we consider the model to be valid.

I selected users who had made more than 10 comments, and then used the TF_IDF model with this data and ended up with a correctness rate of 0.5970149253731343. Later, I improved the dataset by filtering for users who had more than 50 words per comment and who had made more than 10 reviews, and then ran the TF_IDF model again, and the correctness rate increased to 0.7575757575757576.

3. CONCLUSIONS

The experiments illustrate that when we want to recommend a merchant to a user, if this user does not have any past review history or if he does not want to get recommendations based on his preferences (e.g., if he wants to explore new merchants beyond his comfort zone), then we can recommend a merchant to the user based on the merchant's average rating. Though this is a simple model, it works fairly well. We also tried to understand the relationship between the number of reviews a merchant gets and its rating through linear regression, then recommend based on review data, but it did not work. We think there are two possible reasons: 1. Linear regression does not capture the relationship well; 2. Most merchants just have a number of reviews between 0-500, while a small number of merchants get more than 500 reviews. The merchant that gets the most reviews in this dataset gets 3,500 reviews. It can be seen that this dataset is very unbalanced, which may result in linear regression failure. Subsequent research could focus on a dataset where the number of reviews is more balanced with the number of merchants or use a different approach to this dataset than linear regression. If a user wants to get recommended merchants that match his preferences, then the TF_IDF framework, which recommends merchants by comparing the keyword similarity of people's reviews of the merchants, can be used. But this method performs well only when the number of words in the reviews is high. In my study, I manually tuned the parameter to 50, meaning I select reviews which have length of review larger than 50. Subsequent research can focus on if it can recommend even when the number of words in the reviews is small or find more efficient ways to tune the hyperparameter.

REFERENCES

An Yan, Zhankui He, Jiacheng Li, Tianyang Zhang, Julian Mcauley. Personalized Showcases: Generating Multi-Modal Explanations for Recommendations. The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2023

Jiacheng Li, Jingbo Shang, Julian McAuley. 2022. UCTopic: Unsupervised Contrastive Learning for Phrase Representations and Topic Mining. Annual Meeting of the Association for Computational Linguistics (ACL).https://datarepo.eng.ucsd.edu/mcauley_group/gdrive/googlelocal/

R. M. Gomathi, P. Ajitha, G. H. S. Krishna and I. H. Pranay. 2019. Restaurant Recommendation System for User Preference and Services Based on Rating and Amenities. International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862048.

Ramos, J. 2003. Using tf-idf to determine word relevance in document queries. In Proceedings of the first instructional conference on machine learning (Vol. 242, No. 1, pp. 29-48).

Tanisha Jhalani, Vibhor Kant, Pragya Dwivedi. 2016. Data Mining and Big Data. Volume 9714 ISBN: 978-3-319-40972-6