Content based recommendation using sentiment analysis of vacation accommodations

Rodrigo Ogalde, Pontificia Universidad Católica de Chile, Chile, rjogalde@uc.cl

Ignacio Quera, Pontificia Universidad Católica de Chile, Chile, ignacio.quera@uc.cl

Abstract In the following article we present a technique for recommending accomodations using content-based recommendation, while taking into consideration the sentimient and location of reviews of previous occupants, all of this based on Airbnb Listings and Google Local Reviews.

1 Introduction

In the current digital era, online platforms such as Airbnb and Google Maps have revolutionized the way people travel and explore new destinations. Travelers heavily rely on reviews and ratings from other users to make informed decisions about where to stay and which places to visit. However, with millions of reviews and listings available, it can be overwhelming for users to filter and find the most relevant information for their needs.

The relevance of this issue is influenced by a series of factors. With the growth of technology and connectivity, an increasing number of people trust online platforms to plan and book their trips. On the other hand, travelers seek personalized experiences and depend on accurate recommendations based on their preferences and past behaviors. The growth of businesses and options available to travelers makes it challenging to choose destinations and locations that align with their tastes.

2 Concepts

2.1 Sentiment analysis

Sentiment analysis is a technique used to determine the sentiment and intention expressed on text. It works by analysing the words and the context in which they are used to determine if the sentiment is positive, negative or neutral.

2.2 Content based recommendation

Content based recommendation is a type of recommendation system based on the characteristics of the items used by users in the database. This method focuses more on the features of the items to recommend rather on user preferences or other type of filtering. .

3 Methods

3.1 Recommendation algorithm

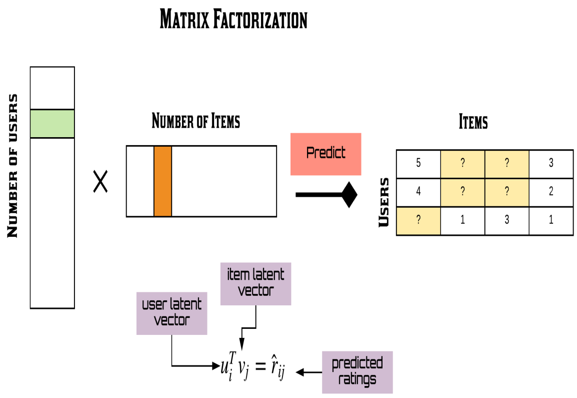

Figure 1: Matrix factorization diagram.

3.1.1 Matrix factorization

The chosen recommendation algorithm is matrix factorization. This method is based on simplifying the matrix of user interactions to predict user preferences. The singular value decomposition (SVD) method or non-negative matrix factorization (NMF) matrix factorization method can be used for this purpose.

3.1.2 Content based recomendation

In this work, the chosen method for analysis is content based filtering or content based recommendation. This method stands out as an approach able to enrich data by taking into account the features and, eventually, intention of reviews of items. With the context of reviews, the system is going to be able to provide better results based on meaningful interactions.

3.1.3 User based Collaborative filtering

As a base for comparison for the content based model, user-based collaborative filtering is used to recommend listings based only on previous user experiences. This method is not enriched by sentiment analysis, so it will be the baseline for improvement.

3.2 Sentiment analysis

In order to extract the intention of each review of each accomodation, sentiment analysis is used. Using natural language processing (NLP), a function is applied to the data to extract the sentiment, which give more information when used in conjunction with a score system. During the experimental procedure, a function which was created to extract the sentiment is applied to the text review of the dataset

3.3 System assembly

By integrating diverse recommendation algorithms, a composite system can be assembled to provide more robust and accurate recommendations. This approach takes advantage of the strengths of the various algorithms, allowing the system to enhance performance. The assembly process involves combining the outputs of the Matrix Factorization and the Content based recommendations, in conjuction with the input from the sentiment analysis function, in order to create a more accurate recommendation system.

3.4 Data

3.4.1 Google Local Reviews

This dataset provides reviews of places on Google Maps, which are essential for understanding user opinions and preferences regarding different places to visit. These reviews can offer information about the popularity of a place, the features that visitors appreciate, and areas for improvement. The nature of this data is both textual and quantitative, as it includes both the content of the review and the ratings given by users. [1]

Reviews	666,324,103
Users	113,643,107
Businesses	4,963,111

Table 1: Google Maps Dataset summary.

For the purposes of this article, only a small portion of the whole database has been used, accounting for around two million reviews of places across the US.

3.4.2 Airbnb Listings and Reviews

This dataset provides information about Airbnb listings in 10 major cities, including details about hosts, prices, location, room type, and over 5 million historical reviews. This data is essential for understanding user preferences when choosing a place to stay. Similar to the previous dataset, the nature of this data is both textual and quantitative. [2]

4 Experimental Procedure

We present an enviroment in which two models where worked with, one which took into consideration the sentiment analysys of places in which the recommendations where from and another which only took user reviews.

Firstly, we utilize matrx factorization in order to extract the data from the reviews. This data does not have text to pass to a sentiment review function, so to achive a metric in order to create a recommendation for users, the mean review scores for places in the same location where calculated. This value is the final score for each location, which will be used for evaluating the recommendation. This recommendation based on review scores is achieved using a user based collaborative filtering method, which was created for this work.

Secondly, we incorporate the data from the Google Maps dataset, in which there are written reviews for several places which reside in the same zone as the Airbnb listings. We take the sentiment analysis from the written reviews and create a content based recommendation system which recommends a location inside the US based on the content of the reviews.

Then, we combine the results from the content based recommendation system with the original user based collaborative filtering system, creating a hybrid system that takes both user scores and user reviews. This new system benefits from the strengths from both methods, making it a robust system capable of recomming items based on different criteria.

Lastly, we compare the results between the first recommendation system and the assembled system which

Model	F1 Score@10	Recall@10	Precision@10
Review Score	0.608	0.587	0.569
Sentiment Analysis	0.608	0.592	0.565

Table 2: Results

incorporates content based recommendation and sentiment analysis.

5 Results

Following the training of both models with their respective dataset and after assembling the hybrid model we did 4 runs of testing to get average results. To measure performance, we used the metrics F1-Score, Recall and Precision to determine how accurate where the recommendations made by the systems. Table 2 shows the results of the collaborative user filtering in comparison with the hybrid model using sentiment analysis. While there is a small improvement in the chosen metrics, this is not a considerable one, showing that the method of sentiment analysis did not improve the accuracy of recommendations made.

6 Limitations

The first limitation we can find was the data available in the dataset. Since only places from the US where present on the Google Maps dataset, a large portion of the Airbnb listings were not able to recieve a sentiment analysis, since no data of reviews in the area where present.

Another limitaion was the ammount of RAM available to load in the datasets. Since the dataset have a large ammount of reviews, we were unable to load the models with enough data from different places around the world. This had a significant impact in our results because we could not generate recommendations outside specific areas, greatly reducing the abilty to compare between models.

7 Conclusion

We proposed a model for recommending acommodations to users based on reviews based in the same area. This is particularly useful for sites such as Airbnb that focus on presenting users with accomodations for vacations, as well as travel agencies that offer destinations to travelers. Current state of the art does not always take into account external factors when recommending items, so this approach enhances the data source from which recommendations are based on.

However, based on the above results, we can conclude that there is not a major difference in taking into consideration the sentiment analysis of places near accomodation listings, at least in the way this article proposed. While there was a minor improvement in the metrics of the hybrid model, on further inspection the recommendations rarely changed, indicating that the new data was not relevant while recommending items. This, however, does not imply that the method is not useful, since the available data given by the content based recommendation system could be useful if paired

8 Future Work

It is possible to improve upon this work in several ways. Future research should explore the integration of multiple recommendation techniques, creating new hybrid models more complex than what was created in this work. By combining collaborative filtering, content-based filtering, and perhaps incorporating newer techniques like deep learning, a recommender system provide more accurate and diverse recommendations. Another way of improvement is using already establish models as a baseline to improve upon, assuring that the base recommendation is solid and accurate. Finally, taking subsets from the large data to better represents the listings in the databse used my further improve the performance of the model and yield better results, and using other dataset to get reviews for place to inject the system analysis might also improve the scope of the work done in this article.

9 References

References

[1] Google Local Reviews: McAuley, J. Google Local Reviews. Recovered on October 13, 2023.

[2] Airbnb Listings Reviews: Bhat, M. A. Airbnb Listings Reviews.

[3] De Pessemier, T., Dhondt, J., Vanhecke, K., Martens, L. (2012). TravelWithFriends: a Hybrid Group Recommender System for Travel Destinations.

[4] Pazzani, M. J., Billsus, D. (2007). Content-based recommendation systems. In The adaptive web (pp. 325-341). Springer Berlin Heidelberg. Xu, W., Liu, X., Gong, Y. (2003)

[5] AL-Ghuribi, S. M., Mohd, S. (2019). A Comprehensive Overview of Recommender System and Sentiment Analysis. https://doi.org/10.48550/arXiv.2109.08794

[6] Dang, C., Moreno, M. (2021). An Approach to Integrating Sentiment Analysis into Recommender Systems. https://doi.org/10.3390/s21165666