Collaborative Filtering and Sentiment Analysis: Basics to Build a Map Recommender System

Zimu Guan

Department of Computer Science, University of California, San Diego, La Jolla CA 92093, USA
ziguan@ucsd.edu

Abstract. Recommendation systems are gaining increasing popularity in today's society, aiming to cater to users' potential preferences for the convenience of users and as part of businesses' commercial strategies. Just as product recommendations have proven effective, location recommendations are also becoming enticing. In our increasingly stable daily lives, more individuals are inclined to receive suggestions for new places that align with their preferences. Even with the popularity of Google Map, recommender system on locations should still be very useful in practice with applications. This paper summarizes some basic models to understand for building recommender systems for maps. Collaborative filtering, sentiment analysis, and feature key-word clustering are all very crucial in constructing the model, with the requirement of user's review, ratings, and location data. Several successful systems to possibly integrate within the system includes categorization labels on users, MERLOT and VADER system, and GeoSCAN system for considering location in recommendations. From researchers in different studies, it is proven possible to mix the analysis of ratings and reviews together to form a better algorithm for recommendation. The result of combination of these systems can overcome each other's shortcomings.

Keywords: Recommender System, Collaborative Filtering, Sentiment Analysis.

1 Introduction

With the wide usage of phones, and the convenience of traveling through cars, public transportation, etc., map applications have become a very crucial component in individuals' daily lives. In fact, current map apps like google map provides wide range of usage to the users, collecting information from all aspects such as traffic to design the best route for the user [1]. For app developers, it is very crucial to keep the user popularity of the application, making sure that users do not turn to their alternatives.

Recommender systems' job is to forecast users' potential future interests and preferences based on information about users and their preferences [2]. With the growth of the internet and the increasing demand from users, recommender systems

B. H. Ahmad (ed.), *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, Advances in Intelligent Systems Research 180,
https://doi.org/10.2991/978-94-6463-370-2_34

are widely used in websites and applications, trying to predict what the user wants from their preferences. Such technique is advanced to a point that it is possible to effectively use the recommender system even without a large amount of data [2]. There can be a wide range of studies and methods within the recommender system topic. For instance, two of the more common types, collaborative filtering and context-based learning, take different types of datasets and have different approaches with the data. Collaborative filtering takes a matrix as users and items as rows and columns, and ratings of users (with a customizable scale) as the elements within the matrix. With realistic expectations of users not rating on every item or most of the items within a database, the filtering should be able to deal with sparse data, having many empty entries for a single row/column [3]. Memory-based technique and model-based technique are two of the usual models to run on such database, and hybrid approach, combing methods together, is also a viable and effective approach to avoid shortcomings [3]. On the other hand, context-based learning focuses on text analysis, which corresponds to analyzing product reviews within the scope of recommender system. One of the main types of text analysis, sentiment analysis, can be done to distinguish positive and negative reviews and differentiate different elements within the review [4].

As sentiment analysis can yield a number based on how "positive" it is from a review, it can also be used in collaborative filtering as a scale for the rating. With such connection, many studies focus on integrating sentiment analysis with collaborative filtering [5-7]. The aim for such combination is to solve problems with different dataset, which may or may not include an explicit ratings from users. Since a number to rate the product might be hard to collect [5], it can be crucial to predict user preferences just with their textual reviews. This review aims to present a comprehensive exploration of various components, ranging from collaborative filtering to sentiment analysis. Additionally, it will provide a summary of how certain research works have integrated these elements into a unified framework.

The paper following this introduction will be divided into 3 parts: Section 2 will focus on introduction and explanation of some of the basic models used for recommender system, and how they combine. Section 3 will dive into other more complicated considerations with the recommender system built for maps. And the conclusion is made with a summary of the study in Section 4.

2 Models and Approaches for Building Recommender System

2.1 Collaborative Filtering

Introduction of Collaborative Filtering. Collaborative filtering is a popular technique used in the recommender system, trying to predict the preference of a user by their past reviews and ratings on other products. There are two main types of collaborative filtering, namely user-based and item-based, representing either evaluating the similarity between users or between items [3]. For instance, user-based approach tries to match User A with another User B, as they have both gone to places such as fast-food restaurants. If User B also went to a steak house and left high ratings and positive reviews, then User A is also likely to enjoy this steak house as well. Conversely, the item-based approach endeavors to group similar locations or establishments together. For instance, if User A displays a preference for fast-food restaurants, this approach will suggest additional fast-food establishments, aligning with User A's tastes. However, this method has certain limitations, including its inability to accommodate new users or recently introduced, uncategorized items. Nevertheless, various techniques can be employed in combination to address and mitigate these challenges.

Methods for Collaborative Filtering. To be more specific in calculating with matrices containing information of users and items, there are a handful of ways. Here are some of the examples on different approach for collaborative filtering containing calculations.

Single Value Decomposition is one of the common method for collaborative filtering and can be used on a numerical scale for the ratings of products by users. SVD splits the original user-item matrix in three parts:

$$A = U\Sigma V^T \tag{1}$$

In (1), U and V are m*m and n*n orthogonal matrices (the original matrix A is m*n) and the middle matrix, Σ, contains values only on the diagonal. The singular values, which are the diagonal elements of Σ, are ordered in a decreasing manner. These values indicate how significant the related singular vectors in both U and V^T are in terms of importance [8]. After a reconstruction after multiplying, it can be observed that similar matrix with before but with some patterns to make recommendations. This method is also integrated in libraries such as the surprise library in Python to provide easy access for calculation [9].

Another example of collaborative filtering technique with matrices is Alternating Least Square, where the factorization will result in two matrices, corresponding to the user or to the items. ALS will optimize the two matrices in alternate, fixing the user matrix with the item matrix, and then fixing the item matrix with the user matrix, etc. As the study from Gábor Takács and Domonkos Tikk points out and presented in their system called RankALS [10], this approach can be modified to deal with implicit feedback, which is a user's preference without explicit description such as ratings or reviews (Simply, if the user has chosen the item or not).

There are also methods that do not consider matrices factorization, like the K-Nearest Neighbor (KNN) Approach. This approach is designed to observe clusters for data points in some sort of graph, but in recommender system's case, an abstraction is made for all the data-to-data points and then KNN can be performed. The simple way to think and understand KNN is to think about finding nearest data points with the target, and there can be lots of different mathematical ways to describe "nearest" (E.g, Euclidean distance, Manhattan distance).The parameter of K can also be determined to either reduce noise or reduce bias.

Introduction to Sentiment Analysis and Connection to Collaborative Filtering. The analysis of text is a well-developed and complicated field, aiming to free humans from tedious tasks to scan through different texts to determine their meanings or emotions. One of the fields of studies from the analysis of text is called sentiment analysis, which focuses on extracting the semantics of a paragraph of text. There are two main approaches of sentiment analysis: Machine Learning and Lexion-based. While the machine learning approach focused on featuring key words from a paragraph of text for clustering, lexicon-based approach tries to understand the semantics of the text [4]. Such types of sentiment analysis can be either conducted polarized, giving a "positive" or "negative" label, or conducted in a more fine-grained way, distinguishing emotions such as happiness or anger. The process of lexicon-based sentiment analysis starts with tokenization and cleaning of the text, removing unnecessary words (such as the, a and this) and punctuations, and then extract the feature of the text with their semantic meanings by bag-of-words representation or word embeddings [4].

One of the successful and effective methods to solidarize the intensity is VADER, which gives a sentiment score to the representations to evaluate the intensity [11]. With the intensity of positiveness and negative-ness summarized in a numerical scale, it is also possible to conduct collaborative filtering on these values to try to predict user preferences, which makes combining sentiment analysis with collaborative filtering possible (See Fig. 1).

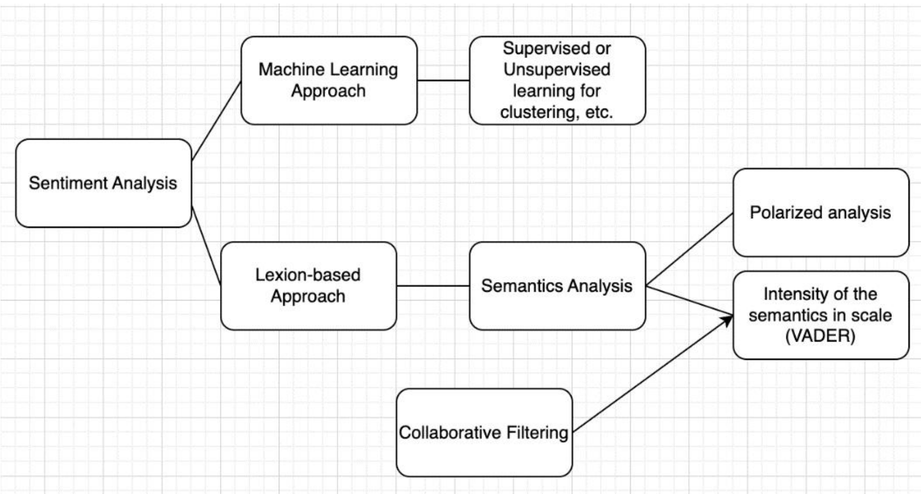

Fig. 1. Relationship between sentiment analysis and collaborative filtering (Photo/Picture credit : Original)

2.2 Key Word Feature and Clustering

As the categories of places on a map are not usually explicitly labeled, another important technique that might be conducted is the clustering of key word withing descriptions or reviews. There are two factors when considering the main focus of a paragraph of text: the term frequency (TF), indicating the frequency of words appearing in the text, and the inverse document frequency (IDF), indicating the inverse of how many times this word appears in other documents. Due to the existence of IDF, the algorithm can easily filter out frequent words without meanings such as a, the, they, etc.

The approach of TF-IDF is very simple and effective, but still have some weaknesses to improve on. In the case of location review databases, the data may or may not contain categorical labels, leaving the reviews vastly different from each other, and hence affect the result of TF-IDF negatively. For example, if reviews from restaurants and from supermarkets are mixed, unsupervised learning, or leaving the algorithm to cluster categories without human intervention, might lead to confusing results since both business places have key words related to food. A paper by Hichem Frigui and Olfa Nasraoui proposes to solve such problem [12], for which they developed a simultaneous approach for clustering categories and doing TF-IDF calculation. Such approach can be very essential when developing algorithms to both prompt key-word search, and also made categorical recommendation possible.

3 Applications and Discussion

3.1 Possible Dataset Selection in Map Recommender System

To construct recommender system models with the map data, several elements are crucial: the user, the corresponding review/rating from the user to the business, the longtitude and latitude of the business, and the category of the business. Having a category of business can be very useful to recommend the correct place when searching for a key word, and help the system better understand what types of places the user prefers to go. An example of such database come from the Google Local Places Review, collected by Tianyang Zhang and Jiacheng Li from University of California, San Diego [13, 14]. This dataset contains information separated by states, which also narrows possible mistakes from different state laws and culture.

3.2 Applications for Sentiment Analysis in Map Recommender System

Lots of other studies realize this possibility and aim to improve the performance of collaborative filtering by some combination with sentiment analysis. An example of this is a study by Miguel Á. García-Cumbreras, et el. to integrate sentiment analysis by giving users a label of “pessimists” and “optimists” [6]. In this study, the first step is to validate the possibility of predicting ratings both from text reviews and from collaborative filtering. Then the study took the ratings from the users, giving some users a label based on their average ratings of all their reviews. In the case of this study, they utilized a movie dataset with a rating scale of 1-10 stars, therefore giving an average rating above 6 to be “optimists” and average rating below 3 to be “pessimists”. The study proceeded to produce rating prediction with the label on users, and got slightly improved results. This approach leaves some users to not be labeled, so sentiment analysis comes into action. In the end, the study performs sentiment analysis on the text reviews and polarizes all the users into either optimists or pessimists, and yields an even better result.

Another investigation by Pythagoras Karampiperis, et al. delves into the impact of text review and rating alignment on the overall performance of the recommender system [15]. This study utilizes another sentiment analysis system called MERLOT to evaluate the semantics of the text, and is more focused on the accuracy of this system to produce a correct numerical value for further collaborative filtering. What adds an intriguing dimension to this study is its categorization of reviews, not only based on their positive or negative sentiments, but also by their structural characteristics. That is, professional reviewers tend to write more structured reviews to analyze both the advantages and disadvantages of a product, providing a more neutral and objective review, while regular user reviews can be much more subjective and emotional. The study shows a slightly higher accuracy for professional reviews to predict their semantics. Study of such division, though not very relevant with review on map apps like Google, still provides interesting context on possible improvements for the system.

3.3 Other Modules in Map Recommender System

When dealing with map location recommendations, another component stands out as an obstacle: the location of the user and businesses. Without considering the location, the system can make unrealistic recommendations to places that are hundreds of miles away from the user. An example to try to address the problem is the GeoSAN system, developed by Defu Lian et al., which divides the map to different grids, granting similar “quadkeys” to closer grids to alternate the weight of places in the recommender system. [16] This system can also deal with sparsity issues, for which some number of users only visit limited number of places. For a successful location recommender system, such location-aware methods are necessary to justify the recommended places, creating realistic suggestions.

Moreover, as map users might not have the tendency to leave reviews and ratings to places, they went, implicit information is a very tempting data source to further improve the recommender system. If user’s information of where they have been every day is accessible, the developer of map apps can collect the frequency of visit of certain place by a user to understand which types of places this user prefers to go, giving a even more accurate classification of users. However, reviews and ratings are voluntary input for other users to see publicly, but it is not necessarily the case for user location. Data privacy should always be considered when conducting implicit data like this, as users might not want their daily to be leaked or utilized in any way.

4 Conclusion

This paper provides concise insights into the fundamental elements constituting a recommender system for mapping locations and businesses. While conventional recommendation systems primarily rely on numerical scales in collaborative filtering, the paper delves into the utilization of text reviews through sentiment analysis as an additional avenue to enhance the overall system. Several researchers present their strategies for elevating the accuracy of sentiment analysis and collaborative filtering, suggesting innovative approaches such as associating users with the semantic content of their reviews or categorizing distinct review types to achieve greater precision. Locations of places also need to be considered to keep the recommendations realistic. Though the market lack space for new map applications with how popular Google Map is, such recommender system has wide range of applications, such as recommending places on travelling apps. When dealing with location data, it is crucial not to leak privacy of users for more accuracy in the system, since such data can be very easily accessed.

References

1. Xiang, Y.: Research on Influencing Factors of Mobile map Software users' continuous use Intention (in Chinese). Beijing Jiaotong University, MA thesis (2015).
2. Linyuan, L., et al.: Physics Reports, Volume 519, Issue 1, Pages 1-49, ISSN 0370-1573, https://doi.org/10.1016/j.physrep (2012).
3. Su, X., et al.: A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, vol. 2009, pp. 1–19, https://doi.org/10.1155/2009/421425 (2009).
4. Medhat, W., et al.: Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, https://doi.org/10.1016/j.asej.2014.04.011 (2014).
5. Leung, C W., et al.: Evaluation of a rating inference approach to utilizing textual reviews for collaborative recommendation. *Cooperative Internet Computing*, https://doi.org/10.1142/9789812811103_0007 (2008).
6. García-Cumbreras, M., et al.: Pessimists and optimists: Improving collaborative filtering through sentiment analysis. *Expert Systems with Applications*, vol. 40, no. 17, pp. 6758–6765, https://doi.org/10.1016/j.eswa.2013.06.049 (2013).
7. Kawamae, N.: Predicting future reviews. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, https://doi.org/10.1145/1935826.1935911 (2011).
8. Dan, K.: A Singularly Valuable Decomposition: The SVD of a Matrix, The College Mathematics Journal, 27:1, 2-23, DOI: 10.1080/07468342.1996.11973744 (1996).
9. Hug, N.: Surprise: A python library for Recommender Systems. *Journal of Open Source Software*, vol. 5, no. 52, 2020, p. 2174, https://doi.org/10.21105/joss.02174 (2020).
10. Takács, G., et al.: Alternating least squares for personalized ranking. *Proceedings of the Sixth ACM Conference on Recommender Systems*, https://doi.org/10.1145/2365952.2365972 (2012).

11. Hutto, C., and Eric G.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216–225, https://doi.org/10.1609/icwsm.v8i1.14550 (2014).
12. Frigui, H., and Olfa N.: Simultaneous clustering and dynamic keyword weighting for text documents. *Survey of Text Mining*, pp. 45–72, https://doi.org/10.1007/978-1-4757-4305-0_3 (2004).
13. Jiacheng L., Jingbo S., Julian M.: UCTopic: Unsupervised Contrastive Learning for Phrase Representations and Topic Mining, Annual Meeting of the Association for Computational Linguistics (ACL) (2022)
14. An Y., et al.: Personalized Showcases: Generating Multi-Modal Explanations for Recommendations, The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) (2023)
15. Karampiperis, P., et al.: Collaborative filtering recommendation of educational content in social environments utilizing sentiment analysis techniques. *Recommender Systems for Technology Enhanced Learning*, pp. 3–23, https://doi.org/10.1007/978-1-4939-0530-0_1 (2014)
16. Lian, Defu, et al. “Geography-aware sequential location recommendation.” *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, https://doi.org/10.1145/3394486.3403252.

.

.