WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A REVIEW PAPER ON PG RECOMMENDATION SYSTEM

Mohit Sharma , Hritik Baliyan , Akhil Chikara , Govind Sharma, Kirti Kushwah

*Department of Computer Science & Engineering, Inderprastha Engineering College, Ghaziabad,*Uttar Pradesh, India

ABSTRACT

When you plan of going to study far away from your comfort zone, one of the first things to do is to book a good place to stay. Booking a PG online can be an overwhelming task with thousands of PG to choose from Inspired by the importance of these situations, we made a decision to apply our skills on the task of recommending PG to Students. We used a hotel recommendation dataset, which gives us with a variety of features that helped us gain a better understanding of the process that makes a user choose certain PG. The aim of this PG recommendation project is to predict and recommend best PG to a user that he/she is more likely to book from the given hundred distinct choices.

1. INTRODUCTION

The pandemic has affected almost each and every sector across the world and the hotel industry is one of the hardest hit. The pandemic has changed the way we work, learn and communicate but the major thing that has been affected is a decent place to live. As everyone wants to live in a safe and sanitized environment but due this pandemic, the shortage of homes and proper maintenance has decreased. When you plan to study far away from your comfort zone, one of the first things to do is to book a good place to stay. Booking a PG online can be an overwhelming task with thousands of PG to choose from. Inspired by the importance of these situations, we made a decision to apply our skills on the task of recommending PG to students. This project will help students as well as people who are living far from their homes for work. Our target audience will include two groups of people i.e. landlords with houses for rent and Individuals with demand for a rental place. After getting the desirable place the students/working professional can select the place he/she is interested in. The objective of this project is to develop an Interactive website for the Booking of the PG especially for the students and provide recommendation system for the ease of the user and also provide interface for the owner to add/delete property. The background about the project idea is that whenever anyone is in need of pg they can get it through our website. Our website provides every detail about the pg and the pricing and features of the pg. Also it tells about the availability of the rooms.

Background

A Diversity of strategies has been used to Supply advice like collaborative filtering, content material primarily based totally and hybrid technique. remote Algorithms and strategies are there to offer advice that could use rating, but collaborative filtering and content material primarily based totally approach be afflicted by equal boundaries. A lot of researchers have attempted to conquer those boundaries via way of means of combining each collaborative filtering and content material primarily based totally approach as a hybrid technique that blended valuation in addition to content material information. Recommendation gadget will usually abide lively seek place for researchers.

2. APPROACHES OF RECOMMENDATION SYSTEM

Recommendation system is usually classified on rating estimation

- Collaborative Filtering system
- Content based system
- Hybrid system

In content-based approach, similar items to the ones the user preferred in past will be recommended to the user while in collaborative filtering, items that similar group people with similar tastes and preferences like will be recommended. In order to overcome the limitations of both approach hybrid systems are proposed that combines both approaches in some manner .

3. COLLABORATIVE FILTERING SYSTEM

In Collaborative filtering system The user can categorized based on the attributes of their similar rating preference in order to find user with similar features. Collaborative filtering is the process of predicting user preferences by identifying the interests and information of multiple users. This is done by filtering data for informations using techniques involving collaboration among multiple agents, data sources, etc. The technique then recommends items that are preferred by these similar user .

Advantage

The main advantage of using collaborative filtering models is the simplicity of implementation and the high level of coverage they provide. It also has the advantage that it captures subtle features (very true of latent factor models) and doesn't require a clear understanding of the item's contents.

Disadvantage

The main disadvantage to this model is that it's not friendly for recommending new items, this is because there has user/item interaction with it. This is referred to as the cold start problem.

EXAMPLE: Example of collaborative filtering algorithms:

Recommend YouTube content to users - recommend videos to you based on other users who have subscribed/watched similar videos to you.

Although thereare many collaborative filtering techniques, they can be divided into two major categories-

- Memory Based approaches
- Model Based approaches

4. MEMORY BASED APPROACH

Memory Based Approach regularly called neighborhood collaborative filtering. Important, rankings of consumer-object mixtures are expected on the premise in their neighborhoods. This may be similarly cut up into consumer primarily based totally collaborative filtering and object primarily based totally collaborative filtering. User primarily based totally basically way those likeminded customers are going to yield sturdy and comparable recommendations. Item primarily based totally collaborative filtering recommends objects primarily based totally at the similarity among objects calculated the usage of consumer rankings of these objects.

The prediction technique in memory-primarily based totally CF consists of Three steps. They are similarity assessment, era of nearest neighborhoods and rating prediction. For assessment of the overall performance, the CF device considers the imply absolute error (MAE), precision and bear in mind. The CF overall performance varies consistent with the processing approach of every step.

Existing Similarity Measures

The maximum vital first step in memory-primarily based totally CF is similarity assessment. The CF device on this step evaluates the similarity among the goal consumer and different customers for not unusual place score objects. The similarity is used as a weight for predicting the desire rating. Various similarity metrics were proposed in preceding studies. These are as follows-

- **Tanimoto coefficient** It is similarity among sets. It is a ratio of intersections. Assume that set X is and set Y is . The Tanimoto coefficient T of set A and B is zero.5. This metric doesn"t recollect the consumer score however the case of a totally sparse facts set is efficient.

$$T(X,Y) = \frac{X \cap U}{(X+Y)-(X \cap Y)}$$

- **Cosine similarity.** The Cosine similarity is called the Vector similarity or Cosine coefficient. This metric assumes that not unusual place score objects of customers are factors in a vector area version, after which calculates cosΘ among the 2 factors..

$$COS(U1,U2) = \frac{\sum rU1IirU2Ii}{||U1||\ ||U2||}$$

- **Person's Correlation** In Equation, SU1 is the usual deviation of consumer U1. The Pearson Correlation measures the energy of the linear dating among variables. It is generally signified with the aid of using r, and has values withinside the range [-1.0,1.0]. Where -1.zero is a great terrible correlation, isn't any correlation, and 1.zero is a great advantageous correlation..

$$r(U1, U2) = \frac{\sum(rU1Ii-U1)(rU2Ii-U2)}{SU1SU2}$$

Formation of Nearest Neighbor

The second step after the similarity evaluation is generation of nearest neighborhoods. To improve performance, many methodshave been proposed by CF researchers. The methods for selecting nearest neighborhoods include classification using K-means, a threshold for the number of common rating items and a graph algorithm. In general, it selects similar users greater than a given threshold or high rank users.

Prediction of Preference Score

The last step in memory-based CF is to predict the preference score of the target user for non-rating items. It predicts the preference score of non-rating items for the target user, based on the rating of nearest neighborhoods. Various methods have been proposed, and Weighted Mean is used as most general algorithm. PSU1,Ii is the predicted score of item i for U1 , and NNUi is the nearest neighbor

$$PSU1, Ii = \frac{\sum sim(U1,NNUi)rNNUi,Ii}{sim(U1,NNUi)}$$

5. PERFORMANCE EVALUATION

In the CF system, there are two types of measure for the performance evaluation. The first type is prediction accuracy, which isevaluated by MAE. Pi is the real preference score of item i and qi is the predicted score of item i

$$MAE = \frac{\sum i-1|pi-qi|}{N}$$

The second category is the quality of the recommendation, as measured by accuracy and recall. Accuracy is the percentage of movies that are rated as premium and recall is the percentage of premium movies that are rated higher. In addition, Fmeasure is also used. The measurement method is proposed as a way to visually represent the two measurements and overcome the inverse ratio of accuracy and recall. In the equation, p is the precision and r is the recovery.

$$F - Measure = \frac{2rp}{r+p}$$

A user-based rating prediction can be formalized as an aggregation of the ratings that the different neighbors suggest to the target item, denoted by *fA*(*V*, *t*). These suggestions are combined by weighting the contribution of each neighbor by its similarity with respect to the active user[8].

$$ra, = \frac{\sum V \in (a) sim(A,V) fA(V,t)}{\sum V \in (a) sim(A,V)}$$

6. MODEL BASED APPROACH

Model based approach is predictive models using machine learning. Features associated to the datas are parameterized as inputs of the model to try to solve an optimization related problem.

Techniques of Model Based Approach

K-MEANS CF: *k*-means clustering is applied to define the segments. *k*- means is a clustering method that has found wide application in data mining, statistics and machine learning. The input value for *k*-means is the pair-wise distance between the items to be clustered, where the distance means the dissimilarity of the items. The number of clusters, *k* is also an input parameter. It is an iterative algorithm and starts with a random partitioning of the items into *k* clusters. Each iteration, the centroids of the clusters is computed and each item is reassigned to the cluster whose centroid is closest. The Algorithm is Described Below:

Algorithm k-means clustering [7]

1. Input: R = r1
   …
   r
   m
2. Function kmeans(R; k)
3. c*i* = r*pi* ; ∀r*pi* ∈ R; ∀c*i* ∈ C; ∀*i* = 1; : : : ; *k*;
4. While(*k* C*0*
¡ C *k6*= 0)5. C*0*
= C;
6. *Ci* = {*j* : *sj;i* ≥ *sj;i** ; ∀*i** = 1; : : : ; *k*}; ∀*i* = 1; : : : ; *k*;
7. $\underline{c\text{i}} = \frac{\sum r\text{j}}{|c\text{i}|}$; ∀*j* ∈ *Ci;* ∀*i* = 1; : : : ; *k*;
8. End While
9. return C*0*.

7. CLUSTER MODEL

To find customers similar to users, cluster models divide customers into segments and treat the task as a classification problem. The goal of the algorithm is to assign users to the segment that contains the most similar customers. To find customers similar to users, cluster models divide customers into multiple segments and treat the task as a classification problem.

The goal of the algorithm is to assign users to a segment containing many customers. most similar products. It then uses the purchases and customer reviews in the segment to make recommendations. Segments are typically created using clustering or another unsupervised learning algorithm, although some applications use manually defined segments. Using the similarity metric, the clustering algorithm groups the most similar clients together to form clusters or shards. Since optimal clustering on large data sets is impractical, most applications use different forms of greedy clustering. They then continually link customers to existing segments, often by planning to create new segments or merge existing ones. When the algorithm creates segments, it calculates how similar the users are to the vectors summarizing each segment, then selects the segment with the highest similarity and ranks the users accordingly. Some algorithms classify users into multiple segments and describe the strength of each relationship. Cluster models supply better scalability and online performance as compare to collaborative filtering because they compare users to a controlled number of segments rather than the entire customer base. Complex and expensive clustering calculations are performed offline. However, the quality of the recommendations is low. together in a segment, match users to a segment, then treat all customers in the segment as similar customers for recommendation purposes. Because the similar customers found by the cluster models are not the most similar, the recommendations they make are less relevant.

CONTENT BASED APPROACH

The content-based system generates recommendations based on the user's interests and profiles. They try to compare users with items they liked before. The degree of similarity between the elements is generally established according to the attributes of the elements that are highly rated by the user.

Advantages

Content-based models are most beneficial for recommending items when there is not enough rating data available. Indeed, other items with similar attributes may have been rated by users. Therefore, a model must be able to leverage ratings with item attributes to generate recommendations even without a lot of data.

Disadvantages

The recommendations provided are “obvious” based on the items / content the user has consumed. This is a disadvantage because if the user has never interacted with a particular type of item, that item will never be recommended to the user. For example, if you’ve never read mystery books, then through this approach, you will never be recommended mystery books. This is because the model is user specific and doesn’t leverage knowledge from similar users. This reduces the diversity of the recommendations, this is a negative outcome for many businesses.

Example:

Some examples of content based systems are:

- Amazon product feed (you’re being recommended products similar to what you’ve previously purchased)
- Spotify music recommendations.

8. METHODS FOR CONTENT BASED FEATURE SELECTION

1) **Wrapper methods** evaluate different subsets of features by training a model for each subset and then evaluating each subset's contribution on a validation dataset. As the number of all possible subsets is factorial in the number of featuredifferent heuristics are used to choose “promising" subsets (forward-selection, backward-elimination, tree-induction, etc.). Wrapper methods are independent of the prediction algorithm.

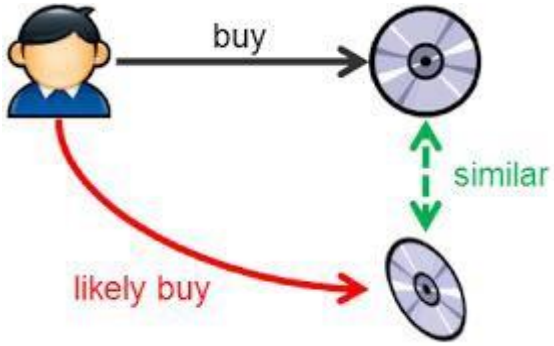
   FILTER METHODS is based on heuristic measures, such as mutual information or Pearson correlation, to score features based on their information content for the prediction task. Similar to wrapper methods, filter methods are also independent of the algorithm used. does not require training many models and thus scales well for large datasets. However, filtering methods cannot be natively extended to recommendation systems, where the prediction goal varies and depends on both the user's history and the factor under consideration. This work proposes a framework and algorithms to solve the above difficulties.

2) **Embedded methods** are a family of algorithms in which feature selection is performed in the learning phase.Unlike wrapper methods, they are not cross-validation based and therefore scale to the size of the data. However, since feature selection is an inherent property of the algorithm, an integrated approach is tightly coupled to the particular model: if the proposed algorithm is substituted, the feature selection reason must be reviewed.

9. APPROACH OF CONTENT BASED RECOMMENDATION SYSTEM

Approach 1: Analyzing the Description of Content Only

Based on my understanding, approach 1 is similar to item-based collaborative filtering. In short, the system will recommend anything similar to an item you like before.



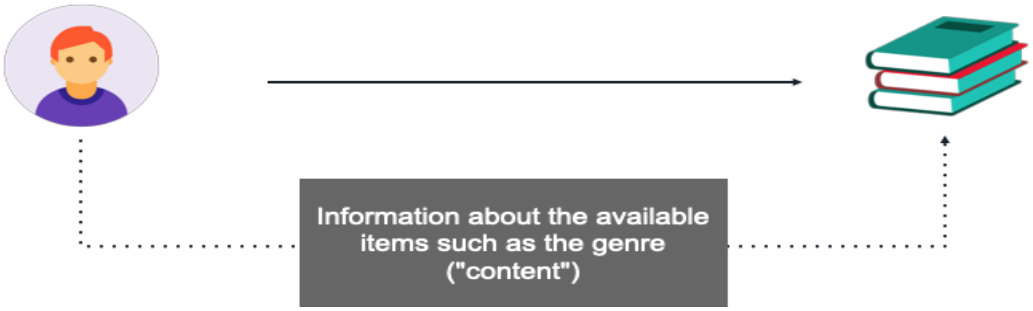
In model-building stage, the system first find the similarity between all pairs of items, then it uses the most similar items to a user’s already-rated items to generate a list of recommendations in recommendation stage.

For example, if someone watches Edge of Tomorrow, system may recommend Looper based on similarity .

Approach 2: Building User Profile and Item Profile from User Rated Content

Approach 2 leverages description or attributes from items the user has interacted to recommend similar items. It depends only on the user previous choices, making this method robust to avoid the cold-start problem. For textual items, like articles, news and books, it is simple to use the article category or raw text to build item profiles and user profiles.

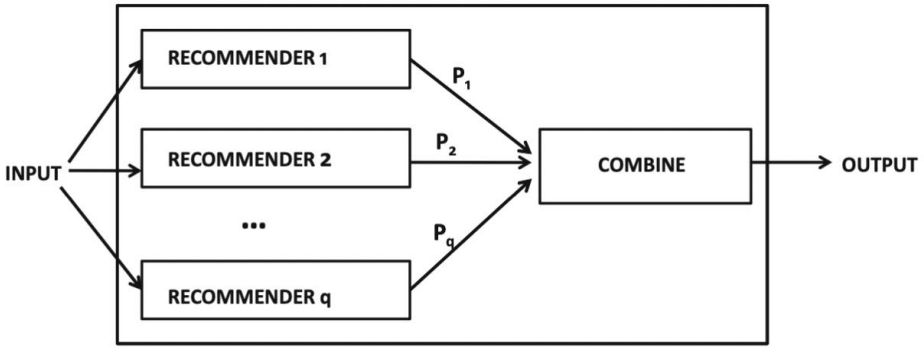
Suppose I watch a particular genre movie I will be recommended movies with respective to that specific genre. The Title, Year of Release, Director, Cast are also helpful in identifying similar movie content.



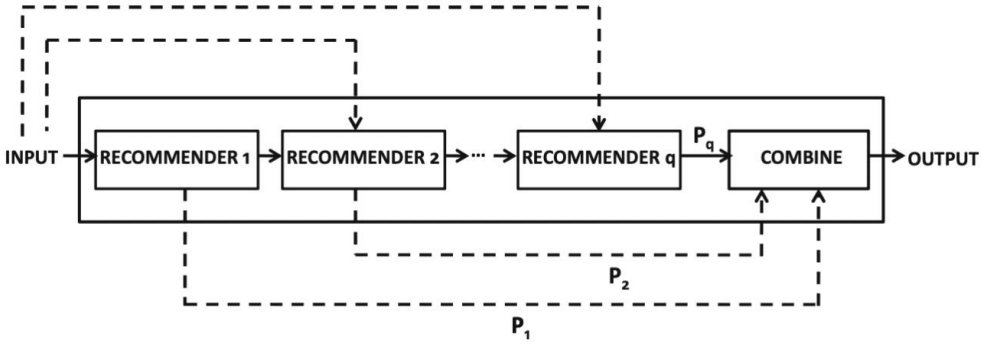
10. HYBRID APPROACH

Hybrid recommendation systems are designed to use different available data sources to generate robust inferences..

Hybrid recommendation systems have two predominant designs, parallel and sequential. The parallel design provides the input to multiple recommendation systems, each of those recommendations is combined to generate one output. The sequential design provides the input parameters to a single recommendation engine, the output is passed on to the following recommender in a sequence. Refer to the figure below for a visual representation of both designs.



(a) Parallel design



(b) Sequential design

Advantages

Hybrid systems combine different models to combat the disadvantages of one model with another. This overall reduces the weaknesses of using individual models and aids in generating more robust recommendations. This yields more robust and personalized recommendations for users.

Disadvantages

These types of models often have high computational complexity and require a large database of ratings and other attributes to stay up to date. Without up to date metrics (user engagement, ratings, etc.) it makes it difficult to retrain and provide new recommendations with updated items and ratings from various users.

Example

Netflix is a company that uses a hybrid recommendation system, it generates recommendations for users based on the viewing and search styles of similar users combined with movies that share similar characteristics that have been reviewed by other users

11. TYPES OF HYBRID

- **Weighted Hybrid** In the weighted recommendation system, we can identify a few models that can interpret the data set well. The weighted recommendation system takes the outputs of each model and combines the results into a static weight, the weight of which does not change over the training and testing sets.

- **Mixed Hybrid** First, the union method will take the user's profile and characteristics to create different candidate datasets. Thus, the recommendation system will import different candidate groups into the recommendation model and combine the predictions to produce the resulting recommendation.

12. CONCLUSION

The recommendation system opens up new possibilities for retrieving personalized information from the Internet. They also help alleviate the problem of information overload, which is very common with systems that retrieve information and allow users to access products and services that are not available to the user. on the system. This article discussed two traditional recommendation techniques and highlighted their strengths and challenges with different types of matching strategies used to improve their performance. Recommended algorithms have been discussed. This knowledge will empower researchers and serve as a road map for improving modern recommendation techniques.

REFERENCES

[1] Alexandrin Popescu and Lyle H. Ungar, David M. Pennock and Steve Lawrence," Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments", POPESCUL ET Al,2001

[2] Greg Linden, Brent Smith, and Jeremy York,"Amazon.com Recommendations Item-to-Item Collaborative Filtering",IEEE Computer Society 2003.

[3] Jun Wang, Arjen P. de Vries, Marcel J.T. Reinders," Unifying Userbased and Itembased Collaborative FilteringApproaches by Similarity Fusion", 2006 ACM.

[4] Panagiotis Symeonidis *, Alexandros Nanopoulos, Apostolos N. Papadopoulos,Yannis Manolopoulos," Collaborative recommender systems: Combining effectiveness and efficiency", 2007 Elsevier Ltd.

[5] Kazuyoshi Yoshii, Masataka Goto, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno," An Efficient Hybrid Music Recommender System Using an Incrementally TrainableProbabilistic Generative Model",IEEE 2008

[6] Akhmed Umyarov, Alexander Tuzhilin,"Improving Collaborative Filtering Recommendations Using External Data",2008 IEEE.

[7] Zunping Cheng, Neil Hurley," Effective Diverse and Obfuscated Attacks on Model-based Recommender Systems" 2009ACM.

[8] Hendrik Drachsler, Hans G.K. Hummeland Rob Koper, "Personal recommender systems for learners in lifelong learning networks: the requirements,techniques and model".

[9] Mohammad Yahya H. Al-Shamri, Nagi H. Al-Ashwal," Fuzzy-Weighted Similarity Measures for Memory-Based Collaborative Recommender Systems", Journal of Intelligent Learning Systems and Applications, 2014.

[10] Michael J. Pazzani and Daniel Billsus," Content-based Recommendation Systems".

[11] Robin van Meteren and Maarten van Someren," Using Content-Based Filtering for Recommendation".