Research on Cosine Similarity and Pearson Correlation Based Recommendation Models

Suja Cherukullapurath Mana\textsuperscript{1}\textsuperscript{*} and T.Sasipraba\textsuperscript{2}

\textsuperscript{1} Research Scholar, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India

\textsuperscript{2} Vice Chancellor, Sathyabama Institute of Science and Technology, Chennai, India

Email: cmsuja@gmail.com

Abstract. Recommendation models are based on data mining and machine learning techniques to suggest product and services to users. Advancement in the machine learning field results in the growth of more efficient recommendation models. Recommendation systems can contribute a lot to increase the business of internet based companies. Techniques for making recommendations generally fall into two categories namely collaborative filtering based models and content based recommendation models. The first techniques explores similarity based approaches like cosine similarity, Pearson correlation similarity etc. The second approaches uses predictions based on explicit and implicit ratings given by users. Recommendation systems try to suggest products to the users based on their previous purchases or general nature. A successful recommendation system is capable of improving the product sales to a large extend. This paper reviews about various recommendation models and performs a comparison of these models.

Keywords: Recommendation system, collaborative filtering, cosine similarity

1. Introduction

In this modern era most of the business models are internet based. Customer satisfaction is a very important aspect for their business success. Companies are trying to keep their customer base satisfied for increasing their profit. Analyzing the shopping pattern and recommending products to customer which they may like to purchase in near future is an important way of increasing the customer satisfaction. Recommendation systems play an important role in this aspect. Recommendation models tries to identify user interest and recommend them products based on their prior choices and interest. Growth of new technologies paves the way for the growth of advanced recommendation models which can identify user preferences with great accuracy. Majority of the recommendation framework depends on techniques like collaborative and content based filtering techniques. [1]

Collaborative filtering works on various similarity measures like cosine similarity and Pearson correlation methods [2]. It is mainly based on user to user similarities and correlation. It foresee thing to clients dependent on the decisions made by comparable users. Various collaborative filtering methods include user based approach in which the model tries to identify similarity between users in which it identifies the choices made by similar users and recommend products to similar users. It also includes item based approach in which the system tries to identify similarities among items and try to recommend similar items to the users. Fig 1 .shows a pictorial representation of content based filtering
model. In the following sections authors will go over various collaborative filtering methods including cosine based similarity method and Pearson correlation method in detail. Paper also studies about content based recommendation models and perform comparison between two models.

![Collaborative Filtering Diagram](image)

Fig 1: A diagram showing the collaborative filtering

2. **Collaborative filtering based recommendation model.**

The key concept of collaborative filtering is the similarity. Both item based similarity and user based similarity measures will be consider for collaborative filtering methods. There are two different types of collaborative filtering namely user based and item based. In user based approach similarity between two users will be the key measure. In item based approach based on the similarity among items, recommendations will be made.
Fig 2: A diagram indicating the user based approach

Fig 2 shows a representation of user based collaborative filtering. This method of filtering based on choices made by users who are similar to the targeted user and recommend the product to user. A few similarity measures which will play key role in collaborative filtering algorithms are discussed below

2.1 Cosine Similarity

Cosine similarity will be calculated as the dot product of two vectors divided by the magnitude of each vector. Generally if the cosine similarity value between two items is 1, then the items are considered to be highly similar. Cosine similarity is utilized to quantify comparability between two things of any type. Consider two items a and b, the cosine similarity between these items are calculated by using the formula given below. Cosine similarity generally checks the similarity in orientation not in size.

\[
\text{Cosine similarity (a,b)} = \frac{(a.b)}{||a||.||b||}
\]

2.2 Pearson Correlation

It is a measure to predict the correlation between two items. It is a popularly used method in recommendation systems to find the likeness between things. It uses Pearson formula to calculate the correlation between two items. The Pearson formula is given by

\[
r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}
\]

r = 1 means a strong positive correlation.

r = 0 means no correlation

r = -1 means a strong negative correlation
2.3. Similarity based on conditional probability

It is a probability that a purchaser will prefer an item given that the user has liked a similar item. In this measure if an item p is purchased by a user and he also purchased another item q, then these two items are considered to be similar. It is also interpreted in a way that if a customer is buying the item p then there is a high chance that he will buy q also as both items are similar.

\[ \text{Sim}(p,q) = P(p|q) \times C, \text{ where } C \text{ is a constant.} \]

Collaborative filtering algorithms are comprehensively delegated as user based approach and item based approach.

3. User Based Collaborative Filtering

This approach is based on the idea of categorizing users into different groups. It works under the assumption that users of a group are having same interest on items. Users are divided into groups having similar interests based on various measures. Items will be recommended on the basis of the items purchased by other similar users. Then general steps involved in a user based collaborative filtering are given below.

Step 1: Analyze user profiles and find out users which are similar interests as the targeted user.
Step 2: Identify items purchased by users and assign weights to each item so as to determine the items capable of recommending.
Step 2: Analyze the shopping trend of targeted user and recommend the item with high probability of liking by the user.

4. Content Based Filtering Method

In content based filtering method only the likeness between items are identified. This algorithm depends on the assumption that the user will buy an item which is colder to one that he has already purchased. This method tries to does not have problems of huge number of users need to be grouped and computational complexity of
greedy algorithms used for user based approach. The general steps involved in item based filtering can be summarized as follows.

Step 1: Identify the items already purchased by user
Step 2: Find the items which are similar to already purchased item
Step 3: Recommend this item to user.

Similarity can be calculated using any of the similarity calculation algorithms. Cosine similarity and Pearson correlation can be used for it.

The table given below shows the comparison between two algorithms.

| Collaborative Filtering | Content Based Filtering |
|-------------------------|-------------------------|
| It is based on how similar user buy items | It is based on the features of items purchased by users |
| It needs data about other users and their item priorities | No need for data about other users |
| It is not that efficient in making recommendation to users with unique tastes | Efficient in making recommendations to users with unique tastes |
| It is popularity biased | Not popularity biased |
| Has cold start and sparcity problem | Does not have cold start or sparcity problem |
| Can accommodate quality judgments of users. | Unable to accommodate quality judgments of other users |

Table 1: Comparison between two algorithms

5. Conclusion and Future Work

This paper reviews about various recommendation models. It will go in detail about various similarity based models like cosine similarity, Pearson correlation etc. In later sections paper discusses the steps in content based similarity model and item based similarity model algorithms. Each of these methods have its pros and cons. Content based model is subject to problems like cold start and sparcity . Similarly item based model has problems like its inability of including quality judgments from user . A proposed approach is the hybrid model which will combine content based and item based model together. This approach can address the drawbacks of both content based and item based approaches. The studies done in this paper about various approaches is an initial step towards building a hybrid recommendation model. In the future work we are trying to develop an efficient recommendation model which is capable to solving some of the problems discussed for the content based and item based models.

6. References

1. Singh, A., & Sharma, A. (2018). MAICBR: A Multi-agent IntelligentContent-Based Recommendation System. In Information and Communication Technology for Sustainable Development (pp. 399-411). Springer, Singapore.

2. H. Thanh-Nhan, H. Nguyen and N. Thai-Nghe, "Methods for building course recommendation systems", in 2016 Eighth International Conference on Knowledge and Systems Engineering (KSE), 2016, pp. 163-168.

3. Ehsan Aslanian, Mohammadreza Radmanesh, and Mahdi Jalili. 2016. Hybrid recommender systems based on content feature relationship. IEEE Transactions on Industrial Informatics (2016).

4. Alejandro Bellogin, Pablo Castells, and Ivan Cantador. 2011. Precision-oriented evaluation of recommender systems: an algorithmic comparison. In Proceedings of RECSYS. ACM, 333–336.

5. Joseph Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. User Modeling and User-Adapted Interaction 22 (2012), 101–123. Issue 1.
1. R. Devika and V. Subramaniyaswamy, "A novel model for hospital recommender system using hybrid filtering and big data techniques," 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2018, pp. 267-271.

2. M. Dhamecha, K. Dobaria and T. Patalia, "A Survey on Recommendation System for Bigdata using MapReduce Technology," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 54-58.

3. Y. Afoudi, M. Lazaar and M. Al Achhab, "Impact of Feature selection on content-based recommendation system," 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, Morocco, 2019, pp. 1-6.

4. S. R. Gandhi and J. Gheewala, "A survey on recommendation system with collaborative filtering using big data," 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, 2017, pp. 457-460.

5. Almazro D, Shahatah G, Albdulkarim L, et al. “A survey paper on recommender systems.” arXiv:10065278 2010.

6. F. Liu and W. Guo, "Research on House Recommendation Model Based on Cosine Similarity in Deep Learning Mode in Grid Environment," 2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), Jishou, China, 2019, pp. 121-12

7. J. E. Purnomo and S. N. Endah, "Rating Prediction on Movie Recommendation System: Collaborative Filtering Algorithm (CFA) vs. Dissymetrical Percentage Collaborative Filtering Algorithm (DSPCFA)," 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 2019, pp. 1-6.

8. N. Sivaramakrishnan and V. Subramaniyaswamy, "GPU-based Collaborative Filtering Recommendation System using Task parallelism approach," 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2018, pp. 111-116.