Scientific Journal of Impact Factor(SJIF): 3.134
e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2,Issue 3, March -2015

COLLABORATIVE FILTERING APPROACHES FOR MOVIE RECOMMENDATION SYSTEM USING PROBABILISTIC RELATIONAL MODEL

Asst. Prof. Hiren M. Patel [1], Asst. Prof Jayna B. Shah [2]

[1] *Computer Engineering Department, SVIT-Vasad*
[2] *Computer Engineering Department, SVIT-Vasad*

ABSTRACT:- *Recently, it has become more and more difficult for the existing web based systems to locate or retrieve any kind of relevant information, due to the rapid growth of the World Wide Web in terms of the information space and the amount of the users in that space. However, in today's world, many systems and approaches make it possible for the users to be guided by the recommendations that they provide about new items such as articles, news, books, music, and movies. However, a lot of traditional recommender systems result in failure when the data to be used throughout the recommendation process is sparse. This Paper focuses on the development and evaluation of a web based movie recommendation system.*

Keywords: *Recommender Systems, Personalization, User Modeling, Collaborative Filtering, Content Based Filtering, Information Extraction, Pearson Correlation Coefficient, Sparsity Problem, Probabilistic Relational Model.*

I. INTRODUCTION

In today's world, many systems and approaches make it possible for the users to be guided by the recommendations they provide about new items such as news, web pages, articles, books, music, and movies. In addition, recommender systems are being used in an ever-increasing number of e-commerce sites such as Amazon[1] for books, IMDb[2], MovieLens[3], and MovieFinder[4] for movies, Pandora[5] for music domains, in order to assist buyers in finding suitable precuts[7] .At the online world, RSs act as vendors who connect the items to customers by predicting their preferences. Such systems not only reduce the searching time of the user, but also enhance the selling rates.

For movie recommendation system One features are integrated in order to handle the sparsity[11]. In addition, the cultural metadata of the movies are exploited by the CF approach in order to make more successful and realistic predictions. The first fact that is being focused on Information Based collaborative filtering methods may not achieve success when the training data is sparse. Thus, a synthesis of two methods and Effective Missing Data Prediction, which have been proven to handle this problem in a reasonable way, is used to handle the data sparsity problem with a stronger approach. RSs require any kind of information about the preferences of the users and a method to decide if an item is interesting for a specific user.

II.RECOMMENDATION TECHNIQUES

Since one of the most important components of every recommender system is making predictions, it is logical to classify them according to the prediction technique. Based on the prediction technique it classifies the RSs into three main groups:

A. Information-based prediction techniques.
B. Social-based prediction techniques.
C. Combination of the first two techniques (Hybrid Techniques).

A. Information-based Prediction Techniques

In this method the actual user is taken into consideration and the information related to other users is not processed. So these techniques are considered domain specific you have to analyze the information stored in the metadata associated with them. Examples of these techniques are: case-based reasoning or content-based recommendation (CBR), information filtering, attribute-based techniques [8].

The main assumption under case-based reasoning or content-based recommendation techniques is that a user has similar preferences over similar items. The more similar the items, the more equal the preferences of the user on those items. In content-based recommender systems, various candidate items are compared with items previously rated by the user and the best matching items are recommended. The strengths of CBR can be listed as below

1. It can recommend completely new items.
2. Simple to understand, it‘s easy to explain a process of recommendations to users.
3. Since the content is usually constant, an item should be analyzed only once.

The weaknesses of CBR

1 This weakness can be shortly named as limited content analysis. Most of the time, it is really hard work to extract the truly relevant and significant features from the content.

B. Social-based Prediction Techniques

In this techniques, both the information about the actual user and the whole set of users are analyzed. Since item information is not used, these techniques are domain independent. Some examples of social-based prediction techniques are collaborative filtering (CF), item-item filtering, stereotypes and demographics, popularity, average. Popularity and average are simple prediction techniques that recommend items based on their popularity among all users, or on the average of all the set of ratings of an item.

1. Collaborative Filtering Method

CF is based on the assumption that people who rate items in a similar way probably have similar preferences and tastes. The use of stereo types and demographics are related with the fact that people who comes from the same background (i.e. age, gender, occupation, education, demographic data, etc.) usually exhibits exchangeable preferences.

2. Classification of Collaborative Filtering Method

Algorithms for collaborative recommendations can be grouped into two main classes.

2.1Memory-Based Algorithms

These are heuristics that make rating predictions based on the entire collection of previously rated items by the users. That is, the value of the unknown rating for a user and an item is usually computed as an aggregate of the ratings of some other (usually the N most similar) users for the same item.

User-based CF predicts an active user‘s interest in a particular item based on rating information from similar user profiles.

Item-based Approaches use the similarity between items instead of users. After, the similarity of items are calculated, unknown ratings can be predicted by averaging the ratings of other similar items rated by the active user.

The main advantage of item-based CF over user-based CF is its scalability.

2.2. Model-Based Algorithms

In contrast to memory-based algorithms, model-based algorithms use the collection of ratings as a training dataset to learn a *model*, which is then used to make rating predictions.

3. Algorithms for Collaborative Filtering Method

As mentioned before, one important aspect to be considered in CF method is the way similarity between the profiles of users is computed. There will be explained in detail below.

3.1 Mean Squared Differences Algorithm

This algorithm estimates the degree of dissimilarity between user profiles by calculating the mean squared difference between them.

The formula used for calculating the similarity between two users is given below:

$$D_{x,y} = \frac{\sum_{\mathrm{n}}^{\mathrm{Na}} C_{xn} . C_{yn} . (S_{xn} - S_{yn})^2}{\sum_{\mathrm{n}}^{\mathrm{Na}} C_{xn} . C_{yn}}$$

where C_{xn} = [1,0]: depending on whether item n is rated by user x or not

C_{yn} = [1,0]: depending on whether item n is rated by user x or not

S_{xn} is the rate of item n given by user x

S_{yn} is the rate of item n given by user y

3.2 Vector Similarity (VS)

This algorithm looks at the arrays of user ratings as vectors, and uses the cosine of the angle between the vectors as an index of similarity. The equation for similarity between users in the vector similarity algorithm becomes the normalized dot product of the two vectors, or the cosine of the angle between them.

$$w_{up,uq} = \frac{\sum_{\{i|r_{p,i}\,\&\,r_{q,i}!=0\}} r_{p,i}\ r_{q,i}}{\sqrt{\sum_{\{i|r_{p,i}!=0} r_{p,i}^2} \cdot \sqrt{\sum_{\{i|r_{q,i}!=0} r_{q,i}^2}}$$

where :

$w_{up,uq}$: Similarity between users u and q

$r_{p,i}$: The rate user p gave item i

$r_{q,i}$: The rate user q gave item i

3.3 Pearson Correlation Coefficient (PCC)

User-based CFF engaging PCC is used in a number of RSs, since it can be easily implemented and can achieve higher accuracy than other similarity computation methods. In user-based CFF, PCC is employed to define the similarity between user a and user u based on the items that they rated in common. The related formula, where Sim(a,u) denotes the similarity between users a and u, and i belongs to the subset of items which were rated by both of the users, is the rate user a gave item i, and avg(r_a) represents the average rating of user a, is given below :

$$\mathrm{Sim}(a,u) = \frac{\sum_{i\in I(a)\cap I(u)}(r_{a,i}-avg(r_a)).\,(r_{u,i}-avg(r_u))}{\sqrt{\sum_{i\in I(a)\cap I(u)}(r_{a,i}-avg(r_a))^2}\ .\sqrt{\sum_{i\in I(a)\cap I(u)}(r_{u,i}-avg(r_u))^2}}$$

On the other hand, the basic idea in similarity computation between two items i and j by using PCC is to first isolate the users who have rated both of these items and then apply a similarity computation technique to determine the similarity Sim(i,j). The related formula, where Sim(i,j) denotes the similarity between items i and j, and u belongs to the subset of users who rated both of the items is the rate user u gave to item i, and avg(r_i) represents the average rating of item i, is given below:

$$\mathrm{Sim}(i,j) = \frac{\sum_{u\in U(i)\cap U(j)}(r_{u,i}-avg(r_i)).\,(r_{u,j}-avg(r_j))}{\sqrt{\sum_{u\in U(i)\cap U(j)}(r_{u,i}-avg(r_i))^2}\ .\sqrt{\sum_{u\in U(i)\cap U(j)}(r_{u,j}-avg(r_j))^2}}$$

→Strengths of the CF method

1. No electronic representation of the items in order to be analyzed by the computer is required

2. The content of the recommendations might be very different from the original preferences of the user so that the user can also taste items other than his/her previous likes.

3. CF techniques are domain independent and can work perfectly in domains where it‘s hard to extract content information from the items or where there is not content at all associated with the items.

→Weaknesses of CF Method

1. Most users rate just a few of the items in the collection, which causes the user-item rating matrix to become very sparse. And this leads to a reduced probability of finding a set of users with many ratings in common. This is called the sparsity problem.

2. Collaborative RSs tend to fail when there exist little information about preferences or when a user has quite uncommon interests.

C. Hybrid Techniques

The main idea behind hybrid recommendation techniques is that a combination of algorithms can provide more accurate recommendations than a single algorithm and disadvantages of one algorithm can be overcome by other algorithms. In order to exploit the advantages of the CB and CF recommendation methods that were mentioned in the previous sections, several hybrid approaches have been proposed, concerning combinations of CB and CF.

III. PROBABILISTIC RELATIONAL MODEL FOR MOVIE RECOMMENDATION SYSTEM

Collective approach using Probabilistic Relational Model(PRMs) and Information Based Prediction technique to recommend a movie and solve the problem of sparsity in Movie Recommendation System. A probabilistic relational model(PRM)[6] for a relational schema S is defined as follows. For each class X €*X* and each descriptive attribute A€ A (X), we have a set of parents Pa(X:A), and a conditional probability distribution (CPD)[6] that represents

P(*I* /σ,Θs)= P(X.A/ Pa(X.A)).

For example, consider a schema describing a domain describing votes on movies. This schema has three classes called Vote, Person, and Movie. For the Vote class, the descriptive attribute is Score with values {0,1,2,3,4,5}, for Person the descriptive attributes are Age and Gender, which take on values {young; middle-aged; old} and {Male; Female} respectively. For example, the Vote class would be associated with two reference slots: Vote.ofPerson, which describes how to link Vote objects to a specific Person; and Vote.ofMovie, which describes how to link Vote objects to a specific Movie object. If new user register with system then his age is 35 and gender is Male then using PRM it find the best middle age male user interest probability and give best recommendation based on that probability.

IV. RESULT

The experimental evaluation of Movie Recommendation System was conducted using the MovieLens[3] dataset maintained by the GroupLens[6] Research group. That datasets containing 100,000 ratings on a scale of 1 to 5 for 1682 movies by 943 users, where each user has rated at least 20 movies, was preferred in order to make the evaluation results. The density of the user-item matrix created from the MovieLens dataset is:

$$\frac{100{,}000}{943 \times 1682} = 6.30\ \%$$

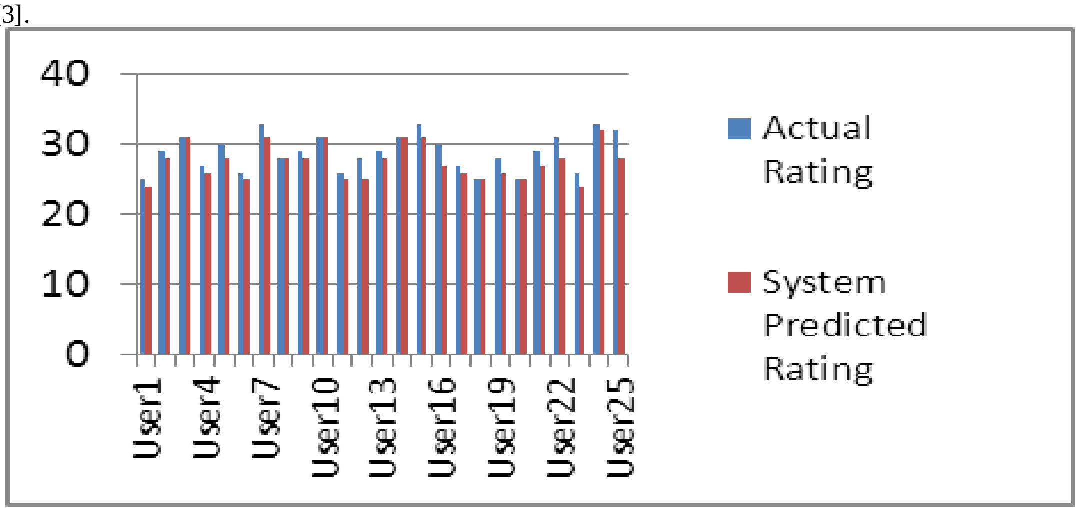
which can be considered to be appropriate enough in terms of sparsity for the evaluation of the system.

A. Testing and Result

For the purpose of measuring the prediction quality of the proposed approach and comparing with other CF methods, Mean Absolute Error (MAE) metrics was used. Testing and result are taken on the MovieLens[3] dataset. The MAE[12] is computed by first summing the absolute errors of the N corresponding ratings prediction pairs and then averaging the sum. And it can be more formally defined as:

$$\text{MAE} = \frac{\sum_{i=1}^{N} | r_i - r'_i |}{N}$$

where r_i denotes the actual rating that the related user gave for item i, and r'_i denotes the rating predicted by System approach, and N denotes the number of tested ratings. As can be observed, a larger MAE indicates a lower accuracy [3].



For experimental result we test different User. Comparison shows that our algorithm has a relatively better performance than the existing method[2]. Proposed algorithms show better result for MAE.

Algorithm	Mean Absolute Error(MAE)
Social Based Prediction Technique	1.059
Pure Collaborative filtering	1.002
Naïve Hybrid	1.011
Information Based Prediction Technique	0.962
Proposed Approach	**0.917**

V. CONCLUSION

In this paper various methods have been summarized for collaborative filtering. There is an abundance of practical applications that help users to deal with information overload and provide personalized recommendations, content, and services to them. We outlined a framework for modeling the collaborative filtering problem with PRMs. We improve the expressiveness and context-sensitivity of other methods using standard PRM.

REFERENCES

[1] Amazon.com, www.amazon.com.
[2] The Internet Movie Database (IMDb), http://www.imdb.com/.
[3] MovieLens, www.movielens.umn.edu.
[4] E! Online, www.moviefinder.com.
[5] PANDORA, www.pandora.com.
[6] http://www.grouplens.org
[7] J. Ben Schafer, Joseph Konstan, and John Riedl, "Recommender systems in e-commerce," ACM Conference on Electronic Commerce (EC-99), Denver, CO, November 1999.
[8] Mark Van Setten, "Supporting people in finding information, Hybrid recommender systems and goal-based structuring," Telematica Instituut Fundamental Research Series, vol. 016. Enschede, the Netherlands: Telematica Instituut, 2005.
[9] Kazuyoshi Yoshii, Masataka Goto, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno, "Hybrid collaborative and content-based movie recommendation using probabilistic model with latent user preferences," ISMIR, pp. 296-301, 2006.
[10] Hugo Siles Del Castillo, Hybrid Content-Based Collaborative-Filtering Music Recommendations, Master Thesis, University of Twente, November 2007.
[11] Lise Getoor, Nir Friedman, Daphne Koller, Benjamin Taskar "Learning Probabilistic Models of Relational Structure " in Journal of Machine Learning Research, **3**:679-707, 2002.
[12] Prem Melville, Raymond J. Mooney, Ramadass Nagaraja, Department of Computer Sciences University of Texas :"Content-Boosted Collaborative Filtering for Improved Recommendations" in American Association for Artificial Intelli-gence 2002.
[13] J. Newton. and R. Greiner. Hierarchical probabilistic relational models for collaborative filtering. In Proc. Workshop on Statistical Relational Learning, 21st International Conference on Machine Learning. Banff, Alberta, Canada. 2006.
[14] Gediminas Adomavicius, and Alexander Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Transactions on Knowledge and Data Engineering,VOL. 17, NO.6, June 2005.
[15] Shinhyun Ahn, Chung-Kon Shi, "Exploring Movie Recommendation System Using Cultural Metadata," International Conference on Cyberworlds, 2008.
[16] Heng Luo, Changyong Niu, Ruimin Shen, Carsten Ullrich, "A collaborative filtering framework based on both local user similarity and global user similarity," in Proc. 2008 European Conference on Machine Learning and Knowledge Discovery in Databases, Part I, 2008.
[17] Ma, H., King, I., and Lyu, M. R., "Effective missing data prediction for collaborative filtering," in Proc. 30th annual international ACM SIGIR conference on research and development in information retrieval, pp. 39-46, NewYork,

2007.
[18] George Lekakos, Petros Caravelas, “A hybrid approach for movie recommendation,” Springer Science + Business Media, LLC, December 2006.
[19] ChoiceStream Technology Brief, Review of Personalization Technologies: Collaborative Filtering vs. ChoiceStream‘s Attributized Bayesian Choice Modeling.

