

# Collaborative Filtering for Movie Recommendation using RapidMiner

Arpita Jain

Research Scholar

Department of Computer Science & Engineering,  
GGITS, Jabalpur

Santosh K. Vishwakarma

Associate Professor

Department of Computer Science & Engineering,  
GGITS, Jabalpur

## ABSTRACT

Recommender System is a special type of information filtering system that provides a prediction which helps the user to evaluate items from a huge collection that the user is likely to find interesting or useful. Recommender System is used to produce meaningful suggestions about new items for particular consumers. These recommendations facilitate the users to make decisions in multiple contexts, such as what items to buy, what online news to read or what music to listen to. Recommender Systems have become important in information and decision overloaded in the world. Recommender Systems helped their founders to increase profits. This paper, presents a brief overview of collaborative filtering based movie recommender system and their implementation using rapid miner.

## Keywords

Recommender System, collaborative filtering, utility matrix, rapidminer operators.

## 1. INTRODUCTION

The access growth of e-commerce and online environments have made problems in information search and selection increasingly serious, information are overloaded by options to consider and users do not have knowledge or the time to personally evaluate these options. The purpose of developing recommender systems is to reduce information overloaded by fetching the most important information and services from a huge amount of data providing personalized services. Recommender systems can be defined as programs which attempt to recommend the most suitable items to particular users by predicting a user's interest in an item based on related information about the item, the users and the interactions between items and users [3]. The most important feature of a recommender system is its ability to guess a user's preferences and interests by analyzing the behavior of this user and /or the behavior of other users to generate personalized recommendations [4].

The recommender systems pertain data mining [5][6] techniques and prediction algorithms to predict user's interest on information, product and users. There are three basic types of recommender systems: content based, collaborating filtering and hybrid systems. Content based approach recommends items based on a similarity between the content of the items and a user profile. If social information is available, i.e. interaction of other users with the items, then it is known as collaborative filtering. By combining both the content based and the collaborative filtering recommendation a hybrid system is constructed.

Online social networks make available new openings as to improve the accuracy of recommender systems. In real life, people often takes an advice from their friends or relatives in

their social networks before purchasing a product. People are more enthusiastic to share their personal opinions with their friends and belief recommendations from their friends more than those from strangers and sellers. Popular online social networks like Facebook suggests us to make more friends using 'You may know' section, similarly LinkedIn advises you to connect with people you may know and YouTube recommends you relevant videos based on previous browsing history. Now a days a number of e-shopping recommender systems such as Amazon, Flip cart and e-Bay have been developed to help their customers for finding products to purchase. With the wide usage of mobile devices, a fast growth in movie, music and video resources has taken place. For this many movie recommender systems such as Pocket Lens[7] and Cinema Screen[8], and music recommender systems such as Fly casing[9], Smart Radio[10] have been developed. Following figure 1 shows Recommender system.

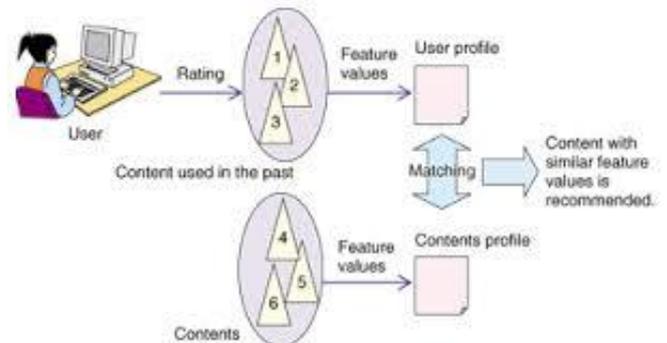


Fig 1: Recommender system

Collaborative filtering filters information by using the recommendations of other people. It is built on the knowledge that people who decided in their evaluation of certain items in the earlier are probable to decided again in the future. For example, a person who wants to watch a movie, may inquire for recommendations from friends. The recommendations of specific friends who have similar comforts are trustworthy more than recommendations from others. This tactic produce user specific recommendations of items founded on patterns of ratings without requisite for external information about either items or users[11]. The performance of the users are open :the users are requested to rate the items i.e. 1-5 star measure. But in real life, it is very emblematic for a user to rate all the items, so the matrix of all ratings users × items is sparse shown in fig. 2.

	Item 1	Item 2	Item 3	...	Item n
User 1	2	3	?	...	5
User 2	?	4	3	...	?
User 3	3	2	?	...	3

...	...	...	...	...	...
User n	1	?	5	...	4

Fig 2: Utility matrix

Collaborative Filtering systems can be categorized as User-based and Item-based. Among each of the user-based and item-based collaborative tactic there can be further sub-categorized depending upon the similarity degree designated, Pearson or Cosine. In user-based and item-based, the effort is to predict user's ratings for items that has not rated and at that point suggest to him the item(s) with the uppermost rating(s). The design of Recommender systems in RapidMiner has been moreover simplified over the Recommender Extension.

## 2. REVIEW OF LITERATURE

We have studied various research papers and know about recommender systems and collaborative filtering. We discuss here some related work in the field of Recommender System with collaborative approach by the name of authors and their respective title.

By Guangping Zhuo, Jingyu Sun and Xueli Yu [12] "A Framework for Multi-Type Recommendations" treaties in the ground of web mining anxiety on certain shortcomings in collaborative filtering and also on multi type recommendation. Collaborative filtering (CF) is an operative method of recommender systems (RS) has been broadly used in online goods. Still, CF discomfords from some flaws such as difficulties with new users (cold start), data sparseness, and difficulty in noticing "malicious" or "untrustworthy" users and so on.

By Qian Wang, Xianhu Yuan, Min Sun [13] "Collaborative Filtering Recommendation Algorithm based on Hybrid User Model". Collaborative filtering confronts difficulties of adaptability and furthermore suggestion precision so the paper proposes a cross breed user model to expel some of its downsides. The recommender framework in light of this model not just holds the benefit of suggestion accuracy in memory-based strategy, additionally has the scalability in the same class as model-based technique.

By Chuanguang Huang and Jian Yin [14] "Effective Association clusters Filtering to Cold-Start Recommendations". This paper concentrates on the best way to defeat cold start issue in the conventional research of Recommendations System (RS). The famous procedure of RS is Collaborative Filtering (CF). While in genuine online RS, CF can't basically take care of cold start issue for the sparsity ratings dataset. The paper proposed a fresh effectually association clusters filtering (ACF) algorithm. It also tells an effective personalized collaborative filtering scheme for Network service recommendation.

## 3. RAPIDMINER EXTENSION FOR COLLABORATIVE RECOMMENDER ENGINE

Recommender engine in RapidMiner Extension has been prepared by e-LICO, an e-Laboratory for Interdisciplinary Collaborative Research in Data –Intensive Science and Data mining. In the Recommender extension there are essentially three sorts of operators as Item Recommendation, Item Rating Prediction, and Recommender Performance. We usage operators relating to Collaborative Filtering in 'Item Rating Prediction'. Collaborative grounded operators' proceeds an example set as input having training data however recurring a

trained model and unaffected training data. We also formed the Apply Model and the Performance operators for every type. Apply Model Rating prediction operator retrieve a trained model and a test as input. Output of the Apply Model is practice to compute performance using the Performance operator. The Performance operator compute the assessment of rating prediction error methods: Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Normalized Mean Absolute Error (NMAE) [15]. These error measure values are returned as a performance vector and an example set.

Here using the data set from Group Lens Research Project. Ratings are on a gauge of 1 to 5 (whole-star ratings only). There are three files: 'ratings.dat', 'users.dat' and 'movie.dat'. For prediction of user ratings using collaborative filtering we requisite data with three fields: userid, movieid and ratings. This information is obtainable in file 'ratings.dat' which shows in figure 3. File 'ratings.dat' contains the user ratings in the format: UserID, MovieID, Rating, Timestamp. Each user has at least 20 ratings and surely ratings by users are sparse.

ExampleSet (100004 examples, 1 special attribute, 2 regular)			
Row No.	rating	userid	movieid
1	2.500	1	31
2	3	1	1029
3	3	1	1061
4	2	1	1129
5	4	1	1172
6	2	1	1263
7	2	1	1287
8	?	1	1293
9	?	1	1339
10	?	1	1343
11	?	1	1371
12	?	1	1405
13	?	1	1953
14	?	1	2105
15	?	1	2150
16	?	1	2193
17	?	1	2294
18	?	1	2455
19	?	1	2968
20	?	1	3671
21	?	2	10
22	?	2	17
23	?	2	39
24	4	2	47
25	4	2	50

Fig 3: Initial datasets from GroupLens

## 4. CREATE WORKFLOW OF ITEM RECOMMENDATION IN RAPID MINER

### 4.1 Using Single Model

In RapidMiner, using AML operators to read sparse datasets. For this purpose creates two attribute description files (AML files), training.aml and testing.aml to describe the attributes of our two '.dat' files. Collaborative recommender operators practice the user-item matrix to figure a recommendation model. The example set having data for that user whose ratings for certain movies we want to predict. Item recommendation systems purpose to rank dissimilar items for

each user in the database. Item based recommender first produce a matrix that processes the similarities between all duos of items (which in our sequence are the different movies). For every particular user, an unranked item is formerly existing with a short-list (k-nearest neighbors) of

"utmost" similar items. The ending predicted for that item will be a weighted average of the ratings delivered by the user for the k-similar items. Then we build the following workflow Figure 4 in Rapid Miner and run it.

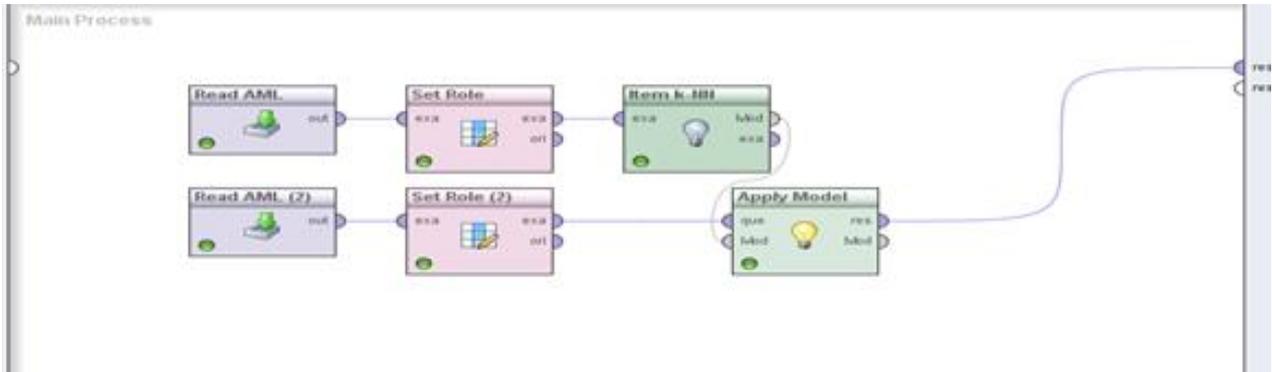


Fig 4: Workflow to find users rating for movies for which no ratings have been given by the user.

The recommended results shown in following Figure 5.

Data View  Meta Data View  Plot View  Advanced Charts

ExampleSet (682 examples, 3 special attributes, 1 regular attribute)

Row No.	rating	movieId	userId	prediction
1	?	1293	1	3.194
2	?	1339	1	3.409
3	?	1343	1	3.227
4	?	1371	1	3.073
5	?	1405	1	3.488
6	?	1953	1	3.079
7	?	2105	1	3.179
8	?	2150	1	3.528
9	?	2193	1	2.932
10	?	2294	1	3.172
11	?	2455	1	3.277
12	?	2968	1	2.644
13	?	3671	1	3.297
14	?	10	2	3.467
15	?	17	2	3.970
16	?	39	2	4.047

Fig. 5: Ratings have been predicted for movies that the users had not rated

In the item recommendation workflow in fig 4, Read AML operator two times in workspace to delivered training and testing data. Then set roles for the three attributes. Roles will be as given below:

- UserId: user identification
- ItemId: item identification
- Rating: label

Then drag Item k-NN operator, from in Recommender Systems→Item Rating Prediction and moreover Apply Model (Rating Prediction). It gives an example set having predicted ratings for unrated movies for a particular user in a testing set. Fig. 5 shows predicted unrated movies for a particular user. The following figure 6 shows how to measure performance of a recommendation model. The workflow for evaluating recommender model performance in fig. 6 is similar as in fig. 4. Here using Read AML operators two times to load the input data and set roles for userID and itemID as user identification and item identification respectively. Next drag Item k-NN operator, from in Recommender System under Item Rating Prediction and also drag Apply Model under Rating Prediction together with the Performance operator under Rating Prediction and run it. Then getting result in terms of RMSE, MAE and Normalized Mean Absolute Error (NMAE) are shown in Figure 7. Lesser values of both RMSE and MAE shows better accuracy. RMSE is usually larger than or equal to the MAE. If both metrics are equal, then all errors have the same measure.

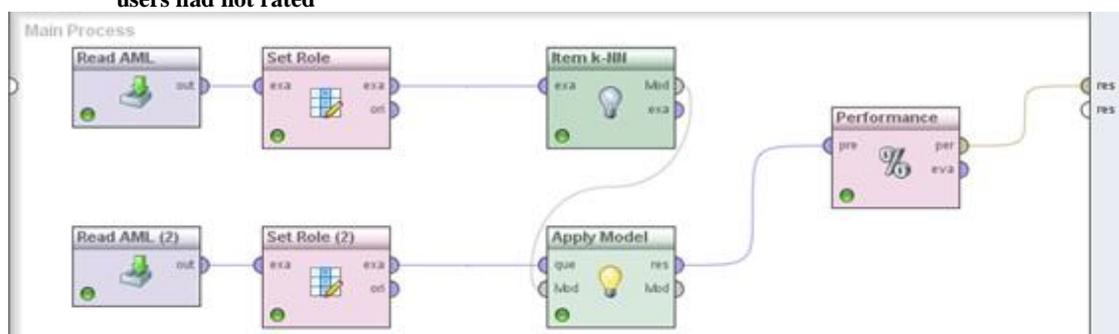


Fig 6: The complete model with Item k-NN as model builder

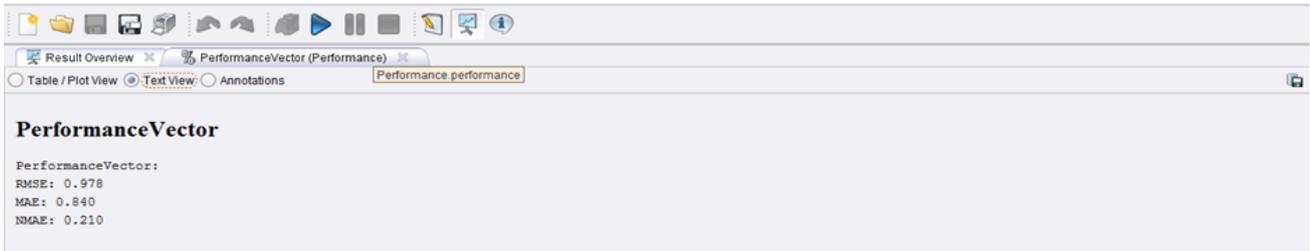


Fig 7: Performance of the Recommender System using single model

## 4.2 Ensemble of models

Numerous item ratings algorithms can be associated in parallel and their outcome pooled in a weighted manner to build a model for recommender. For this purpose usage

twofold operators Multiply and Model Combiner (Rating Prediction). Here also uses Item k-NN and Matrix Factorization algorithms in parallel. The workflow for this is shown in following figure 8.

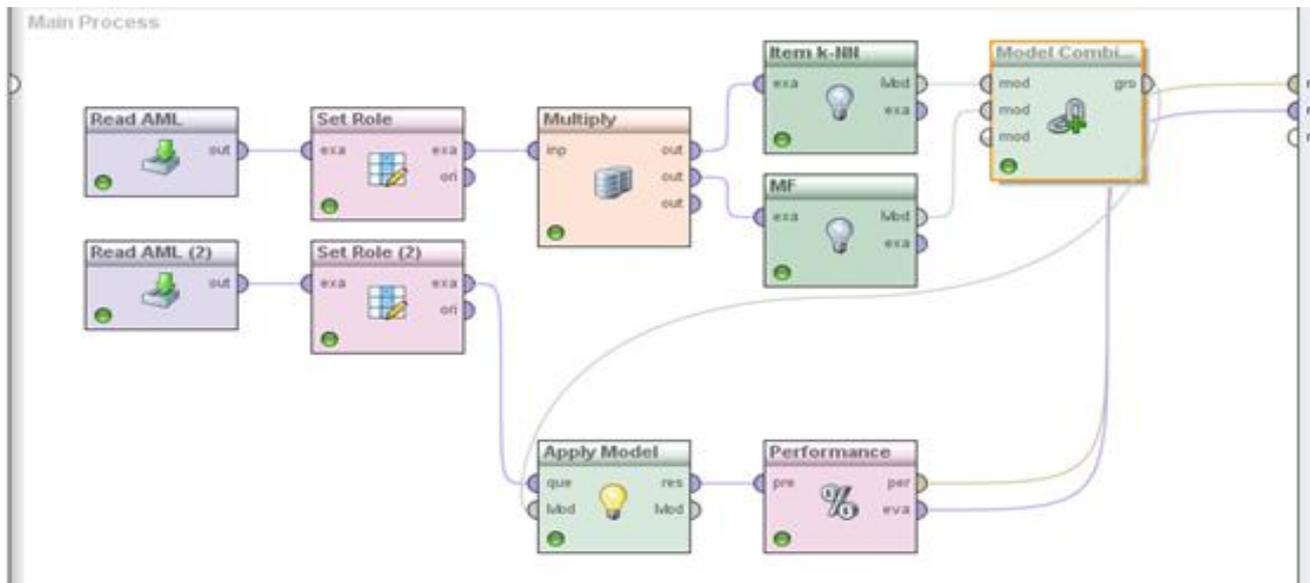


Fig 8: Using Multiply operator, Model Combiner (Rating Prediction) to put on multiple models to a set of data.

On successively running this multiple model process, with defaulting parameter, there is a minor improvement in the

performance of a recommender system. The result is shown below in figure 9.

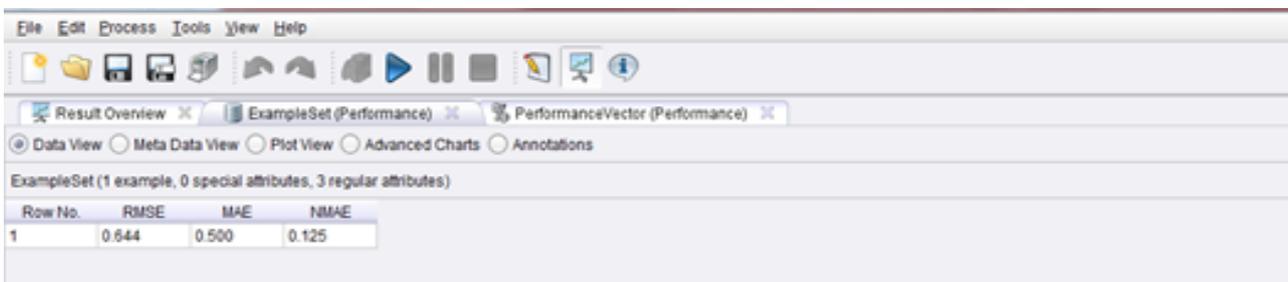


Fig 9: Performance of the Recommender System using multiple model

## 5. CONCLUSION

This paper presented recommender systems and implementation of collaborative movie recommendation using RapidMiner Recommender Extension. The benefit of using a device like RapidMiner is that it permits persons innovative to the mode an easy overview to the system and delivers a very well-organized means to speedily experimentation with dissimilar algorithms until you discover the precise

arrangement that workings for your dataset. This paper measured the performance of recommender system using a single model (i.e. item k-NN) and using multiple models (i.e. item k-NN and Matrix Factorization). However using multiple models for recommender system, give better performance than a single model. In future using the knowledge gained from this paper in other recommendation problem applications and business will help to getting quality and knowledgeable decision.

## 6. REFERENCES

- [1] Google: <http://www.google.com/>.
- [2] Social Computing Research at the University of Minnesota <http://www.grouplens.org/>.
- [3] D J.Bobadilla, F. Ortega, A.Hernando, A. Gutierrez, Recommender systems survey, *Knowledge-Based Systems*, 46(2013) 109-132.
- [4] P.Resnick,H.R. Varian, Recommender systems, *Communications of the ACM*, 40(1997) 56-58.
- [5] Deshpande, M., and Karypis, G. Item-based top-n recommendation algorithms. *ACM Trans. Inf. Syst.* 22, 1 (2004), 143-177.
- [6] Breese, J., Heckerman, D., and Kadie, C., Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence*, page 4352, 1998.
- [7] B.N. Miller, J.A. Konstan, J. Riedl, PocketLens: Toward a personal recommender system, *ACM Transactions on Information Systems (TOIS)*, 22 (2004) 437-476.
- [8] B.N. Miller, J.A. Konstan, J. Riedl, PocketLens: Toward a personal recommender system, *ACM Transactions on Information Systems (TOIS)*, 22 (2004) 437-476.
- [9] Chuanguang Huang and Jian Yin “Effective Association Clusters Filtering to Cold-Start Recommendations”, 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery
- [10] J. Salter, N. Antonopoulos, CinemaScreen recommender agent: combining collaborative and content-based filtering, *Intelligent Systems, IEEE*, 21 (2006) 35-41.
- [11] C. Hayes, P. Cunningham, Smart radio—community based music radio, *Knowledge-Based Systems*, 14 (2001).
- [12] Ricci F, Rokach L, Shapira B, Kantor PB (eds) (2011) *Recommender Systems Handbook*. Springer
- [13] Yechun Jiang, Jianxun Liu, Mingdong Tang and Xiaoqing (Frank) Liu “An Effective Web Service Recommendation Method based on Personalized Collaborative Filtering”, 2011 IEEE International Conference on Web Services.
- [14] Guangping Zhuo, Jingyu Sun and Xueli Yu “A Framework for Multi- Type Recommendations”, Eighth International Conference on Fuzzy Systems and Knowledge Discovery, 2007.
- [15] Qian Wang, Xianhu Yuan, Min Sun “Collaborative Filtering Recommendation Algorithm based on Hybrid User Model”, FSKD, 2010.
- [16] Z. Gantner, S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. Mymedialite: a free recommender system library. In *Proc. of the fifth ACM conference on Recommender systems, RecSys '11*, pages 305–308, 2011.