ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 9(31), DOI: 10.17485/ijst/2016/v9i31/96291, August 2016*

MOVBOK: A Personalized Social Network Based Cross Domain Recommender System

Sharu Vinayak, Richa Sharma and Rahul Singh

Department of Computer Science and Engineering, Chandigarh University, India;
sharuvinayak@gmail.com, richas30@yahoo.com, rahulsinghcse25@gmail.com

Abstract

Objective: We propose a novel idea for resolving research issues like cross domain recommendations and recommendations using social networks in the emerging research field recommender systems. **Methods/Analysis:** According to this idea user will be recommended with the list of books that belong to the genre that is most liked by the user in terms of movies. **Findings**: Here we will collect user's tastes in movies from his social network profile and extract out the most liked genre by him and using an appropriate collaborative filtering algorithm will recommend him with the books that may interest him. **Improvement**: The proposed idea is expected to resolve research problems like cold start problem and sparsity. Our proposed methodology gives more competent results than the traditional.

Keywords: Cross Domain Recommendations, F1 Score, Precision, Recall, Recommender Systems, Root Mean Square Error

1. Introduction

Recommender system is one of the trending research fields in the present era. Many researches have been done and are still being done in this field. In literature maximum of the recommender systems make recommendations for a single domain[1-6]. For example, Movie Lens compute recommendation for movies only i.e. movies recommended to the user are closely related to those that he/she already likes.

On the other hand[4,7] aim to recommend news, recommending music albums is the goal of[8,9] is to recommend restaurants etc. Proceeding with this paper a domain refers to set of those items that are similar to one another in characteristics that can be easily distinguished for instance, movies, TV programs, music, games etc. In actual we do not stick to a fixed definition and would use it in a more supple form, as a domain can be split up into more precise ones like books into textbooks and novels.

Despite being useful a single domain recommendation is not enough in some of the cases for instance, Amit likes romantic movies and enjoys watching the movie The Fault in Our Stars. Single domain recommender systems will recommend him with the movies that are similar to the one liked by him. However, for an online shopping site it may constitute of items like movie DVDs, Books, Music CDs etc. Therefore, as per our belief the recommender system would be more effective if it can recommend Amit with other items apart from movies. For example, book with the same title can be recommended to Amit or some other related books too can be recommended to him. The biggest dearth of current recommender systems is that

**Author for correspondence*

they focus on getting input from user about their interests or tastes in an area in order to produce recommendations in the same area.

The core perception that inspires our study i.e. the cross domain recommendation is highlighted in the above example. In spite of being applied in the real world this concept remains a rarely studied research issue. On one hand where census scrutiny of popular items that abandon the personalization of recommendation are counted upon by current recommender systems , this paper tries to fix this gap through the study of cross domain and personalized recommendation. Initial studies on cross domain were presented[10], where the authors hypothesized that the precision of recommendations of a cross domain recommender systems may be less as compared to that of a single domain recommender system. The topic has become oh so trending and vastly popular these days due to the arising sparsity issue[10]. In their work authors have generated a new algorithm namely the Code Book Transfer (CBT) algorithm to user movie ratings given by the user from source domain for recommending books to the users in the target domain. The focus is not to have users or items in the two domains to be identical or even overlap. Empirical tests are performed in order to explore how accurate recommendations are produced by CBT as compared to the existing algorithms[12]. In their work have presented an approach in which they recommend items in one domain based on the user ratings in other domain, where the domains are completely disjoint and auxiliary domains. The main purpose of authors in this paper is to deal with the cold start issue by exploiting cross domain recommendation[13], in their paper have presented a method for increasing the effectiveness of recommendations by assimilating information from social media into collaborative filtering. The results of the paper indicate that more accuracy in making predictions can be achieved by engulfing information from social networking websites into collaborative filtering.

2. Traditional Recommender Systems

2.1 Content-Based Filtering

Content-based recommender frameworks work with profiles of clients that are made toward the beginning. A profile has information around a client and his taste. Taste relies on upon how the client evaluated things. All things considered, while making a profile, recommender frameworks make a Recall, to get beginning information around a client in order to keep up a key separation from the new-client issue. 1, 2 in the recommendation method, the engine considers the things that were by then determinedly assessed by the client with the things he didn't rate and hunt down resemblances. Those things that are generally similar to the decidedly assessed ones, will be endorsed to the client. Figure 1 exhibits an instance of a client profile with the films he/she has seen and the evaluations the client made. Figure 2 exhibits the summary of films and their quality qualities. A substance based recommender framework would find movies from the once-over (Figure 2) that the client has starting now saw and insistently evaluated. By then, it would differentiate those movies and the straggling leftovers of the films from the once-over (Figure 2) and quest for comparable qualities. Similar movies would be

Movies	Green Lantern	Source Code	American Pie	Hangover 2
Ratings	8	7	9	10

Figure 1. The films the client has saw

Movies	Comedy	Violence	Horror	Exploit Content
American Pie	10	3	1	9
Scary Movie	8	8	4	9
Saw	2	10	10	7

Figure 2. The films list

recommended the client. In the present representation we can see that there is a film "Terrifying Movie" like the film "American Pie" that the client decidedly assessed. The client hasn't evaluated "frightening Movie" so it will be recommended him/her.

There are particular computations of measuring similarities among things in data base and those in client's profile[3] one of such procedures is cosine likeness. Addressing things as vectors on a cordinate space it gages edges amongst vectors and gives out their cosine regard. Vectors ~wc and ~ws of two things with properties are considered in cosine closeness limit as takes after[4]:

$$u(c,s) = cos(\vec{w_c}, \vec{w_s}) = \frac{\vec{w_c} \cdot \vec{w_s}}{||\vec{w_c}|| \times ||\vec{w_s}||} =$$

$$= \frac{\sum_{i=1}^{K} \vec{w_{ic}} \vec{w_{is}}}{\sqrt{\sum_{i=1}^{K} \vec{w_{ic}^2}} \sqrt{\sum_{i=1}^{K} \vec{w_{is}^2}}}$$

The more comparable two things are, littler the point between their vectors[5].

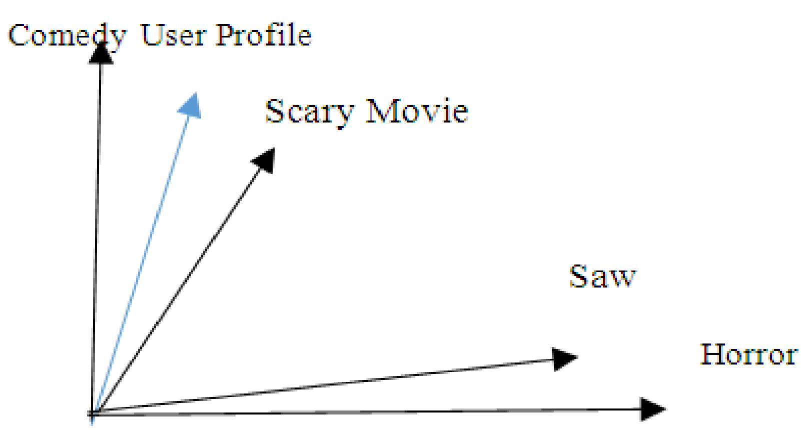

Figure 3. Cosine likeness on a coordinate plane Similarity look requires point by point information about the things. Better delineated things lead to more exact proposition.

2.2 Collaborative Filtering

Collaborative filtering got the chance to be a champion amongst the most asked about procedures of recommender frameworks since this philosophy was said and portrayed by Paul Resnick and Hal Varian in 1997. 1The plausibility of helpful separating is in finding clients in a gathering that offer thanks6. In case two clients have same or skirting on same evaluated things in like way, then they have equivalent tastes. Such clients fabricate a social occasion or an assumed neighborhood. A client gets proposition to those things that he/she hasn't assessed some time as of late, yet that were by then positively assessed by clients in his/her neighborhood. Figure 4 exhibits that each one of the three clients rate the movies earnestly and with tantamount engravings. That suggests that they have relative taste and create a region. The client A hasn't evaluated the film "TRON: Legacy", which doubtlessly infer that he hasn't watched it yet. As the film was quite assessed by interchange clients, he will get this thing recommended. Instead of less troublesome recommender frameworks where proposition base on the most assessed thing and the most well-known thing procedures, aggregate recommender frameworks consider the pith of client. The taste is thought to be enduring or potentially change bit by bit.

Collaborative filtering is for the most part used as a piece of e-exchange. Clients can rate books, songs, movies and after that get proposals concerning those issues in future. Other than Collaborative Filtering separating is utilized as a part of seeking of particular records (e.g. reports among investigative works, articles, and magazines).

3. Proposed System

In the literature of recommender systems only 25% of the work has been done on cross domain recommendations. Out of which maximum of the proposed methods make use of ratings of one domain to recommend items in another domain.

In this paper ,we propose a novel approach for cross domain recommender systems where genre of items of data set of one domain will be used to recommend items based on the most liked genre for other domain(i.e. most liked genre in terms of movies will be basis for making recommendations about the books).

To collect information for the first data set i.e. data about movies we will use social networking site *Facebook.com* from where we will fetch user profile which includes username, profile picture, Facebook ID and list of movies liked by the user using Facebook APIs and use this information further to fetch data about the movies for IMDB using IMDB APIs and will then store this into a database from where we will find out the most liked genre by the user. Along with this on other side we will evaluate four different algorithms namely Log Likelihood, Tanimoto, Eucledian Distance and Pearson correlation by calculating *root mean square error* and using a sample dataset of MovieLens.com.

The algorithm with lowest *root mean square error* will be used to calculate similarity between the most liked

Movies Users	Titanic	Gladiator	Black swan	The Fighter	Legacy
A	8	7	9	10	-
A	9	7	9	9	10
A	9	8	9	8	9

Figure 4. Collaborative recommender framework case

movie genre and the genre of the books. The top 9 books with highest similarity results will be recommended to the user.

Precision, recall and F1 score are the parameters that we will use to evaluate our recommender system[12].

3.1 Root Mean Square Error (RMSE)

A commonly used plumb to find out the difference between the predicted values and the actual observed values is called as *root mean square error or root mean square deviation.*

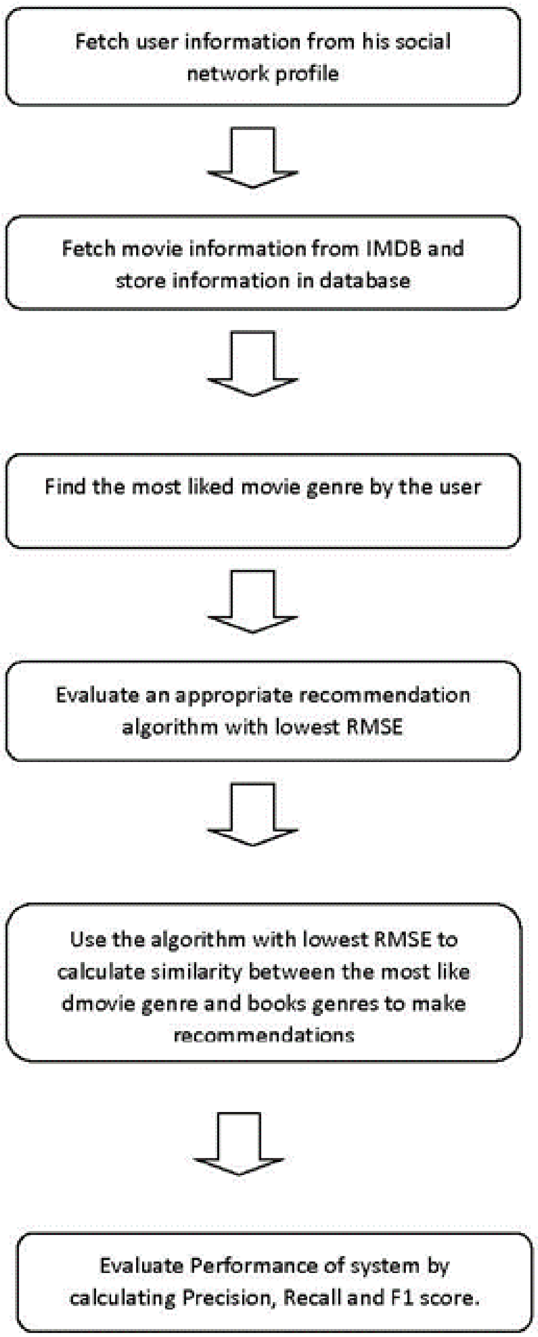

Figure 5. Flow Chart of Proposed Methodology.

The calculated differences are known as residuals, which are further aggregated into a single unit of predictive power[13].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_{obs,i} - X_{model,i})^2}{n}}$$

where X_{obs} is observed values and X_{model} is modelled values at time/place i

The calculated values are used to differentiate between the performances of model in a calibration period with respect to that of a validation periods. These values can also be used to compare performance of an individual model with other predictive models.

4. Proposed Algorithm

4.1 Client Based Collaborative Filtering Algorithm

Client based CF Algorithm produces recommendation list for the client according to the viewpoint of various clients. The suspicions are if the examinations of a couple of things assessed by a couple of clients are practically identical, the rating of various things assessed by these clients will moreover be comparative[3]. CF recommendation structure uses genuine frameworks to look the nearest neighbors of the article client and subsequently basing on the thing rating assessed by the nearest neighbors to predict the thing rating assessed by the thing client, and a short time later convey contrasting proposition list. Shared Filtering part that uses a territory based figuring is according to the accompanying. In neighborhood based figuring's, a subset of clients are picked in perspective of their closeness to the dynamic client, and a weighted mix of their evaluations is used to convey conjectures for the dynamic client.

The Algorithm can be condensed in the accompanying strides:

Step: 1. all clients are weighted regarding closeness with the dynamic client. Correspondence between clients is measured as the Pearson relationship between's their evaluations vectors.

Step: 2. Select n dynamic clients that have the most astounding closeness.

Step: 3. Process a forecast, Pa,u from a weighted mix. Comparability between two clients is processed utilizing the Pearson connection coefficient

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_{obs,i} - X_{model,i})^2}{n}}$$

Where ra,i is the rating given to thing i by client a;

What's more, ra is the mean rating given by client a.

In step 3, expectations are processed as the weighted normal of deviations from the neighbor's mean:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_{obs,i} - X_{model,i})^2}{n}}$$

Where Pa,i is the expectation for the dynamic client a for thing i. Pa,u is the closeness between clients an and u. n is the quantity of clients in the area.

5. Results and Discussion

5.1 Precision, Recall and F1 Score

Precision, recall and f1 score are three commonly used parameters for evaluating the performance of recommender systems.

Precision in simple term can be explained as the percentage of retrieved items that are relevant. Precision is also called as positive predictive value.

Recall unlike precision is the measure that a randomly selected item out of the items retrieved in search is relevant.

F1 score is calculated as the harmonic mean of both precision and recall.

Recommender frameworks research has used a couple sorts of measures for surveying the way of a recommender framework. We have gotten estimations consistently used as a part of the information recuperation bunch specifically audit, exactness and F1 (van Rijsbergen, C. A. 1979). These estimations are moreover customary for the appraisal of recommender frameworks (Breese, J. S. et al. 1998), (Sarwar, B. et al. 2000), (Jorge, A. et al. 2002).

Recall is overall measure for the whole game plan of wicker container in the test set. It thinks about to the degree of right answers and is an evaluation of the probability of having no short of what one germane proposition. It tends to increase with N.

$$Recall = \frac{|Hidden \cap Rec|}{|Hidden|}$$

Precision is moreover a typical for all the test wicker canister. Give us the way of individual proposal. As N grows, the nature of each proposal lessens.

$$Precision = \frac{|Hidden \cap Rec|}{|Rec|}$$

F1 has been proposed as a measure that joins Recall and precision with a proportionate weight. It ranges from 0 to 1 and higher qualities exhibit better recommendations. It is useful as an outline of the other two measures.

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

The data used for these tests insinuates the period between September 2001 and November 2002. For this period we have 290 resources and 26234 boxes. The typical number of advantages per carton is 2,68. With the train and test split we got 20987 wicker holders for train set and 5247 carton for test set.

To manufacture the course of action of connection standards we endeavored unmistakable mixes of slightest sponsorship and minimum assurance. Table 1 has the results for Recall, precision and F1, for different N values. The best results for Recall were refined with slightest support = 0,003 and with minimum sureness = 0,1. For these parameters, the amount of benchmarks in the model was 8957.

Review is around 15% when one and just recommendation is made (N = 1) – this suggests we have a 15% chance that the proposition is pertinent. In case we differentiate this Recall regard and the evaluated result for an advantage sporadic supposition (Rnd portion), we see that we would get a Recall rate around 49 times as high. This shows the estimation of Collaborative Filtering recommendations when appeared differently in relation to discretionary hypothesis. These subjective qualities were procured by parceling N by the total number of advantages (290).

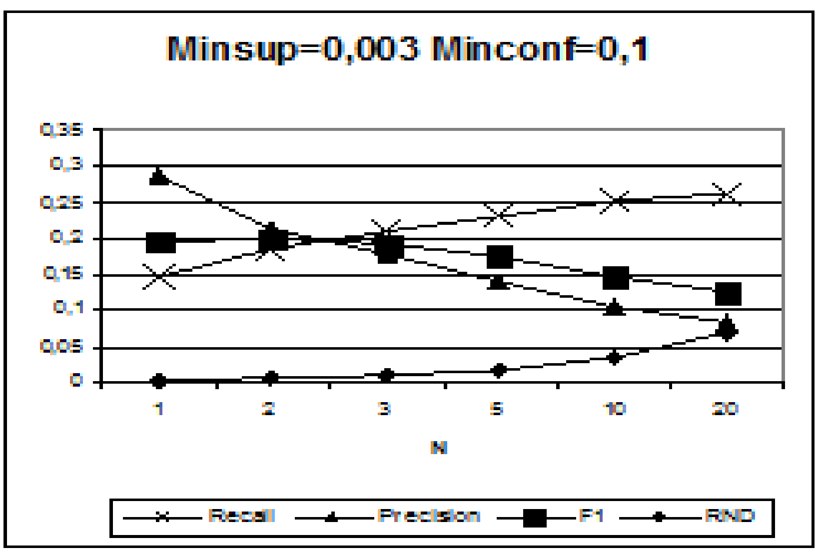

Figure 6. Results for recall, precision and F1, for different N values – minimum support=0.003.

Table 1. Results for recall, precision and F1, for various N values. Review values for arbitrary speculation and in addition review and exactness for default supposition are likewise appeared

	Minsup=0,003 Minconf=0,1			Minsup=0, 005 Minconf=0,1			Minsup=0,003			Minsup=0,005			(Recall) RND	Default	
	Recall	Prec	F1	Recall	Prec	F1	Recall	Prec	F1	Recall	Prec	F1		Recall	Prec
1	0.147	0.287	0.194	0.089	0.277	0.134	0.091	0.425	0.160	0.064	0.360	0.093	0.003	0.011	0.019
2	0.194	0.288	0.201	0.106	0.213	0.142	0.100	0.362	0.166	0.069	0.299	0.098	0.007	0.017	0.016
3	0.217	0.168	0.189	0.115	0.189	0.143	0.103	0.318	0.155	0.061	0.269	0.099	0.010	0.029	0.013
5	0.240	0.129	0.168	0.121	0.163	0.139	0.105	0.282	0.153	0.063	0.241	0.100	0.017	0.034	0.012
10	0.261	0.096	0.014	0.125	0.130	0.131	0.109	0.262	0.152	0.065	0.218	0.101	0.034	0.057	0.010
20	0.272	0.076	0.119	0.127	0.128	0.128	0.111	0.236	0.161	0.066	0.210	0.101	0.069	0.108	0.009

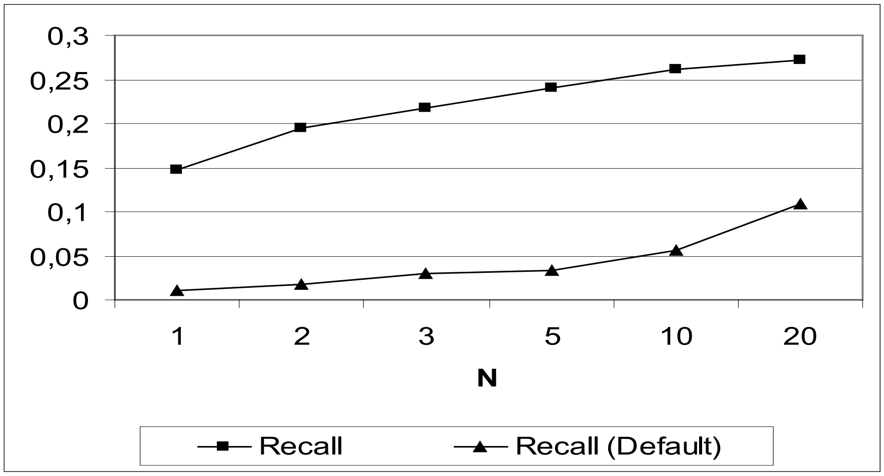

Figure 7. Comparison of collaborative oriented separating results with default suggestions (Recall).

We have furthermore taken a gander at the farsighted precision of our model with the default recommendations (the no doubt resources from the prior). Right when N = 1, the default recommendation for every wicker container in the detectable set is the advantage with the most significant sponsorship in the planning set; when N = 2, the default proposition for every bushel in the unmistakable set are the two resources with the most hoisted support in the train set, and so forth. In Figure 5 we can see the connection of Recall qualities between our model and default recommendations, for different N values.

6. Future Scope

In this paper we have proposed an innovative recommendation. In future we aim to implement this idea successfully and introduce the world with not so popular research issues in the field of recommender systems and also resolve problems that a recommender systems faces in making recommendations to a new user.

7. References

1. Berkovsky S, Kuflik T, Ricci F. Entertainment Personalization Mechanism through Cross-domain User Modeling. In: Proc. 1st Int'l Conf. Intelligent Technologies for Interactive Entertainment, LNAI 3814; 2005. p. 215-219.
2. Herlocker J, Konstan J, Borchers A, Riedl J, An Algorithmic Framework for Performing Collaborative Filtering. In: Proc. 22nd Annual Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR'99); 1999. p. 230-237.
3. Joachims T, Freitag D. Mitchell T. WebWatcher: A Tour Guide for the World Wide Web. In: Proc. 15th Int'l Joint Conf. on AI (IJCAI'97). 1997. p. 770-775.
4. Konstan JA, Miller BN, Maltz D, Herlocker JL, Gordon LR, Riedl J. GroupLens: Applying Collaborative Filtering to Usenet News, CACM. 1997; 40(3):77-87,
5. Li Y, Liu L, Li X. A Hybrid Collaborative Filtering Method for Multiple-interests and Multiple-content Recommendation in E-Commerce. Expert Sys. with Applications.2005; 28:67-77.
6. Tang TY, Winoto P, Chan KCC. Scaling Down Candidate

Sets Based on the Temporal Feature of Items for Improved Hybrid Recommendations. In: Intelligent Techniques in Web Personalization (B. Mobasher and S. S. Anand Eds.), LNAI 3169; 2004. p. 169-185.
7. Billsus D, Pazzani M. User Modeling for Adaptive News Access. User Modeling and User-Adaptive Interaction. 2000; 10(2-3):147-180.
8. Burke R. Hybrid Recommender Systems: Survey and Experiments. User Modeling and User Adaptive Interaction. 2002; 12(4):331-370.
9. Shardanand U, Maes P. Social Information Filtering: Algorithms for Automating 'Word of Mouth," In: Proc. ACM SIGCHI Conf. Human Factors in Computing Sys., (ACM CHI'1995); 1995. p. 210-217.
10. Winoto P, Tang T. If You Like the Devil Wears Prada the Book, Will You also Enjoy the Devil Wears Prada the Movie? A Study of Cross-Domain Recommendations. New Generation Computing. 2008; 26:209–225.
11. Li B, Yang Q, Xue X. Transfer learning for collaborative filtering via a rating-matrix generative model. In: Proc. of ICML 2009; 2009. p. 617-624.
12. Enrich Manuel, Matthias Braunhofer, Francesco Ricci. Cold-start management with cross-domain collaborative filtering and tags. In: E-Commerce and Web Technologies. Springer Berlin Heidelberg, 2013. p. 101-112.
13. Liu Fengkun, Hong Joo Lee. Use of social network information to enhance collaborative filtering performance. Expert Systems with Applications. 2010; 37(7):4772-4778.