www.ijecs.in
*International Journal Of Engineering And Computer Science ISSN: 2319-7242*
*Volume 4 Issue 10 Oct 2015, Page No. 14899-14904*

# Recommender System: Revolution in E-Commerce

***[1]Pratibha Yadav, [2] Kumari Seema Rani, [3]Sonia Kumari***

[1, 2, 3]Shyama Prasad Mukherji College (For Women), University Of Delhi, India

[1]pratibhamcadu2011@gmail.com, [2]raniseemabca@gmail.com, [3]Soniakumari.ducs@gmail.com

**Abstract**

With the growing expansion of information on World Wide Web, web sites are facing challenges to meet their customers' needs to present them with the information they are interested in. Recommender systems have emerged as a solution to this issue. Recommender system makes predictions for the users based on the analysis of their past behaviour. It is majorly classified in three categories which include: content based collaborative filtering and hybrid recommender system. Recommender systems have become an integral part of internet. They are becoming popular in the area of data mining, information filtering and e-commerce. In this paper, we have presented our study of various recommender techniques. We have also described the limitations of various recommendation techniques.

**Keywords:**

Recommender system, Content Based, Collaborative filtering, Hybrid Recommender System

## 1. Introduction

World Wide Web is an excellent medium of information exchange in this modern era exchange. Everyday billions of users interact with the internet. Now a day's internet is being flooded with abundant of information. It is becoming extremely difficult for the users to find the information they are interested in. To alleviate this problem, recommender systems (RS) have been incorporated by the web sites. RS is an intelligent tool which makes recommendations for the users [1][10][23]. RS analyses the users past behaviour and then present user with the personalised information. RS have become popular in the research area after the publication of first paper on collaborative filtering in mid-1990's[4][5] . RS are applied in various applications which include music, books[24] , movies, articles, news, web pages and many more. For example, Amazon.com [6] site uses RS to help the customers which product to purchase, CDNOW is another RS which help its users to buy CD's, MovieFinder.com recommends movies to its customers. RS have become an integral part of e-commerce [2][14] RSs have boosted the sales of e-commerce websites as they help customers to purchase the products they desire for. . RS is divided into various categories which includes content based, collaborative filtering, hybrid, demographic and knowledge based [22] approaches. These are described in the following section

## 2. Recommendation techniques

There are three main categories of RS which includes: Content based, Collaborative filtering and hybrid recommendation method. Content based recommender method makes predictions based on users past profile. Collaborative Filtering recommender method makes predictions based on the other users who share similarity with the user. Hybrid recommender method is the combination of content based and collaborative filtering techniques.

### 2.1 Content Based RS

Content based RS analyse users' profile and generates predictions based on the users' past preferences [3][15][17].

For example, in a movie RS if user has liked horror genre movies in past, Content based RS will recommends horror movies to the user which the user has not seen yet. Content based RS analyses the item features and user past behaviour for recommendations.

Let $U$ be the set of users and $I$ be the set of items. Let $ut(u,i)$ be the item $i \in I$ for user $u \in U$.e utility function which measures the utility of. In content based RS, $ut(u,i)$ is based on $ut(u,i_s)$ where $i_s$ i similar to item $i$.s the set of items which are similar to item $i$. In Content based recommendation technique, item features are presented in text format. To extract item features, various algorithms have been employed. The most popular algorithm is vector space model (VSM) [8][9]. It is also known as TF-IDF (term frequency – inverse document frequency). VSM is vector representation of terms. Terms are the keywords which are extracted from the set of documents. Each term is assigned a weight factor, which signifies the degree of importance of the terms. Each document is represented as a vector of n–dimensional space. Let $T = \{t_1, t_2, t_3, \ldots, t_n\}$ represents the set of terms and $D = \{d_1, d_2, d_3, \ldots, d_n\}$ represents a set of documents where each document $d_i$ is represented as an n – dimensional vector space, $d_i\ \{w_{1i} w_{2i}, w_{3i}, \ldots, w_{ni}\}$ where, $w_{1i}$ represents the weight of term $t_1$ in document $d_i$. For weighting computation, Term Frequency – Inverse Document Frequency (TF-IDF) [1][8] is used. TF represents the frequency of a keyword i.e. number of times a keyword appears in a document. Thus, for document $d_i$ and term $t_k$, the Term Frequency, denoted by $TF_{k,i}$ is computed as: $TF_{k,i} = \frac{freq_{k,i}}{max_j freq_{j,i}}$

Where, $freq_{k,i}$ represents the frequency of term $t_k$ in document $d_i$ and $max_j freq_{j,i}$ is the maximum frequency of term $t_j$ in document $d_i$. By dividing $freq_{k,i}$ by $max_j freq_{j,i}$ ,we ensures that longer documents will not be given preference over short documents.

IDF measures the importance of a term in document**.** It acknowledges the fact that the terms that appear less frequently in a document should be given more importance. $IDF_k$ For term $t_k$is defined as:

$$IDF_k = \log \frac{N}{n_k}$$

Where, $N$ denotes the number of documents and $n_k$ denotes the number of document in which term $t_k$ appears.

TF-IDF weight for term $t_k$ in document $d_i$ is defined as:

$$TF - IDF_{k,i} = TF_{k,i} * IDF_k$$

For learning user profile, various learning methods are being adopted in literature. The most popular and common user learning methods are Bayesian classifier and Relevance feedback.

Contents based RS suffers from several limitations [1][3]. The limitations are described below:

- **Over specialization**: Since content based RS's analysis is based on user past behaviour, it restricts RS to recommend only those items that matches the users' past preferences.
- **New User Problem**: When a new user joins the system, RS does not have adequate information to analyze user's profile. Therefore, content based RS will the not be able to predict accurately items for the new user.

**Limited Content Analysis**: Content based RS extract item features using various extraction algorithms. But it is always not possible to extract item features using the feature extraction algorithms on certain item sets like images, video, audio etc.

### 2.2 Collaborative Filtering RS

Collaborative Filtering technique is one of the most widely used techniques of the RS [7][10][13][25]. Collaborating Filtering (CF) predicts items for a user based on the analysis of choices of other users who have similar profile as the user has. In CF RS, $ut(u,i)$ is based on $ut(u_s,i)$ where, $i \in I$ and $u_s \in U$ is the set of user which are similar to user $u \in U$. CF is classified into three categories which are described below:

**Model based CF**: Model based CF analyses the dataset and extract useful information to generate a model [19]. This model is then used for prediction by RS. It uses various methods for model learning which includes Bayesian classifiers, clustering techniques, CF using dimensionality reduction techniques and many more.

**Memory based CF**: Memory based CF uses entire dataset for analysis of user profiles and prediction of items for users [23]. It uses various measures like Cosine correlation, Pearson correlation, k-Nearest neighbours and many more.

**Hybrid CF**: Hybrid CF is the combination of model and memory based CF techniques [20].

### 2.2.1 Similarity Evaluation

For computation of similarity between users or items, several measures have been proposed. The most commonly used and popular measures are: Pearson correlation coefficient and cosine-based.

**(i) Correlation-based Approach**

In correlation-based similarity computation method, Pearson correlation coefficient is used to compute similarity [1][13]. Similarity computation can be item based or user based. In item based approach [12], similarity between items is evaluated. In user based approach, similarity between users is computed. Let $S_{u_1u_2}$ be the set of items co-rated by user $u_1$ and $u_2$. $r_{u,i}$ is the rating user $u$ has given to item $i$ and $\bar{r}_u$ is the average rating of user $u$. The similarity between two users $u_1$ and $u_2$ is computed as:

$$Sim(u_1, u_2) = \frac{\sum_{s\in S_{u_1u_2}}(r_{u_1,s} - \bar{r}_{u_1})(r_{u_2,s} - \bar{r}_{u_2})}{\sqrt{\sum_{s\in S_{u_1u_2}}(r_{u_1,s} - \bar{r}_{u_1})^2}\sqrt{\sum_{s\in S_{uv}}(r_{u_2,s} - \bar{r}_{u_2})^2}}$$

Let $S_{ij}$ be the set of users who rated both the items $i$ and $j$. $r_{u,i}$ Denotes the rating given by user $u$ to item $i$ and $\bar{r}_i$ is the average rating of item $i$.The similarity between two items $i$ and $j$ is computed as:

$$Sim(i, j) = \frac{\sum_{u\in S_{ij}}(r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u\in S_{ij}}(r_{u,i} - \bar{r}_i)^2}\sqrt{\sum_{u\in S_{ij}}(r_{u,j} - \bar{r}_j)^2}}$$

Similarity ranges from [0,1].

**(ii) Cosine-based Approach**

In cosine-based approach [1][13], similarity computation is based on the cosine angle between the two items or users. Let $S_{uv}$ be the set of items co-rated by user $u$ and$v$. $r_{u,s}$ Denotes the rating user $u$ has given to item $s$ and $\bar{r}_u$ average rating of user $u$. The similarity between two users $u$ and $v$ is computed as

$$Sim(u, v) = \cos(\vec{u}, \vec{v})$$

Where, $\cos(\vec{u}, \vec{v}) = \frac{\vec{u}\,.\vec{v}}{\|\vec{u}\|_2 \, X \, \|\vec{v}\|_2}$

$$= \frac{\sum_{s\in S_{uv}} r_{u,s}\, r_{v,s}}{\sqrt{\sum_{s\in S_{uv}}(r_{u,s}^2)}\sqrt{\sum_{s\in S_{uv}}(r_{v,s}^2)}}$$

Let $S_{ij}$ be the set of users who have co-rated items $i$ and $j$. $r_{u,i}$ Denotes the rating user $u$ has given to item $i$ and $\bar{r}_i$ average rating of item $i$. The similarity between two items $i$ and $j$ is computed as follows:

$$Sim(i, j) = \cos(\vec{i}, \vec{j})$$

Where, $\cos(\vec{i}, \vec{j}) = \frac{\vec{i}\,.\vec{j}}{\|\vec{i}\|_2 \, X \, \|\vec{j}\|_2}$

$$= \frac{\sum_{s\in S_{ij}} r_{i,s}\, r_{j,s}}{\sqrt{\sum_{s\in S_{ij}}(r_{i,s}^2)}\sqrt{\sum_{s\in S_{ij}}(r_{j,s}^2)}}$$

### 2.2.2 Prediction Computation

After evaluation of similarity between users or items, CF RS makes recommendations based on the similarity computed [7][10]. Let $P(u, i)$ denotes the rating predicted of item $i$ by RS for user$u$. Let $S_u$denotes the set of users which shares similarity with user $u$. In user based similarity approach [26], $P(u, i)$ is evaluated as:

$P(u, i)$ =

$$\bar{r}_u + \frac{\sum_{v\in S_u} Sim(u, v)\,.\,(r_{v,i} - \bar{r}_v)}{\sum_{v\in S_u} Sim(u, v)}$$

Let $S_i$ denotes the set of items which are similar to item$i$. In item based similarity approach [12], $P(u, i)$ is evaluated as:

$P(u, i)$=

$$\bar{r}_i + \frac{\sum_{j\in S_i} Sim(i, j)\,.\,(r_{u,j} - \bar{r}_j)}{\sum_{j\in S_i} Sim(i, j)}$$

There is another approach which computes prediction based on the linear combination of user based and item based similarity approaches [26]. $P(u, i)$ is computed as follows:

$$P(u, i) = \alpha * P(u, i)_{item} + (1 - \alpha) * P(u, i)_{user}$$

Where $\alpha$ denotes the weightage which determines to which amount prediction depends on item based similarity and user based similarity. $P(u, i)_{item}$ represents the prediction based on item based similarity approach and $P(u, i)_{user}$ denotes the prediction based on user based similarity approach.

If $\alpha = 1$ then $P(u, i) = P(u, i)_{item}$ and if $\alpha = 0$ then $P(u, i) = P(u, i)_{user}$.

Although CF is one of the most successful approaches of RS, it suffers from several limitations. The limitations are described as follows:

- **New Item Problem**: For similarity computation between items, CF RS requires adequate information. CF RS suggests an item to a user only when it has been liked by sufficient users. Since new item does not have enough likings, it will not be recommended to users.
- **New User Problem**: Like new item problem, for similarity computation of users also CF RS requires

sufficient information. For new user, it is difficult for RS to find users with similar preferences as the new user has.

- **Sparsity**: Datasets in CF RS are represented in form of user and item matrix. Sparsity refers to the problem where items are large but ratings for the item are few. In this situation, CF RS may not adequately makes predictions.

### 2.3 Hybrid RS

It combines Content based and CF approaches. For instance, Fab RS maintains user profile using content based RS and find other users with similar profile using CF RS [16][18][21]. Hybrid RS alleviates certain limitations of the two above mentioned approaches. There are various hybridisation techniques that have been proposed.

These are described as follows:

- The content based and CF based approaches are implemented independently and then the output of content based and CF based approaches are combined linearly to generate the final output.
- The Content based features are blended into CF approach. Fab RS uses this technique.
- The features of CF are blended into content based approach. This technique is used for dimensionality reduction.
- Construction of a general model which includes features of both content based and CF based approaches.

## 3. Evaluation Metrics

Evaluation metrics are used to evaluate the performance of RS [11][23]. There are various evaluation measures that have been proposed to measure the performance of carious approaches of RS. The most common and widely used evaluation metrics are: Mean Absolute Error, Root Mean Square Error, Normalised Mean Absolute Error, Precision, Recall and F-measure. Precision, Recall and F-measure evaluates RS based on the classification matrix which is described in figure 1

**Classification Matrix**

| | Recommended | Not Recommended |
|---|---|---|
| Liked | True Positive(TP) | True Negative(TN) |
| Not Liked | False Positive(FP) | False Negative(FN) |

Figure 1

The evaluation metrics are described as follows:

- **Mean Absolute Error**: Mean Absolute Error (MAE) calculates the difference between the actual rating and predicted rating of an item.

  Let $r(u,i)$ denotes the actual rating of an item $i$ given by user $u$ and $p(u,i)$ is the rating of item $i$ predicted by RS for user $u$. MAE is defined as follows:

$$MAE = \frac{\sum_{k=0}^{n} |r(u,i) - p(u,i)|}{n}$$

Where $n$ is the total predictions.

- **Normalised Mean Absolute Error**: Normalised MAE (NMAE) normalizes the deviation between the actual and predicted ratings of an item. It is defined as:

$$NMAE = \frac{MAE}{(r_1 - r_2)}$$

  Where $r_1$ and $r_2$ denotes the maximum and minimum possible value of rating.

- **Root Mean Square Error**: Root mean square error (RMSE) is defined as the square root of the average of the variation between the actual and predicted rating. It is defined as:

$$RMSE = \sqrt{\frac{\sum_{k=0}^{n} \left(r(u,i) - p(u,i)\right)^2}{n}}$$

- **Precision**: Precision refers to the fraction of recommended items liked by the user out of the total suggestions made by RS to the user.

$$Precision = \frac{TP}{TP + FP}$$

- **Recall**: Recall refers to the fraction of recommended items liked by the user out of the total relevant items.

$$Recall = \frac{TP}{TP + FN}$$

- **F-measure:** F-measure is a harmonic mean of precision and recall.

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall}$$

## 4. Conclusion

Recommender Systems are emerging as efficient tool in e-commerce. Broadly they are classified in three categories namely content based, collaborative filtering and hybrid approaches. Collaborative filtering is considered as one of the most widely used approach of recommender system. It is further classified in three categories which include model based, memory based and hybrid collaborative filtering approach. Recommendation techniques also suffer from various limitations. To assess the performance of the approaches, evaluative measures are used. These mainly include mean absolute error, root mean square, precision, recall and f-measure.

## References


1. G. Adomavicius and A. 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 17 (6) 734–749, 2005
2. J. Ben Schafer, Joseph Konstan and John Riedl, “Recommender systems in e-commerce”, EC '99 Proceedings of the 1st ACM conference on Electronic commerce, 1999
3. G. Adomavicius, A. Tuzhilin, “Context-Aware recommender Systems”, in: F.Ricci, et al. (Ed.), Recommender Systems Handbook, 2011, pp. 217–253
4. U. Shardanand and P. Maes, “Social Information Filtering: Algorithms for Automating ‘Word of Mouth’,”Proc. Conf. Human Factors in Computing Systems, 1995
5. P. Resnick, N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl, “GroupLens: An Open Architecture for Collaborative Filtering of Netnews,”Proc. 1994 Computer Supported Cooperative Work Conf., 1994
6. Linden, G. , Smith, B. and York, J. , ”Amazon.com recommendations: item-to-item collaborative filtering” Internet Computing, IEEE Volume 7 Issue 1, pp. 76 – 80, 2003
7. Su, Xiaoyuan, and Taghi M. Khoshgoftaar, "A survey of collaborative filtering techniques", Advances in Artificial Intelligence Vol 3, pp 1-20, 2009
8. Gerard Salton and Christopher Buckley,” Term-weighting approaches in automatic text retrieval”,Information Processing & Management, Vol. 24,Issue 5, pp. 513-523, 1988
9. G. Salton , A. Wong and C. S. Yang,” A vector space model for automatic indexing”, Communications of the ACM, Vol.18, Issue11, pp.613-620,1975
10. L Lü, M Medo, CH Yeung, YC Zhang, ZK Zhang and Tao Zhou,” Recommender systems “, Physics Reports, Vol. 519, Issue 1, pp. 1–49, October 2012
11. Asela Gunawardana and Guy Shani,” A Survey of Accuracy Evaluation Metrics of Recommendation Tasks”, The Journal of Machine Learning Research, Vol. 10, pp. 2935-2962, 2009
12. Badrul Sarwar, George Karypis, Joseph Konstan and John Riedl, ”Item-based collaborative filtering recommendation algorithms”, WWW '01 Proceedings of the 10th international conference on World Wide Web, pp. 285-295, 2001
13. Michael D. Ekstrand, John T. Riedl and Joseph A. Konstan, “Collaborative Filtering Recommender systems”, Foundations and Trends in Human-Computer Interaction Volume 4 Issue 2, pp. 81-173, 2011.
14. J.B. Schafer, J.A. Konstan, and J. Riedl, “E-Commerce Recommendation Applications ,”Data Mining and Knowledge Discovery,vol. 5,nos. 1/2, pp. 115-153, 2001.
15. Pasquale Lops, Marco de Gemmis and Giovanni Semeraro, “Content-based Recommender Systems: State of the Art and Trends” pp. 73-105, 2011
16. M. Balabanovic and Y. Shoham, "Fab: Content-Based, Collaborative Recommendation," Comm. ACM, vol. 40, no. 3, pp. 66-72, 1997
17. R.J. Mooney and L. Roy, "Content-Based Book Recommending Using Learning for Text

Categorization," Proc. Fifth ACM Conf. Digital Libaries, pp. 195-204, 2000

18. Ana Belén Barragáns-Martínez, Enrique Costa-Montenegro, Juan C. Burguillo, Marta Rey-López, Fernando A. Mikic-Fonte and Ana Peleteiro ”A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition”, Information Sciences: an International Journal , Vol 180 Issue 22, pp. 4290-4311 November, 2010
19. K. Miyahara and M.J. Pazzani, "Collaborative Filtering with the Simple Bayesian Classifier," Proc. Sixth Pacific Rim Int',l Conf. Artificial Intelligence PRICAI 2000, pp. 679-689, 2000.
20. K. Yu, A. Schwaighofer, V. Tresp, X. Xu, and H.-P. Kriegel,” Probabilistic memory-based collaborative filtering” IEEE Transactions on Knowledge and Data Engineering, 16(l):56-69, 2004.
21. Robin Burke,” Hybrid Recommender Systems: Survey and Experiments” User Modeling and User-Adapted Interaction, Vol 12, Issue 4, pp 331-370, November 2002
22. Burke, R., “Knowledge-based Recommender Systems”. In: A. Kent (ed.): Encyclopedia of Library and Information Systems. Vol. 69, Supplement 32, 2000.
23. J. Bobadilla , F. Ortega , A. Hernando and A. GutiéRrez,” Recommender systems survey”, Knowledge-Based Systems, Vol. 46, pp.109-132, July, 2013
24. E.R. Núñez-Valdéz, J.M. Cueva-Lovelle, O. Sanjuán-Martı´nez, V. Garcı´a-Dı´az, P.Ordoñez and C.E. Montenegro-Marı´n, “Implicit feedback techniques on recommender systems applied to electronic books”, Computers in Human Behavior 28 (4) 1186–1193, 2012
25. Breese, J.S., Heckerman, D., Kadie and C.M.,” Empirical analysis of predictive algorithms for collaborative filtering”, in: 14th Conference on Uncertainty in Artificial Intelligence,, pp. 43–52, 1998
26. H. Ma, I. King, and M. R. Lyu,” Effective missing data prediction for collaborative filtering”, In SIGIR ’07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 39–46, New York, NY, USA, 200

DOI: 10.18535/ijecs/v4i10.48