2018 4th International Conference on Computing Communication and Automation (ICCCA)

Principles and Methods For Recommendation Framework

Kavitha S,Ranjana Rajesh Badre
School of Computer Engineering and Technology
MIT Academy of Engineering
Pune, India
kavithas@comp.maepune.ac.in,rrbadre@comp.maepune.ac.in

***Abstract*—Structured, semi-structured and unstructured data is growing exponentially. The exponential increase in the data resulted in providing a plethora of alternatives for items, services and products. In e-commerce, recommender systems have become the most powerful tool to avoid choice overload. People who lack effective proficiency to evaluate the various alternatives can use Recommender systems. Poor decision making is due to the exponential growth of information and the convention of new e-business services like comparison of products, auction, purchase of products etc. This paper outline the key concepts used in recommendation systems and elucidate different recommendation approaches using content based, collaborative and hybrid methods. The paper reviews the drawbacks of various recommendation methods and the possible future scope to enhance the performance of recommendation systems. Several approaches are addressed in this paper to deal with the common cold start and data sparsity problems. The paper also proposes a hybrid recommendation system, which provides a recommendation based on the score for each scenario.**

***Index Terms*—Choice Overload, Decision Making, Hybrid methods,Item, Recommender systems**

I. INTRODUCTION

The growth of E-commerce technology is exponential with the increasing number of customers daily. Recommendation systems are extensively used in E-commerce for increasing the customer satisfaction as well as for increasing the profit of the company.

Following facts substantiate the fact that, nowadays the use of recommendation systems has tremendously increased. 1.One million dollar prize money for the competition to predict the user enjoyment of movies based only on their preferences for films [1]. 2.Dedicated symposiums, conferences, workshops organized by IEEE and ACM in the fields of/related to recommender systems [1].

Stages of Recommendation process are 1. Information Gathering Phase 2. Learning Phase 3. Recommendation or Suggestion Phase [2].

A. Information Gathering Phase

This phase is related to the capturing of related and relevant information of the user.This information is used for making a user profile, which is then used for the prediction task. The prediction task includes predicting the items the user accesses, predicting users characteristic etc. Recommendation systems depend on the various types of quality inputs such as distinct direct inputs and indirect inputs. The distinct direct inputs are referred to as Explicit Feedback and indirect inputs as Implicit Feedback. The system should be modeled in such a way that it should get quality information from the explicit and implicit feedback to provide an accurate recommendation. In certain situations, the combination of explicit and implicit feedback may be combined to get hybrid input or feedback.

The outcome of the Recommendation system depends on the ability of the system to point out and represent the current interest of the user. The accuracy of the recommendation depends on the how efficiently the information is captured and the models being built [2].

1) Explicit Feedback: For improving the model, the user can give ratings for the items, through the user interface.It needs some efforts from the user and the user may not be ready always to provide the ratings. Recommendation accuracy depends upon the ratings of the user. Explicit feedback provides definitive data. It does not require extracting information from the actions and provides quality recommendations. Example the movie p has y ratings ie. explicitly rated the movie with the value y [2].

2) Implicit Feedback: By observing and tracking the users various actions, the system infers the user preferences accordingly. For example by tracking the browsing history, links which the user tracks, number of times a particular web page is tracked etc. user preferences are captured. Data obtained from implicit feedback is less accurate [2].Implicit feedback data are typically noisy as they are tacit from the user actions. Implicit feedback helps in overcoming unstable preferences as the user preferences may change with time.

3) Hybrid Feedback: For the better performance of the systems, the strengths of Implicit and Explicit feedback are combined to get Hybrid feedback. By allowing the user to provide explicit feedback, whenever he wishes to do and by using the implicit data as a check on the explicit data, hybrid feedback may be obtained. By using different techniques in a combined manner, like Nave Bayes and content based filtering better application optimization is obtained.

B. LearningPhase

For extracting the users characteristics, learning algorithm is applied to refine the information gathered in the information phase [2]. Learning algorithms generate patterns that are most

978-1-5386-6947-1/18/$31.00 ©2018 IEEE

appropriate for the situation in the application. Learning algorithms broadly fall into three types Supervised, Unsupervised and Reinforcement Learning.

C. Recommendation or Suggestion Phase

The Recommendation algorithms are grouped into two types-memory-based and model-based algorithms. Depending upon the data collected in the information phase of the memory based or model based model, the application recommends or suggests what type of items the user will like. Different methods used for recommendation algorithms are: 1. Content Based Recommendation Systems 2. Collaborative Filtering 3. Hybrid Systems [2].

II. Types of Recommendation System

A. Content Based Recommendation System

Content based recommendation system is based on the item descriptions. It identifies items that are of particular interest to the user, based on users interest for the items in the past. The content of the item suggested is provided as features, and a model is built. The model recommends the likelihood that the user would prefer the item based on its content. When a movie recommendation system is considered, train a classifier to represent the user liking for the movie, by making use of an actor as the characteristic attribute. Based on the actor, the system will suggest the user would like the unfamiliar movie or not. Content based filtering focuses on the reasoning of attributes of the item to provide recommendations and is area/domain dependent. Recommendations for the news and publication etc. generally use content based filtering [2] [3].

B. Collaborative Recommendation System

This is the process of extracting information based on the similarity, in collaboration among data sources, items etc. Collaborative filtering makes use of user-item matrix preferences for predictions. Collaborative filtering make use of neighborhood paradigm and cannot be described by the metadata of music, movie etc. The recommendation is based on the positive rating given for an item in the neighborhood of user, even though the user has not rated the item [3]. Collaborative filtering is divided into two categories-Memory Based Techniques and Model Based Techniques.

1) Memory Based Techniques: Memory based collaborative filtering is accomplished by user-based and item-based techniques. These techniques use data ie. ratings, no of likes etc.to establish closeness or correlation or similarity, between users or items to predict the liking of a particular user to a particular item. Cosine similarity, Jaccard Coefficient, Pearson correlation are some of the commonly used similarity measures. The weighted average of the recommendation of several users are considered for recommendation [4].

Pearsons correlation coefficient can be used as a method for correlation. Considering, user1 and user2 are in the same neighborhood, the weight for user1s recommendation depends upon the correlation between user1 and user2 (to whom the recommendation is to be made) [5].Pearsons correlation coefficient measures the likelihood of users to rate items a and b similarly.

Cosine Similarity is based on linear algebra using the vector-space model. In cosine similarity, based on the angle between two m-dimensional vectors, the similarity is measured. Cosine similarity is widely used for comparing text documents in information retrieval [2].Overlying vectors (zero angle)with cosine value is equal to one indicates full similarity (ie a user gives an identical rating for all items). No similarity is indicated by Cosine of zero ie. an angle of ninety degrees.

2) Model Based Techniques: Machine Learning and Data mining Techniques are used for model building. Model based techniques use earlier ratings to learn the model for collaborative filtering. The machine learning and data mining techniques employed recommend items to the user based on the neighborhood similarity. Common learning algorithms used in Model Based Techniques are:

Regression Analysis estimates the interrelationship between two or more variables. It is commonly used for predicting the relationship between dependent variable and independent variable [2].

Bayesian Classifiers are probabilistic classifiers used for classification, based on Bayes theorem and conditional probability. Naive Bayes classifier is the most commonly used Bayesian Classifier [2].

Matrix Completion Techniques -Unknown values within the user-item matrices is predicted with Matrix Completion Techniques. Co-relation based K-nearest neighbor is one of the common techniques used [2].

Clustering segregates the given dataset into small clusters, each cluster having some common attribute. K-means and Self Organizing Map(SOM) are the common clustering Techniques used.

Association Rule defines rules that predict the occurrence of an item based on the presence of other items [2].

C. Hybrid Recommendation System

Various different recommendation techniques may be combined to overcome the disadvantage of recommendation systems and to get improved optimized results [6]. Weaknesses of the individual recommendation system may be suppressed by combining different recommendation techniques.

III. Performance Estimation Metrics

The excellence of the Recommendation Systems depends on the outcome of the estimation. As per the functionality of the applications metrics used may vary. Accuracy metrics determines to what extent the actual value/rank differs from the predicted value/rank.

Mean Absolute Error is the absolute difference between the actual rating and the predicted rating. The low mean absolute error indicates more accuracy [7].

Root Mean Square Error is the square root of the Mean square error. Low values of the root mean square error indicates more accuracy [1].

Precision and Recall are the metrics used for decision support. Precision is the ratio of rightly recommended items to the total recommended items. The recall is the ratio of rightly recommended items to the total appropriately recommended items [1].

A. Model Evaluation

Model evaluation is done by considering True Positive(TP), True Negative(TN), False Positive(FP) and False Negative(FN) [1]. Accuracy is given by the ratio of the sum of TP and TN to the sum of TP, TN, FP and FN.

TABLE I
CONFUSION MATRIX

		In Reality	
		Truly Good	Truly Bad
Prediction	Good Rating	TP	FP
Prediction	Bad Rating	FN	FP

IV. LITERATURE REVIEW

This section explores the survey of various recommendation approaches used for decision making.

Yang et al. [8], addressed cold start and data sparsity problems in collaborative filtering method by incorporating trust. Trust plays a role in generating user preferences. Trust among each other also plays a crucial role in generating the review. Authors proposed a Truster Model and Trustee model. These two models are joined to TrustMF model. Further, the authors proposed TrustPMF model. TrustPMF model contributed a likelihood of understanding the proposed models. Through a series of validations on Epinions and Douban data sets, authors could prove that TrustPMF model gave a better recommendation than the collaborative filtering approaches. Limitation of the approach is that the model works well only with trust information in binary form.

Faridani et al. [9], stated that reliable neighbors of the currently active user are combined to complement and denote the options for the active user. The importance of each user to provide recommendation is decided by differentiating different users. Similar users are aggregated based on their weighted importance. Predictions are calculated by finding the average of the ratings given by similar users. The disadvantage is that the model works only on explicit trust. Privacy constraints restrict the users to share the information. The model may be modified for implied trust.

Dou et al. [10], discussed user based collaborative filtering and item based collaborative filtering. Pearson based similarity, cosine based similarity and revised cosine based similarity were discussed in the paper. The challenges like new user or new item referred as cold start problem, sparsity problem, the weights of different items and the best neighbor finding cost are addressed. For the new user problem, the solution provided was getting information from the login profile of the user account. If the profile of the user is not available, user details may be retrieved indirectly through the histories of blog and browser. The drawback is that as a precondition, access privacy of user has to be ensured.

Yang et al. [11], proposed a model that combined the viewpoints(opinion) and preferences on various aspects for the general rating prediction. Data sparsity is removed by aspect weights, which are obtained by Tensor Factorization Method. Tensor is computed by the overall general ratings and the weighted aspect ratings.

In the survey paper by Yang et al. [11], two ways are provided to deal with sparsity-1)By making use of filling and dimension reduction approach. The filling method places a default value as the value of the unrated item. But this method has a low reliability. Another way of the filling method is to put the blanks with the similarity of users and items. Machine learning approaches like neural network [12], bayes method [13] are also used for filling the matrix.2)By increasing the accuracy and efficiency without altering the sparsity of the matrix by using singular value decomposition [14] and by using clustering algorithms.

In the review paper, Isinkaye et al. [2], discussed the three phases of recommendation process-Information gathering phase, Learning phase, and Recommendation phase. The paper reviewed various recommendation techniques and the metrics for evaluating the performance of recommendation algorithms. The paper also exposed various features and capabilities of forecasting techniques in recommendation systems.

Wang et al. [15], proposed that when the hierarchy is assimilated in the user preference as well as in items the capability of the recommender system is improved. But generally hierarchical structures are not specifically and clearly available. A framework is proposed which looks into the inferred hierarchies of items and user preferences. The proposed framework acquires implicit hierarchical structures when they are not available readily and combines with the available explicit frameworks for the recommendation. The framework used weighted non-negative matrix factorization to get the user preferences. Incorporating pervasive social media data is a challenge for this method.

For improving the performance of recommendation system, Cai et al. [16], proposed that customer/users neighbors are clustered.ie in user groups, instead of similarly rated things, typical users who rate the similar things are found by making use of the fuzzy method. Accuracy is determined by using Mean Absolute error.

Limitation of this approach is the additional prototype building using clustering, to group the customers depending on the typicality.

A hybrid recommendation system was proposed by Wang, et al. [17], by making use of model based collaborative filtering. The model takes care of sparsity problem and high dimensionality problem. To enhance the correctness and precision for the recommendation of movies, the authors developed an enhanced k-means clustering based hybrid model. Off and online modules are combined to get the right recommendation for the movie.

The offline part includes dimensionality reduction with principal component analysis and applying k-means combined with genetic algorithms. Users having similar preferences are grouped into the same cluster by the PCA-GAKM framework. Once the user set is identified, in the online phase, by making use of the TOP N recommendation list, movies of interest are extracted from cluster neighborhood rather than searching the complete userspace. K-means involves arbitrary selection of initial seeds and the final result may fall to local optima. To overcome this, the authors themselves combined genetic algorithm to k-means clustering.

V. Proposed System

Gray sheep, sparsity, shilling attacks, synonymy, protection of privacy etc. are concerns in the existing algorithms [7] [18].Machine learning accord a crucial role in recommender systems.

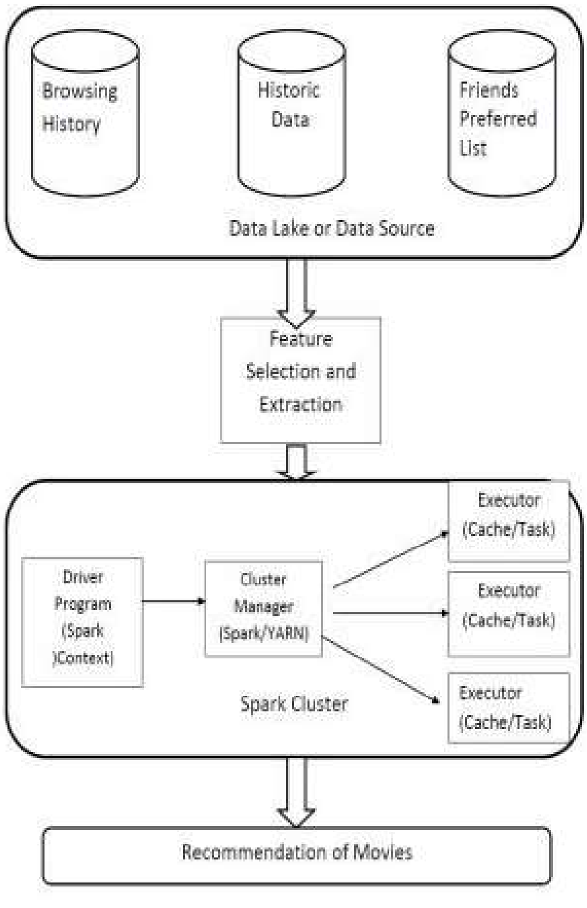

Fig. 1. Architecture of Proposed System.

Proposed recommendation system combines content and collaborative based techniques. The accuracy of the recommendation will be higher in hybrid systems. The integration of content based and collaborative filtering results in an increase in knowledge of the domain, which results in the improved recommendation.

The proposed method has two elements.one-predictions are made by recognizing similarities among rating profiles by making use of the collaborative technique. The second component utilizes semantic knowledge regarding the movie features, makes use of latent semantic analysis and recommends the movies that are semantically identical to user preferences. The paper proposes Wilcoxon signed rank test for classifier comparison.

The approach uses adaptive coding suggested by Pelckmann [14]. Movie recommendation is considered for examining the relevance of adaptive coding for the recommendation Step-1 For each user q in the system, p is the rating vector of the user, such that pq= rq, i , user q rated item i with r.

Step-2 Related to an active user, weight all users based on similarity. Among the rating vectors, the Pearson correlation is used to find similarity between users.

Step-3 N users with the highest resemblance with the active user are selected to form the neighborhood.

Step-4 From the selected neighbors calculate prediction based on weight.

Pearson correlation coefficient used to compute similarity, defined below:

$$P_x, u = \frac{\sum_{i=1}^{n}(r_x, i * -\overline{\mathrm{r_x}}) * (r_u, i * -\overline{\mathrm{r_u}})}{\sum_{i=1}^{n}(r_x, i - \overline{\mathrm{r_x}})^2 * \sum_{i=1}^{m}(r_u, i * -\overline{\mathrm{r_u}})^2} \tag{1}$$

where

$$r_x \tag{2}$$

is the item i rating given by user x ,

$$\overline{\mathrm{r_x}} \tag{3}$$

is the user xs mean rating, n is the total number of items.

Predictions are computed by

$$P_x, i = \frac{r_x + \sum_{u=1}^{m}(r_x, i - \overline{\mathrm{r_x}}) * P_x, u}{\sum_{u=1}^{m} P_x, u} \tag{4}$$

where m is the number of users in the neighborhood,

$$P_(x, i) \tag{5}$$

refers to the item is prediction of the active user and

$$P_(x, u) \tag{6}$$

is the similarity between user x and u.

VI. Open Issues

Recommendation Systems provides an adequate opening for researchers, where they can look over in various orientations depending upon the necessity.

Recommendation Systems should intelligently handle user data privacy-preserving mechanisms so that user data is not accessed maliciously [9].

Nowadays cloud-based, parallel systems are available for a variety of applications. Under this scenario recommendation applications on big data, using parallel environment is required [19].

Users require recommendations as they do not have sufficient acquaintance with a particular item to make the judgment. The system should ideally persuade users to purchase new items, to listen or watch new songs or movies etc. This depends only on how the specific intercommunication between human-computer is backed by the system when the items are correlated and made clear. As the user is directed towards new inexperienced items, the system must provide recommendations with pictures and there must be some means to improve recommendations. Support for user control through an interactive user interface and handles are expected [20].

VII. Discussion

TrustPMF framework proposed by Yang et al. [8], carried out a recommendation in a remarkably better manner as compared with the general other collaboration filtering techniques. The framework appreciably addressed the cold start problem with limited ratings. The framework incorporated Trust in giving Recommendation by understanding the fact that trust between users will affect each other in providing a recommendation. Experiments worked out on Epinion and Douban dataset are listed below.

TABLE II
Performance Evaluation Metrics on Epinions Dataset [8]

Metrics Considered	Social Recommendation Method [21]	RSTE Method [22]	TrustPMF	Metrics for
Recall	0.40	0.25	0.68	Users with cold start
Precision	0.44	0.28	0.71	Users with cold start
Recall	0.60	0.42	0.61	Every User(All)
Precision	0.67	0.51	0.78	Every User(All)

TABLE III
Performance Evaluation Metrics on Epinions Dataset [8]

Metrics Considered	Social Recommendation Method [21]	RSTE Method [22]	TrustPMF	Metrics for
Recall	0.77	0.54	0.80	Users with cold start
Precision	0.77	0.56	0.80	Users with cold start
Recall	0.85	0.68	0.86	Every User(All)
Precision	0.79	0.68	0.79	Every User(All)

Experiments conducted on the framework proposed by Faridani et al. [9], showed noteworthy improvements in the correctness of recommendations. The method combined trust, resemblance or similarity to achieve recommendation. The method relied on explicit trust. Yang et al. [11], experimented with their framework on two data sets-hotel and movie. The experiments revealed that the Tensor factorization outshined the matrix factorization and the correctness in prediction is improved by weights.

Experiments performed on the openly available Douban, MovieLens100K, MovieLens1M and Ciao datasets through the model proposed by Wang et al. [15], showed the significance of explicit and implicit data structures in boosting the performance of recommendation. Root mean square error is used to assess the performance for predicting the ratings.

In the TyCo framework proposed by Cai et al. [16] , the authors have not mentioned the use of parallel computing to speed up the recommendation process while dealing with large datasets.

VIII. Conclusion

Much research has been done on recommendation systems in the recent past. A wide range of techniques including machine learning, information extraction etc. have been used for comparison in content based and collaborative filtering. Many real time business applications demand more features or characteristics for considering the recommendation. Recommendation Systems for variegated applications requires that various techniques mentioned needs to be combined.

References

[1] F. Ricci, L. Rokach, and B. Shapira, "Recommender systems: introduction and challenges," in *Recommender systems handbook*. Springer, 2015, pp. 1–34.

[2] F. Isinkaye, Y. Folajimi, and B. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egyptian Informatics Journal*, vol. 16, no. 3, pp. 261–273, nov 2015.

[3] 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 & Data Engineering*, no. 6, pp. 734–749, 2005.

[4] S. Badsha, X. Yi, and I. Khalil, "A practical privacy-preserving recommender system," *Data Science and Engineering*, vol. 1, no. 3, pp. 161–177, sep 2016.

[5] M. Chen, "Performance evaluation of recommender systems," *International Journal of Performability Engineering*, 2017.

[6] R. Burke, "Hybrid web recommender systems," in *The Adaptive Web*. Springer Berlin Heidelberg, pp. 377–408.

[7] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in Artificial Intelligence*, vol. 2009, pp. 1–19, 2009.

[8] B. Yang, Y. Lei, J. Liu, and W. Li, "Social collaborative filtering by trust," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 8, pp. 1633–1647, aug 2017.

[9] V. Faridani, M. Jalali, and M. V. Jahan, "Collaborative filtering-based recommender systems by effective trust," *International Journal of Data Science and Analytics*, vol. 3, no. 4, pp. 297–307, mar 2017.

[10] Y. Dou, H. Yang, and X. Deng, "A survey of collaborative filtering algorithms for social recommender systems," in *2016 12th International Conference on Semantics, Knowledge and Grids (SKG)*. IEEE, aug 2016.

[11] C. Yang, X. Yu, Y. Liu, Y. Nie, and Y. Wang, "Collaborative filtering with weighted opinion aspects," *Neurocomputing*, vol. 210, pp. 185–196, oct 2016.

[12] F.-Z. Zhang, J.-F. Chang, and D. Wang, "Multi-criteria recommendation algorithm based on widrow-hoff neural network," *Pattern Recognition and Artificial Intelligence*, vol. 2, 2011.

[13] K.-Y. Jung, H.-J. Hwang, and U.-G. Kang, "Constructing full matrix through nave bayesian for collaborative filtering," in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2006, pp. 1210–1215.

[14] M. Vozalis and K. Margaritis, "Applying SVD on item-based filtering," in *5th International Conference on Intelligent Systems Design and Applications (ISDA'05)*. IEEE, 2005.

[15] S. Wang, J. Tang, Y. Wang, and H. Liu, "Exploring hierarchical structures for recommender systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 6, pp. 1022–1035, jun 2018.

[16] Y. Cai, H. fung Leung, Q. Li, H. Min, J. Tang, and J. Li, "Typicality-based collaborative filtering recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 3, pp. 766–779, mar 2014.

[17] Z. Wang, X. Yu, N. Feng, and Z. Wang, "An improved collaborative movie recommendation system using computational intelligence," *Journal of Visual Languages & Computing*, vol. 25, no. 6, pp. 667–675, dec 2014.

[18] B. M. Sarwar, J. A. Konstan, A. Borchers, J. Herlocker, B. Miller, and J. Riedl, "Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system," in *Proceedings of the 1998 ACM conference on Computer supported cooperative work - CSCW '98*. ACM Press, 1998.

[19] H. Su, C. Wang, Y. Zhu, B. Yan, and H. Zheng, "Parallel collaborative filtering recommendation model based on expand-vector," in *2014 International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI)*. IEEE, sep 2014.

[20] C. He, D. Parra, and K. Verbert, "Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities," *Expert Systems with Applications*, vol. 56, pp. 9–27, 2016.

[21] H. Ma, T. C. Zhou, M. R. Lyu, and I. King, "Improving recommender systems by incorporating social contextual information," *ACM Transactions on Information Systems*, vol. 29, no. 2, pp. 1–23, apr 2011.

[22] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with explicit and implicit social relations," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–19, apr 2011.