Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019)
IEEE Xplore Part Number: CFP19K34-ART; ISBN: 978-1-5386-8113-8

Machine Learning Algorithms for building Recommender Systems

[1]Richa Sharma
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
richa.sharma@chitkara.edu.in

[2]*Shalli Rani
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
Shalli.rani@chitkara.edu.in

[3]Sarvesh Tanwar
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
Sarvesh.tanwar@chitkara.edu.in

***Abstract*— Over the years, Recommender systems have emerged as a means to provide relevant content to the users, be it in the field of entertainment, social- network, health, education, travel, food or tourism. Till date several recommendation approaches have been introduced, the most popular being Content-based filtering, Collaborative filtering, Hybrid and Knowledge based systems. Hybrid systems combine multiple recommendation techniques to enhance the performance of a single recommendation approach and to do so, they follow several hybrid models. This article presents an overview of the state-of-the-art Recommender systems with the prime focus on hybrid recommender systems. Further, different categories of hybridization models are studied, and the existing work is classified categorically based on the hybrid model they follow, and the Machine learning algorithm used.**

Keywords*— *Recommender Systems, Collaborative Filtering, Hybrid Recommender Systems

I. INTRODUCTION

The problem of information overload across the world wide web is growing extensively with the exponential rise in the volume of digital data. This problem further challenges the decision-making process on the user-end [1]. Moreover, it is a challenge for service providers also, to deliver only relevant and suitable content to a user, keeping in mind his personal choices and preferences [2]. Therefore, to tackle this problem and meet user requirements in the most convenient yet effective way, most of the e-commerce sites or service providers use the concept of Recommender systems. The term Recommender originated from the Latin word Recommendare, which means to approve, endorse, entrust or acclaim something or someone as being worthy or desirable. Such systems improve the decision-making process by presenting the users with the most suitable suggestions [3]. Therefore, Recommender systems are often referred to as software tools or techniques that narrow down user preferences according to their requirements or likings [4].

The basic concept that led to the birth of recommender systems is that people often seek recommendations from their peers, family members or experts for the purpose of decision-making. This can further be related with the fact that users now have a plethora of options available in front of them and they often struggle to decide which product or service will actually meet their requirements. Social networking sites, e-commerce sites, entertainment sites, digital libraries etc. are some of the major application areas of recommender systems. The number of likes or dislikes given to a product, the number of times a video is viewed or the frequency with which a website is visited, determines how popular a product or service is.

The very first recommender system called Tapestry, was developed in 1992, by Xerox Palo Alto Research Centre [6]. Later as Tapestry gained success, recommender systems for multiple applications were developed, including Amazon, CiteSeer, Docear, E-bay, Hulu, Jester, LinkedIn, Netflix, Pandora, Trip advisor, Yahoo, YouTube etc. Recommendations for these applications can be viewed in terms of “You may like”, “People you may know”, “Suggested videos”, “Trending now”, “Similar products”, “Suggestions for you”, “People who bought this also bought this” etc.

Till date numerous recommendation techniques have been developed, the most popular being Content-based filtering, Collaborative filtering, Knowledge based systems and Hybrid systems [7-8]. While the first two make recommendations based on how popular a product is amid a similar group of people and what type of products does a user prefer, knowledge-based systems make recommendations as per domain knowledge or system expertise. Hybrid systems as the name suggests combine multiple recommendation approaches such that the limitations of one system are removed by the other [3-4]. Further categorization of recommender systems is based on user Demography, Community and Context. This work studies the various recommendation techniques introduced till date. Recommender systems are based on Machine learning algorithms, to learn about user interests, find similar users, find hidden patterns between users or items etc. Further, the article moves towards a detailed study of Hybrid recommender systems, various hybridization techniques and the existing systems using hybridization.

The major contributions of this study are:

1. We provide an overview of the various categories of recommendation approaches.
2. We provide an insight into the state-of-the art hybrid recommendation models.
3. Based on the study conducted, we categorized the existing hybrid systems on the basis of the hybridization model they follow, Machine learning algorithm used and the recommended services or products.

The rest of the article is ordered as follows: Section II provides an overview of existing recommendation approaches along with their pros and cons. Section III focusses on the detailed study of hybrid recommender systems. The related work is discussed in Section IV. Finally, the article is concluded in Section V.

II. Popular Recommendation Approaches

A. *Collaborative filtering (CF)*

CF refers to user-to-user association [8-9]. It follows the concept that if two or more folks have identical interests in one area then there is a likelihood that they will get attracted towards similar products or items of some other category as well [3-4]. Implicit and explicit user ratings are considered to compute similarity between two or more users. While implicit ratings are derived from user browsing pattern and click-through rate, explicit ratings are delivered by the user himself. The options of people you may know, suggested posts, similar pages you may like, suggested pokes, displayed on Facebook, are the examples of collaborative filtering. These are nothing but recommendations, based on features like number of mutual friends, similar pages liked or number of mutual groups, locations a user have been to or belongs to etc. For example, if two users have mutual friends, then the possibility is that they two may know each other as well.

B. *Content-based filtering (CBF)*

CBF follows the idea “Show me more of what I have liked” [8]. These systems recommend only those products to a user, which are alike the ones they enjoyed in the past [3-4]. The similarity between two or more products is computed based on the features they have in common. While browsing through videos on YouTube, the browsing pattern of the user is observed to find out the types of videos he or she prefers and based on that, he or she is recommended similar kind of content under the feature suggested videos. Therefore, CBF based systems assume that if a user likes an item from some specific category, probability is that he or she might show his interest in another item from that category as well.

C. *Knowledge based systems (KBS)*

KBS follows the notion “Tell me what fits my needs” [8]. These systems generate recommendations based on a particular domain knowledge or domain expert [3-4]. The user specifies his needs to the system which further compares those needs with its knowledge base and provide the most relevant suggestions accordingly. While buying anything from an e-commerce site, the users are expected to specify the desired features of the product they wish to buy, like the price range, color, size etc. and based on the resemblance between the features specified by the user and the properties of the product, the most suitable products are recommended.

D. *Hybrid recommender systems*

Hybrid systems are the combinational systems, the inspiration following which is to merge the characteristics of two or more recommendation techniques in such a way that the limitations of a lone recommendation approach are conquered by the other. Netflix is the most popular hybrid recommender based on collaborative and content-based approach. It suggests movies or series to a user as per his interests, view history and the similarity between him and other Netflix users. Consider a user likes *PS I Love you, The Notebook* and *The Fault in our stars* then as he subsequently uses Netflix, he is recommended movies belonging to Romantic genre. Similarly, if two users have liked or viewed similar content on Netflix then each of them would be suggested what the other views next.

E. *Demographic systems*

These systems consider user demography (age, gender, region) to provide recommendations to a user [3-4]. e-commerce sites like Amazon, eBay, Flipkart etc. are the examples. There is Amazon.com, Amazon. In, and Amazon.uk and similar is the case with eBay. Therefore, for each particular region, only the items available in that region would be recommended and as an add-on the price would be specified in the relevant currency only. Further, users have the option to select the category (girl, boy, kids, men, women etc.) for which they want to buy a product for.

F. *Community-based systems*

These systems trail the notion, “Tell me who your friends are, and I will tell you who you are [8]”. Such systems consider interests or likings of user acquaintances while making the recommendations. Such systems can be considered similar to the concept of collaborative filtering except the fact that the later considers anonymous or acquainted users to generate the recommendations, while community-based systems generated the recommendations based on user’s acquaintances only.

A comparison of afore-mentioned recommendation approaches based on the parameters considered for generating recommendations is given in Table 1.

TABLE I. COMPARISON OF RECOMMENDATION APPROACHES

Approach	**Parameters considered**						
	User profile	***Explicit ratings***	***Implicit ratings***	***Knowledge-base***	***Community***	***Context***	***Demography***
CF	✓	✓	✓				
CBF	✓	✓	✓				
HS	✓	✓	✓	✓	✓	✓	✓
KBS	✓			✓			
CBS	✓	✓	✓		✓		
DBS	✓	✓	✓		✓		✓

CF= Collaborative filtering approach, CBF= Content based filtering systems, HS = Hybrid systems, KBS= Knowledge based systems, CBS= Community based systems, DBS= Demography based systems

III. HYBRID RECOMMENDER SYSTEMS

Hybrid recommender system associates multiple recommendation approaches together for better results. The basic definition of Hybrid recommender states that, "Given two recommendation approaches X and Y, a Hybrid system XY combines the two in such a way that the limitations of one are overcome by the other." One of the most popular examples of a Hybrid system is Netflix [10]. Netflix uses both Content-based and Collaborative filtering approach to recommend online media content to the users.

Further classification of Hybrid recommender systems:

A. Monolithic hybrid systems

Monolithic systems consist of only a particular recommendation module that combines numerous recommendation approaches after pre-processing and assimilating diverse knowledge sources [11]. Therefore, a built-in amendment of any algorithm envisioned for managing and pre-processing input data leads to the development of a monolithic hybrid system. It has been further categorized as: Feature augmentation and Feature combination hybrid systems. The prior integrates compound input statistics to a solo recommendation algorithm. Whilst in case of later, the output of individual approach is considered as the input for the other algorithm.

B. Parallelized hybrid systems

These systems are the combination of various recommender systems engaged concurrently and some hybridization mechanism that combines the outputs obtained from those systems. These have been further categorized as: Switching, Mixed and Weighted hybrid systems. Weighted systems provide the recommendations by combining the weighted sum of the ranks of multiple recommendation approaches into a distinct recommendation list. Mixed hybrid recommender systems are comparable to weighted systems, and in these systems, the results obtained by various recommendation approaches are pooled together and offered to the user as one. Switching systems are based on the superiority of recommendation outcomes obtained and the preferences of the user. These systems switch among recommendation approaches so as to offer the most suitable results to the user.

C. Pipelined hybrid systems

These systems generate the user recommendations by adapting a staged progression wherein compound recommendation approaches are constructed one above the other. These systems are further categorized as: Meta-level systems and Cascade systems. Cascade systems follow the notion of successor (descendent) and predecessor (antecedent) where the yield of the predecessor after it gets polished by the successor yields the concluding recommendations. Meta-level systems construct a model derived from just one recommendation algorithm and further use that model as an input value for some other recommendation algorithm.

Based on a survey conducted by Robin Burke, Hybrid systems can be classified into seven categories [11-12]:

- Weighted: The system combines score of several recommendation approaches together to provide a higher weighted recommendation.
- Switching: The system picks amid recommendation approaches and administer the chosen one.
- Mixed: The system combines recommendations generated by multiple recommenders and present them together.
- Feature Combination: In this system, features extracted from multiple knowledge sources are integrated and a single recommendation algorithm is applied to it.

- Feature Augmentation: In this system, a feature or a set of features are computed by a single recommendation approach, and this result is further used as an input to the next approach.
- Cascade: In this system, the recommendations generated by one recommender are further refined by another recommender system.
- Meta-level: This system creates a model based on a single recommendation technique and uses this model further as an input for the other approach.

Table II summarizes the existing hybrid recommender systems based on the hybridization technique they follow. From the summary as given in Table II, it can be observed that Weighted, Mixed, Feature combination and Switching hybrid models are used most frequently while Feature augmentation, Cascade and Meta-level models are not deployed much and are yet to be explored further. Moreover, Meta-level has so far hybridized Collaborative filtering with Content-based filtering only. Therefore, other possible combinations also need to be hybridized, explored further and evaluated for their future scope.

TABLE II. EXISTING HYBRID MODELS [11-12]

	Hybridization Techniques						
	Weighted	***Switching***	***Mixed***	***FC***	***FA***	***Cascade***	***Meta-level***
CF- CBF	✓	✓	✓	✓	✓	✓	
CF- DBS	✓						
CF- KBS	✓	✓					
CBF- CF	✓	✓	✓	✓			✓
CBF- DBS	✓						
CBF- KBS				✓			
DBS- CF	✓	✓	✓	✓			
DBS- CBF	✓	✓	✓	✓			
DBS- KBS							
KBS- CF	✓	✓	✓	✓	✓	✓	
KBS- CBF	✓	✓	✓	✓			
KBS- DBS	✓	✓	✓	✓			

FC= Feature combination, FA= Feature augmentation,

IV. RELATED WORK

Hybrid recommender systems collaborate multiple recommendation strategies to reinforce their advantages and minimize their shortcomings. One of the earliest Hybrid systems was Fab, a meta-level-based recommender that suggested websites to the users [13]. It combined Collaborative filtering with Content based filtering to find users with identical website likings and to find websites with similar content respectively. Similar work was done for combining content-based system with collaborative filtering system of GroupLens to address the sparsity and early-rater problem [14]. Another work based on the same grounds combined information filtering agents with collaborative filtering to minimize information overload by restricting the recommendations to the ones only which may interest the user [15]. Similarly, the potential of hybridization was further explored on Movie Lens and IMDB datasets using clustering and Bayesian network model respectively [16-17].

Most of the earlier hybrid models were based on collaborative and content-based filtering. But soon the researchers started exploring other combinations as well. One of such works was based on an interactive hybrid music recommender system that considered user semantics from Facebook, Twitter and Wikipedia [18]. Further, the work contributed towards a novel cross-source hybridization model. Hybridization is not just limited to combining two approaches, rather three or more recommendation techniques can also be hybridized. Following the concept, the authors introduced a novel recommender integrating CBF, KBS with Context-aware systems for e-tourism to generate recommendations based on user-location [19]. The work received positive feedback from the users given its ability to satisfy the users while they are moving around. Similarly, the authors proposed a hybrid fuzzy linguistic model to recommend interesting and relevant research resources to the users [20]. Another work proposed a unique cascading hybrid model merging user features, rating, and demographic information and compared existing machine learning algorithms with the proposed approach based on Boosted similarity [21].

Recent works do not solely depend on user ratings or profile, rather other parameters namely social network profiles, user semantics, user feedback and current location are also considered. The authors introduced GeoSRS, a location-based social network (LBSN) hybrid model that considers user location and text reviews to recommend places of interest to a user [22]. Furthermore, a hybrid restaurant recommendation model based on text reviews and images posted on blogs is introduced to study the influence of visual data in attracting customers [23]. Further, a hybrid model following enhanced fuzzy multi-criteria collaborative filtering is proposed that integrates demographic data into an item based ontological semantic filtering for suggesting movies [24]. Similarly, a mixed hybrid model based on content and collaborative strategies is introduced to provide group recommendations [25]. Furthermore, a novel healthcare recommender system called iDoctor, following hybrid matrix factorization approach is proposed, that considers doctor features, user preferences and user reviews as extracted using sentiment analysis [26].

An overview of the existing hybrid systems as described above is given in Table III along with the hybridization model, Machine learning algorithm followed and their respective application areas.

TABLE III. EXISTING SYSTEMS BASED ON HYBRID MODEL

System	Hybrid of	Model	Machine Learning Algorithm	Recommends
[13]	CF, CBF	Cascade	-	Websites
[16]	CF, CBF	Cascade	Clustering	Movies
[17]	CF, CBF	Feature Augmentation	Bayesian Network	Movies
[18]	CBF, Context -based	Weighted, Mixed	-	Music
[19]	CBF, KBS, Context-based	Feature Augmentation	Distance based re-ranking	e- Tourism
[20]	CF, CBF	Mixed	Multi-granular Fuzzy linguistic	Research resources
[21]	CF, CBF, DBS	Cascade	Boosted Similarity	Movies
[22]	CF, CBF	Weighted	Cosine similarity, Data Mining	Places of interest
[23]	CF, CBF	Mixed	Factorization Machine	Restaurants
[24]	CF, DBS	Mixed	Fuzzy Cosine, Fuzzy Jaccard	Movies
[25]	CF, CBF	Mixed	-	Movies
[26]	CF	Weighted	Hybrid Matrix factorization	Healthcare

V. CONCLUSION

With the dawn of computer era, access to information became too handy. In the interim, the outburst of information led to the problem of overload. As a result, Recommender system was subsequently developed as an information filtering system. Recommender systems are referred as software tools that aid in reducing the options accessible and endow with the most apposite suggestions as per the desires. Based on the recommending mechanisms, recommender systems are mainly classified into: Content-based approach, Collaborative Filtering, Knowledge-based recommendation, Hybrid approach, Community-based and Demography based systems.

In this paper, we studied and compared the various recommendation approaches developed till date. The prime focus of this study is on Hybrid recommender systems. Hybrid systems are the combinational systems, which integrate two or more recommendation techniques in a fashion that the limitations of a lone recommendation method are conquered by the other. We described the various hybridization models as proposed by Burke and based on his work, we further classified the existing hybrid models according to the hybridization technique, Machine learning algorithm used and their respective application domains. Based on the literature, it is observed that most of the hybrid models developed till date primarily combine Content-based and Collaborative filtering while the other approaches are seldom used. We also observed that there are various combinations of hybrid models which still remain unexplored, like Meta-level which is the least used hybrid model and so far, only the combination of Collaborative and Content-based filtering has been hybridized based on Meta-level. Furthermore, Feature-augmentation and Cascade models are also not used as frequently and have only been used to combine Collaborative filtering with Content-based and Knowledge-based systems respectively. Therefore, further research is required to explore the other remaining possible hybrid models using various Machine learning algorithms.

REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of [1] Çano, E. and Morisio, M., 2017. Hybrid recommender systems: A systematic literature review. Intelligent Data Analysis, 21(6), pp.1487-1524.

[2] Mcginty, L. and Smyth, B., 2006. Adaptive selection: An analysis of critiquing and preference-based feedback in conversational recommender systems. International Journal of Electronic Commerce, 11(2), pp.35-57.

[3] Ricci, F., Rokach, L. and Shapira, B., 2011. Introduction to recommender systems handbook. In Recommender systems handbook (pp. 1-35). Springer, Boston, MA.

[4] Jannach, D., Zanker, M., Felfernig, A. and Friedrich, G., 2010. Recommender systems: an introduction. Cambridge University Press.

[5] Terry, D.B., 1993, December. A tour through tapestry. In Proceedings of the conference on Organizational computing systems, ACM, pp. 21-30.

[6] Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T., 2004. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems, 22, pp. 5-53.

[7] Jannach, D. and Friedrich, G., 2013, August. Tutorial: recommender systems. In International Joint Conference on Artificial Intelligence, Beijing, China. Association for the Advancement of Artificial Intelligence Press, pp. 1-67.

[8] Shani, G. and Gunawardana, A., 2011. Evaluating recommendation systems. In Recommender systems handbook (pp. 257-297). Springer, Boston, MA.

[9] Thorat, P.B., Goudar, R.M. and Barve, S., 2015. Survey on collaborative filtering, content-based filtering and hybrid recommendation system. International Journal of Computer Applications, 110(4), pp.31-36.

[10] Burke, R., 2002. Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction, 12(4), pp.331-370.

[11] Burke, R., 2007. Hybrid web recommender systems. In The adaptive web (pp. 377-408). Springer, Berlin, Heidelberg.

[12] Shoham, Y., 1997. Combining content-based and collaborative recommendation. Communications of the ACM.

[13] Sarwar, B.M., Konstan, J.A., Borchers, A., Herlocker, J., Miller, B. and Riedl, J., 1998. Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system. In in the GroupLens Research Collaborative Filtering System. Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW).

[14] Good, N., Schafer, J.B., Konstan, J.A., Borchers, A., Sarwar, B., Herlocker, J. and Riedl, J., 1999. Combining collaborative filtering with personal agents for better recommendations. AAAI/IAAI, 439.

[15] Li, Q. and Kim, B.M., 2003, October. Clustering approach for hybrid recommender system. In Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003) (pp. 33-38). IEEE.

[16] De Campos, L.M., Fernández-Luna, J.M., Huete, J.F. and Rueda-Morales, M.A., 2010. Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks. International Journal of Approximate Reasoning, 51(7), pp.785-799.

[17] Bostandjiev, S., O'Donovan, J. and Höllerer, T., 2012, September. TasteWeights: a visual interactive hybrid recommender system. In Proceedings of the sixth ACM conference on Recommender systems (pp. 35-42). ACM.

[18] Noguera, J.M., Barranco, M.J., Segura, R.J. and MartíNez, L., 2012. A mobile 3D-GIS hybrid recommender system for tourism. Information Sciences, 215, pp.37-52.

[19] Porcel, C., Tejeda-Lorente, A., Martínez, M.A. and Herrera-Viedma, E., 2012. A hybrid recommender system for the selective dissemination of research resources in a technology transfer office. Information Sciences, 184(1), pp.1-19.

[20] Ghazanfar, M.A. and Prugel-Bennett, A., 2010, January. A scalable, accurate hybrid recommender system. In 2010 Third International Conference on Knowledge Discovery and Data Mining (pp. 94-98). IEEE.

[21] Capdevila, J., Arias, M. and Arratia, A., 2016. GeoSRS: A hybrid social recommender system for geolocated data. Information Systems, 57, pp.111-128.

[22] Chu, W.T. and Tsai, Y.L., 2017. A hybrid recommendation system considering visual information for predicting favorite restaurants. World Wide Web, 20(6), pp.1313-1331.

[23] Kermany, N.R. and Alizadeh, S.H., 2017. A hybrid multi-criteria recommender system using ontology and neuro-fuzzy techniques. Electronic Commerce Research and Applications, 21, pp.50-64.

[24] Kaššák, O., Kompan, M. and Bieliková, M., 2016. Personalized hybrid recommendation for group of users: Top-N multimedia recommender. Information Processing & Management, 52(3), pp.459-477.

[25] Zhang, Y., Chen, M., Huang, D., Wu, D. and Li, Y., 2017. iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization. Future Generation Computer Systems, 66, pp.30-35.