International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Employing Community Detection into Recommender System: A Review

Mandeep Kumbhar[1], Jidhanyasa Kolhe[2], Devesh Kumawat[3], Prof. (Mrs.) S.P. Bansu[4]

[1-4]*Department of Computer Engineering, A.C.Patil College of Engineering, Navi Mumbai, India*

---***---

Abstract - *The recommendation system aims to predict user's preferences for new items based on the preferences for old items. As no of internet users has increased exponentially, the corporate sectors, especially the E-commerce industry is incorporating 'evolved' recommendation system to retain customers. Recommendation system has evolved from traditional Collaborative-based Filtering or Content-based Filtering to hybrid models which combines two or more strategies to gain accurate results. Both these traditional methods has its strengths and weakness, thus there is a need to make use of different strategies from different domains to build more advance recommender system. One of the strategies is Community Detection in social network analysis which is being used to make hybrid models. Community detection aims to find a highly connected group of nodes (users) in social networks. A social network may contain many hidden information that can be used to know 'preferences' or 'interests' of users. This paper provides a review on recommender systems which incorporates community detection, metrics used for evaluation purpose, its performance against old traditional methods.*

Key Words: Recommender sytem, Collaborative filtering, Content filtering, Community detection, Hybrid, Social network.

1.INTRODUCTION

The Recommendation has been part of human practice from a very early date. The most simple and used, but yet the most effective recommendation method is 'Word of mouth'. This informal way of passing information orally from person to person is the most practiced way of recommendation. With the rapid growth in internet users, the method of advertisements of products has changed dramatically. Now companies are experimenting with new methods to recommend their products to users. This experimentation has made the 'Recommendation system' a new standalone research area. According to the article published in the journal Mckinsey & Company, 35% of the purchases on Amazon are due to their recommendation system and about 75% of what people are watching on Netflix is the result of their recommender system [2]. This data shows the prominence of recommender systems in the online-based services. The traditional methods of recommender systems have been studied extensively by researchers from past decades. It can be divided into many categories depending on the information they are using. Collaborative filtering (CF), Content-based filtering (CBF), Knowledge-based filtering (KF), Demographic filtering (DF), and hybrid methods are major categories of recommender systems.

1.1 Collaborative-based filtering (CF)

CF relies on the past interactions of users and items to predict preferences. This data, stored in the 'user-item' matrix, is used to find similarity between users or items to make predictions. CF is again divided into two subcategories - Model-based approach and Memory-based approach. The model-based approach uses Machine learning techniques like PCA, SVD, Matrix factorization, Neural networks, etc. to find the user's rating of unrated items. The memory-based approach uses Pearson correlation or Cosine similarity functions to find similarities in users or items and then computes weighted average ratings of unrated items. It has two subcategories - User-item filtering and Item-item filtering. The user-item filtering approach aims to find similarities between users while the item-item filtering approach relies on similarities between items. The major problem that Memory-based approach suffers is Data sparsity, Non-scalability, and Cold start. As Model-based approaches don't solely depend on the user's rating data, it doesn't suffer from data sparsity and non-scalability problems. Overall, CF is preferred for diverse recommendations as it considers the data and preferences of other users. It also provides flexibility in the user's perspective and preferences due to its ability to capture the change in user's interest over time.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

1.2 Content-based filtering (CBF)

CBF examines the description of each item and constructs a personalized user profile. The description of each item is defined by tags (eg. In the movie dataset, *director name*, *actor name*, *genre*, etc. will be the tags.). The user profile is built up with the same tags by analysing the description of each item preferred by the user in the past. CBF doesn't need any data from other users and their preferences for other items making it a highly personalized recommendation method. This makes it scalable to large no. of users even in the sparse dataset. However, it fails to deliver diverse recommendations as it is specific to the existing interests of the user.

1.3 Knowledge-based filtering (KF)

KF uses knowledge about items such as features, requirements of users, and recommendation constraints. This approach is used in those domains in which historical data of users is not available and thus traditional CF or CBF cannot be applied. For eg, people rarely buy cars or houses and therefore a prior knowledge about the user's requirements is needed.

1.4 Demographic filtering (DF)

DF aims to classify the users based on demographic attributes of users like age, sex, occupation, race, etc. A user of a particular category will be recommended items that are preferred by other users of the same category. It doesn't suffer from the cold start problem as it doesn't consider ratings for recommendations. However, it becomes a difficult task to extract demographic attributes from users. Moreover, classification based on demography can be too general and vague leading to inaccurate recommendations.

2. SOCIAL NETWORK ANALYSIS

Social Network Analysis (SNA) is the systematic examination of the relationship between individuals, groups or any other social structure. SNA primarily focuses on relations between the actors rather than the attributes of actors [3]. SNA has emerged from the fields of Network analysis and Graph theory. Network theory deals with the composition and representation of the problem into the graph and tries to discover the solution with the methods or algorithms provided by graph theory. Euler's Seven Bridges of Konigsberg problem is considered to be the earliest known problem in Network Analysis.

As humans are social animals, many hidden information can be inferred about individuals just by examining their social relations. With the increase in Internet users, the study of virtual online social relations has gained importance. Han, *et al* [4] presented a comprehensive study on large Facebook data to find correlations between the user's similarity and social features. They concluded that the peoples are more likely to have similar tastes if they have similar demographic information (e.g., age, location) or share more common friends. The one such relationship that has been widely studied is the Trust relationship between users in social media sites for recommendation [5],[6],[7],[8]. Chen, *et al* [9] proposed a factor analysis approach that explicitly and implicitly used social trust relationships simultaneously to overcome the limitations of traditional RS's. Qin, *et al* [10] proposed a recommender system for YouTube based on the network of its reviewers. They created a network of videos called YouTube Recommender Network (YRN) from the comments left in the comment section by users. The weights on the edges are the no. of common comments left by the user. They find the group of nodes that are highly connected, called communities with an intent to find similarities in videos. Their approach has more diversity in recommendations than traditional RS's. Fields, *et al* [11] have analyzed the network of music artists on myspace social networking website and used community detection algorithms to partition the network.

2.1 Community Detection (CD)

Because of human tendency to get associated with peoples having similar interests, they often form communities of similar likings and tastes. The detection of communities in social media helps us to find groups of like-minded users for marketing and recommendations. Community Detection (CD) aims to find community structure or simply groups in graphs. The community structure refers to the organization of vertices in groups, with many edges joining vertices of the same group and comparatively few edges joining vertices of different groups [12]. CD is one of the key tools in SNA to partition the graph into clusters having densely connected vertices.

Based on the structure, Communities can be Overlapping and Non-overlapping. Overlapping communities contains nodes that are part of two or more communities i.e they do not form disjoint communities. Non-overlapping communities do not contain common nodes, they form disjoint communities [13]. Most of the CD algorithms are static i.e they can work only on the snapshot of data, but in recent years algorithms for the dynamic network have been introduced [14]. The two widely known algorithms for CD are Girvan–Newman method and Louvain method.

2.1.1 Girvan–Newman algorithm (GN)

It was proposed by Girvan & Newman [15].It is based on the graph-theoretic measure called as edge betweenness centrality. The betweenness centrality of an edge e is defined as the number of shortest paths between pairs of vertices that pass through the edge e. The edges connecting communities have high edge betweenness. By removing these edges, groups from one another are separated and the underlying community structure of the graph is detected.

The GN has following steps:-

1) Calculate the betweenness for all edges in the network.
2) Remove the edge with the highest betweenness.
3) Recalculate betweennesses for all edges affected by the removal.
4) Repeat from step 2 until no edges remain.

The entire algorithm runs in worst-case time $O(m^2n)$ in a graph having m edges and n vertices.

2.1.2 Louvain algorithm

It was proposed by Blondel, *et al* [16]. It is based on the optimization of modularity. Modularity measures the density of edges inside communities to edges outside communities. Its value lies between -1 and 1. For a weighted graph, modularity is defined as:

$$Q = \frac{1}{2m} \sum \left[A_{ij} - \frac{k_i k_j}{2m} \right] \boldsymbol{\delta}(c_i, c_j)$$

Where

- A_{ij} represents the edge weight between nodes i and j.
- k_i and k_j are the sum of the weights of the edges attached to nodes i and j, respectively.
- m is the sum of all of the edge weight in the graph.
- c_i and c_j are the communities of the nodes.
- δ is Kronecker delta function ($\delta_{x,y} = 1\ if\ x = y, 0\ \ otherwise$).

Louvain method has two steps: In the first step, small communities are detected by optimizing modularity locally on all nodes and, in the second step, each detected small communities are treated as a single node and the first step is repeated. It runs in $O(n^2)$ time in graph having n vertices.

3. RESEARCH METHODOLOGY

The purpose of this study is to understand the trend of recommender systems that are using community detection as a tool and to provide future directions in the fields of recommender systems and social network analysis. To review the literature systematically, Kithchenam's systematic review procedure was employed [17]. Following steps were taken for the literature review:-

1) Determining the research topic.
2) Searching keywords to form query strings.
3) Extracting literature using query strings, considering exclusion and inclusion criteria.
4) Analysing the literature.
5) Documenting the results

The topic of our research was on how community detection is incorporated into various recommender systems. Thus keywords like "Community detection", "Recommendation", "Social network analysis", "community structure" and "overlapping communities" was used to form query strings. Keywords like "recommender", "recommendation" was common in all the query strings. The search was performed on four databases, namely IEEE Xplore, Science Direct, JOSTOR, and arXiv. In total, 120 papers were retrieved from all databases.

3.1 Inclusion and Exclusion criteria

Following criteria were used to include papers:-
1) Papers published between Jan 1st, 2001 and Jan 1st, 2020.
2) Papers having both keywords "recommender" and "community detection" in either title or abstract.

Following criteria were used to exclude papers:-
1) Duplicate reports of the same study.
2) Papers having ambiguity in their proposed solution or results.
Based on the above criteria, 120 papers were analyzed and thus refined to 48 papers. Further, papers with no citations were eliminated to improve the quality of papers. In total, the no. of papers was reduced to 27 by applying the inclusion and exclusion criteria.

3.2 Quality assessment

Each paper was evaluated using the York University, Centre for Reviews and Dissemination (CDR) Database of Abstracts of Reviews of Effects (DARE) criteria [18]. The criteria are based on the following quality assessment (QA) questions:-
1) Question/objective sufficiently described?
2) Study design evident and appropriate?
3) Context for the study clear?
4) Connection to a theoretical framework / wider body of knowledge?
5) Sampling strategy described, relevant and justified?
6) Data collection methods clearly described and systematic?
7) Data analysis clearly described and systematic?
8) Use of verification procedure(s) to establish credibility?
9) Conclusions supported by the results?
10) Reflexivity of the account?
Depending on the degree to which each specific criteria were met, each paper was given scores ("Yes" = 2, "Partial" = 1, "No" = 0). The papers having mean score less than or equal to 0.5 were eliminated. (see Table 1). In total, 9 papers were considered for litreture review. The detailed quality assessments of selected papers is given in Table 2.

Table -1: Quality scores of selected papers

Author's name	**Total score**	**Mean score**
Qiu, *et al* [19]	12	0.6
Ying, *et al* [20]	18	0.9
Rohit & Doina [21]	19	0.95
Hou & Gai [22]	17	0.85

Feng, *et al* [23]	19	0.95
Xin, *et al* [24]	15	0.75
Zheng, *et al* [25]	20	1
Maliheh, *et al* [26]	16	0.8
Qin, *et al* [10]	14	0.7

4. RESULTS AND CONCLUSION

Altogether,9 papers were selected for review after analysing the quality of papers based on the above mentioned criteria. The metadata of 9 papers are given in Table 3. The main focus of all studies was generally on the exploitation of the user's social network through community detection. The problem of diverse recommendation, capturing change in user's interest, and more personalized recommendation was some of the issues that papers have put more emphasis. Data sparsity was the less addressed problem in all the studies [19],[20],[21]. New approaches in community detection was also given importance [22], [23], [25], [26]. Hou & Gai [22] has introduced the Multi Label Propagation Algorithm(MLPA), an extension of LPA, to detect overlapping community structure. Zheng, *et al* [25] has proposed a new approach, called HIOC, for clustering. A community detection method based on the PageRank and Fuzzy c-mean method was also studied by Maliheh, *et al* [26]. Qin, et al [10] proposed an RS for YouTube based on the user's comments. The metrics used for evaluation were Precision [19], [20], [23], [25], Recall [19], [23], [25], F1 measure [20], [23], Mean Average Precision (MAP) [21], Mean Absolute Error (MAE) [22], [24], [26], Root Mean Square Error(RMSE) [24], [26]. The results of the analysis of each paper are given in Table 4.

This paper presents the review of the recommender system which exploits the social network of users with the help of community detection methods. The network of users can provide enough information about user's preferences and taste, thus recommender systems can be more personalized, dynamic and diverse if studied together with analysis of the user's social network.

REFERENCES

[1] J. E. Solsman, "Youtube's ai is the puppet master over most of what you watch," CNET, January, vol. 10, 2018.

[2] I. MacKenzie, C. Meyer, and S. Noble, "How retailers can keep up with consumers," McKinsey & Company, vol. 18, 2013.

[3] E. Otte and R. Rousseau, "Social network analysis: a powerful strategy, also for the information sciences," Journal of information Science, vol. 28, no. 6, pp. 441–453, 2002.

[4] X. Han, L. Wang, S. Park, A. Cuevas, and N. Crespi, "Alike people, alike interests? a large-scale study on interest similarity in social networks," in 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014). IEEE, 2014, pp. 491–496.

[5] M. Eirinaki, M. D. Louta, and I. Varlamis, "A trust-aware system for personalized user recommendations in social networks," IEEE transactions on systems, man, and cybernetics: systems, vol. 44, no. 4, pp. 409–421, 2013.

[6] L. Guo, C. Zhang, and Y. Fang, "A trust-based privacy-preserving friend recommendation scheme for online social networks," ieee transactions on dependable and secure computing, vol. 12, no. 4, pp. 413–427, 2014.

[7] P. Victor, C. Cornelis, M. De Cock, and A. Teredesai, "Trust-and distrust-based recommendations for controversial reviews," in Web Science Conference (WebSci'09: Society On-Line), no. 161, 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, 2016.

[9] C. Chen, J. Zeng, X. Zheng, and D. Chen, "Recommender system based on social trust relationships," in 2013 IEEE 10th International Conference on e-Business Engineering. IEEE, 2013, pp. 32–37.

[10] S. Qin, R. Menezes, and M. Silaghi, "A recommender system for youtube based on its network of reviewers," in 2010 IEEE Second International Conference on Social Computing. IEEE, 2010, pp. 323–328.

[11] B. Fields, K. Jacobson, C. Rhodes, M. d'Inverno, M. Sandler, and M. Casey, "Analysis and exploitation of musician social networks for recommendation and discovery," IEEE Transactions on Multimedia, vol. 13, no. 4, pp. 674–686, 2011.

[12] S. Fortunato, "Community detection in graphs," Physics reports, vol. 486, no. 3-5, pp. 75–174, 2010.

[13] P. Bedi and C. Sharma, "Community detection in social networks," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 6, no. 3, pp. 115–135, 2016.

[14] T. Aynaud and J.-L. Guillaume, "Static community detection algorithms for evolving networks," in 8th International symposium on modeling and optimization in mobile, Ad Hoc, and wireless networks. IEEE, 2010, pp. 513–519.

[15] M. Girvan and M. E. Newman, "Community structure in social and biological networks," Proceedings of the national academy of sciences, vol. 99, no. 12, pp. 7821–7826, 2002.

[16] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of statistical mechanics: theory and experiment, vol. 2008, no. 10, p. P10008, 2008.

[17] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering–a systematic literature review," Information and software technology, vol. 51, no. 1, pp. 7–15, 2009.

[18] L. M. Kmet, L. S. Cook, and R. C. Lee, "Standard quality assessment criteria for evaluating primary research papers from a variety of fields," 2004.

[19] H. H. Qiu, Y. Liu, Z. J. Zhang, and G. X. Luo, "An improved collaborative filtering recommendation algorithm for microblog based on community detection," in 2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE, 2014, pp. 876–879.

[20] J.-C. Ying, B.-N. Shi, V. S. Tseng, H.-W. Tsai, K. H. Cheng, and S.-C. Lin, "Preference-aware community detection for item recommendation," in 2013 conference on technologies and applications of artificial intelligence. IEEE, 2013, pp. 49–54.

[21] R.ParimiandD.Caragea,"Community detection on large graph datasets for recommender systems," in 2014 IEEE international conference on data mining workshop. IEEE, 2014, pp. 589–596.

[22] H. Qiang and G. Yan, "A method of personalized recommendation based on multi-label propagation for overlapping community detection," in 2012 3rd International Conference on System Science, Engineering Design and Manufacturing Informatization, vol. 1. IEEE, 2012, pp. 360–364.

[23] H. Feng, J. Tian, H. J. Wang, and M. Li, "Personalized recommendations based on time-weighted overlapping community detection," Information & Management, vol. 52, no. 7, pp. 789–800, 2015.

[24] L. Xin, E. Hai-Hong, J.-j. TONG, and M.-n. SONG, "Collaborative recommendation based on social community detection," The Journal of China Universities of Posts and Telecommunications, vol. 21, pp. 20–45, 2014.

[25] J. Zheng, S. Wang, D. Li, and B. Zhang, "Personalized recommendation based on hierarchical interest overlapping community," Information Sciences, vol. 479, pp. 55–75, 2019.

[26] M. Goliforoushani, R. H. Rad, and M. A. Haeri, "A fuzzy community-based recommender system using pagerank," arXiv preprint arXiv:1812.09380, 2018.

Table - 2: Quality assessment of selected papers

Author	QA1	QA2	QA3	QA4	QA5	QA6	QA7	QA8	QA9	QA10
Qiu, *et al* [19]	Y	P	Y	P	P	N	P	Y	Y	Y
Ying, *et al* [20]	Y	Y	Y	Y	Y	P	Y	Y	Y	P
Rohit & Doina [21]	Y	Y	Y	Y	Y	Y	Y	Y	Y	P
Hou & Gai [22]	Y	P	Y	Y	P	P	P	P	P	P
Feng, *et al* [23]	Y	Y	Y	P	Y	Y	Y	Y	Y	Y
Xin, *et al* [24]	Y	Y	Y	Y	P	N	P	Y	Y	P
Zheng, *et al* [25]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Maliheh, *et al* [26]	Y	Y	P	Y	Y	P	Y	Y	P	P
Qin, *et al* [10]	Y	Y	Y	Y	P	N	P	N	P	P

Table - 3: Metadata of selected papers

Author	Title	Year	Conference/Journal
Qiu, *et al* [19]	An improved collaborative filtering recommendation algorithm for microblog based on community detection	2014	International Conference on Intelligent Information Hiding and Multimedia Signal Processing
Ying, *et al* [20]	Preference-aware community detection for item recommendation	2013	Conference on technologies and applications of artificial intelligence
Rohit & Doina [21]	Community detection on large graph datasets for recommender systems	2014	IEEE international conference on data mining workshop
Hou & Gai [22]	A method of personalized recommendation based on multi-label propagation for overlapping community detection	2012	3rd International Conference on System Science, Engineering Design and Manufacturing Informatization
Feng, *et al* [23]	Personalized recommendations based on time-weighted overlapping community detectio	2015	Information & Management
Xin, *et al* [24]	Collaborative recommendation based on social community detection	2014	The Journal of China Universities of Posts and Telecommunications
Zheng, *et al* [25]	Personalized recommendation based on hierarchical interest overlapping community	2019	Information Sciences
Maliheh, *et al* [26]	A Fuzzy Community-Based Recommender System Using PageRank	2018	N/A
Qin, *et al* [10]	A recommender system for youtube based on its network of reviewers	2010	IEEE Second International Conference on Social Computing

Table - 4: Findings of literature review

Author	Used methods/algorithms	Result of the work	Evaluation metrics	Proposed future works
Qiu, *et al* [19]	Girvan and Newman method of CD; Person coefficient of correlation.	The proposed approach tackles the issues of data sparsity and user relationship influence in microblog recommendation	Precision; Recall.	N/A
Ying, *et al* [20]	Cosine similarity; and other similarity measures	The proposed approach of PCRS and PCD tackles the problem of mining user's preference in community detection and provides a method for evaluation of ratings of each user-item pair.	Precision; Coverage; F1 measure	N/A
Rohit & Doina [21]	Louvian method for CD; Adsorption algorithm	The proposed approach is effective for sparse graphs	Mean Average Precision (MAP).	Integrating CD with popularity-based recommender system and using matrix factorization; Applying the proposed approach in domains like music..
Hou & Gai [22]	Label Propogation Algorithm (LPA).	An extension of LPA, Multi Label Propagation algorithm (MLPA),	Mean Absolute	Modifying MLPAO to detect overlapping

		is introduced to detect overlapping communities. The efficiency of the proposed RS's is better than traditional CF's while accuracy is unchanged	Error(MAE); Running Time.	community structure in the bipartite network.
Feng, *et al* [23]	Association rule mining	The proposed recommender system, TOTAR uses Association rule mining for recommendations. It also considers the time effects of changes in the user's interest.	Precision; Recall; F1 measure; Diversity	The selection strategy of start node in community detection; The long tail effects in user interests; consideration of similar dislikes among users.
Xin, *et al* [24]	Pearson coefficient correlation; Matrix factorization	The proposed method uses two factors, namely Community similarity and Community affection in modified RS's.	Mean Absolute Error (MAE); Root Mean Sqaure Error(RMSE).	N/A
Zheng, *et al* [25]	Term Frequency–Inverse Document Frequency (TF-IDF); Regularization; Similarity meausres.	A multi-granularity similarity calculation method, which describes similar relationships between users, is introduced. A new approach, called HIOC, is proposed to detect clusters of similar users.	Precision; Recall; F1 measure.	Effects of different granular communities in recommendation; Recommendation model which captures the change in user's taste
Maliheh, *et al* [26]	PageRank; Principal Component Analysis (PCA); Fuzzy C-means clustering	A new Fuzzy community detection, using PageRank metaphor, is introduced	Mean Absolute Error(MAE); Root Mean Sqaure Error(RMSE).	N/A
Qin, *et al* [10]	Clique-percolation; Utility value.	A network of users who left comments is created to find similar clusters of users. The proposed approach provides diversity in recommendations	N/A	N/A