ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **9** (2020), no.6, 3741–3750
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.6.52 Spec Issiue on ICAML-2020

USING COMMUNITY DETECTION TECHNIQUE IN RECOMMENDER SYSTEM

MINI SINGH AHUJA

ABSTRACT. Recommender systems play a very important role in e-commerce sites. The goal of a recommender system is to predict the customer's interest. Recommender systems are widely used to recommend products/items to the customers that are most relevant to their needs. Recommender systems make use of various data sources, in order to collect the characteristics of items, users, and their transactions. Traditional recommender systems have many issues which are still unresolved. Social network based recommender systems are based on communities. By forming user communities we can overcome problems of classical recommender systems and also enhance recommendations diversity. Finding communities is crucial because these communities in a network can help to classify users who share common interest. Community detection is one of the most active fields in complex networks which is used to find communities. This property can be used in various applications such as to study the spread of disease in social networks, product recommendation etc.

1. INTRODUCTION

Recommender Systems are software techniques which provide suggestions to customers in various decision-making processes. Developing a recommender system implies the skills of various fields such as Artificial intelligence, Human Computer Interaction, Data Mining, Statistics, Decision Support Systems etc.

2010 *Mathematics Subject Classification.* 90B50, 91A35.
Key words and phrases. recommender systems, communities, community detection.

These recommender systems are playing crucial role in the business world. Everybody is busy, so everyone prefers to search/shop online. Many e-commerce sites like flipkart, Amazon have million of products. Customers get confused while searching for a specific product from a wide range of applications, products, services etc. To overcome this problem, many e-commerce sites are using recommender systems to recommend specific products and services to their customers. Recommender systems have been widely used in recommending various items such as recommending movies, songs, videos, property etc. Recommender systems (RS) are used to solve the problem of information overload. There are many different types of recommender systems and also different ways to implement them. Recently, many recommender systems have been developed for variety of applications. A good recommender system should suggest the most relevant and useful product/services to its customers. Recommender systems are categorized according to their technique they use to rate user preference values. Most common types of traditional recommender systems are:

(i) Content-based recommendation systems [12] recommend products similar to those which a given user has liked/preferred in the past. It collects the information about the items. Limitations: Limited content analysis, over specialization (it prevents customers from exploring new products).
(ii) Collaborative filtering [1] uses user behavior for recommending products. They use behaviour of other users and products in the form of transaction history, ratings, selection and purchase information. It is the most popular and widely used technique in recommender systems. Limitations: new item problem (cold start problem), new user problem. When new items /products are introduced in the system, they need to be rated by few customers before they can be recommended to customers who have similar interests with the ones who rated them.
(iii) Hybrid systems [2] combine both the techniques: collaborative filtering and content-based methods in order to remove their individual disadvantages and improve overall performance while recommending products/services. Netflix uses hybrid technique for making recommendations.
(iv) Demographic based recommender systems use demographic information such as age, gender and country to identify the types of customers

that prefer a certain product. Demographic technique uses "people-to-people" correlations. The main advantage of demographic technique is that it does not require a history of customer ratings as those in collaborative and content based recommender systems. Limitation: cold start problem, grey sheep problem, need demographic information.

The main aim of any recommender system is to provide meaningful recommendations to the users for the product and services in which they are interested. These classical recommender systems either use the content of the product or the interaction between the user and the product or both. Due to the large amount of data about the customers and their product preferences, traditional recommender systems are time consuming. Moreover due to the type of data they use, they have to deal with performance issues during data processing. These traditional recommender systems are not competent enough in providing accurate recommendations to the users. There are many weaknesses in these recommender systems such as cold start problem, sparsity of data, prediction accuracy, grey sheep (when a user profile doesn't match any other user's profile), trust etc. This paper proposes the use of community detection technique in designing a recommender system in order to overcome the various issues of traditional recommender systems. By forming user communities we can overcome problems of classical recommender systems and also enhance recommendations diversity

2. COMMUNITY DETECTION

Communities reflect topological relationship between elements. So Community detection is the branch of Complex Network science dealing with the characteristics, definition, extraction and identification of such close-knit nodes called communities. The community detection is an interesting problem and has provided solutions to many real world problems. Its usage has benefited several fields such as e-commerce, sociology, medicine, chemistry, telecommunication, computer science, biology, physics etc. For example, online social communities around the world have changed the method of information propagation. Well organized arrangement of an electric circuit is formed by finding the community structure of electrical networks. Spread of epidemic or disease around the world depends on the network of migrating birds or human beings. Failure

propagation in an electrical supply system can be forecasted by finding community structure in these networks. Arenas, Fernandez, and Gomez [4] studied a network of bottlenose dolphins to uncover their social strata. Krings, Dabin, and Blondel [5] studied criminal networks to reveal the structure of large organizations of criminals. Usage of community detection in social networks can help in to revealing the fraud events [6] and other suspicious leakages of money by generating a network of customers using the text messages and telephone communications between the individuals and identifying community structures. Community detection can help in easy visualization of complex graphs as few communities might be a small version of the whole graph. Community discovery in World Wide Web can help in finding group of web sites which hyperlink to every other website in the group (link farms). Hu, Ronhovde, and Nussinov [7] applied community extraction in image processing to reveal the contour of objects and to track moving bodies in a video. Some other applications of community structure are product recommendation, friend suggestion and link inference. The basic assumption of any Community detection algorithms is that the nodes in the network have a tendency to form communities which have to be identified. Finding network communities can be called as a clustering problem where we have to cluster a set of nodes or vertices into communities. Here a node or a vertex can be a member of multiple communities simultaneously. Nodes in communities share common characteristics or attributes, so they have many relationships among themselves. People form friendship groups, proteins interact, and authors work together. Many efficient techniques been proposed by many researchers for detecting both disjoint and overlapping communities on static networks. Few algorithms also exist for dynamic community detection

3. Classification of community detection algorithms

Community discovery problem is very similar to clustering problem of data mining. Community discovery is a clustering task of data mining which is done on the graphs. Till date many algorithms have been proposed by researchers which are based on different techniques. An excellent review which explores the most popular types of community detection techniques has been presented by Fortunato [8].

(i) Graph Partioning: A Pothen [9] early methods of community detection relied on graph partitioning. Graph partitioning is useful in understanding the detailed formation networks i.e. for the networks with a small size. Graph partitioning divides the nodes into different clusters of predetermined size such that there is minimum number of links between the clusters. The number of links between the groups or clusters is termed as cut size. Various measures are cut size, ratio cut, conductance, normalized cut etc. Network partitioning problem is a NP-hard problem. Many efficient heuristic techniques have been developed over years to solve the problem. Significant applications of graph partitioning are in scientific computing, VLSI design, image processing, task scheduling in multi-processor systems, clustering and detection of clusters in social, transportation and biological networks. Algorithms based on this approach are not fit for community detection as they need to know information about the whole composition of the network and also the number of communities and size of the communities in advance.

(ii) Hierarchical Clustering Techniques: Most community detection algorithms are based on hierarchical clustering techniques, which extract clusters based on similarity metrics. Hierarchical clustering is useful in social network analysis, biology and marketing because along with identification of communities in the network, it also provides a hierarchical structure in the communities. In the Hierarchical clustering techniques we don't require any prior knowledge of the number and size of the clusters. The result of the method depends on the similarity metric chosen. The hierarchical community detection algorithms are based on the assumption that the community structure of a network is hierarchical but this type of assumption is valid for only some type of networks for e.g. organizational networks. Hierarchical clustering techniques are broadly divided into two categories: Agglomerative (bottom up) and Divisive (top down) algorithms.

In Agglomerative clustering, firstly every node belongs to its own community. Edges are added step by step on the basis of the value of a similarity metric. The nodes are grouped into bigger and bigger communities. This process can be halted at any point and the resulting clusters

in the network are taken to be the communities. Numbers of node-independent paths, edge-independent paths or paths that run between the vertices are some of the examples of metrics considered in a hierarchical agglomerative algorithm. These algorithms are good at finding strongly connected cores of communities [10]. As the edges are added, a tree from an empty set of vertices is formed, which is represented by a Dendrogram.

Divisive Algorithms detects the links that join nodes of different communities and removes them in such a way that communities get disconnected from each other. In divisive algorithms, we start with the full network and iteratively links are removed, dividing the network into smaller and smaller separate networks with passage of time. These smaller networks are classified as communities. The most critical point in the Divisive Algorithms is to decide if the links are at the position of between-communities. Girvan and Newman is the most accepted divisive algorithm which has laid the foundation of a new period in the area of community detection.

(iii) Modularity Based Algorithms: A metric is necessary to measure how well the communities are detected. Newman and Girvan proposed modularity a measure for evaluating the overall quality of a graph partition. Modularity evaluates internal connectivity, with respect to a randomized null model. Modularity based algorithms try to identify the perfect community structure with reference to modularity. Now day's modularity is the most widely used objective function for partitioning and to compare the quality of the partitions obtained by different algorithms. Obtaining high value of modularity has been the main objective of all modularity maximizing algorithms. Modularity is a numerical value which tells how good a particular partition is and is very significant in community detection studies. Initially, all the nodes in the network are considered to be in a community of its own. Then, the pairs of communities are merged which will give a greatest increase or smallest decrease to the modularity value of the network. Modularity optimization is a NP complete problem, so it is not possible to identify the true maximum value of the modularity in polynomial time but then also there are several algorithms which use modularity as an optimization criterion.

4. COMMUNITY BASED RECOMMENDER SYSTEM

These days we are surrounded with networks like social networks, biological networks, technological networks etc. They exist almost everywhere. People form groups within their work place, family and friends (on twitter, Facebook etc). These groups are called communities in which they discuss various issues, share their experiences, gossip etc. People in one community perform same function and share some common interests. They form close trust bond. While shopping, people mostly rely on the recommendations given to them by their friends rather than recommendations from similar but unknown users. So these social interactions can be very helpful in creating accurate predictions in recommender systems. Community detection can help in understanding user's collective behavior. This paper proposes the use of community detection technique in designing a recommender system. Traditional recommender systems don't provide realistic results. The social interaction information can be used in overcoming the weaknesses of traditional recommender systems. Community based recommender systems (CRS) [3] recommend products and services based on the preferences of the user's friends. These recommendations are based on the ratings that are provided by the user's friends. Community based recommender system (CRS) provides more accuracy and can also help in solving cold start and overspecialized recommendation problems. CRS can generate recommendations even if we have little amount of data, thereby overcoming the sparsity issue. Moreover, these recommender systems can also recommend products to a group of users. Such community based recommender systems make use of user relations which are stored in the form of a graph where nodes are users and edges represent relationship between the users. Using community detection technique in recommender system is a two step method. In step one, we use community detection algorithm on the data set to identify communities in which users have similar interests. In the second step, we use particular community recommendations and recommend those to the new user of that community. This method helps in providing a more diverse list of products for the particular customer.

Steps in the proposed Community based recommender system:

(1) Collecting the network dataset for preprocessing.
(2) Applying any community detection algorithm to detect various communities.
(3) Collect recommendations of the community.
(4) Recommending the recommendations of the community to the particular user of that community.

5. Simulation and results

Here we have used Amazon online purchasing network for experimentation. The data set is available at SNAP (network dataset). The nodes in Amazon online purchasing network are customers and edges are present between commonly co-purchased products. We have used CESNA algorithm [11] for finding communities. It combines the information from the node attributes as well as the network for detecting communities. It is a scalable and accurate algorithm for detecting communities in the networks with respect to node attributes. Communities from edge structure and node attribute help to detect communities by identifying the relevant node attributes for the community. Moreover, CESNA also helps to identify the interaction between the node attributes and network structure which in turn detects the communities accurately. CESNA takes linear time in terms of number of edges and attributes. Simulation is done in Network Simulator (NS-2) for social network in a 1000x1000 m2 environment with the simulation time of 200 seconds. The simulated traffic is constant bit rate (CBR). Figure 1 shows the detected communities from Amazon online purchasing network from 0th simulation time to 200th simulation time. Input network split into five groups at the simulation time 50 second (as shown in third snap), after detecting the communities we collect the recommendation of the community and later we use them to recommend the recommendations of the community to the particular user of that community.

6. Conclusion

Recommender systems have been used widely in the information retrieval, forecasting theories etc. Inspite of all the advancements in recommender systems, they still require further development of new algorithms and techniques

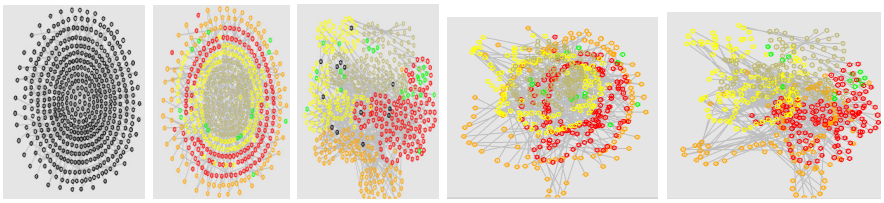

FIGURE 1. Detecting communities in Amazon online purchase network from 0th simulation time to 200th simulation time.

for their better use in the real-life applications: Community detection is one of the most active fields in complex networks because it has many practical applications. Community detection techniques can be used to improve the performance of recommender system. Till date very work has been done in this area. Here we have applied community detection in recommender systems and achieved fruitful results.

REFERENCES

[1] X. SU, T. M. KHOSHGOFTAAR: *A survey of collaborative filtering techniques*, Adv. Artif. Intell., **2009** (2009), 1–19.

[2] R. BURKE: *Hybrid recommender systems: Survey and experiments*, User Model. User-adapt. Interact., **12**(4) (2002), 331–370.

[3] G. ADOMAVICIUS, 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) (2005), 734–749.

[4] A. ARENAS, A. FERNANDEZ, S. GOMEZ: *Analysis of the structure of complex networks at different resolution levels*, New Journal of Physics, **10** (2008), 1–22.

[5] G. KRINGS, D. DABIN, V. BLONDEL: *Communities in a crime network*, Proceedings of NetSci2011 International School and Conference on Network Science, Budapest (Hungary), 2011.

[6] C. YU, S. LIN: *Fuzzy rule optimization for online auction frauds detection based on genetic algorithm*, Electronic Commerce Research, **13**(2) (2013), 169–182.

[7] D. HU, P. RONHOVDE, Z. NUSSINOV: *Replica inference approach to unsupervised multiscale image segmentation*, Phys Rev E Stat Nonlin Soft Matter Phys., **85** (2012), 1–26.

[8] S. FORTUNATO: *Community detection in graphs*, Physics Reports, **486**(3-5) (2010), 75–174.

[9] A. POTHEN: *Graph partitioning algorithms with applications to scientific computing*, Technical report, National Science Foundation, 1997.

[10] M. E. J. NEWMAN: *Finding and evaluating community structure in networks*, Phys. Rev. E., **69**(2) (2004), 1–15.

[11] J. YANG, J. MCAULEY, J. LESKOVEC: *Community Detection in Networks with Node Attributes*, arXiv preprint, (2014), 1–10.

[12] M. BALABANOVIC: *Fab: content-based, collaborative recommendation*, Commun. ACM, **40**(3) (1997), 66–72.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GURU NANAK DEV UNIVERSITY REGIONAL CAMPUS
ADDRESS
Email address: miniahuja06@gmail.com