

## A Community Detection and Recommendation System

Dipika Deshmukh<sup>1</sup>, Dr. D. R. Ingle<sup>2</sup>

<sup>1</sup>ME(Comp.) Student of Bharati Vidyapeeth College of Engineering, Kharghar, Navi Mumbai, India

<sup>2</sup>Professor, Computer Dept., Bharati Vidyapeeth College of Engineering, Kharghar, India

\*\*\*

**Abstract** - Recommendation systems play an important role in suggesting relevant information to users. Community-wise social interactions form a new dimension for recommendations. A social recommendation system using community detection approaches is proposed. I will be using community detection algorithm to extract friendship relations among users by analysing user-user social graph. I will be developing the approach using MapReduce framework. This approach will improve scalability, coverage and cold start issue of collaborative filtering based recommendation system.

**Key Words:** Community Detection, Recommendation System; Cold-start; E-commerce;

### 1. INTRODUCTION

E-Commerce sites are gaining popularity across the world. People visit them not just to shop products but also to know the opinion of other buyers and users of products. Online customer reviews are helping consumers to decide which products to buy and also companies to understand the buying behaviour of consumers. Collaboration, interaction and information sharing are the main driving forces of the current generation of web applications referred to as Web 2.0. Well-known examples of this emerging trend include weblogs (online diaries or journals for sharing ideas instantly), Friend-Of-A-Friend (FOAF) files (machine-readable documents describing basic properties of a person, including links between the person and objects / people they interact with), wikis (web applications such as Wikipedia that allow people to add and edit content collectively) and social networking sites (virtual communities where people with common interests can interact, such as Facebook, dating sites, car addict forums, etc.). We focus on one specific set of Web 2.0 applications, namely social recommender systems. These recommender systems generate predictions (recommendations) that are based on information about users' profiles and relationships between users. Nowadays, such online relationships can be found virtually everywhere, think for instance of the very popular social networking sites Facebook, LinkedIn and MSN. Research has pointed out that people tend to rely more on recommendations from people they trust (friends) than on online recommender systems which generate recommendations based on anonymous people similar to them. This observation, combined with the growing popularity of open social networks and the trend to integrate e-commerce applications with recommender systems, has generated a rising interest in trust-enhanced recommendation systems. The recommendations generated

by these systems are based on information coming from an (online) social network which expresses how much the members of the community trust each other. Augmenting a recommender system by including trust relations can help solving the sparsity problem. Moreover, a trust-enhanced system also alleviates the cold start problem: it has been shown that by issuing a few trust statements, compared to a same amount of rating information, the system can generate more, and more accurate, recommendations.

### 2. LITERATURE SURVEY

Recommender Systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user [2]. The suggestions relate to various decision-making processes, such as what items to buy, what music to listen to, or what online news to read. „Item“ is the general term used to denote what the system recommends to users. A RS normally focuses on a specific type of item (e.g., CDs, or news) and accordingly its design, its graphical user interface, and the core recommendation technique used to generate the recommendations are all customized to provide useful and effective suggestions for that specific type of item. RSs are primarily directed towards individuals who lack sufficient personal experience or competence to evaluate the potentially overwhelming number of alternative items that a Web site, for example, may offer. A case in point is a book recommender system that assists users to select a book to read. In the popular Website, Amazon.com, the site employs a RS to personalize the online store for each customer. Since recommendations are usually confidence  $c$  in the transaction set  $D$  if  $c$  is the percentage of transactions in  $D$  containing  $A$  that also personalized, different users or user groups receive diverse suggestions. In addition there are also non-personalized recommendations. These are much simpler to generate and are normally featured in magazines or newspapers. Typical examples include the top ten selections of books, CDs etc. While they may be useful and effective in certain situations, these types of non-personalized recommendations are not typically addressed by RS research. RS can play a range of possible roles. First of all, we must distinguish between the roles played by the RS on behalf of the service provider from that of the user of the RS. In fact, there are various reasons as to why service providers may want to exploit this technology:

- Increase the number of items sold.
- Sell more diverse items.
- Increase the user satisfaction.

- Increase user fidelity.
  - Better understand what the user wants.
- Herlocker et al. [2], in a paper that has become a classical reference in this field, define eleven popular tasks that a RS can assist in implementing. Some may be considered as the main or core tasks that are normally associated with a RS, i.e., to offer suggestions for items that may be useful to a user. Others might be considered as more “opportunistic” ways to exploit a RS. As a matter of fact, this task differentiation is very similar to what happens with a search engine, Its primary function is to locate documents that are relevant to the user’s information need, but it can also be used to check the importance of a Web page or to discover the various usages of a word in a collection of documents.
- Find Some Good Items
  - Find all good items
  - Annotation in context
  - Recommend a sequence
  - Recommend a bundle
  - Just browsing
  - Find credible recommender
  - Improve the profile
  - Express self
  - Help others
  - Influence others

As these various points indicate, the role of a RS within an information system can be quite diverse. This diversity calls for the exploitation of a range of different knowledge sources and techniques used to identify the right recommendations. Several different types of recommender systems that vary in terms of the addressed domain, the knowledge used, but especially in regard to the recommendation algorithm, i.e., how the prediction of the utility of a recommendation is made. Other differences relate to how the recommendations are finally assembled and presented to the user in response to user requests. A taxonomy provided by [2] that has become a classical way of distinguishing between recommender systems and referring to them. [2] Distinguishes between six different classes of recommendation approaches:

### 2.1 Content-based Recommender Systems

The system learns to recommend items that are similar to the ones that the user liked in the past. The similarity of items is calculated based on the features associated with the compared items. For example, if a user has positively rated a movie that belongs to the comedy genre, then the system can learn to recommend other movies from this genre.

### 2.2 Collaborative filtering based Recommender Systems

The simplest and original implementation of this approach recommends to the active user the items that other users with similar tastes liked in the past. The similarity in taste of two users is calculated based on the similarity in the rating history of the users. This is the reason why collaborative filtering is referred to as “people-to-people correlation.” Collaborative filtering is considered to be the most popular and widely implemented technique in RS.

### 2.3 Hybrid Recommender Systems

These RSs are based on the combination of the mentioned techniques. A hybrid system combining techniques A and B tries to use the advantages of A to fix the disadvantages of B. For instance, CF methods suffer from new-item problems, i.e., they cannot recommend items that have no ratings. This does not limit content-based approaches since the prediction for new items is based on their description (features) that are typically easily available. Given two (or more) basic RSs techniques, several ways have been proposed for combining them to create a new hybrid system [2].

### 2.4 Demographic Recommender Systems

This type of system recommends items based on the demographic profile of the user. The assumption is that different recommendations should be generated for different demographic niches. Many Web sites adopt simple and effective personalization solutions based on demographics. For example, users are dispatched to particular Web sites based on their language or country. Or suggestions may be customized according to the age of the user. While these approaches have been quite popular in the marketing literature, there has been relatively little proper RS research into demographic systems.

### 2.5 Knowledge-based Recommender Systems

Knowledge-based systems recommend items based on specific domain knowledge about how certain item features meet users needs and preferences and, ultimately, how the item is useful for the user. Notable knowledge based recommender systems are case-based [2]. In these systems a similarity function estimates how much the user needs (problem description) match the recommendations (solutions of the problem). Here the similarity score can be directly interpreted as the utility of the recommendation for the user. Knowledge-based systems tend to work better than others at the beginning of their deployment but if they are not equipped with learning components they may be surpassed by other shallow methods that can exploit the logs of the human/computer interaction (as in CF).

## 2.6 Community-based Recommender Systems

This type of system recommends items based on the preferences of the users friends. This technique follows the epigram "Tell me who your friends are, and I will tell you who you are" [13, 14]. Evidence suggests that people tend to rely more on recommendations from their friends than on recommendations from similar but anonymous individuals [10]. This observation, combined with the growing popularity of open social networks, is generating a rising interest in community-based systems or, as or as they usually referred to, social recommender systems [2]. This type of RSs models and acquires information about the social relations of the users and the preferences of the user's friends. The recommendation is based on ratings that were provided by the user's friends. In fact these RSs are following the rise of social-networks and enable a simple and comprehensive acquisition of data related to the social relations of the users. The research in this area is still in its early phase and results about the systems performance are mixed. For example, overall social-network based recommendations are no more accurate than those derived from traditional CF approaches, except in special cases, such as when user ratings of a specific item are highly varied (i.e. controversial items) or for cold-start situations, i.e., where the users did not provide enough ratings to compute similarity to other users. Others have showed that in some cases social-network data yields better recommendations than profile similarity data and that adding social network data to traditional CF improves recommendation results [2].

## 3. Cold Start Problem and Community Detection in Recommendation Systems

The "Cold start" problem [8] happens in recommendation systems due to the lack of information, on users or items. Usage-based recommendation systems work based on the similarity of taste of user to other users and content based recommendations take into account the similarity of items user has been consumed to other existing items. When a user is a newcomer in a system or he/she has not yet rated enough number of items. So, there is not enough evidence for the recommendation system to build the user profile based on his/her taste and the user profile will not be comparable to other users or items. As a result, the recommendation system cannot recommend any items to such a user. Regarding the cold start problem for items, when an item is new in the usage based recommendation systems; no users have rated that item. So, it does not exist in any user profile. Since in collaborative filtering the items consumed in similar user profiles are recommended to the user, this new item cannot be considered for recommendation to anyone. Here, we concentrate on cold start problem for new users. We propose that if a user is new in one system, but has a history in another system, we can use his/her external profile to recommend relevant items, in the new system, to this user.

## 4. Importance Community detection techniques

Past research has shown that corporations benefit from using community-based recommender systems. Through them, they create digitized word-of-mouth that helps consumers make purchase decisions. Since the advent of the Internet, organizations have increasingly recognized its possibilities as a primary medium for advertising their products and services, and for supporting new and effective means of communication with consumer. Emblematic of this recognition, Amazon.com, which has mostly focused on the Internet, eliminated its entire budget for television and print advertising in 2003 [15]. The firm's management team came to believe it is better served by digitized word-of-mouth. Digitized word-of-mouth has become an important source of information to consumers and firms, allowing consumers to easily share their opinions and experiences about the quality of various products and sellers [15]. A community-based recommender system (or simply, a recommender system in this research) is a system that makes use of digitized word-of-mouth to build up a community of individuals who share personal opinions and experiences related to their recommendations for products and seller reputations [15]. Such systems present or aggregate user-generated opinions and ratings in an organized format. Consumers consult this information before making purchase decisions.

## 5. Community detection techniques

Community detection techniques aim to find subgroups where the amount of interactions inside the group is more than the interaction outside it, and this can help to understand the collective behaviour of users. The community identification process depends on the nature of networks, either static or dynamic. Static networks are basically constructed by aggregating all observed interactions over a period of time and representing it as a single graph. Dynamic networks, also called time varying graphs can be either a set of independent snapshots taken at different time steps or a temporal network that represents sequences of structural modifications over time [16]. In what follows, we present both static and dynamic community detection.

### 5.1 Static community detection

Panoply of community detection algorithms exists in literature. The first idea using static networks was proposed by Girvan and Newman. It is based on a modularity function representing a stopping criterion, aiming to obtain the optimum partitioning of communities. In the same context, Guillaume et al. have proposed Louvain algorithm to detect communities using the greedy optimization principle that attempts to optimize the gain of modularity. Rosvall and Bergstrom have presented Infomap, considered as a solution to the simplest problem of static and non-overlapping community detection. The mentioned algorithms are not

able to detect overlapping communities where a node can belong to more than one community in the same time. To ensure this basic property, Palla et al. have proposed the Clique-Percolation Method (CPM) to extract communities based on finding all possible k-cliques in the graph. This method requires the size of the cliques in input.

## 5.2 Dynamic community detection

Several researchers explored the dynamic aspect of networks to identify community's structure and their development over time. Hopcroft et al. have proposed the first work on dynamic community detection which consists in decomposing the dynamic network into a set of snapshots where each snapshot corresponds to a single point of time. The authors applied an agglomerative hierarchical method to detect communities in each snapshot and then they matched these extracted ones in order to track their evolution over time. Palla et al. have used the (CPM) method of static community detection to extract communities from different snapshots. Then, they tried to look for a matching link between them to detect their structural changes over time. Methods applying static algorithms on snapshots cannot cover the real evolution of community's structures over time because it seems harder for these methods to recognize the same community from two different time steps of network. To overcome this problem, new studies have exploited another representation of data that takes into account all temporal changes of the network in the same graph. We cite, in particular, the intrinsic Longitudinal Community Detection (iLCD) algorithm proposed by Cazabet et al. The algorithm uses a longitudinal detection of communities in the whole network presented in form of a succession of structural changes. Its basic idea was inspired from multi-agent systems. In the same context, Nguyen et al. have proposed AFOCS algorithm to detect overlapping communities in a dynamic network  $N$  composed of the input network structure  $N_0$  and a set of network topology changes  $\{N_1, N_2, \dots, N_n\}$ .

## 6. Related Works

There has recently been much research on merging community detection and recommender systems in order to provide more personalized recommendations related to users belonging to the same community. In fact, community-based recommendation is a two-step approach. The first step consists in identifying groups in which users should share similar properties and the second step uses the community into which the target user pertains to recommend new items. Using static community detection algorithms, Kamahara et al. have proposed a community-based approach for recommender systems which can reveal unexpected user's interests based on a clustering model and a hybrid recommendation approach. In the same context, Qin et al. have applied CPM method on the YouTube Recommendation Network of reviewers to detect communities of videos.

These latter are used to provide the target user by local a recommendation which consists in recommending videos pertaining to the same community of the video watched by him. This approach aims to propose a more diverse list of items for target user. Another aim behind incorporating community detection to recommendation is to provide a solution to the cold start problem, and this idea was proposed by Sahebi et al. while applying Principal Modularity Maximisation method to extract communities from different dimensions of social networks. Based on these latent communities of users, the recommender system is able to propose relevant recommendations for new users. Qiang et al. have defined a new method of personalized recommendation based on multi-label Propagation algorithm for static community detection. The idea consists in using the overlapping community structures to recommend items using collaborative filtering. More recently, Zhao et al. proposed the Community-based Matrix Factorization (CB-MF) method based on communities extracted using Latent Dirichlet Allocation method (LDA) on twitter social networks. In [1], authors focused on community-based recommendation of both individuals and groups. They used the Louvain community detection method on the social network of movies building from the Internet Movie Database (IMDb) in order to provide personalized recommendations based on the constructed communities. These methods only deal with static networks, derived from aggregating data over all time, or taken at a particular time. The accumulation of an important mass of data in the same time and in the same graph can lead to illegible graphs, not able to deal with the dynamic aspect of real-world networks. To take into account the evolution of user's behaviors' over time using a kind of community-based dynamic recommendation, a first attempt was established by Lin et al. The main idea consists in providing a dynamic user modelling method to make recommendations by taking into consideration the dynamic users' patterns and the users' communities. This approach is limited since it uses a manual method to identify communities, which is not efficient especially when we deal with strongly evolving and large networks. More recently, About et al. proposed to use the fuzzy k-means clustering from time to time to dynamically detect the users' interests over time. Then, they exploited these formed communities to determine user's preference for new items with regard to the updated users' ratings. Another author proposed an article recommender system to recommend documents for users based on the same members of communities which are identified according to their interests while browsing the web. The detection of users' interests is repeated continuously in subsequent time intervals in order to deal with the dynamic aspect of new portals. In both the previously presented methods, applying clustering techniques for time to time cannot cover the real evolution of community structure over time. In fact, several structural changes may occur and get lost without being detected. Besides, the temporal complexity of these methods increases in large networks. Aiming to benefit from the

whole advantages of the community detection process as part of recommender systems, we propose an architecture allowing ensuring this combination.

## 7. PROPOSED SYSTEM

### 7.1 Community Detection

With the growth of social network web sites, the number of subjects within these networks has been growing rapidly. Community detection in social media analysis helps to understand more of users' collective behaviour. The community detection techniques aim to find subgroups among subjects such that the amount of interaction within group is more than the interaction outside it. Multiple statistical and graph-based methods have been used recently for the community detection purposes. Bayesian generative models, graph clustering approaches, hierarchical clustering, and modularity-based methods are a few examples. By considering one of these dimensions, e.g. connections network, we propose a system to detect communities using Bron-Kerbosch algorithm.

### 7.2 Modules in the Proposed System

#### Module 1 - Upload of Dataset

In this module Administrator will upload the dataset which will be analyzed further for data preprocessing and community detection mechanism. The administrator will also upload file containing users and their favorite movies and another file containing movie links.

#### Module 2 - Community Analysis

In this module Input Attributes from data will be calculated for preprocessing and values related to community detection will be stored.

#### Module 3 - Community Detection

In this module we are going to detect the communities according to the data given in dataset using neighboring node linkage analysis implemented using MapReduce. The identified communities will be displayed to user based on above processing.

#### Module 4 - Community and Movie Links Recommendation

In this module we are going to recommend a community to a particular user and based on the members in the community and their movie preferences, movie links will be recommended to that particular user who belongs to the same community.

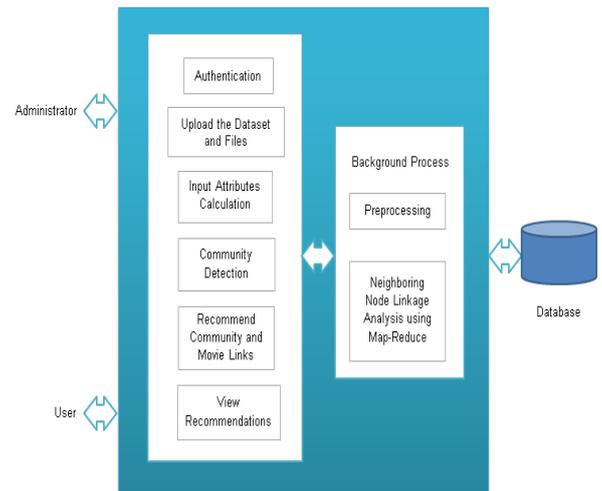


Figure 1: System Architecture

## 8. RESULTS AND DISCUSSION

### A. System Modules - For YouTube Dataset

- Upload YouTube Dataset File, User Favourite Movie File, Movie Links File
- View YouTube Dataset File
- View YouTube Community Recommendations
- YouTube Community Recommendation for a particular user

### B. For Facebook Dataset

- Upload Facebook Dataset File
- View Facebook Dataset File
- View Facebook Community Recommendations
- Facebook Community Recommendation for a particular user

### C. Datasets Used

There are two Datasets are used in the project:

- YouTube Dataset (consisting of 500 entries)
- Facebook Dataset (consisting of 700 entries)

### D. Community Detection

- We over here taken a YouTube dataset and we will detect a community present in dataset using Bron-Kerbosch algorithm (without pivot).

## 9. CONCLUSION AND FUTURE SCOPE

A convincing quantitative indicator of recommender system effectiveness is the corresponding sales of the recommended products. The premise is that by acquiring additional information from other consumers, the uncertainty discount

associated with a sale item can be reduced. This impacts consumer decisions, increasing the likelihood of making a sale. On merging community detection and recommender systems, we aspire to provide more personalized recommendations related to users belonging to the same community. In fact, community-based recommendation is a two-step approach. The first step consists in identifying groups in which users should share similar properties and the second step uses the community into which the target user pertains to recommend new items. Another aim behind incorporating community detection to recommendation is to provide a solution to the cold start problem. The network formed connects all the users in the system, from which we form communities based on their friends. We attempt to form user communities by incorporating friendship relations and recommend movies based on collective interests. People from the same community are provided with the same recommendations so that new users can be assigned a community profile and benefit from the experience of other users. Recommender systems are important in strategic marketing. They modelled how firms should take advantage of community-based recommender systems by combining them with pricing and conventional advertising strategies. Recommender systems are also beneficial to attract and retain users. A careful examination of the relationship between user loyalty and recommender systems in an empirical study of data from Amazon.com suggests that providing consumer reviews increases perceived usefulness and the quality of a user's psychological connection with a web site. This, in turn, influences consumer attraction and customer retention.

## 10. ACKNOWLEDGEMENT

I wish to express my sincere thanks and deep sense of gratitude to respected mentor and guide Dr. D.R.Ingle, Assistant Professor BVCOE for his depth and enlightening support with all of his kindness. His constant guidance and willingness to share his vast knowledge made me understand this topic and its manifestations in great depths.

## 11. REFERENCES

- [1] DeepikaLalwani, D. V. L. N. Somayajulu, P. Radha Krishna, "A community driven social recommendation system", IEEE International Conference on Big Data (Big Data), Year: 2015, Pages: 821-826, DOI: 10.1109/BigData.2015.7363828.
- [2] Francesco Ricci, LiorRokach, BrachaShapira, Paul B. Kantor., "Recommender Systems Handbook", Vol. 1, New York: Springer, 2011, ISBN 978-0-387-85819-7, e-ISBN 978-0-387-85820-3, DOI 10.1007/978-0-387-85820-3.
- [3] Venkata Rajeev P., SmrithiRekha V., "Recommending Products to Customers using Opinion Mining of Online Product Reviews and Features", 2015 Int. Conf. on Circuit, Power and Computing Technologies [ICCPCT].
- [4] AymenElkhelifi, Firas Ben Kharrat, Rim Faiz, "Recommendation Systems Based on Online User's Action", DOI 10.1109/CIT/IUCC/DASC/PICOM 2015.
- [5] AkshayChadha, PreetiKaur, "Comparitive Analysis of Recommendation System", 2015 4th Int. Symposium on Emerging Trends and Technologies in Libraries and Information Services.
- [6] S. Chen, S. Owusu and L. Zhou, "Social Network Based Recommendation Systems: A Short Survey", 2013 Int. Conf. on Social Computing (SocialCom), pp. 882-885, 2013.
- [7] M. Fatemi and L. Tokarchuk, "A Community Based Social Recommender System for Individuals & Groups", IEEE Int. Conf. on Social Computing, pp. 351-356, 2013.
- [8] S. Sahebi and W. Cohen, "Community Based Recommendations: a Solution to the Cold Start Problem", RSWEB Workshop, 2011.
- [9] J. Tang, X. Hu and H. Liu, "Social Recommendation: A Review", Social Network Analysis and Mining, vol. 3(4), pp. 1113-1133, 2013.
- [10] Sinha R., Swearingen, K.: Comparing recommendations made by online systems and friends. Proc. of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries (2001).
- [11] Mahmood, T., Ricci, F.: Improving recommender systems with adaptive conversational strategies. In: C. Cattuto, G. Ruffo, F. Menczer (eds.) Hypertext, pp. 73-82. ACM (2009).
- [12] Resnick, P., Varian, H.R.: Recommender systems. Communications of the ACM 40(3), 56-58 (1997).
- [13] Arazy, O., Kumar, N., Shapira, B.: Improving social recommender systems. IT Professional 11(4), 38-44 (2009).
- [14] Ben-Shimon, D., Tsikinovsky, A., Rokach, L., Meisels, A., Shani, G., Naamani, L.: Recommender system from personal social networks. In: K. Wegrzyn-Wolska, P.S. Szczepaniak (eds.) AWIC, Advances in Soft Computing, vol. 43, pp. 47-55. Springer (2007).
- [15] Pei-Yu Chen, Yen Chun Chou, Robert J. Kauffman, "Community-Based Recommender Systems: Analyzing Business Models from a System Operator's Perspective", Proceedings of the 42nd Hawaii International Conference on System Sciences - 2009, 978-0-7695-3450-3/09 © 2009 IEEE.
- [16] Sabrina Ben Abdrabbah, RaouiaAyachi and Nahla Ben Amor, "Collaborative Filtering based on Dynamic Community Detection".