A Study on Clustering Techniques in Recommender system

Kamna Sharma, Taruna Sharma
Computer Science, Kurukshetra Institute of Technology and Management, Kurukshetra

***Abstract*- A recommender system produces a list of suggestions for users based on their preferences. It is an intelligent system that can help users to come across their interesting items. Recommender systems have attracted increasing attention lately due to the booming applications in online advertising. Recent studies demonstrate that information from social networks can be exploited to improve accuracy of recommendations. For organizing information, the recommender system incorporates data mining techniques into their recommendations using knowledge learned from the actions and attributes of the users. In this paper we study different Clustering techniques used in users profiling which enables us to get a quality recommendation.**

***Index Terms*- recommender system, collaborative filtering, similarity, Clustering, K-Means, SOM.**

I. INTRODUCTION

Recommender systems are software tools and techniques providing suggestions for items to be of use to a user. With the development of web 2.0, it had found a wide application in both social media websites (e.g. Youtube, Facebook, Myspace etc) and commercial applications. However the accuracy of prediction remains to be further improved [1]. Accurate recommendations enable users to quickly locate desirable items without being overwhelmed by irrelevant information. It is also of great interest for vendors to recommend those products that match the interest of each of the visitors of their websites and hopefully turn them into satisfied and returning customers.

Obtaining recommendations from trusted sources is a critical component of the natural process of human decision making. With burgeoning consumerism buoyed by the emergence of the web, buyers ae being presented with an increasing range of choices while sellers are being faced with the challenge of personalizing their advertising efforts. In parallel, it has become common for enterprises to collect large volumes of transactional data that allows for deeper analysis of how a customer base interacts with the space of product offering. Recommender systems have evolved to fulfill the natural dual deal of buyers and sellers by automating the generation of recommendations based on data analysis. Recommender system roots back to several related research disciplines, such as Cognitive science, approximation theory and information retrieval etc. Due to the increasing importance of recommendation, it has become an independent research field since the mid 1990s .

II. BACKGROUND WORK RECOMMENDER SYSTEM

The goal of a Recommender System is to generate meaningful recommendations to a collection of users for items or products that might interest them. Suggestions for books on Amazon, or movies on Netflix, are real world examples of the operation of industry strength recommender systems [4]. Recommender systems are classified into three categories [2][3] on the basis of how recommendations are made:

1. Content Based Recommender System
2. Collaborative Filtering Based Recommender System
3. Hybrid Recommender System

. In content based approach, the users get recommended items similar to the ones preferred in the past. The system recommend items to users based on correlation between the content of items and the user preferencs [6]. In

collaborative approach, users get recommended items that people with similar tastes and preferences liked in the past [7]. The third category of recommender system combines both collaborative filtering and content based approaches [7].

III. RELATED RESEARCHES IN CLUSTERING USERS/ITEMS IN RECOMMENDER SYSTEMS

In this section, we discuss significant work in the areas of Clustering techniques in recommender systems.

Clustering is a division of data into groups of similar objects. Each group, called Cluster, consist of objects that are similar between themselves and dissimilar to objects of other groups. Representing data by fewer clusters necessarily loses certain fine details, but achieves simplification. Clustering is often one of the first steps in data mining. It identifies groups of related records that can be used as a starting point for exploring further relations. For example, the buying habits of multiple population segments might be compared to determine which segments to target for a new sales campaign.

Collaborative Filtering system is the most powerful recommendation technique and used in various web applications like in recommending books, movies etc. Despite its wide spread adoption, it suffers from various limitations, including sparsity, scalability and problem in predicting the rating for the yet unrated item [sarwar et al 2000].

Clustering methods for CF have been extensively studied by several studies. Ungar [Ungar and Foster, 1998] proposed a repeated K-means and Gibb sampling clustering techniques that group users into clusters with similar items and group items into clusters which tend to be liked by the same users [8]. Kohrs [Kohrs and Merialdo, 1999] used a hierarchical clustering algorithm to independently cluster users and items into two cluster hierarchies. The recommendation is made by the weighted sum of the defined centers of all nodes in the cluster hierarchies on the path from the roots to the particular leaves [9]. Sarwar et al [Sarwar et al, 2002] evaluated a clustering approach that showed, clustering provides comparable recommendation quality as traditional CF, while significantly improving the online performance [10].

The research work [George and Merugu,2005., xue et al,2005, Truong et al,2007., nathanson et al,2007., Rashid et al,2006., Zhang and Hurley,2009], clusters users and items using the rating data while ignoring the additional information, for example the relationship between items [11] [12] [13] [14] [15] [16]. Kavitha et al [17] the authors have used Self Organizing Map (SOM) which is a clustering algorithm based on unsupervised neural network model. SOM is based on two layers, an input layer and computation layer. This has a feed-forward structure with a single computational layers of neurons arranged in rows and columns. Each neuron is fully connected to all the source units in the input layer. The stages of the SOM can be summaries as follows:

1. Initialization – Choose random values for the initial weight vectors **w**j.

2. **Sampling** – Draw a sample training input vector **x** from the input space.

3. **Matching** – Find the winning neuron I(**x**) that has weight vector closest to the

input vector, i.e. the minimum value of dj i xi wji

D (**x**) = (-) = å 2

4. **Updating** – Apply the weight update equation Dwji t Tj I t xi wji =h( ) , (**x**)( ) (-)

where Tj,I (**x**)(t) is a Gaussian neighborhood and h(t) is the learning rate.

5. Continuation – keep returning to step 2 until the feature map stops changing.

Kyounge-jae kim [18] presented Genetic Algorithm based K-Means algorithm. K-Means clustering is a method of cluster analysis which aims at portioning of n observations into k clusters. Each of the observation belongs to a cluster with the minium distance between cluster centre and the observation point. It uses an iterative hill climbing method. The process of genetics based K-Means is as follows:

1. The initial seeds with the chosen no. of clusters, K, are selected and an initial partition is built by using the seeds as the centroids of initial clusters.
2. Each record is assigned to the centroid that is nearest, thus forming a cluster.
3. Keeping the same no. of clusters, the new centroid of each cluster is calculated.
4. Iterate step (2) and (3) until the clusters stop changing or stop conditions.

Fig-1 shows the whole framework of GA K-Means clustering.

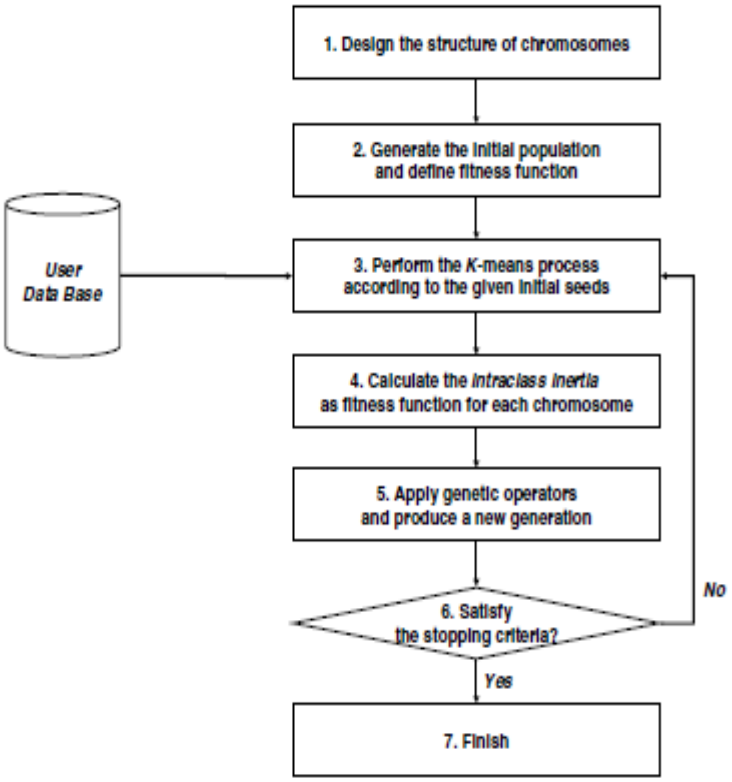

Fig. 1. The overall framework of GA *K*-means.

The study suggests that GA K-Means clustering performed better than the SOM, K-Means in terms of the segmentation.

IV. RELATE RESEARCHES IN CLUSTERING WEB USERS IN RECOMMENDER SYSTEM

In this section, we discuss significant work in the areas of clustering techniques in web usage mining based recommender system.

Web Usage Mining (WUM) is the process of extracting knowledge from web user's access data by exploiting data mining technologies. It can be used for different purposes such as personalization, recommendation system improvement etc.

Georgios Paliourus et al [19] have used two clustering methods namely Auto class and cluster mining [Perkowitze et al].

a) Autoclass

Autoclass is a general purpose clustering algorithm developed by Bayesian Learning group since the 1980. It is an unsupervised Bayesian classification system based upon the finite mixture model supplemented by Bayesian method for determining the optimal

class. The algorithm considers that each class Cj has ita own probability distribution Tj. The model deals with two main probabilities : 1)The interclass probability of an instance Xi being the member of class Cj, P(Xi € Cj) and 2)The class probability of observing the instance attribute values Xik conditional on the assumption that Xi is a member of Cj,P(Xik| Xi € Cj).

b) Cluster Mining

Cluster ming algorithm is a simple graph based clustering method. It discovers patterns of common behavior, by looking all fully connected sub graphs of a graph that represents the user's characteristics attributes.

Santosh Rangarajan Et al [20] presented a novel approach to group users according to their Web access patterns. They used ART 1 neural network for grouping users. There results shows that ART 1 Clustering algorithm performs better than the K means algorithm.

Partitioning Algorithm

Mehrdad Jalali Et al[22] advanced an architecture presented in [21] by using Graph partitioning algorithm .which finds groups of strongly correlated pages by partitioning the graph according to its connected components. The Experimental results showed that quality of clustering is better than K means.

Norwati Mustapha Et al [23] used EM clustering algorithm for mining user modeling web navigation pattern behavior. EM algorithm is based on finding maximum likelihood estimates of parameters in probabilistic models where the model depends on unobserved latent variables. The experimental results showed that by decreasing the number of clusters, the log likelihood converges towards the lower values and probability of the largest cluster will be decreased while the number of the clusters increases in each treatment.

Choochart Haruechaiyasak [24] proposed a dynamic frame work for maintaining customer profiles in E commerce recommender system. They used hierarchical clustering algorithm for maintaining customer profiles. Since recommender systems suffer from scalability they have proposed a dynamic way of maintaining customer profiles.

V. CONCLUSION

Typical recommender systems suffer from poor scalability and the lack of ability to handle dynamic changes in the user profiles. In this paper, we have discussed different clustering technique s which helps in maintaining customer profiles dynamically. Future work would be to find out the best of these algorithms.

REFERENCES

[1] Bell R., Bennett J., Koren Y. et al. The million dollar programming prize . Spectrum IEEE, 2009. 46(5): 28-33.

[2] G. Adomavicius and A. Tuzhilin. "Toward the Next Generation of Recommender
Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, June
2005, doi:10.1109/TKDE.2005.99

[3] F. Ricci, L. Rokach, B. Shapira and P. B. (Eds.) Kantor "Recommender Systems Handbook", 1st Edition., 2011, XXX, 842 p. 20 illus.

[4] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. on Knowl. and Data Eng., 17(6):734–749, 2005.

[5] X. Su and T. M. Khoshgoftaar. "A Survey of Collaborative Filtering Techniques", Advances in Artificial Intelligence Volume 2009 (2009), Article ID 421425, 19 pages doi:10.1155/2009/421425

[6] M. Balabanovic andY. Shoham, "Fab: Content-Based,Collaborative Recommendation", Communications of the ACM, vol. 40, no. 3, pp. 66-72, 1997.

[7] G. Adomavicius andA. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Transaction on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734-749, 2005.

[8] [Ungar and Foster, 1998] Ungar, L. and Foster, D. (1998). Clustering methods for

collaborative filtering. In Proceedings of the Workshop on Recommendation Systems.
AAAI Press, Menlo Park California.
[9] [Kohrs and Merialdo, 1999] Kohrs, A. and Merialdo, B. (1999). Clustering for collaborative filtering applications. In In Proceedings of CIMCA'99. IOS Press.
[10] [Sarwar et al., 2002] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2002). Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering. In Proceedings of the Fifth International Conference on Computer and Information Technology.
[11] George and Merugu, 2005] George, T. and Merugu, S. (2005). A scalable collaborative filtering framework based on co-clustering. In ICDM '05: Proceedings of the Fifth IEEE International Conference on Data Mining, pages 625–628, Washington, DC, USA. IEEE Computer Society.
[12] [Xue et al., 2005] Xue, G.-R., Lin, C., Yang, Q., Xi, W., Zeng, H.-J., Yu, Y., and Chen, Z. (2005). Scalable collaborative filtering using cluster-based smoothing. In SIGIR
'05: Proceedings of the 28th annual international ACM SIGIR conference on Researchand development in information retrieval, pages 114–121, New York, NY, USA. ACM.
[13] [Truong et al., 2007] Truong, K., Ishikawa, F., and Honiden, S. (2007). Improving
accuracy of recommender system by item clustering. IEICE - Trans. Inf. Syst., E90-D (9):1363–1373.
[14] [Nathanson et al., 2007] Nathanson, T., Bitton, E., and Goldberg, K. (2007). Eigentaste 5.0: constant-time adaptability in a recommender system using item clustering.
In RecSys '07: Proceedings of the 2007 ACM conference on Recommender systems, pages 149–152, New York, NY, USA. ACM.
[15] [Rashid et al., 2006] Rashid, A. M., Shyong, Karypis, G., and Riedl, J. (2006).
Clustknn: A highly scalable hybrid model- & memory-based cf algorithm. In WEBKDD2006, Philadelphia, Pennsylvania, USA.
[16] [Zhang and Hurley, 2009] Zhang, M. and Hurley, N. (2009). Novel item recommendation by user profile partitioning. In WI-IAT '09: Proceedings of the 2009
IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, pages 508–515, Washington, DC, USA. IEEE Computer Society.
[17] M K Kavitha Devi, P Venkatesh "Kernel Based Collaborative Recommneder Systems for E-purchasing" Sadhana Academy Proceedings in Engineering Sciences Vol 35 Part 5 October 2010 Pg 513-524.
[18] Kyoung-jae Kim,Hyunchal Ahn " A Recommender System using GA Kmeans clustering in an online shopping market", Elsevier ,2007 doi:10.1016/j.eswa.2006.12.025
[19] Geogious Paliouras, Christos Papatheodorou " Clustering the Users of large web sites into communities" Data & Knowledge Engineering, Elsevier, 2003, pp.304-330.
[20] Santosh K. Rangarajan, VirV. Phoha, Kiran Balagani, S. S. Iyengar "Web User Clustering and Its Application to Prefetching Using ART Neural Networks" citiseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.138.497
[21] R.Baraglia, F.Silvestri," An online recommender system for large Web sites", Web Intelligence, IEEE/WIC/ACM, 2004, pp. 20–24.
[22] Mehrdad Jalali, Norwati Mustapha, Md. Nasir B Sulaiman, Ali Mamat, "A Web Usage Mining Approach Based on LCS Algorithm in Online Predicting Recommendation Systems" 12th International Conference Information Visualisation IEEE 2008.
[23] Norwati Mustapha Et al "Expectation Maximization Clustering Algorithm for User Modeling in Web Usage Mining Systems", European Journal of Scientific Research ISSN 1450-216X Vol.32 No.4 (2009), pp.467-476
[24]Choochart Haruechaiyasak, Chatchawal Tipnoe, Sarawoot Kongyoung, Chaianun Damrongrat. Niran Angkawattanawit "A Dynamic Framework for Maintaining Customer Profiles in E-Commerce Recommender Systems" Proceedings published in IEEE International Conference on E technology, E commerce and E service doi=10.1109/EEE-2005.8