Australian Journal of Basic and Applied Sciences, 8(18) December 2014, Pages: 367-373

AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

An Enhanced Prediction Mechanism for Recommendation System using Power Graph

[1]M.Parvathy, [2]Dr.K Sundarakantham, [3]Dr.S.Geetha and [4]Dr.S.Mercy Shalinie

[1]*Ph.D Research Scholar, Department of Computer Science and Engineering, Thiagarajar College of [2]Engineering, Madurai, India*
[2]*Assistant Professor, Department of Computer Science and Engineering, Thiagarajar College of Engineering, Madurai, India*
[3]*Professor, School of Computer Science and Engineering, VIT-University,Chennai Campus, Chennai, India*
[4]*Associate Professor, Department of Computer Science and Engineering, Thiagarajar College of Engineering, Madurai, India*

ARTICLE INFO

Article history:
Received 19 August 2014
Received in revised form
19 September 2014
Accepted 29 September 2014
Available online 3 December 2014

Keywords:
Power graph analysis, recommendation system, scale free network, rating predictions

ABSTRACT

The unpredictable growth in internet led to a development of an effective tool for coping with information overload. Recommendation system is one of the most widely adopted and perceptible technology for solving information overload issue. Collaborative filtering is the traditional technique which mainly depends on the ratings provided by their neighbors. It is a very complicated task to predict the individual user preference in a large scale network. In this paper, a scale free network model is proposed to convert the complex network into lossless compressed network. Based on the constructed network, features of each and every user are identified from the ratings provided by their direct and indirect neighbors. Using the extracted feature values, recommended rating is predicted using conditional distributions. The proposed model is evaluated experimentally using Movie Lens dataset and promises a better rating prediction for all users in the network.

To Cite This Article: M.Parvathy, Dr.K Sundarakantham, Dr.S.Geetha and Dr.S.Mercy Shalinie., An Enhanced Prediction Mechanism for Recommendation System using Power Graph. ***Aust. J. Basic & Appl. Sci.,*** *8(18): 367-373, 2014*

INTRODUCTION

The massive and fast development of web makes it an inevitable factor for the people in day-to-day life. By making use of various social networking tools, people can attain and share knowledge. This will reduce the time with huge content availability and people are highly induced towards the fanatic online search, discovery and e-commerce sites for their ease shopping. Recommender system is a vital research area to analyze and provide better results for the users to deal with information overload problem(Adomavicius & Tuzhilin, 2005). With the help of recommender systems, people can unearth the pertinent products/items, information and experts to upgrade themselves.

The popular approaches used in recommender systems are collaborative filtering (CF), content-based filtering and hybrid based recommendation system. Content-based filtering technique analyse the content of the items and the ratings the users provided to that items to determine the preferences of the user (Mooney & Roy ,2000 & Billus & Pazzani, 2000). The major disadvantage of content-based filtering is difficult to show the isolation between average and high quality information (Yu *et al.*, 2004). Unlike content-based filtering, collaborative filtering based on the preferences of the users to predict the ratings of unrated items (Sarwar *et al.*, 2001). Moreover, CF technique have pertained applications in many areas such as Grouplens, the first automated CF algorithm (Resnick *et al.*, 1994), Movielens (Dahlen *et al.*, 1998) etc. The existing CF algorithms suffer from the cold start and sparsity problems. Better recommendations can be provided by ensembling social relations with the CF approaches.

Most recommender systems depend on the user preferences and the ratings they provided for items. The interests of the users have to be extracted from multiple heterogeneous data sources. The integration of such heterogeneous data is a difficult task in many applications (Jing Gao *et al.*, 2013). To represent various types of entities as nodes and edges, a heterogeneous graph based model is used (Lee *et al*, 2013). For information retrieval problem, graph based model have been built for operating large scale databases (Xu *et al.*, 2007).

The remaining paper comprises of the related work in this particular area. Proposed power graph based recommendation method is explained in recommended system design, construction of power graph and Features evaluation and Prediction Inference. Experimental results are evaluated using Movie Lens dataset in performance analysis. Finally, the paper is concluded.

Corresponding Author: M.Parvathy, Ph.D Research Scholar, Department of Computer Science and Engineering, Thiagarajar College of Engineering, Madurai, India.
E-mail: parvathydurai2002@gmail.com

Related Work:

Nowadays, searching for the vital information from a massive quantity of data and advancing the new technologies in the advanced web 2.0 era is a challenging task. A way out to solve this issue is the advancement of recommender systems to make this complicated task more efficient and ease. Usually the graph structure is very helpful to represent the large repositories such as digital libraries, web information and large databases (Fouss *et al*., 2007). In this paper, proposed a recommendation scheme that fuses graph based approach with collaborative filtering recommendations. Identifying the individual user preference at any time is a very difficult task because of the growing information and service provisions. To reduce the computational complexity, large scale networks must be converted to smaller networks.

The representation and analysis of complex network is important to extract information how the components in the network interact with one another. The study of complex networks is the ongoing research area in real world networks like computer networks and social networks. In social networks, to identify communities the information about the connectivity between nodes should be analysed. Several algorithms exist to identify the communities in the network (Lancichinetti & Fortunato, 2009). Power Graph Analysis is successfully applied to various domains like bioinformatics, bibliographical databases, and community detection algorithms. This motivated to apply Power Graph analysis to recommendation system for efficient analysis of user-item graph.

Methodology:

A recommendation algorithm model for E-commerce applications by compacting the large scale networks into power graphs is proposed. To alleviate the scalability problem and for efficient recommendation prediction, the power graph for the given social network is computed. The features of the user are evaluated from the power graph using which rating of unrated item for the user is predicted.

Preliminaries of Power graph:

A power graph is a representation of complex networks like online social network without loss of information. Power graph is constructed by clustering the nodes based on their connectivity and neighbourhood similarity. Also this graph preserves the connectivity information between the clusters.

The important property of networks, defined as statistically significant subpatterns of graphs is called as network Motifs. Power graph can be built by utilizing three basic network motifs. They are the star, the clique and the biclique which are the basic forms used for transforming the original graph into a power graph. The set of nodes which have a common neighbour node, star motif, can be represented by a power edge between a regular node and a power node. Clique motif in which an edge exists between every pair of nodes in the set can `be represented as a self loop in power graph. The two set of nodes with an edge between every node in one set and node in other set is called Biclique motif. It can be indicated as a power edge between two power nodes. The original network structure is converted into power graph with maximal compression for efficient visualization.

Recommended system Design:

Our social power graph recommendation system works in the following flow model

- Users in a social networks are represented as a graph structure G=(V,E) where V is the set of nodes which corresponds to user and E is the set of edges that represents the relationship between users.
- The large complex network is converted into a lossless compression power graph based on common interest
- Based on the domain search, identify the features of each node in the power graph
- Upon identifying the feature set of nodes, measure the rating by a set of conditional distributions, missed users rating is predicted by other user’s rating distribution and trust value of the particular user

The overall architecture of the proposed system is shown in Fig.1. The proposed methodology mainly consists of the following three steps: (1) construction of the power graph based on the common interest of the users. (2) Computation of the user features such as their ratings history, impact of their ratings on other users. (3) Prediction of rating for unrated products based on the set of conditional distributions of ratings provided by other users and their trust value.

Construction of Power graph:

Let the user-item network be represented as graph $G = (V, E)$ where V is the set of nodes and $E \in v \times v$ is the set of edges, then a power graph $G^{'} = (V^{'}, E^{'})$ is a graph defined on the power set of nodes $V^{'} \in P(V)$ which are connected by power edges $E^{'} \in V' \times V^{'}$. The semantics of power graphs are two folds: (i) there exists a power edge in G’ if and only if all nodes in first power node is connected to all other nodes in second power node. (ii) if all the nodes in the power node are connected to each other, then it is represented as a self power edge.

Australian Journal of Basic and Applied Sciences, 8(18) December 2014, Pages: 367-373

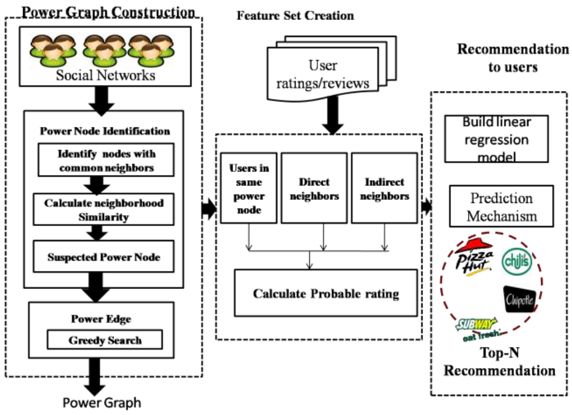

Fig. 1: Overall Architecture of the Proposed System.

Identifying the Power Node:

If any two nodes have neighbours in common then such set of nodes are categorized as power nodes. The power nodes are identified using a hierarchical clustering algorithm based on the jaccard index. To identify the basic motifs, two sets are attached to the power nodes. One set is the neighbourhood set of power node and another set is the common neighbours in neighbourhood set.

Searching the power edge:

The edge reduction to generate the power graph is an optimization problem of searching the power edge. An edge that connects each pair of power nodes is assumed to be an aspirant power edge. From those edges, the one which abstracts many edges are selected as actual power edge to the power graph (Ahnert, 2014).

The power edge which represents the information between two power nodes U and V is defined as

$$PI_{UV} = 2(I_U + I_V)\log_2 T \qquad (1)$$

Where I_U and I_V are the number of nodes in power node U and V and T is the total number of nodes.

All the edges between the nodes present in the two power nodes can be defined with the information requirement as follows

$$I_{UV} = 2I_U I_V \log_2 T \qquad (2)$$

The edges between the nodes in U and V can be reduced into one power edge and amount of information compressed can be defined as

$$CI_{UV} = I_{UV} - PI_{UV} \qquad (3)$$

$$CI_{UV} = 2(I_U I_V - I_U - I_V)\log_2 T \qquad (4)$$

If $CI_{UV} > 0$, then such pair on node sets U and V are said to be compressible component of the graph.

Sample power graph construction from MovieLens dataset:

The Table.1 shows the sample records found in the Movie Lens dataset. Based on the genres, users are categorized and the resulting power graph for original graph is shown in Fig. 2 and Fig.3. Here we consider five categories of genres and the first five users id who have interested in that genre. User u1 is interested in all types of movies and so it is connected all other nodes in the power graph. Users 1, 2, 5, 6 and 10 are interested in g1 genre. Thus they form a clique motif and represented as a power node with self loop. Users 1,5are interested in the genre which are liked by users 13, 22 and 30. This is illustrated with the bi-clique form of the power graph. Similarly other users are categorized and represented with the help of other forms and reduced into a power graph.

Table 1: Sample data from Movie Lens dataset.

Item id	User id
g1	1,2,5,6,10
g2	1,5,13,22,30
g3	1,43,49
g4	1,7,10,12,13
g5	1,13,21,28,43,44

The conversion of network graph into power graph leads to three vital benefits (i) it clusters the users based on the relationship between them (ii) it reduces the large scale graph into small graph structure without lossless of information (iii) it improves the visualization of the social network information. Finally, using the power graph and ratings of all users, compute the final recommendations ratings for the users.

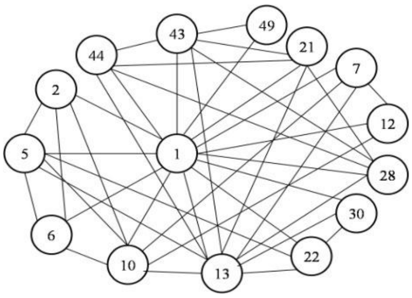

Fig. 2: Original Graph.

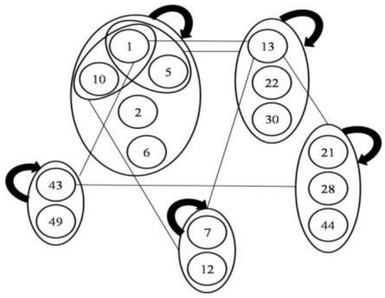

Fig. 3: Power Graph.

Features evaluation and Prediction Inference:

After the successful construction of power graph, evaluate the features set of each user to predict the recommendation rating. For the accurate prediction of the rating, users are categorized into three types: cold start users, who rate exactly twenty items; intense raters, who rate more than twenty items; and malicious users who rate more than twenty item but the difference between any rating of a particular item is greater than one. For the given power graph G', the following features are measured for each category of the users to infer the rating of unrated item for user X : (1)number of users belonging to each category in the same power node to which user X fit in (2) Compute the impact of ratings of users based on the number of common interest between them. Likewise, the above features are measured for the users in the directly connected power node and neighbours friends in the other power nodes.

Based on the constructed power graph, feature set for each user is developed for prediction. The following table Table.2 shows the feature set measured for each user

Table 2: Feature Set of User.

	Cold start users		Intense raters		Malicious users		All users	
	#	Rating Prediction	#	Rating Prediction	#	Rating Prediction	#	Rating Prediction
Power Node –X fit in	N_{cs}	R_{cs}	N_{is}	R_{is}	N_{ms}	R_{ms}	N_{as}	R_{as}
Direct Neighbours	N_{cd}	R_{cd}	N_{id}	R_{id}	N_{md}	R_{md}	N_{ad}	R_{ad}
FOAF	N_{ci}	R_{ci}	N_{ii}	R_{ii}	N_{mi}	R_{mi}	N_{ai}	R_{ai}

To predict the ratings of direct and indirect neighbours, Bayes theorem is applied to calculate the most probable rating of neighbours. Our objective is to estimate $P(R_x = x|\psi_x)$. For the range of rating $[1..N]$, using the Bayes theorem, we define

$$P(R_x = x|\psi_x) = \mathrm{P}(\psi_x|\mathrm{R_x} = \mathrm{x})\mathrm{P}(\mathrm{R_x} = \mathrm{x}) / \sum_{\mathrm{r}=1}^{\mathrm{N}} \mathrm{P}(\psi_x|\mathrm{R_x} = \mathrm{r})\mathrm{P}(\mathrm{R_x} = \mathrm{r})$$

Where ψ_x is the joint rating of all users in the group such as same power node, direct neighbours and indirect neighbours.

For any user X in the power graph, define the features set using the topology of the power graph, the number of users of specified category and their probable rating is estimated. Then utilizing the regression prediction technique, predict the rating of unrated item for the user using the computed feature set.

Given the data set $\{R_{xi}, X_{i1}, \ldots, X_{ip}\}_{i=1}^{V}$ of V users, a predictive model is developed. The variable R_{xi} denotes the rating of an item which is to be predicted. The set $\{X_{i1}, \ldots, X_{ip}\}_{i=1}^{V}$ represents the feature set of V users used for prediction. Then the model takes the following vector form

$$R_{xi} = X\alpha + \varepsilon$$

Where α represents the trust metric which indicates the influential nature of every user. In the above equation $X\alpha$ dentoes the dot product of two vectors.

Evaluation of trust metric:

In this paper, the trust value obtained using the social strength and influence of the user is utilized. The topology of the social network and social relationship between the users are considered for trust value estimation. It can be defined as follows

$T_v = \gamma W_v + \delta C_B(v)$

Where T_v is the trust value of user v with respect to user u, W_v is the weight of node based on social distance between users u and v and $C_B(v)$ is the centrality measure of node. γ and δ are the factors used to amend the involvement of the two measures. The evaluation of factors W_v and $C_B(v)$ are detailed in Parvathy *et al*., (2014).

Experimental Results:

The proposed method is evaluated using Movie Lens and Epinion dataset. Both the data sets are recommendation site that allow its users to rate the items and can also view the predicted rating value for the items of their interested category. The characteristic of datasets are shown in Table.3

Table 3: Description of Datasets.

Dataset	Users	Items	Ratings	Range
Movie Lens	6040	3952	1000209	1-5
Epinion	49290	139738	664824	1-5

Power graph Analysis:

The efficiency of our method lies in the compression of the original graph without any loss of information. A quantitative measure used to show that is edge reduction rate. Edge reduction rate is defined as the ratio of compressed edges to actual number of edges in the network. It can be computed as

$$Edge\ reduction\ rate = {n(E) - n(P)}\Big/{n(E)}$$

Where $n(E)$ is the number of edges in the actual network and $n(P)$ is the number of power edges. Table.4 shows the number of edges in the original graph, number of power edges in power graph and edge reduction rate obtained during this conversion.

Table 4: Edge reduction rate of power graph.

Nodes	#edges	#Power edges	Edge reduction rate
500	4595	2889	0.3712
1000	9837	5396	0.4514
1500	15245	6098	0.6322
2000	19645	5893	0.7271

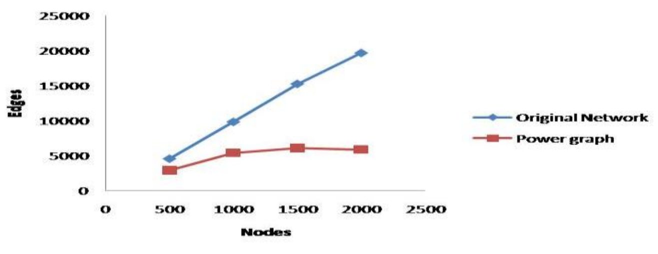

Fig. 4: Comparison of Edge Reduction.

As the edge density, ratio of number of edges in the network to the maximum number of edges, increases edge reduction also increases. Since the constructed power graph contains a very less number of edges compared to the original graph, it can be applied to application which involves very large graph

Performance Analysis:

The recommendation system quality depends on the basis of evaluation. Depending upon the applications, different metrics can be used for evaluation. The accuracy of our approach is evaluated using MAE. MAE is defined as the average absolute difference between predicted rating and the true rating value and is defined as:

$$MAE = {1}/{n} \sum_{i=1}^{n} |r_{pi} - r_{ti}|$$

Where n is the total number of ratings of all users, r_{pi} is the predicted rating and r_{ti} is the actual rating. Another type of metric which is used to measure recommendation quantity, how much the system can predict, is coverage. The performance of power graph recommendation prediction algorithm is compared with the existing trust based recommendation methods Tidal Trust and Mole trust and KNN using MAE and Coverage metrics.

The Table.5 shows the results obtained MAE and coverage for KNN and Power graph based recommendation. The KNN method depends upon the number of common items rated between the users. Hence KNN does not provide better recommendations due to the sparsity of rating matrix. Also the coverage of power graph based recommendation is larger than that of KNN as depicted in Fig.5 and Fig.6

Table 5: Comparision of KNN and Power graph.

	KNN		Powergraph	
	Mean Absolute error	Coverage	Mean Absolute error	Coverage
All users	0.78	55.8%	0.635	75.34%
Cold Users	0.93	31.42%	0.714	55.81%
Intense raters	0.82	63.19%	0.618	79.48%
Malicious users	1.27	52.48%	1.012	70.27%

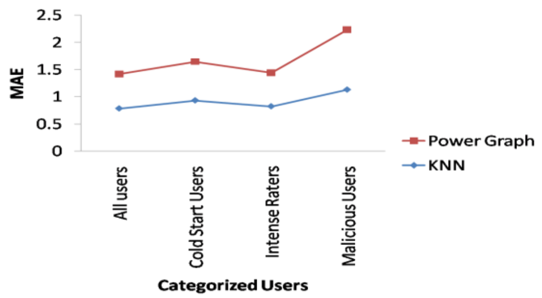

Fig. 5: Performance Analysis using MAE.

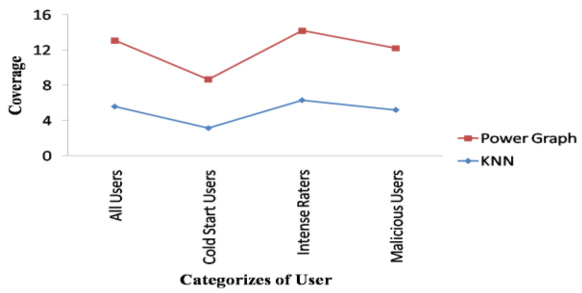

Fig. 6: Performance Analysis using Coverage.

The accuracy metric MAE obtained for Epinion dataset is compared with the existing trust based recommendations methods such as TidalTrust(Golbeck & Hendler, 2006) and Mole trust (Massa &Avesani, 2007) in Table.6. Compared with these existing trust based recommendation systems our approach outperforms by approximately 10 to 15 percent.

Table 6: Comparison of power graph with existing algorithms.

	Mean Absolute Error		
	Power Graph	TidalTrust	Mole trust
All users	0.653	0.833	0.829
Cold Users	0.715	0.878	0.854
Intense raters	0.706	0.828	0.846
Malicious users	1.049	1.237	1.242

Conclusion:

In this paper, a mechanism for recommendation prediction, by transferring the idea of power graph conversion from the bioinformatics protein network to domain of recommendation system is proposed. The major benefits of the proposed approach are 1. Any large scale graph can be better analysed as it compresses the original network to approximately 70%. 2. Accurate prediction of recommendation using the compressed power graph. Our experimental evaluation shows that the proposed methodology outperforms the existing KNN method and trust based recommendation methods, Mole Trust and Tidal Trust. As a future work, our plan to analyse the same mechanism for directed and weighted graph.

REFERENCES

Adomavicius, G. and A. Tuzhilin, 2005. Towards the next Generation of Recommender Systems: A Survey of the state-of –the-Art and Possible Extensions. IEEE Trans. Knowledge and Data Engg, 17(6): 734-749.

Ahnert, E., Sebastian, 2014. Generalised power graph compression reveals dominant relationship patterns in complex networks, Scientific Reports 4, Article number:4385, doi:10.1038/srep04385.

Billus, D. and M. Pazzani, 2000. User modelling for Adaptive News Access. User Modelling and User Adapted Interaction, 10: 147-180.

Dahlen, B.J., J.A. Konstan, J.L. Herlocker and J. Riedel, 1998. Jump-Starting Movie Lens: User Benefits of starting a Collaborative filtering System with dead data. Technical Report. 7, University of Minnesota.

Fouss, F., A. Pirotte, J.M. Renders and M. Saerens, 2007. Random-Walk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation. IEEE Transactions on Knowledge and Data Engineering, 19(3): 355-369.

Golbeck, J. and J. Hendler, 2006. FilmTrust: Movie Recommendations Using Trust in Web-Based Social Networks. Proc. IEEE Consumer Comm. and Networking Conf, 282-286.

Jing Gao, Feng Liang, Wei Fan, Yizhou Sun and Jiawei Han, 2013. A graph-based consensus Maximization Approach for Combining Multiple Supervised and Unsupervised Models. IEEE Transactions on Knowledge and Data Engg, 25(1): 15-28.

Lancichinetti, A. and S. Fortunato, 2009. Community detection algorithms: a comparative analysis. Phys Rev E Stat Nonlin Soft Matter Phys, 80(5pt 2): 056117.

Lee, S., S. Park, M. Kahng and S. Lee, 2013. PathRank: Ranking nodes on a heterogeneous graph for flexible hybrid Recommender system. Expert Systems with Applications, 40(2): 684-697.

Mooney, R..J. and L. Roy, 2000. Content Based Book Recommending Using Learning for Text Categorization. Proceedings of the Fifth ACM Conference on Digital Libraries, ACM Press, New York, 195-240.

Massa, P. and P. Avesani, 2007. Trust-Aware Recommender Systems. Proc. ACM Conf. Recommender Systems.

Massa, P. and P. Avesani, “Controversial Users Demand Local.

Parvathy, M., R. Ramya, K. Sundarakatham and S. Mercy Shalinie, 2014. TNet:A Trust Network for predicting trust in Online Social Networks. Information:An International Interdisciplinary Journal, 17(4): 1383-1394.

Resnick, P., N. Iacovou, M. Sushak, P. Bergstrom and J. Riedl, 1994. Grouplens: An open architecture for Collaborative Filtering of Netnews. Proceedings of ACM Conference on Computer Supported Cooperative Work, Chapel Hill, North Carolina, 175-186.

Sarwar, B., G. Karypis, J. Konstan and J. Reidl, 2001. Item-based Collaborative filtering Recommendation Algorithms. Proc. of the 10^{th} International conference on World Wide Web, 285-295.

Xu, B., J. Bu, C. Chen, C. Wang, D. Cai and X. He, 2007. EMR: A scalable Graph-based Ranking Model for Content-based Image Retrieval. IEEE Transactions on Knowledge and Data Engineering, 6(1): 1-14.

Yu, K., A. Schwaighofer, V. Tresp, Xiaowei Xu and H. Kriegel, 2004. Probabilistic Memory-Based Collaborative Filtering. IEEE Transactions on Knowledge and Data Engineering, 16(1): 56-69.