Je-LKS
PEER REVIEWED PAPERS
FOCUS ON RECOMMENDER SYSTEMS
Vol. 10, n.1, 2014
ISSN: 1826-6223 | eISSN: 1971-8829
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal

COMMUNITY BASED RECOMMENDATION IN E-LEARNING SYSTEMS

V Senthil kumaran[1], A Sankar[2], K Kiruthikaa[1]

[1]Department of Applied Mathematics and Computational Sciences, PSG College of Technology, Coimbatore – India, vsk.mca@gapps.psgtech.ac.in
[2]Department of Computer Applications, PSG College of Technology, Coimbatore – India
Keywords: community based recommendation; e-learning; recommender system in learning

The success of any e-learning system depends on quality and the quantity of assistance provided to its students, in the learning process. Hence, it is essential to analyze a student's academic skills in order to personalize the education provided both vertically and horizontally. This paper proposes a novel approach through which initially students are grouped based on several factors including their academic interests and further motivate the students to enhance their knowledge by providing appropriate recommendations made based on students belonging to their group. It has been proved that neither link information nor content information individually is sufficient to form student communities (Rabbany *et al.*, 2011). Therefore, the approach includes both together for community detection. Further, the approach also intends to recommend courses based on the ratings for courses given by other students with similar skill sets in the same group. Experimental results

for citations:
Senthil kumaran V., Sankar A., Kiruthikaa K. (2014), *Community based recommendation in e-learning systems,* Journal of e-Learning and Knowledge Society, v.10, n.1, 51-61.
ISSN: 1826-6223, e-ISSN:1971-8829

highlight the quality or relevance of the recommendations made within communities, which in turn reflects on the accuracy of the proposed community detection method.

1 Introduction

In an e-learning environment, students get motivated by their peers, apart from their teacher. Group learning is observed to be more effective than traditional methods, provided the group is compatible and also motivate the students to focus solving on potential problems in various related areas. The students are grouped based on common interest, but they can also individually possess various other expertise or interests in different courses or fields. On observing the career path of students with similar interests, we can recommend related courses to the students. These recommendations aim at creating awareness and widen the scope learning. This is usually addressed as personalized and collaborative e-learning, which is proved to improve the learning efficiency. By data exchange or any other form of communication, a student gets to interact with many other students in the e-learning system by means of a discussion forum. Such an interaction can be modelled as a graph. Before joining the e-learning environment, a student is tested and graded for various qualitative attributes like capability, performance etc. Based on the interactions represented in the graph and students personal abilities, we form communities. Further, for every community, we keenly analyze and recommend courses.

2 Related Work

2.1 Social Network Analysis and community detection

Social networks are formally defined as a set of actors or network members, connected by one or more type of relations (Marin & Wellman, 2009). There can be different types of relationships such as collaborations, friendships, hyperlinks, citations etc. (Marin & Wellman, 2009). Social network analysis is the mapping and measuring of relationships and flow between entities.

In social network analysis, the links can be visualized as graph, consisting of agents (students in our study) and their respective connections are called sociograms. These graphs can be either directed or undirected. In an e-learning forum, if a student posts a question / notes and another student responds to it, then that is a directed interaction and the graph will be directed: the second student has responded to the first, but the first has not responded to the second, then that is an undirected interaction and the graph will be undirected (Rabbany *et al.*, 2011).

There has been a considerable amount of work done to detect communities in social networks (Ruixuan Li, 2012). Klamma has significantly contributed to the field of social network analysis which focuses on learners network (Klamma *et al.*,2006; Klamma *et al.*, 2009).

2.2 Recommendation Systems

A learning system, be it online or face-to-face learning, takes responsibility of teaching the required course. It can also focus on increasing the knowledge of the student horizontally, i.e. motivating the student to learn new courses which are related to them or similar to their primary interests. The system responsible for making recommendations is called recommender system. Many such recommendation systems have been proposed by various researchers, which are well appreciated by the application users.

Sandy El Helou with many other researchers have collectively worked to propose various recommender systems including Trust-based recommender system (Li *et al.*, 2010), Federated recommender system (Zhou *et al.*, 2012) and a 3A personalized, contextual and relation-based recommender system (El Helou *et al.*, 2010). Few other recommender systems relavant to the current research include, graph-based recommender system (Wang *et al.*, 2010; Ruixuan Li, 2012) and Context-aware recommender systems (Verbert *et al.*, 2011).

Recommendation systems are highly used in e-commerce and social network applications. Researchers have also taken abundant interest in applying these recommender systems to education (Manouselis *et al.*, 2011). The recommendations have also been personalized in order recommend the best fit to the learners (Drachsler *et al.*, 2008). (Verbert *et al.*, 2011) focuses on improving the efficiency of recommender systems. (Pham *et al.*, 2011) worked on introducing clustering of nodes in a social network to provide recommendations. Also, several researchers have focussed on comparing and evaluating the performance of various recommender systems.

3 Community Detection: A novel approach in e-learning

E-learning can be viewed as social network, where students interact with each other, share their interests or materials, posting questions, answering the questions, following other students, from which community structures can be detected based on their performance and interests. Each node in the social network represents a student. The link between the two nodes (students) is established when one student interacts with another student. Community detection can divide users into several subgroups with similar interest and study ability. In this approach, students belong to a same community will perform alike and they can share the proper learning contents easily. Fig. 1 shows the schematic representation of a student network consisting of 9 nodes and 2 communities.

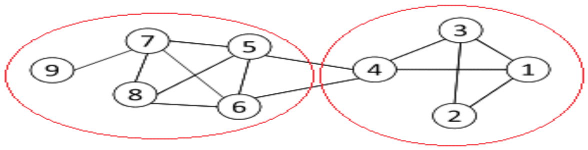

Fig. 1 - A schematic representation of a student network with 9 nodes (student) with two community structures.

The closeness between two nodes (students) is differentiated using the edge weight, which is calculated based on the interaction in the e-learning forum. The edge weight (W_{ij}) between two students S_i and S_j is given Eq.(1).

$$Wij = \frac{W_{cur}^{j} * a + M}{a + 1} \tag{1}$$

For each pair of students, there are three variables namely,

- i. W^{ij}_{cur}, a current edge weight, which is initialized to 0 in the first iteration. The edge weight is recomputed for every new interaction.
- M, a mark denoting the level of interaction currently happening between a pair of nodes. The value of M is calculated based on the type of participation of the student in the forum.
- α (alpha), an aging factor. This paper uses a constant value, 0.5, for the aging factor. It is possible for each pair of nodes to have their own aging factor α, and they can be updated in response to changes that occur in the interaction pattern.

In order to compute the similarity between the students and group them based on the similarity, we use both structural and contextual information. Contextual information includes personal details like performance assessment, interests, qualifications, capability levels etc. The structural information of a student is collected by observing interactions of students posting and answering questions in discussion forums. In this section, we present formulae that are employed to compute structural and contextual similarity and finally a similarity formula including both structural and contextual similarity.

The structural similarity is calculated as given in Eq. (2).

$$Sim_{struct}(S_i,S_j)=\begin{cases}\left(\sum_{z\in N_i\cap N_j}\frac{1}{DEG(z)}\right)*W_{ij} & \text{if } S_i\leftrightarrow S_j\\ 0 & otherwise\end{cases} \tag{2}$$

In the above formula, N_i and N_j are the set of neighbours of Si and Sj respectively, z is the common neighbour of S_i and S_j, DEG(z) is the degree of the vertex and W_{ij} is the normalized edge weight calculated based on the interaction between S_i and S_j.

The contextual similarity is defined in Eq. (3).

$$S=\left\{x\in P_i\cap P_j \middle| V(x_{P_i})=V(x_{P_j})\right\}$$

$$Sim_{context}(S_i,S_j)=\frac{|S|}{|P_i\cap P_j|} \tag{3}$$

In the above formula, P_i and P_j are the set of properties of S_i and S_j respectively, V (x_{Pi}) and V (x_{Pj}) are the interpret mapping on P_i and P_j respectively.

In the network, we might have one of the three situations in the student network.

If there is a direct link between two learners, then both structural and context similarity will be considered.

If there is no direct link, then the similarity will be the product of the similarity of all intermediate learners. S_p is on the path between S_i and S_j. There may be more than one path between two learners, we consider the shortest path and used the shortest path algorithm proposed in (Fredman & Tarjan, 1987).

If the learner is isolated, that is a leaner has not interacted with others, only the context similarity will be considered. On taking into account, the above scenarios, we have formulated the similarity measure which is given in Eq. (4).

$$Sim(S_i,S_j)=\begin{cases}\alpha*Sim_{struct}(S_i,S_j)+(1-\alpha)*Sim_{context}(S_i,S_j), & if\ S_i\leftrightarrow S_j\\ \prod_{a=1}^{q}Sim(S_{pa}+S_{pa+1}), & if\ S_i\ldots\ldots S_j,\ \text{Sp is on path between Si and Sj}\\ (1-\alpha)*Sim_{context}(S_i,S_j), & otherwise\end{cases} \tag{4}$$

Once we compute the similairity between nodes in the student graph, we apply k-Means, a clustering algorithm to form student communities. The value k, in k- Means is an arbitrary value and hence we can form multiple communities by varying the value of k. There might be a k value for which there will be no eligible students to be a part of it i.e., the community will be non-empty for similarity based clustering, but when we compute the frequency it is possible that few communities might be empty. This scenario occurs when the range of each community narrows down when we increase k. Hence, the size of the community reduces. So, when we compute frequencies we find that the communities with the least size end up with no students. Therefore, we do no further clustering. Also, mathematically the alpha value also plays a role in this situation. The range of similarity values will depend on the alpha value, and when the shortest path contains many intermediate nodes, the product of the intermediate similarity values reduces the final similarity value eventually. Hence, there might be very less values under a particular range after a level of clustering. These factors shall be considered before deciding the values of k.

4 Community Based Recommendation Systems

A student is expected to rate a course as a feedback to the e-learning system based on their personal experience. These ratings highly contribute to the recommendation process. Each community has a group of students similar to each other and taking up a set of courses together. A student may also take up other courses which can be recommended to other students in the community having the similar career paths. The system developed for recommendation, takes into consideration, the similarity values between the student to whom recommendations has to be made and everyother student in the community having common qualifications in terms of the courses completed, and the ratings given by students in the community for courses which are learnt by them. The recommendation value is computed by using these two parameters and the ageing factor, α, as chosen in the previous section for community detection. This recommendation value denotes its proximity to the student's current interest and capabilities. So the higher the value, the better chances of the course being recommended. We take the similarity value computed for community detection, in order to recommend courses taken by students who are highly similar to the current student and proceed till the least similar student. We then combine it with the ratings given by the student to that course. As the similarity values range only between 0 and 1, the ratings also should fall within the same range. Otherwise, they are normalized. The recommendation value is computed for all the similar students and all the courses they have taken. The recommendation value is computed using the formula,

$$R_{i, j, k} = (\alpha * Sim(S_i, S_j)) + ((1- \alpha) * (rating(S_j, C_k))$$

In the above formula, $R_{i,j,k}$ represents the recommendation value for the course C_k to student S_i based on the rating given by student S_j. A status is maintained recording whether the recommended course is accepted or rejected by the student. As and when students respond to the recommendations, the system computes he recommendation values for other courses and updates the list containing recommendable courses. The list is refreshed based on the students response.

5 Experiment Set up, Results and Discussion

The system developed was made available to a group of students in our college. The system offers few major courses related to computer science namely, Semantic web, Information retrieval, Soft Computing, Robotics, Machine learning, Data mining, Java programming, C++, Matlab, and Object Oriented Programming. The students initially entered their areas of interest in the enrolment section of the e-learning system, using which recommendations were made. The academic qualities such as performance, capabilities etc., of the students were assessed based on their performance in a series of tests of varying levels conducted in the courses they are intrested in or have learnt. The interactions made by the students in the discussion forum was monitored to form communities. Fig 1 shows a network of nine students who used the system and two communities were formed based on their structural and contextual similarities. Based on the individual interests of students, their similarity to other members and the ratings given by other students to the courses, the recommendations were made. The students were allowed to accept or reject the recommendation. The recommended list of courses was refreshed based on the response given by the student to the recommendations made previously. Table 1 tabulates the student communties, intrests of students in each community and recommendations made to them student.

Table 1
TWO STUDENT COMMUNITIES, THEIR INDIVIDUAL INTERESTS AND RECOMMENDED COURSES

	Student ID	Interests	Recommendations made by the system
Cluster 1	S9	Information Retrieval	Matlab, Soft computing, Robotics
	S7	Machine Learning, Information Retrieval, Matlab	Soft Computing, Java
	S5	Data Mining, Java	Machine Learning, Information Retrieval
	S8	Data Mining, Machine Learning	Soft Computing, Matlab
	S6	Machine Learning, Soft Computing, Information retrieval, Robotics	Matlab, Semantic Web, Robotics
Cluster 2	S4	Object Oriented Programming	C++
	S3	Information Retrieval, Soft computing, Java, Matlab	Object Oriented Programming, Machine Learning
	S1	Object Oriented Programming, Semantic web	Java, Information Retrieval
	S2	Object Oriented Programming, C++, Data Mining	Java, Semantic Web, Machine learning

Recommender systems researchers use a number of different measures for evaluating the success of the recommendation or prediction algorithms (Hernández *et al.*, 2008). For our experiments, we use the widely popular traditional metrics accuracy and average Mean Absolute Error (MAE). Interactive recommender systems usually consider only the accuracy/precision of the displayed recommendations (Hernández *et al.*, 2008). The possible decisions are either recommend the item or not. Next, the learners can either follow the recommendation or skip it.

An optimal recommender system designed for an e-learning environment is the one that accurately predicts the preferences of individual learners and recommend them, when required. So, the best way of evaluating a recommendation system is to evaluate the quality of its predicted preference values against the actual preferences of the user. Hence, to evaluate recommender systems, we apply classic information retrieval metrics Precision, Recall, F-measure Accuracy and MAE. Recommender system in e-learning environment should be able to return as many relevant courses as possible. We can say that the 'Precision' in terms of e-learning recommender systems is the proportion of highly recommended courses that are relevant. So, for example, 'Precision at 10' would be the proportion computed from the top 10 results. The 'Recall' measures the proportion of relevant courses appearing in the top recommended course list. The F-measure is a measure of a statistic test's accuracy that considers both precision and recall measures of the test to compute the score. It is the weighted average of the precision and recall. The measures and their values are given in Table 2.

Table 2
MEASURES AND THEIR VALUES

Measure	Value
Precision@10	0.88
Recall@10	0.87
F-Measure	0.9
Accuracy %	92.14

The Precision at 10 is 0.88, implying, on an average about 9 of 10 recommendations were accepted/ relevant. Recall at 10 is 0.87 implies that on an average of 9 accepted/relevant recommendations are present amongst the top recommended course list. Accuracy is the overall correctness of the model and the accuracy of our proposed model is 92.14%. The F-measure of our model is 0.9. MAE is the measure of the deviation of recommendations from their expected learner specified values. The MAE is defined as the average difference between the predicted ratings and the real user ratings, as defined within the test sets. Formally, MAE can be mathematically defined as in Eq.(5).

$$MAE = \frac{\sum_{i=1}^{N} \left| \dot{p} \ , \dot{r} \right|}{N} \tag{5}$$

In the above formula, pi is the predicted value for item i and ri is the learner's rating. It is proved that MAE = 1 – Accuracy. So, the MAE value for k=2, i.e., two clusters formed from the sample graph, is, 0.0786. In cluster analysis, we can use density and entropy to measure the quality of the clustering method based on the clusters formed. Density refers the amount of data each cluster contains. In our case, density denotes the number of nodes or students the clusters contain. Entropy refers to the amount of uncertainties in the clusters. The goal of an efficient clustering system is to identify a k, for which the density is high and entropy is low. Table 3 tabulates the MAE, density and entropy for different values of k in the k-Means algorithm.

Table 3
CLUSTERING RESULTS BY VARYING NUMBER OF CLUSTERS (K), QUALITY IS CALCULATED USING DENSITY, ENTROPY AND MAE

K	Clusters	MAE	Density	Entropy
2	{9,8,6,5,7} {4,3,1,2}	0.0786	0.53	0.142
3	{8,5,6,4},{1,3,2},{7,9}	0.0751	0.22	0.053
4	{1,3},{2,4,5},{8,6}{7,9}	0.0714	0.199	0.053

On observing the implicit and explicit feedback attained from students after working with the system, we are fairly confident about the correctness of the values exhibited by the chosen accuracy measure.

Conclusion

In this paper we attempt to form student communities and enhance their knowledge by incorporating a recommendation system within an e-learning system. The network formed connects all the students in the system, from which we form student communities based on their personal skills and interactions in the discussion forum. The results of the initial tests are processed to obtain a list of personal qualities and the interactions are monitored to obtain social information about the student. The system further works on enhancing the knowledge of its students by recommending courses similar to their interests. From the measured outcomes of the system (MAE), we say that the proposed model is approximately 92% accurate. The scope of this work is to propose a tool that personalizes the e-learning process and aims at enhancing the knowledge gained.

REFERENCES

Marin, A, Wellman, B, (2011), *Social Network Analysis: An Introduction*, The Sage Handbook of Social Network Analysis, eds. Peter Carrington and John Scott, Sage, 11-25.

Rabbany Reihaneh k et al. (2011), *Social Network Analysis and Mining to Support the Assessment of Online Student Participation*, SIGKDD Explorations, 13(2), 20 – 29.

Hernández del Olmo, F., & Gaudioso, E (2008), *Evaluation of recommender systems: A new approach*, Expert Systems with Applications, 35(3), 790-804.

Fredman, M. L., & Tarjan, R. E. (1987), *Fibonacci heaps and their uses in improved network optimization algorithms*, Journal of the ACM (JACM), 34(3), 596-615.

Huang, L., Li, R., Li, Y., Gu, X., Wen, K., & Xu, Z. (2012), *ℓ1-Graph Based Community Detection in Online Social Networks*, Lecture Notes in Computer Science, (7235), (Springer Verlag, 2012), 644–651.

Wang, Z., Tan, Y., & Zhang, M. (2010), *Graph-based recommendation on social networks*, Web Conference (APWEB), 12th International Asia-Pacific. IEEE, 116-122.

Drachsler, H., Hummel, H. G., & Koper, R. (2008), *Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model*, International Journal of Learning Technology, 3(4), 404-423.

Verbert, Katrien, et al. (2011), *Dataset-driven research for improving recommender systems for learning*, Proceedings of the 1st International Conference on Learning Analytics and Knowledge, ACM, 44-53.

Pham, M. C., Cao, Y., Klamma, R., & Jarke, M. (2011), *A Clustering Approach for Collaborative Filtering Recommendation Using Social Network Analysis*, J. UCS, 17(4), 583-604.

Klamma, R., Spaniol, M., Cao, Y., & Jarke, M. (2006), *Pattern-based cross media social network analysis for technology enhanced learning in Europe*, Innovative Approaches for Learning and Knowledge Sharing, Springer Berlin Heidelberg, 242-256.

Klamma, R., Cuong, P. M., & Cao, Y. (2009), *You never walk alone: Recommending academic events based on social network analysis*, Complex Sciences, Springer Berlin Heidelberg, 657-670.

Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R. (2011), *Recommender systems in technology enhanced learning*, In Recommender systems handbook, pp. 387-415. Springer US.

El Helou, S., Salzmann, C., & Gillet, D. (2010), *The 3A Personalized*, Contextual and Relation-based Recommender System, J. UCS, 16(16), 2179-2195.

Li, N., El Helou, S., & Gillet, D. (2010), *Trust-based rating prediction for recommendation in web 2.0 collaborative learning social software*, 9th International Conference on Information Technology Based Higher Education and Training (ITHET), IEEE, 208-212.

Zhou, L., El Helou, S., Moccozet, L., Opprecht, L., Benkacem, O., Salzmann, C., & Gillet, D (2012), *A federated recommender system for online learning environments*, Advances in Web-Based Learning-ICWL 2012, Springer Berlin Heidelberg, 89-98.