COMMUNITY BASED RECOMMENDATION IN E-LEARNING SYSTEMS

V Senthil Kumaran¹, A Sankar², K Kiruthikaa¹

¹Department of Applied Mathematics and Computational Sciences, PSG College of Technology, Coimbatore – India, vsk.mca@gapps.psgtech.ac.in
²Department 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
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 relevant 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.
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).

$$W_{ij} = \frac{W_{ij}^{\text{cur}} \cdot a + M}{a + 1}$$

(1)

For each pair of students, there are three variables namely,
- $W_{ij}^{\text{cur}}$, 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.
- $\alpha$ (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 $\alpha$, 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).
\[
Sim_{\text{struct}}(S_i, S_j) = \begin{cases} 
\sum_{z \in N \cap N} \frac{1}{\text{DEG}(z)} * W_{ij} & \text{if } S_i \leftrightarrow S_j \\
0 & \text{otherwise}
\end{cases}
\]

(2)

In the above formula, \(N_i\) and \(N_j\) are the set of neighbours of \(S_i\) and \(S_j\) respectively, \(z\) is the common neighbour of \(S_i\) and \(S_j\), \(\text{DEG}(z)\) 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).

\[
S = \left\{ x \in P_i \cap P_j \mid V(x_P) = V(x_P) \right\}
\]

\[
Sim_{\text{context}}(S_i, S_j) = \frac{|S|}{|P_i \cap P_j|}
\]

(3)

In the above formula, \(P_i\) and \(P_j\) are the set of properties of \(S_i\) and \(S_j\) respectively, \(V(x_P)\) 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 learner 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).

\[
Sim(S_i, S_j) = \begin{cases} 
\hat{a} * Sim_{\text{struct}}(S_i, S_j) + (1 - \hat{a}) * Sim_{\text{context}}(S_i, S_j), & \text{if } S_i \leftrightarrow S_j \\
\prod_{a=1}^{q} Sim(S_{pa} + S_{pa} + 1), & \text{if } S_{\ldots} S_p \text{ is on path between } S_i \text{ and } S_j \\
(1 - \hat{a}) * Sim_{\text{context}}(S_i, S_j), & \text{otherwise}
\end{cases}
\]

(4)
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,
\[ R_{i,j,k} = (\alpha * \text{Sim}(S_i, S_j)) + ((1 - \alpha) * \text{rating}(S_j, C_k)) \]

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 the 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 interested 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 communities, interests of students in each community and recommendations made to them student.
Table 1
TWO STUDENT COMMUNITIES, THEIR INDIVIDUAL INTERESTS AND RECOMMENDED COURSES

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Interests</th>
<th>Recommendations made by the system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S9</td>
<td>Information Retrieval</td>
<td>Matlab, Soft computing, Robotics</td>
</tr>
<tr>
<td>S7</td>
<td>Machine Learning, Information Retrieval, Matlab</td>
<td>Soft Computing, Java</td>
</tr>
<tr>
<td>S5</td>
<td>Data Mining, Java</td>
<td>Machine Learning, Information Retrieval</td>
</tr>
<tr>
<td>S8</td>
<td>Data Mining, Machine Learning</td>
<td>Soft Computing, Matlab</td>
</tr>
<tr>
<td>Cluster 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>Object Oriented Programming</td>
<td>C++</td>
</tr>
<tr>
<td>S3</td>
<td>Information Retrieval, Soft computing, Java, Matlab</td>
<td>Object Oriented Programming, Machine Learning</td>
</tr>
<tr>
<td>S1</td>
<td>Object Oriented Programming, Semantic web</td>
<td>Java, Information Retrieval</td>
</tr>
<tr>
<td>S2</td>
<td>Object Oriented Programming, C++, Data Mining</td>
<td>Java, Semantic Web, Machine learning</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision@10</td>
<td>0.88</td>
</tr>
<tr>
<td>Recall@10</td>
<td>0.87</td>
</tr>
<tr>
<td>F-Measure</td>
<td>0.9</td>
</tr>
<tr>
<td>Accuracy %</td>
<td>92.14</td>
</tr>
</tbody>
</table>

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).

\[ MAE = \frac{1}{N} \sum_{i=1}^{N} | \hat{r}_i - r_i | \]  

(5)

In the above formula, \( \hat{r}_i \) is the predicted value for item i and \( r_i \) is the learner’s rating. It is proved that \( MAE = 1 - \text{Accuracy} \). 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.
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**


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.


