Personalized Article Recommendation Based on Student’s Rating Mechanism in an Online Discussion Forum

Chuen-He Liou
National Taipei University of Nursing and Health Sciences, Taiwan
chliou@ntunhs.edu.tw

Abstract

Online discussion forum is one of e-learning activities to construct learner’s knowledge and interact with their classmates, which is also a module of Learning Management System (LMS). However, students all have the same latest articles even though their article preferences are different. Traditional LMS could not recommend the personalized articles for individual student based on her or his own preference. In this study, an article recommendation method was proposed based on student’s rating to recommend the personalized articles in an online forum. Experiment was conducted with two hundred nursing students in a university of northern Taiwan. Student could rate the articles which classmate posted by the proposed rating mechanism, which is similar to the Like button in Facebook. Experiment result shows that the proposed method performs well compared to collaborative filtering (CF) method. The proposed method could recommend the better personalized articles than the typical CF method.

1. Introduction

E-learning is an instructing and learning way via Internet. Teacher could instruct, and student could learn at any time anywhere in an e-learning environment. Learner use the e-learning website as a complement to in-classroom learning and obtained higher marks in the examination [1]. Internship is an important learning process for skill training before they are going to work. Students felt e-learning to be useful during their internship in problem solving, knowledge construction and personal reflection [2]. E-learning could be complement to clinical education for the internship to form a part of blended-learning [3]. Most medical students recognized the positive impact of e-learning on their clinical skill training in a blended approach [4].

Online discussion forum is a part of e-learning activities. Learners could construct their knowledge to have the better learning performance. Wu [5] use Google+ as discussion forum to construct nursing students’ knowledge. Nursing student could post articles, share knowledge and interact with each other to strengthen their learning. Pena-Shaff and Nicholls [6] use bulletin board system (BBS) as discussion forum where students could engage in knowledge construction process, such as clarification, elaboration, interpretation, and reflections. Wu and Hiltz [7] propose online discussion which could improve students’ perceived learning. Online discussion could enhance their learning and develop their ability of critical thinking by sharing their ideas to each other [8]. Cheng, Paré, Collimore and Joordens [9] mentioned that student who participated in forum tend to have the better performance in course.

Learning Management System (LMS) is an e-learning system which students could engage in learning activities, such as forum and chat. Moodle has been used as LMS for students to interact with their classmates [10]. However, students all have the same recent articles even though they have the different article preferences in Moodle. Traditional LMS like Moodle could not offer the personal learning environment. Hence, data mining technology is applied to educational systems, which include on-line analytical processing, clustering, association rules, classifying and visualization [11].

Recommendation system is one of data mining applications, which recommend several items like movies and music to users according to their preferences [12]. Generally, recommendation systems are classified as collaborative or content-based filtering technique. Collaborative filtering (CF), which employs ratings given by users with similar interests to recommend items to a target audience [13]. In contrast, content-based filtering (CBF) makes recommendations by matching user profiles with content features [14]. Some studies combined collaborative and content-based filtering techniques as a hybrid recommender system[15].
There are the latest posted articles on the right column of the user interface in Moodle, but every student has the same article list. Since the article preferences of all students are not the same, the recommended article list should be different for every student. Typical Moodle could not provide the personalized discussion articles for individual student based on her or his own preference. In this study, an article recommendation method is proposed based on student’s rating, which also combines clustering and association rules to recommend articles in an online discussion forum. The experiment was conducted with two hundred nursing students in a university of northern Taiwan. Student could rate the articles which classmate posted by the proposed rating mechanism, which is similar to the Like button in Facebook. The interactions in social network between classmates could be observed. First, an online discussion forum is established for a course. Students could post their articles and rate the other classmates’ articles. Second, students are clustered into groups by k-means clustering method. In each student group, the association rules between articles were retrieved. Third, the proposed method recommends articles which are sorted by the corresponding confidences of association rules in descending order. Finally, the method is compared with the traditional CF method.

The rest of the paper is as follows. The related studies are discussed in Section 2. Section 3 describes the proposed approach. In Section 4, the experiment results are illustrated. Finally, the findings are summarize, and the limitations and future research are proposed in Section 5.

2. Related works

2.1. Online forum article recommendation

Online discussion could be divided into synchronous and asynchronous discussion. Synchronous discussion has the characteristic of social contact. Synchronous discussion is usually short, concise and easy to understand, such as chat. On the other hand, asynchronous discussion is usually complicated and deep, such as forum [16]. Synchronous and asynchronous discussion has the different functionality. The former provides the emotional communication and the later provide deep discussion and interaction about course. The participants of synchronous discussion are required to discuss at the same time, but the participants of asynchronous discussion could interact with the others at any time [17]. Schellens and Valcke [18] found that students construct their knowledge through asynchronous discussion.

In asynchronous discussion, there are article recommender systems in the online forum. Pereira Nunes, Mera, Kawase, Fethah, Casanova and de Campos [19] used semantic technology and statistical method to recommend topics in online forums. Albatayneh, Ghauth and Chua [20] proposed a content-based filtering method that recommends relevant post messages to users with latent semantic analysis. Drachslers, Pecceu, Arts, Hutten, Rutledge, Van Rosmalen, Hummel and Koper [21] present a collaborative filtering method based on tag and rating of articles from Web 2.0 social network services to recommend learning resources. Abel, Bittencourt, Costa, Henze, Krause and Vassileva [22] provided a collaborative filtering method to recommend thread or post based on post count or energy. In this study, the proposed method uses student’s rating, clustering technique and association rules to recommend articles, which is compared to the collaborative filtering method. The students are clustered into groups according to their article preferences and the association rules between articles in each group are retrieved. Finally, the system recommends articles by association rules between articles and frequent rated articles in the forum.

2.2. Clustering techniques

In e-learning, clustering techniques seek students with similar learning preferences in a group and students in different groups with the dissimilar preferences. k-means method [23] is a clustering technique which initially chose k data points as cluster centers, and the remaining points are assigned to clusters iteratively by similarity measure based on the minimal distance between the point and cluster mean. Finally, all the data points are assigned to clusters until all the cluster centers remain the same.

2.3. Mining association rules

In this study, association rule mining is used to retrieve the associations between two sets of forum articles in a learning management system. Agrawal et al. [24] proposed that association rules are discovered when the supports and the confidences of these rules are both larger than the user-specified minimum support threshold and minimum confidence threshold. Let I be a set of article items. An association rule is a rule in the form \( X \Rightarrow Y \), where \( X \subseteq I \), \( Y \subseteq I \), and \( X \cap Y = \emptyset \). The support is the percentage of instances that both satisfy \( X \) and \( Y \) and the confidence is the percentage of instances satisfying \( X \) that also satisfy \( Y \).
2.4. Typical collaborative filtering method

In e-learning, typical collaborative filtering (CF) [25] recommend learning resources to a student by using k-nearest neighbors (k-NN). The neighbors are determined by calculating the similarity between students’ learning preferences. The similarity is computed by Pearson’s correlation coefficient (pcc) and its definition is in the following (1).

\[
pcc(s_i, s_j) = \frac{\sum_{A \in I} (r_{s_iA} - \overline{r}_i)(r_{s_jA} - \overline{r}_j)}{\sqrt{\sum_{A \in I} (r_{s_iA} - \overline{r}_i)^2 \sum_{A \in I} (r_{s_jA} - \overline{r}_j)^2}},
\]

where \(\overline{r}_i\) and \(\overline{r}_j\) are the average ratings of all articles from students \(s_i\) and \(s_j\); \(I\) indicates the mix of the set of articles; \(r_{s_iA}\) and \(r_{s_jA}\) are the ratings of article A from students \(s_i\) and \(s_j\).

After the formation of the neighborhood, articles are recommended by k-nearest neighbors in the following steps. First, the frequency count of each article is computed by scanning all the instances from k-nearest neighbors. Second, all articles are sorted by the corresponding frequency count in descending order. Finally, the sorted articles which have not been rated by the student are recommended to the student.

3. Methodology

In this section, an article recommendation method is proposed based on a rating by the student of an article in an online discussion forum, which is shown in Figure 1. Student could rate the articles which classmate posted by the proposed rating mechanism, which is similar to the Like button in Facebook. First, an online discussion forum is established in a course. Students could post their articles and rate the other classmates’ articles in the forum. Because there are articles which were not rated by students and some students did not rate any articles. In data preprocessing phase, these students and unrated articles are filtered in data preprocessing phase.

The nominal rating values, e.g., “like” or “so-so”, are transformed into the discrete value, e.g., 2 or 1. Article preference (AP) is defined as follows.

\[
\text{Article preference (AP)} = \begin{cases} 
2, & \text{"like"} \\
1, & \text{"so-so"} \\
0, & \text{unrated}
\end{cases}
\]

where AP=2 when the student likes the article, AP=1 when the student feels so-so about the article and AP=0 when the student did not rate the article.

With (2), the nominal student-article rating matrix has been transformed into the discrete student-article rating matrix by article preference, which is shown in the following Table 1.

### Table 1. Student–article rating matrix

<table>
<thead>
<tr>
<th>Article ID</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student ID</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.2. Article recommendation engine

The recommendation engine [26] is shown in Figure 2. First, students are clustered into groups by k-means clustering method based on the similarity between students, which is computed by Pearson’s correlation coefficient of student-article rating matrix. The method then retrieves association rules between articles in each group when the supports and confidences of these rules are both larger than the thresholds of supports and confidences set in the experiment.

Second, let $X_u$ be the articles previously rated by the student $u$. For association rule $X^k \rightarrow Y^k$, if $X^k \subseteq X_u$ then articles in $Y^k \cup X_u$ denoted by $Y^k_u$ are the candidate articles for recommendation to the student $u$. Let $Y^u_{AR}$ be the candidate articles generated from association rules which satisfy $X^k \subseteq X_u$. The articles in $Y^u_{AR}$ are ranked by the confidence $c(Y^u_{AR})$ of the association rule $X^k \rightarrow Y^k$.

Third, the number of candidate articles $|Y^u_{AR}|$ is compared to the number of N recommendations. If $|Y^u_{AR}| \geq N$, the engine recommends the top N articles within $Y^u_{AR}$. If $|Y^u_{AR}| < N$, the remaining $N-|Y^u_{AR}|$ articles are selected from $Y^u_{MF}$, which are the articles ranked by the frequency count $f_G(Y^u_{MF})$ in the student $u$’s group. Finally, articles in $Y^u_{MF}$ that have not been included in $Y^u_{AR}$ and have not been rated by the student $u$ are recommended to the student $u$ as the top N article recommendation.

4. Experiment evaluation

4.1. Experiment setup

The experiment data were collected from an online discussion forum of a course in the learning management system (LMS) of a university in the northern Taiwan. There are 200 nursing students who discuss 16 related topics in the first semester of 2014. They post 1,000 articles and rate the other classmates’ articles with 1,000 ratings. Because some students rate few articles and some articles are rated by few students, we set the rating record and the rated article threshold to 3. Besides, the minimum support and confidence of the association rules are set to 0.1 and 0.5 and the cluster number is set to 3. These parameters are set based on student’s online learning behaviors that we observed.

4.2. Evaluation metric

We employed three metrics, precision, recall and F1-metric, to evaluate the recommendation quality of recommender systems. Precision and recall are two commonly used to information retrieval [25]. Precision is defined as the ratio of the rightly recommended articles divided by the total number of the recommended articles, which is shown in (3). The rightly recommended articles are the recommended articles which are provided by system, and are also exactly interesting to students. The recommended articles are provided by recommender system. Recall is defined as the ratio of rightly recommended articles divided by the preferred articles, which is shown in (4). The preferred articles are interesting to students.

$$\text{Precision} = \frac{\text{number of rightly recommended articles}}{\text{number of recommended articles}}$$

(3)

$$\text{Recall} = \frac{\text{number of rightly recommended articles}}{\text{number of preferred articles}}$$

(4)

However, precision and recall are used to evaluate recommendation quality, which often show the different result. For example, when the recommended articles increase, recall will increase but precision will decrease. Since precision and recall are both important metrics to evaluate the experiment result. Therefore, F1-metric is used as the harmonic mean of precision and recall to combine these two metrics with equal weight [25], which is defined as follows.
Finally, precision, recall and F1-metric are calculated for each student. The total average of precision, recall and F1-metric of all students is computed for the top N recommendation.

4.3. Evaluation of the proposed method

Because the average ratings of the students are 9.09, the system recommends 10 articles for students. Table 2 demonstrates the evaluation result of these two article recommendation methods. The proposed method is compared with the typical CF method including precision, recall and F1-metric with top N recommendation. The top N recommendation is the top N ranked articles provided by the recommender systems. The proposed method recommends articles based on student’s rating, clustering and association rules which described in Section 3. The typical CF method recommends articles which are ranked by the frequency count of the articles computed by the users with the similar interests as described in Section 2.4. The proposed method performs well compared to the typical CF method under top N recommendation. The average precision, recall and F1-metric of the proposed method are larger than those of the typical CF method.

Table 2. Evaluation of the methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Proposed method</th>
<th>Typical CF method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>top N</td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>1</td>
<td>0.118</td>
<td>0.069</td>
</tr>
<tr>
<td>2</td>
<td>0.118</td>
<td>0.084</td>
</tr>
<tr>
<td>3</td>
<td>0.137</td>
<td>0.135</td>
</tr>
<tr>
<td>4</td>
<td>0.132</td>
<td>0.156</td>
</tr>
<tr>
<td>5</td>
<td>0.118</td>
<td>0.186</td>
</tr>
<tr>
<td>6</td>
<td>0.098</td>
<td>0.186</td>
</tr>
<tr>
<td>7</td>
<td>0.101</td>
<td>0.212</td>
</tr>
<tr>
<td>8</td>
<td>0.096</td>
<td>0.227</td>
</tr>
<tr>
<td>9</td>
<td>0.092</td>
<td>0.242</td>
</tr>
<tr>
<td>10</td>
<td>0.094</td>
<td>0.268</td>
</tr>
<tr>
<td>Average</td>
<td>0.100</td>
<td>0.160</td>
</tr>
</tbody>
</table>

The experiment result shows that the proposed method performs well compared to the typical CF method under top N recommendation, which is shown in Figure 3.

5. Conclusion

A combination of E-learning and the traditional classroom learning could create a blending-learning environment for students to learn anywhere and anytime. However, students all have the same latest articles even though their article preferences are different in an e-learning environment. Typical learning platform like learning management system (LMS) could not recommend the personalized articles for individual student.

In this paper, an article recommendation method is proposed based on student’s rating, which combines clustering technique and association rules to recommend the personalized articles to students in LMS. Student could rate the articles that classmate posted by the proposed rating mechanism, which is similar to the Like button in Facebook. The experiment result shows that the average precision, recall and F1-metric of the proposed method are larger than the metrics of the typical CF method. The proposed method performs well compared to the typical CF method under top N recommendation. The proposed method could recommend the better personalized articles to students than the typical CF method.

This study has some limitations. First, the rated articles and the rating students were both few. In the future, a reward mechanism will be established for students to increase the number of the rated articles and the rating students. Second, this study only considers nursing students. Besides, the other medical student of practitioner will be further considered. Third, the proposed method recommends the personalized articles based on students with the similar learning preferences. Actually, it may limit the choice of the reading list for students. They could not read the articles which are interesting to the dissimilar but successful students. In the future, the articles are interesting to the dissimilar but successful students are further considered.

Acknowledgement

This research was supported in part by the Ministry of Science and Technology of the Taiwan under Grant MOST 104-2410-H-227-002.
References


