A Learning-Community Recommendation Approach for Web-Based Cooperative Learning

Jian-Wei Li, Yao-Tien Wang and Yi-Chun Chang

Abstract—Cooperative learning has been defined as “learners working together as a team to solve a problem to complete a task or to accomplish a common goal”, which emphasizes the importance of interactions among members to promote the whole learning performance. With the popularity of society networks, cooperative learning is no longer limited to traditional classroom teaching activities. Since society networks facilitate to organize online learners, to establish common shared visions, and to advance learning interaction, the online community and online learning community have triggered the establishment of web-based societies. Numerous research literatures have indicated that the collaborative learning community is a critical issue to enhance learning performance. Hence, this paper proposes a learning community recommendation approach to facilitate that a learner joins the appropriate learning communities, which is based on k-nearest neighbor (kNN) classification. To demonstrate the viability of the proposed approach, the proposed approach is implemented for 117 students to recommend learning communities. The experimental results indicate that the proposed approach can effectively recommend appropriate learning communities for learners.

Keywords—k-nearest neighbor classification; learning community; Cooperative/Collaborative Learning and Environments.

I. INTRODUCTION

COLLABORATION has become an essential competency in the knowledge-based society [1-3]. Cooperative learning has been defined as “learners work together as a team to solve a problem, to complete a task or to accomplish a common goal” [3-5], which emphasizes the importance of interactions among learners to promote the whole learning performance [6, 7]. In other words, the cooperative learning model requires that members need to work together to accomplish a shared or a common goal rather than working alone [7, 8]. One other relative work [9] also indicated that the essential feature of cooperative learning is one learner helping the others to be successful. Furthermore, communication and information exchange among learners are the most important activity among partners to complete group goals [10, 11].

With the popularity of society networks, cooperative learning is no longer limited to traditional classroom teaching activities [12]. Since society networks facilitate to organize online learners, to establish common shared visions, and to advance learning interaction, the online community and online learning community have triggered the establishment of web-based societies [13]. The learning community is social scaffolding, in which peers communicate with each other and share their learning materials and experiences [13, 14]. The web-based learning community focuses on (1) the interaction and collaboration of relative community members, and (2) the sharing and distribution of knowledge and expertise by network technologies to enhance learning performances [1, 15]. Numerous research literatures related to the web-based cooperative learning have demonstrated that working cooperatively leads to better learning achievements than working individually [15-18]. Furthermore, the corresponding literatures have indicated that, the collaborative learning community is a critical issue to promote learning achievement and to share knowledge in web-based cooperative learning environments [3, 19].

However, it is usually difficult for a new participant to discover or identify whether a learning community is appropriate for herself/himself. Hence, in order to guide a new participant for joining a suitable learning community, this paper proposes a learning-community recommendation approach, which is based on k-nearest neighbor (kNN) classification [20, 21], to find the appropriate learning communities for a new community participator. The main reason for adopting kNN classification in this paper includes that, (1) kNN classification requires only a relatively small amount of samples, which corresponds to the initially developing situations of each learning community; and (2) the result recommended by kNN classification is with the priority ranking of learning communities instead of a unique learning community. In addition, the recommendation approach proposed in this paper is fully independent of learning communities and corresponding learning styles. That is, the proposed approach possesses compatibility, which can be adopted for any learning community and learning style according to users’ attributes. To demonstrate the viability of the proposed approach, this paper conducts relative experiments for 117 students to verify the recommendation accuracy.

The rest of this paper is organized as follows. Section II
briefly reviews related works. Section III presents the proposed learning-community recommendation mechanism in detail, while Section V demonstrates the experimental results. Finally, conclusions are drawn in Section V.

II. RELATED WORKS

The organization of learning community is the core of web-based cooperative learning [22, 23], which enhances the acquisition of knowledge and satisfies the learning demands of its members [23]. As stated in section I, a learning community is defined as a community in which members are joined together by common interests and particular themes in order to learn with each other, to exchange certain knowledge, and to construct new knowledge and meanings [3-5, 23]. Numerous relative literatures have been proposed various strategies to build learning communities for web-based cooperative learning environments, which include effective communication strategies, strengthening of social ties, small-team collaboration, social network establishment, collaborative knowledge construction, and so on [24, 25]. Tang and Chan (2002) presented a data-clustering method to classify learners, which considers both the interests and backgrounds of learners to build learning community [26]. Yang et al. (2007) have also considered the interest- and need-based clustering approach to regroup learning communities when the specific course progresses and learning concerns change [23]. In our previously researches, an GA-based aggregation mechanism has been proposed to form learning communities [27].

The aforementioned relative literatures have proposed the approaches to compose and to organize learning communities for the web-based cooperative learning environments. However, these approaches need to gather the related time-consumed data, including learners’ prior learning results, learning features and so on. Besides, for a new participator, how to find appropriate learning communities and to guide them to join learning communities are poorly discussed so far.

III. THE PROPOSED APPROACH

In this section, the learning community recommendation approach based on kNN classification is proposed.

A. Basic Definitions of KNN Classification

The kNN classification is known as one of most popular classification approaches [20, 21]. For kNN classification, a sample can be represented as a point in an n-dimensional space. By general definition, if two similar samples belong to the same class, their distance must be shorter than that of other not-the-same-class samples. For an unclassified data sample, the class label can be determined by computing the distance between it and other known samples. Suppose that an unclassified data sample contains N attributes, which is denoted as \( y_1(x_u), y_2(x_u), \ldots, y_N(x_u) \), 1 \( \leq n \leq N \), where \( y_i(x_u) \), 1 \( \leq i \leq N \), represents the \( i \)-th attribute value for the unclassified data sample \( x_u \). Similarly, a sample \( x_s \) also contains \( N \) attributes, which is denoted as \( y_1(x_s), y_2(x_s), \ldots, y_N(x_s) \), 1 \( \leq n \leq N \), where \( y_i(x_s) \), 1 \( \leq i \leq N \), denotes the \( i \)-th attribute value for the sample \( x_s \). The distance between an unclassified data sample \( x_u \) and a known sample \( x_s \) can be derived by

\[
d(x_u, x_s) = \sqrt{\sum_{i=1}^{N} (y_i(x_u) - y_i(x_s))^2}.
\]

This paper adopts the function of \( k \)-NN classification as

\[
f(x_s) = \arg \max \sum_{i=1}^{M} g(c_i, f(x_s)) = \max \sum_{i=1}^{M} g(c, f(x_s)) |c \in C = \{ c_1, c_2, \ldots, c_m \}; 2 \leq m \leq M,
\]

where \( g(x) \) denotes the class to which \( x \) belongs, \( x_1, 1 \leq i \leq k \), indicates the \( k \) closest samples, the function \( g(a, b) \) is used to determine whether \( a \) and \( b \) belong to the same class. If the result is positive, \( g(a, b) = 1 \). Otherwise, \( g(a, b) = 0 \). \( C \) denotes the set of classes and \( M \) is the number of classes.

After being computed, the \( k \) closest samples for the unclassified data sample \( x_u \) can be listed. A set \( S = \{(c_1, r_1), (c_2, r_2), \ldots, (c_m, r_m)\} \), and \( r_1 + r_2 + \ldots + r_m = k, 2 \leq m \leq M \), can be given, where \( (c_i, r_i), 2 \leq m \leq M \), represents that the class \( c_1 \) contains \( r_1 \) samples. If \( r_1 \geq r_2 \geq \ldots \geq r_m \) the class rank for the unclassified data sample \( x_u \) is \( \{c_1, c_2, \ldots, c_m\}, 2 \leq m \leq M \), it means that \( x_u \) has the higher priority belonging to the class \( c_1 \).

B. KNN Classification-Based Recommendation Approach

Since this paper adopts kNN classification to recommend learning communities, the corresponding parameters are defined as follows:

- \( AT = \{a_1, a_2, \ldots, a_N\}, 1 \leq n \leq N \), denotes the set of attributes for a learner, where \( N \) is the number of attributes and \( at_1 \leq i \leq N \), the \( i \)-th attributes.
- \( P_{\text{new}} \) represents a new participator, who needs to be recommended learning communities. Each learner contains \( N \) attributes, which is expressed as

\[
\overline{AT}(P_{\text{new}}) = \{a_1(P_{\text{new}}), a_2(P_{\text{new}}), \ldots, a_N(P_{\text{new}})\}, 1 \leq n \leq N,
\]

where \( at_{i}(P_{\text{new}}), 1 \leq i \leq N \), depicts the value of the \( i \)-th attribute \( at_i \) for the new participator \( P_{\text{new}} \).
- \( C = \{c_1, c_2, \ldots, c_m\}, 2 \leq m \leq M \), stands for the set of learning communities, where \( M \) is the number of learning communities and \( c_i, 2 \leq i \leq M \), the \( i \)-th learning community.
- \( DB_{\text{sample}} \) denotes a sample database that includes the attributes of learners who have joined a learning community, which is expressed as,

\[
DB_{\text{sample}} = \{L_{c_1}, L_{c_2}, \ldots, L_{c_m}\} = \begin{bmatrix}
\overline{AT}(L_{c_1})c_{1,1} \\
\overline{AT}(L_{c_2})c_{1,2} \\
\ldots \\
\overline{AT}(L_{c_m})c_{1,m}
\end{bmatrix}
\]
where $c_{x,z} \in C, 1 \leq z \leq Z$. $Z$ is the number of learners whom have joined learning communities, $L_{i,j}, L_{x,j}, 1 \leq i \leq Z$, the $i^{th}$ learner who have joined the learning community $c_{i,j}$, where $c_{x,z} \in C \setminus \{c_{x,j}, 1 \leq z \leq Z$ and $at(L_{i,j}), 1 \leq j \leq N$ and $1 \leq i \leq Z$, represents the value $at_i$ of the $j^{th}$ attribute for the learner $L_{i,j}$.

According to Error! Reference source not found., the distance between a new participator $P_{new}$ and a learners $L_{i,j}, 1 \leq i \leq Z$, who have joined learning communities, can be derived by

$$d(P_{new}, L_{i,j}) = \sqrt{\sum_{i=1}^{N} (at_i(P_{new}) - at_i(L_{i,j}))^2}, 1 \leq i \leq Z. \tag{4}$$

Similarly, according to Error! Reference source not found., the function of $k$-NN classification (Mitchell, 1997) in this paper can be redefined as,

$$f(P_{new}) = c_{(P_{new})}, \tag{5}$$

$$\sum_{i=1}^{N} g(c_{i}, L_{i,j}) = \max \sum_{i=1}^{N} g(c_{i}, L_{i,j}) |c_{i} \in C \setminus \{c_{x,j}, 1 \leq z \leq Z, if a = b, then g(a, b) = 1; if a \neq b, then g(a, b) = 0, 1 \leq i \leq Z. \tag{6}$$

Followed the same fashion stated in the previous subsection, the $k$ closest samples for the new participator $P_{new}$ can be calculated and listed. Given $S = \{(c_1, r_1), (c_2, r_2), ..., (c_m, r_m)\}$ and $r_1 + r_2 + ... + r_m = k, 2 \leq m \leq M$, where $(c_i, r_i), 2 \leq i \leq M$, represents the $i^{th}$ learning community $c_i$ contains $r_i$ samples. If $r_1 \geq r_2 \geq ... \geq r_m$, the ranks of recommended learning communities for the new participator $P_{new}$ are listed as $\{c_1, c_2, ..., c_m\}, 2 \leq m \leq M$. It indicates that the new participator $P_{new}$ has the higher priority belonging to the learning community $c_1$.

IV. EXPERIMENTATION

This section demonstrates the relative experimental results for the proposed procedure, which are based on actual student data collected from the learning behavioral features of 117 elementary school students [29]. Relative information of 117 elementary school students are collected and recorded for the following items [29]: (1) the serial number of each record; (2) the student’s identification; (3) the browsed learning unit; (4) the previous browsed unit; (5) the duration time at which the learning unit is browsed; (6) the IP address of the student; (7) the serial number of the e-learning course; and (8) the identification of the learning unit. By the two-stage cluster analysis, 117 elementary school students are classified into three kinds of learning styles [29], which are summarized as follows. (1) Dilatorily type: The students belonging to this type take more time to browse a learning unit than other students. They often review the same learning unit and skip learning units. (2) Transitory type: The students with this type generally spend the least amount of time instance in browsing and have the least browsing depth. In other words, the browsing order is irregular. (3) Persistent type: The students who belong to this type have the highest browsing depth and browsing order is regular. Furthermore, this experiment is divided into the following four phases for observing the corresponding accuracies, (I) Random phase: In order to execute $k$NN classification, the number of samples for each learning community in $DB_{sample}$ must contain at least $\left\lfloor \frac{k}{2} \right\rfloor$ samples. When the number of samples for each learning community is less than this amount, the recommend community for a new participator $P_{new}$ is randomly determined. (II) $k$-NN phase: When the number of samples for each learning community in $DB_{sample}$ is more than specified above, the $k$NN classification is conducted to find the appropriate learning communities for $P_{new}$. For each phase, 50 students are randomly selected to be included as the testing samples, and the remaining 67 students are put into the sample database $DB_{sample}$. For the above two phases, the selected 50 testing samples are classified to observe (A) the comparison of recommendation accuracy, and (B) the comparison of recommendation stability.

A. The Comparison of Recommendation Accuracy

TABLE I exhibits the recommendation accuracies for the two phases, where the recommendation accuracy can be derived as

$$CA_{k-NN}(DB_{sample}, AT) = \frac{T_{correct}}{T_{total}} \times 100 = \frac{T_{correct}}{50} \tag{6}$$

As shown in TABLE I, although that the highest accuracy in the random phase is 0.9 is better than the 0.89 in the $k$NN-based phases, the average accuracy in the proposed approach is 0.812, which outperforms the random phases (i.e. 0.77). Hence, the next experiment is for comparing the recommendation stability between random and the proposed $k$NN-based phase.

B. The Comparison of Recommendation Stability

The standard deviation $\sigma$ that is used to observe the recommendation quality can be derived by

$$\sigma = \sqrt{\frac{\sum (CA_{k-NN}(DB_{sample}, AT) - \frac{\sum CA_{k-NN}(DB_{sample}, AT)}{T_{total}})^2}{T_{total}}} \tag{7}$$
TABLE I

<table>
<thead>
<tr>
<th>Experiment Numbers</th>
<th>Classification Accuracy</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($CA_{i,N} \times DB_{i,NN}$, $\hat{A}$), $1 \leq i \leq 10$)</td>
<td></td>
</tr>
<tr>
<td>i=1</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>i=2</td>
<td>0.85</td>
<td>0.80</td>
</tr>
<tr>
<td>i=3</td>
<td>0.88</td>
<td>0.84</td>
</tr>
<tr>
<td>i=4</td>
<td>0.88</td>
<td>0.76</td>
</tr>
<tr>
<td>i=5</td>
<td>0.65</td>
<td>0.60</td>
</tr>
<tr>
<td>i=6</td>
<td>0.65</td>
<td>0.90</td>
</tr>
<tr>
<td>i=7</td>
<td>0.67</td>
<td>0.61</td>
</tr>
<tr>
<td>i=8</td>
<td>0.67</td>
<td>0.70</td>
</tr>
<tr>
<td>i=9</td>
<td>0.67</td>
<td>0.61</td>
</tr>
<tr>
<td>i=10</td>
<td>0.77</td>
<td>0.70</td>
</tr>
</tbody>
</table>

The KNN-Based Phases

TABLE II

<table>
<thead>
<tr>
<th>Phase</th>
<th>KNN-Based</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard Deviation ($\sigma$)</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>0.096</td>
<td></td>
</tr>
</tbody>
</table>

where $T_{total}$ refers to the total number of the recommendation times, and $CA_{i,N} \times DB_{i,NN}$, $1 \leq i \leq 10$, is the $i$th recommendation accuracy. Table II presents the numerical results.

As shown in TABLE II, the standard deviation in the KNN-based phase is 0.057, which is lower than the one in the random phase. In other words, the recommendation quality in the KNN-based phase is more stable than the random phase.

V. CONCLUSION

This paper proposes a KNN-based approach to suggest appropriate learning communities for a learner, in which the samples do not need to be collected in advance any more. The proposed approach has been implemented to demonstrate its viability. Based on the collected 117-student datasets, the series of conducted experiments have been completely verified the recommendation accuracy and recommendation stability for the proposed approach. The experimental results and corresponding analysis indicate that the proposed approach can effectively recommend learning communities for a new participant.

ACKNOWLEDGMENT

The research is supported by the National Science Council of the Republic of China under the grant number NSC101-2221-E-241-018 and NSC101-2221-E-324-043.

REFERENCES


