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

Jian-Wei Li is with the Department of Information and Communication Engineering, Chaoyang University of Technology, No. 168, Jifong East Road, Wufong Township, Taichung, Taiwan 41349, R.O.C. (phone: 886-4-23323000 Ext. 7833; e-mail: lijw@cyut.edu.tw).

Yao-Tien Wang is with the Department of Computer Science and Information Engineering, Hungkuang University, No.34, Chung-Chie Rd., Sha Lu, Taichung, Taiwan 43302, R.O.C. (phone: 886-4-26318652 Ext. 4206; fax: 886-4-26324084; e-mail: ytwang@sunrise.hk.edu.tw).

Yi-Chun Chang is with the Department of Computer Science and Information Engineering, Hungkuang University, No.34, Chung-Chie Rd., Sha Lu, Taichung, Taiwan 43302, R.O.C. (phone: 886-4-26318652 Ext. 4207; fax: 886-4-26324084; e-mail: changyc@sunrise.hk.edu.tw).

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 participator to discover or identify whether a learning community is appropriate for herself/himself. Hence, in order to guide a new participator 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

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:7, No:5, 2013

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 ith 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 ith 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

$$d(x_u, x_s) = \sqrt{\sum_{n=1}^{N} (y_n(x_u) - y_n(x_s))^2}. \tag{1}$$

This paper adopts the function of k-NN classification as [28]

$$
\begin{aligned}
&f(x_u) \leftarrow c_{x_u}, \\
&\sum_{i=1}^{k} g(c_{x_u}, f(x_i)) = \max\{\sum_{i=1}^{k} g(c_{x_u}, f(x_i)) \mid c_{x_u} \in C = \{c_1, c_2, \ldots c_m\}, 2 \leq m \leq M, \\
&\text{if } a = b, \text{then } g(a,b) = 1; \text{ if } a \neq b, \text{then } g(a,b) = 0\},
\end{aligned}
\tag{2}
$$

where $f(x)$ denotes the class to which x belongs, x_i, $1 \leq i \leq k$, 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_i contains r_i samples. If $r_1 \geq r_2 \geq \ldots \geq r_m$, the class ranks for the unclassified data sample x_u are $\{c_1, c_2 \ldots, c_m\}$, $2 \leq m \leq M$, 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:

- $\overrightarrow{AT} = \{at_1, at_2, \ldots, at_n\}$, $1 \leq n \leq N$, denotes the set of attributes for a learner, where N is the number of attributes and at_i, $1 \leq i \leq N$, the ith attributes.

- P_{new} represents a new participator, who needs to be recommended learning communities. Each learner contains N attributes, which is expressed as $\overrightarrow{AT(P_{new})} = \{at_1(P_{new}), at_2(P_{new}), \ldots, at_n(P_{new})\}$, $1 \leq n \leq N$, $at_i(P_{new})$, $1 \leq i \leq N$, depicts the value of the ith attribute at_i for the new participator P_{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 leaning communities and c_i, $2 \leq i \leq M$, the ith learning community.

- DB_{sample} denotes a sample database that includes the attributes of learners who have joined a learning community, which is expressed as,

$$
DB_{sample} = \{L_{s\_1}, L_{s\_2}, \ldots, L_{s\_z}\} =
\begin{bmatrix}
\overrightarrow{AT(L_{s\_1})}\, c_{s\_1} \\
\overrightarrow{AT(L_{s\_2})}\, c_{s\_2} \\
\ldots \\
\overrightarrow{AT(L_{s\_z})}\, c_{s\_z}
\end{bmatrix}
$$

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:7, No:5, 2013

$$= \begin{bmatrix} at_1(L_{s\_1})\, at_2(L_{s\_1})\ldots at_n(L_{s\_1})\, c_{s\_1} \\ at_1(L_{s\_2})\, at_2(L_{s\_2})\ldots at_n(L_{s\_2})\, c_{s\_2} \\ \ldots \\ at_1(L_{s\_z})\, at_2(L_{s\_z})\ldots at_n(L_{s\_z})\, c_{s\_z} \end{bmatrix}, \qquad (3)$$

where $c_{s\_z} \in C, 1 \leq z \leq Z$. Z is the number of learners whom have joined learning communities, L_{s_i}, L_{s_i}, $1 \leq i \leq Z$, the ith learner who have joined the learning community c_{s_i}, where $c_{s\_i} \in C \mid \forall c_{s\_i}, 1 \leq i \leq Z$ and $at_j(L_{s\_i})$, $1 \leq j \leq N$ and $1 \leq i \leq Z$, represents the value at_j of the jth attribute for the learner L_{s_i}.

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

$$d(P_{new}, L_{s\_i}) = \sqrt{\sum_{n=1}^{N} \left(at_n(P_{new}) - at_n(L_{s\_i}) \right)^2}, 1 \leq i \leq Z. \qquad (4)$$

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

$$f(P_{new}) \leftarrow c_{p_{new}}, \qquad (5)$$

$$\sum_{i=1}^{k} g(c_{p_{new}}, f(L_{s\_i})) = \max\{\sum_{i=1}^{k} g(c_{p_{new}}, f(L_{s\_i})) \mid c_{p_{new}} \in C = \{c_1, c_2, \ldots c_m\}, \; 2 \leq m \leq M,$$

$$if\ a = b, then\ g(a,b) = 1; \; if\ a \neq b, then\ g(a,b) = 0\}, 1 \leq i \leq Z.$$

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 $S = \{(c_1, r_1), (c_2, r_2), \ldots, (c_M, r_m)\}$ and $r_1 + r_2 + \ldots + r_m = k$, $2 \leq m \leq M$, where $(c_i, r_i), 2 \leq i \leq M$, represents that the ith learning community c_i contains r_i samples. If $r_1 \geq r_2 \geq \ldots \geq r_M$, the ranks of recommended learning communities for the new participator P_{new} are listed as $\{c_1, c_2 \ldots, c_m\}$, $2 \leq m \leq M$. 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 following.

(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 kNN classification, the number of samples for each learning community in DB_{sample} must contain at least $\left\lceil \dfrac{k}{2} \right\rceil$ 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 kNN 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

$$CA_{k-NN}(DB_{sample}, \overline{AT}) = \frac{T_{correct}}{T_{total}} \% = \frac{T_{correct}}{50} \qquad (6)$$

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 kNN-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 kNN-based phase.

B. The Comparison of Recommendation Stability

The standard deviation σ that is used to observe the recommendation quality can be derived by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{C_{total}} \left(CA_{k-NN}(DB_{sample}, \overline{AT})_i - \dfrac{\sum_{i=1}^{C_{total}} CA_{k-NN}(DB_{sample}, \overline{AT})_i}{T_{total}} \right)^2}{T_{total}}} \qquad (7)$$

TABLE I
THE LIST OF ACCURACY FOR A RANDOM AND THE PROPOSED APPROACH

Experiment Numbers	Classification Accuracy $(CA_{k\text{-}NN}(DB_{sample}, \overrightarrow{AT})_i, 1 \leq i \leq 10)$										Average
	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	
Random	0.74	0.85	0.80	0.84	0.88	0.76	0.65	0.90	0.67	0.61	0.770
The KNN-Based Phases	0.85	0.89	0.82	0.74	0.76	0.77	0.84	0.87	0.72	0.86	0.812

TABLE II
COMPARISON OF THE STANDARD DEVIATIONS FOR THE FOUR PHASES

Phase	KNN-Based	Random
Standard Deviation (σ)	0.057	0.096

where T_{total} refers to the total number of the recommendation times, and $CA_{k-NN}(DB_{sample}, \overrightarrow{AT})_i$, $1 \leq i \leq 10$, is the ith 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 participator.

ACKNOWLEDGMENT

REFERENCES

[1] Wang, Q., Design and evaluation of a collaborative learning environment, Computers & Education. 2009, 53(4), pp. 1138-1146.
[2] Wang, R., Wang, X. and Kim, M. J., Motivated learning agent model for distributed collaborative systems, Expert Systems with Applications. 2011, 38(2), pp. 1079-1088.
[3] Rae, J., Roberts, C. and Taylor, G. Collaborative Learning: A Connected Community Approach, in proceedings of Issues in Informing Science and Information Technology. 2006, 3, Manchester, Englandpp. 519-528.
[4] Artz, A. F. and Newman, C. M., Cooperative learning, Mathematics Teacher. 1990, 83, pp. 448-449.
[5] Hron, A. and Friedrich, H. F., A review of web-based collaborative learning: factors beyond technology, Journal of Computer Assisted Learning. 2003, 19(1), pp. 70-79.
[6] Hernandez-Leo, D., Bote-Lorenzo, M. L., Asensio-Perez, J. I., Gocmez-Sanchez, E., Villasclaras-Fernandez, E. D., Jorrin-Abellan, I. M. and Dimitriadis, Y. A., Free-and Open-Source Software for a Course on Network Management: Authoring and Enactment of Scripts Based on Collaborative Learning Strategies, IEEE Transactions on Education. 2007, 50(4), pp. 292-301.
[7] Johnson, D. W., Johnson, R. T. and Smith, K. A., Cooperative Learning: Increasing College Faculty Instructional Productivity. 1991: School of Education and Human Development, George Washington University.
[8] Anaya, A. R. and Boticario, J. G., Application of machine learning techniques to analyse student interactions and improve the collaboration process, Expert Systems with Applications. 2011, 38(2), pp. 1171-1181.
[9] Ellis, S. S. and Whalen, S. F., Cooperative Learning: Getting Started. 1990: Scholastic New York.
[10] Johnson, D. W. and Johnson, F. P., Joining together: group theory and group skills (7th ed.). 2000: Allyn & Bacon, Boston.
[11] Wardell, C. S. and Paschetto, G. Small group instruction in real-time over the web, in proceedings of the inter-service industry training simulation and education conference. 2001, Orlando, Florida.pp.
[12] Chuang, P.-J., Chiang, M.-C., Yang, C.-S. and Tsai, C.-W., Social Networks-based Adaptive Pairing Strategy for Cooperative Learning, Educational Technology & Society. 2012, 15(3), pp. 226–239.
[13] Chang, B., Cheng, N.-H., Deng, Y.-C. and Chan, T.-W., Environmental design for a structured network learning society, Computers and Education. 2007, 48(2), pp. 234-249.
[14] Brown, J. S. Leveraging technology for learning in the cyber age - opportunities and pitfalls, in proceedings of International conference on computers in education (ICCE 98) (Invited Speaker). 1998, pp.
[15] Lipponen, L., Hakkarainen, K. and Paavola, S., Practices and orientations of CSCL. In: J.W. Strijbos, P.A. Kirschner and R.L. Martens (Eds.), What we know about CSCL. 2004, Norwell, MA: Kluwer Academic. 31-50.
[16] Barron, B., Achieving coordination in collaborative problem-solving groups, The Journal of the Learning Sciences. 2000, 9(4), pp. 403-436.
[17] Chou, C. and Sun, C. T., A Computer-Network- Supported Cooperative Distance Learning System for Technical Communication Education, IEEE Transactions on Professional Communication. 1996, 39(4), pp. 205-214.
[18] Neo, M., Developing a collaborative learning environment using a web-based design, Journal of Computer Assisted Learning. 2003, 19(4), pp. 462-473.
[19] Lindberg, E. Networked, Problem Based, Collaborative Learning: Building and Sustaining Learning Communities, in proceedings of International Conference On Engineering Education 1999. 1999, Ostrava - Prague, Czech Republicpp.
[20] Chen, J. H., Yen, Z. H. and Ho, S. Y., Design of optimal nearest neighbor classifier using an intelligent multi-objective evolutionary algorithm, Lecture Notes in Computer Science (LNCS). 2004, 3157, pp. 262-271.
[21] Kuncheva, L. I. and Jain, L. C., Nearest neighbor classifier: Simultaneous editing and feature selection, Pattern Recognition Letters. 1999, 20(11-13), pp. 1149-1156.
[22] Chang, C.-K., Refining Collaborative Learning Strategies for Reducing the Technical Requirements of Web-Based Classroom Management, Innovations in Education and Teaching International. 2001, 38(2), pp. 133-143.
[23] Yang, F., Wang, M., Shen, R. and Han, P., Community-organizing agent: An artificial intelligent system for building learning communities among large numbers of learners, Computers and Education. 2007, 49(2), pp. 131-147.
[24] Rovai, A., Building classroom community at a distance: A case study, Educational Technology Research and Development. 2001, 49(4), pp. 33-48.
[25] Preece, J. and Maloney-Krichmar, D., Online Communities: Design, Theory, and Practice, Journal of Computer-Mediated Communication. 2005, 10(4).

[26] Tang, T. and Chan, K. Feature Construction for Student Group Forming Based on Their Browsing Behaviors in an E-learning System in proceedings of PRICAI 2002: Trends in Artificial Intelligence. 2002, 2417, Springer Berlin / Heidelberg, pp. 512-521.

[27] Li, J.-W., Chang, Y.-C. and Chiang, M.-L. A GA-Based Approach for Automatic Aggregation of Learning Community for Network Supported Cooperative Learning, in proceedings of The International Conference on Information Society (i-Society 2012). 2012, London, U.K.pp.

[28] Mitchell, T., Machine Learning. 1997: McGraw-Hill Companies.

[29] Chang, Y. C., Kao, W. Y., Chu, C. P. and Chiu, C. H., A Learning Style Classification Mechanism for E-Learning, Computers & Education. 2009, 53(2), pp. 273-285.