An Improved E-learner Community Construction Algorithm Based on Learning Interest Feature Vectors

FAN YANG1, BERND KRAEMER1, ZHIMEI WANG2, PENG HAN1
1Faculty of Mathematics and Information Technology
FernUniversitaet in Hagen,
Hagen, 58094
GERMANY
fan.yang@fernuni-hagen.de; bernd.kraemer@fernuni-hagen.de; han.peng@fernuni-hagen.de
2 Department of Computer
Wenzhou Vocational and Technical College
Wenzhou, 325035
CHINA

Abstract: - Finding similar e-learners in a distributed and open e-learning environment and help them to learn collaboratively is becoming one of the urgent challenges of personalized e-learning services. Literature shows that current e-learner community building approaches are generated from qualitative studies of small-sized learner-centered classrooms which may need the teacher's participation. However, the findings might not apply to large classes in distributed learning environments, which make the teachers to face hundreds of e-learners in each class. In such situations, teachers also find it impossible to analyze the learning behaviors of each e-learner and divide them into different learning communities accurately. This paper addresses this problem in the adaptive e-learner community self-organizing point of view. Considering both the feature vector of learning resources and a learner's rating value on each resource, this paper firstly defines the learning interest feature vector to model the learner's behavior. Based on this an accurate learning interest feature representation method and, an innovative e-learner community self-organizing algorithm, called IFV- SORC, are proposed in this paper. Experimental results show that this algorithm exhibits good community organizing efficiency and scalability.

Key-Words: - E-Learning, Collaborative Learning, Community Construction, Interest Feature Vector

1 Introduction
Web-based learning (or e-learning) provides an unprecedented flexibility and convenience to both learners and instructors but can also lead to isolated and un-motivated learners. In the past decade, a handful of research (e.g. [1] and [2]) has explored ways to create collaborative learning environments for geographically dispersed learners. However, most of these research findings are generated from a qualitative study of small-sized learner-centered classrooms; therefore, the findings might not apply to large classes (hundreds of e-learners of each class) and do not work in learning environments that feature teacher-centered instruction.

In such a situation, teachers also find it impossible to analyze the learning behaviors of each e-learner and divide them into different learning communities accurately. Thus, there is an urgent need to find similar e-learners in distributed and open e-learning environment and help them to learn collaboratively.

Traditionally the term "community" referred to geographic communities of people living in close proximity or to "communities of interest", those with shared interests such as members of an organization [3]. The successful construction of a community in the traditional sense is a social network marked by participation, trust, shared interests and values, shared responsibility, norms and rituals, and by the ability to embrace differences while forming a group identity. All these markers are transferable to a Virtual Learning Community (VLC), which brings together learners with shared interests to engage in discourse via media that transcend barriers of time and space [4-6]. However, most of these systems stress only the implementation of collaborative learning activities inside that group (e.g. collaborative learning, competitive learning etc.), and little attention is put on the formation of that group [7].

In [8], the authors create a cybergogy model for creating learning experiences that are cognitively, socially, and emotionally engaging for distance learners with diverse backgrounds. At the core of cybergogy is engaged learning, in which learners
establish their own goals, explore appropriate resources, work with others in groups, and construct knowledge in meaningful ways. However, most of the group-forming and community-building strategies derived from this model are created for online courses of reasonable size (20 to 50 learners); moreover, they are designed for courses that involve more constructive learning activities. Thus, these strategies do not always apply to large-size online classes. As an alternative, educators and technologists have resorted to the help of artificial intelligence to support learner collaboration. However, data related to learners’ prior learning and attributes are often unavailable or very time-consuming to gather, especially in large-size classes. In addition, existing grouping methods such as the aforementioned data-clustering often ignore the learners’ various learning behaviors, such as browsing online courses, evaluating and recommending learning materials to others. Each learning behavior reflects learners’ learning interests and needs for resources and information. As a consequence, the formed learning community may inevitably include some learners with similar approaches but different interests. Thus, this paper proposes an alternative method, which groups learners homogeneously, by tracking learners’ learning behaviors and assessing their current learning features, and then regroups them when their learning status changes.

The rest of this paper is organized as follows. In Section 2, the architecture of the e-learner community organizing system based on a multi-agent platform is presented and discussed. The key features and the pseudo-code of our algorithm are given in Section 3, and the experimental results of our system are presented and analyzed in Section 4. Finally a brief conclusion is given in Section 5.

2 Problem Formulation

Community organization here includes three core elements: how to define similar interest, how to identify learners with similar interest and how to re-organize similar e-learners into sub-communities in distributed environment.

Learning resources are the primary media of e-learning systems. These resources can be classified into different categories according to their context of use, such as business, education, law, literature, politics, science and technology. To identify e-learners with similar interest, the concept of recommendation is introduced which is expressed by a rating value an e-learner assigned to a resource.

This paper assumes that each e-learner of the e-learning system is willing to recommend a resource once she finds it interesting or is asked for a recommendation by others. Other e-learners who receive a recommendation will check whether they were also interested in this resource and whether they assigned a similar rating value to it. If their rating is close to that of another e-learner, we interpret this as a shared interest with respect to this resource. The interest sharing increases with every resource two or more e-learners rate similarly. E-learners with a high interest overlap will be grouped into the same sub-community. Based on this model, we define an e-learner community as a group of e-learners who share common interests and are mutually satisfied with resource recommendations received by community members.

To track the behavior of human learners, we propose to assign a learner agent (LA) to each person and let it handle the requests of an e-learner and seek desired recommendations from the community. To exploit the potential of sub-communities composed of e-learners with similar interest and to increase the efficiency of interest matches, we propose another kind of agent, called group agent (GA) that serves as a broker for learner agents. Their specific purpose is to provide a mechanism for appropriately grouping e-learners with matching interest into small sub-communities. In particular, a group agent is responsible for locating learners interested in recommended resources and managing the association of e-learners to communities, which can interact with both the local learner agents in their management domain and any group agent.

2.1 Learning Interest Feature Vector

In our previous research, we have proposed a community self-organizing algorithm called R3H-SORC [9]. In this method, we have taken it for granted that the resource with a higher rating value reflects a learner’s interest to a certain degree, and we gradually exploited the neighborhood relationship between learners who mutually have similar rating values on resources. However, each resource may have many feature keywords which may not be included in the resource title. That is to say, two learners may also give the same rating value on a resource while they are interested in different content.

In such cases, the matching method based on resource names sometimes leads to invalid results. Therefore, an effective method should be discovered to evaluate learners’ interests more precisely and include it in our algorithm. Inspired by resource similarity calculation methods used in content based
filtering technology, a learner modeling method is proposed based on the vector space model [10] by analyzing the keyword feature of the resources owned by the learner and his corresponding rating value.

Let G and U be disjoint countable sets of group and learner names, respectively, with typical elements g, g', g'' and u, u', u'', respectively. Further let R and K be sets of all possible resources and keywords.

For $R \subseteq R$ and $K \subseteq K$ with $|K| = n$, the mapping $\alpha : K \rightarrow R$ models the fact that keyword $k \in K$ exists in resource $r \in R$ if $\alpha(k) = r$. Then $\forall r \in R$, an n-dimensional feature vector $F^r = (f_{w_1}^r, ..., f_{w_j}^r, ..., f_{w_n}^r)$ is defined by equation 1:

$$f_{w_j}^r = \begin{cases} TF^r \cdot IDF^r_i & \text{if } \exists k_i \text{ such that } \alpha(k_i) = r \\ 0 & \text{otherwise} \end{cases}$$

with $i \in \{1, ..., n\}$, $TF^r$ and $IDF^r_i$ represent the term frequency and inverse document frequency of keyword $k_i$ in resource $r$, respectively. Each frequency weight $f_{w_j}^r$ can be calculated through equation 1 and finally each resource can be represented by the feature frequency vector $F^r$.

A learner agent is defined by combining the learner name with the learner’s resources and interest feature vector supporting their resources. A learner agent $A^u$ acting on behalf of learner $u$ is a structure $A^u = (R^u, RV^u, IFV^u)$ with:

- $R^u \subseteq R$ is the set of resources of $u$;
- $RV^u = \{w_{r_1}^u, ..., w_{r_m}^u\}$ is the rating vector of $u$ on the resources $r_1, ..., r_m$ held by learner $u$, where $w_{r_j}^u \in [0, 1]$;
- $IFV^u = \{v_{i_1}^u, ..., v_{i_n}^u\}$ is an n-dimensional vector representing learning features referred to as the learning interest feature vector (IFV), where $v_{i_j}^u$ is the interest strength of learner $u$ on keyword $k_i$ defined by equation 2:

$$v_{i_j}^u = \kappa \sum_{r \in R^u} f_{w_j}^r \cdot w_{r}^u$$

where $f_{w_j}^r$ denotes the corresponding feature value of keyword $k_i$ in resource $r \in R^u$, $w_{r}^u$ is the rating value on resource $r$ given by learner $u$, and $\kappa$ is the normalization factor defined by equation 3:

$$\kappa = \frac{1}{\sum_{i=1}^{n} v_i^u}$$

A self-organizing community $C$ consisting of a set $U$ of learners $u_1, ..., u_n$ and a set $G$ of group agents $g_1, ..., g_l$ is a structure $(G, (A^u)_{u \in U}, m)$ where $m : U \rightarrow G$ is a mapping associating learners (and thus learner agents) with their managing group agents. Thus, if $m(u) = g$, learner agent $A^u$ is a member of the group managed by $g$. The set $U^g$ of all learner agents managed by group agent $g$ is then defined by equation 4:

$$U^g = \{A^u \in U \mid m(u) = g\}$$

And the set of resources maintained by the community of learner agents managed by $g$ is defined by equation 5:

$$R^g = \bigcup_{A^u \in U^g} R^u$$

3 Community Self-organizing Algorithm based on Interest Feature Vector

Based on the discussion above, this paper introduces the self-organized community construction algorithm IFV-SORC according to the learner preference feature vector. The main improvements of this algorithm compared with the R3H-SORC algorithm defined in [9] are described in the following subsections.

3.1 Recommendation Request

In the R3H-SORC algorithm [9], once the learner has accessed and evaluated a resource, the learner agent will track this behavior and send a corresponding recommendation request to its group agent in order to locate similar learners based on the rating similarity on the recommended resource $r$. Since each learner may have rated many resources which they are not at all interested in, this paper introduces a pre-defined rating threshold $PreT$ to control the generation of recommendation request and avoid the extra communication traffic. That is, the learner agent submits a recommendation request only when the learner has voted a resource with a value higher than $PreT$. Here the recommendation request has the form $(w_r^u, IFV^u)$, where $w_r^u$ and $IFV^u$ are the recommended resource’s rating value and learner interest feature vector, respectively.

Lines 3-6 of the pseudo-code in Figure 1 model the recommendation request process, while Lines 7-23
considering the rating value. The improved similarity between learners instead of simply resource feature vector synthetically to calculate the similarity between learners instead of simply considering the rating value. The improved recommendation scheme is defined as follows:

1. **Local recommendation by a group agent:**
   When learner \( u \) sends a recommendation request to its group agent \( g \), \( g \) will forward this recommendation request to other local learner agents \( u \in U_g \) (Lines 5 and 7 in Figure 1). Each local learner agent \( u' \) will then verify whether the resource is in its local list.

   - If learner \( u' \) has evaluated the resource, the system will calculate the similarity \( \text{Sim}(w^{u}, w^{u'}) \) between the rating values of these two learners through equation 6:
     \[
     \text{Diff}(w^{u}, w^{u'}) = \frac{|w^{u} - w^{u'}|}{W_{\text{max}} - W_{\text{min}}} \tag{6}
     \]
     where \( w^{u} \) and \( w^{u'} \) are the rating values of the recommendation sender \( u \) and acceptor \( u' \) on resource \( r \), respectively. \(|w^{u} - w^{u'}| \) represents the maximum span of rating values used to normalize different rating schemes.

   - If \( \text{Diff}(w^{u}, w^{u'}) < \text{DiffT} \), where \( \text{DiffT} \) is a pre-defined different threshold, it means that these two learners have similar rating values on the recommended resource \( r \) (Line 11 in Figure 1). The learner agent \( u' \) will then calculate the interest similarity of \( u \) and \( u' \) based on the learner preference vector \( IFV^{u} \) and learner preference feature vector \( IFV^{u'} \) defined by equation 7:
     \[
     \text{Sim}(IFV^{u}, IFV^{u'}) = \frac{\sum_{k=1}^{n} W_{ik} \cdot W_{jk}}{\sqrt{\left(\sum_{k=1}^{n} W_{ik}^2\right)\left(\sum_{k=1}^{n} W_{jk}^2\right)}} \tag{7}
     \]
     where \( n \) is the dimension of the feature vector and \( W_{ik} \) is the \( k^{th} \) dimension of the vector. If \( \text{Sim}(IFV^{u}, IFV^{u'}) > \text{SimT} \), where \( \text{SimT} \) is a pre-defined similarity threshold between two learners, these two learners are considered with similar interest (Line 13 in Figure 1). Lines 9-18 model the local recommendation process.

2. **Global recommendation by group agents:**
   In R3H-SORC algorithm, the group agent \( g \) will only perform global recommendation requests when it fails locally. However, under actual conditions, when the community forms gradually, the success ratio of local recommendation requests continues to increase. This will prevent the formation of small communities. So in the IFV-SORC algorithm, whether its local recommendation request is successful or not, the system also requires the group agent \( g \) to perform global recommendations and deliver the recommendation request to other \( s < \text{topSearch} \) group agents \( g' \neq g \in G \) when it launches the local recommendation. Each \( g' \neq g \in G \) will perform the local recommendation process (Lines 20-23 in Figure 1).

### 3.3 Dynamic Switch Scheme Based on Recommendation Acceptance

In the R3H-SORC algorithm, the group membership is used to estimate the learner's interest strength belonging to the primary interest of its group. The system reorganizes the learner agent with similar ratings on the recommended resource according to the membership award.

In the IFV-SORC algorithm, it precisely evaluates the learner's interest based on its learner interest feature vector instead of the rating value on resources, and calculate their similarity based on the learner preference vector \( IFV^{u} \) and learner preference feature vector \( IFV^{u'} \). Furthermore, the recommendation acceptance is introduces to evaluate the learner's interest strength if it belongs to its current group. Here, recommendation acceptance is defined as how many learner agents in a common group accept this recommendation and have similar interests with the requesting learner agent.

In order to evaluate the acceptance strength of one group, a new parameter acceptance coefficient (AC) is proposed to calculate the number of accepted recommendations, where \( AC^{g}_{u} \) is the accepted numbers that the learner agent \( u \) received by the members of group agent \( g \). As discussed above, the requesting group agent will not only deliver the recommendation request to the learner agents in its own group, but also to several neighbor group agents. Thus, each learner agent who received the recommendation request will calculate the similarity \( \text{Sim}(IFV^{u}, IFV^{u'}) \).

If \( \text{Sim}(IFV^{u}, IFV^{u'}) > \text{SimT} \), it will return a positive feedback to its group agent. Each group agent who received the request including the requesting group agent will increase the AC by 1 once it receives a
positive feedback from its local learner agents and return the final value to the requesting group agent (Line 15 in Figure 1).

Based on the acceptance coefficient feedbacks from different group agents, the current group agent will decide whether to keep the learner agent in its group or switch it to another group. If the current group has the highest $AC$ value, then no adjustment will be made. Otherwise, the learner agent should be switched to the group with the highest $AC$ value. A corresponding adjustment on the learner list, resource list and degree of membership will then be made in consequence (Lines 24-30 in Figure 1).

4 Pseudo-Code of IFV-SORC Algorithm

Based on the improved similar learner matching and dynamic switch scheme, the detailed IFV-SORC algorithm is shown in Figure 1.

**Algorithm IFV-SORC**

1. let $m(u) := g. AC^n = 0$
2. while $e ≠ stop$
3.   if $w^n \gt Pmt$
4.     then {
5.       $u$ sends message $msg := (u, r, w^n, IFV^n)$ to $g$
6.     } endif
7. upon receipt of $msg$, $g$ broadcasts $msg$ to all $u' ∈ C$
8. for all $u' ∈ C$
9.   if exists rating $w^n'$
10.      then {
11.        if \[ \frac{|w^n - w^n'|}{|w^n|} ≤ DiffT \]
12.          then {
13.            if $sim(IFV^n, IFV^{n'}) > SimT$
14.              then {
15.                $AC^n = AC^n + 1$
16.              } endif
17.          } endif
18.      } endfor
19. $g$ broadcasts message $(u, r, w^n, IFV^n)$ to all $g' ≠ g ∈ G$
20. for all $g' ≠ g ∈ G$ received the message {
21.   perform Step 7 to 19 by replacing $g$ by $g'$
22. } endfor
23. if $AC^n'$ is the maximum acceptance coefficient
24.   then {
25.     if $AC^n ≠ AC^n'$
26.       then {
27.         $m := m - \{(u, g) \cup \{(u, g')\}$
28.       } endif
29.   } endif
30. endwhile

**Fig.1.** Community self-organization algorithm IFV-SORC

5 Experimental Design

In order to evaluate the performance of the IFV-SORC algorithm, we collected 100 web-page resources from the English Express web site of the Network College at Shanghai Jiaotong University in China, and extracted the feature frequency vector manually based on the Vector Space Model to construct a resource library. In the initialization process, 100 learners’ data is chosen and constructed the corresponding learner agents for them. Based on the history log data, the system could then obtain the learners’ evaluation vectors on the selected resources so as to initialize their read and unread resource list.

Furthermore, the preference feature vector of each learner could also be calculated from the data. The system constructed 15 group agents from which each learner agent should choose one to register in. After simulation, the system will simulate the community organization according to the IFV-SORC algorithm. Each learner agent will recommend the resources they prefer and the group agent will perform local or global recommendation and adjust the community structure accordingly.

As the community forms gradually, the inter-community learner switch times should drop and finally stabilize. So we define how to evaluate the community construction efficiency, which is the ratio of learner switch times compared to all recommendation requests, called as Exchange rate. The lower the exchange rate is, the more successful the community organization is. Figure 2 gives the corresponding analysis on the community construction efficiency. Similarly, on initialization, the inter-community switch is very frequent (approx. 0.57). After each learner sent 20 requests, the exchange rate will decrease to 0.13 and further to under 0.1 after 40 requests.

**Fig.2.** Graph of the community organization efficiency

Finally, in order to evaluate the scalability of the system, the system repeated the above experiment with 500, 1000 and 1500 learners. The corresponding
Exchange Rates under different learner scales are shown in Figure 3. From this figure we can see that under different scales of learner numbers the Exchange Rate stabilized on a small constant around zero, which proved the good scalability performance of our algorithm.

![Figure 3: Community organization efficiency under different learner scales](image)

**Fig.3.** Community organization efficiency under different learner scales

### 6 Conclusion

This paper firstly introduced a learning interest feature vector (IFV) considering both the feature vector of resources and learner's rating value on each resource. The IFV gives the evaluation of learner's interest precisely according to keywords of the content instead of categories of resources. Based on this accurate learner preference representation method, a new community self-organizing algorithm is proposed, called IFV-SORC, which provides an effective way to model the learner interest precisely so as to better measure the similarity between different learners. Furthermore, the scalability of this algorithm is also proven by the stabilization of the exchange rate under different scales of learner numbers.

### 7 Acknowledgement

The authors would like to thank the state government of North-Rhine Westfalia that awarded the first author with a Lise-Meitner fellowship allowing her to deepen her research on innovative personalized recommendation and advanced e-learning technologies.

### References:


