Discovery of Action Patterns and User Correlations in Task-Oriented Processes for Goal-Driven Learning Recommendation

Xiaokang Zhou, *Member, IEEE*, Jian Chen, *Member, IEEE*, Bo Wu, and Qun Jin, *Member, IEEE*

Abstract—With the high development of social networks, collaborations in a socialized web-based learning environment has become increasing important, which means people can learn through interactions and collaborations in communities across social networks. In this study, in order to support the enhanced collaborative learning, two important factors, user behavior patterns and user correlations, are taken into account to facilitate the information and knowledge sharing in a task-oriented learning process. Following a hierarchical graph model for enhanced collaborative learning within a task-oriented learning process, which describes relations of learning actions, activities, sub-tasks and tasks in communities, the learning action pattern and Goal-driven Learning Group, as well as their formal definitions and related algorithms, are introduced to extract and analyze users' learning behaviors in both personal and cooperative ways. In addition, a User Networking Model, which is used to represent the dynamical user relationships, is proposed to calculate user correlations in accordance with their interactions in a social community. Based on these, an integrated mechanism is developed to utilize both user behavior patterns and user correlations for the recommendation of individualized learning actions. The system architecture is described finally, and the experiment results are presented and discussed to demonstrate the practicability and usefulness of our methods.

Index Terms—Learning pattern, user correlation, learning action recommendation, collaborative learning, learning task, enhanced social learning

1 Introduction

With the rapid development of emerging computing paradigms, such as ubiquitous computing, cloud computing, and social computing, we have been continuously experiencing a tremendous change in the web-based learning environment. Learning in a technology-enhanced e-learning system no longer means students always have to follow the defined instruction and assessment procedure to conduct the pre-planned curriculum within a fixed knowledge domain. Meanwhile, since more and more people have been increasingly accustomed to sharing feelings, experience, and knowledge through the online social networking with each other due to the popularity of ubiquitous devices and the high accessibility of communication systems, people no longer only learn from instructors or study by themselves, but can also learn more from the communications with others in a community.

Social learning, which indicates people can learn by observing the behavior of others and the outcomes of those behaviors [1], focuses more on the learning collaborations that occurs within a social context [2]. That is, social learning puts more emphasis on learning through interactions and collaborations in a community or across a social network. Following this way, user behaviors, as well as the interactions with others, should be viewed as an important recourse to form the social knowledge in a rapid-changing social environment. Moreover, differing from the traditional e-learning paradigm, social learning, enhanced by these emerging technologies, can be more flexible to deal with the delivering and exchanging of the potential knowledge that is dynamically generated anytime and anywhere. Vassileva [3] pointed out the fundamental issue of social learning relies on the delivery of right information to appropriate social groups in favor of the stated perspectives. Rienties et al.'s work [4] indicates that both motivation and performance can be improved within the collaborative works in a learning process. Thus, in addition to the pre-defined instructions, the social knowledge, including the information behaviors along with users' learning experience, should be appropriately delivered to the suitable group of users, in order to pursue a more productive and cost-effective learning and education process for both students and instructors.

To cope with this situation, in this study, two important factors, user learning behaviors and their interactions, which are leveraged for discoveries of action patterns and user correlations, will be taken into account to better facilitate the learning collaborations in the web-based learning environment. In details, this paper makes following concentrations on:

- A hierarchical model to describe the relations among learning actions, activities, sub-tasks and

- *X. Zhou and B. Wu are with the Graduate School of Human Sciences, Waseda University, Tokorozawa-shi, Saitama, Japan. E-mail: {xkzhou, wubo}@ruri.waseda.jp.*
- *J. Chen is with the College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan, Shanxi, China. E-mail: chenjian02@tyut.edu.cn.*
- *Q. Jin is with the Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University, Tokorozawa-shi, Saitama, Japan. E-mail: jin@waseda.jp.*

Manuscript received 9 June 2013; revised 20 Oct. 2013; accepted 9 Dec. 2013. Date of publication 1 Jan. 2014; date of current version 12 Sept. 2014.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TLT.2013.2297701

tasks in a user community for the task-oriented learning process.

- The learning behavior modeling which includes the learning action pattern (LA-Pattern) to discover and represent an individual user's learning behavior patterns extracted from sequences of learning actions, and the Goal-driven Learning Group to analyze and describe the similarities of learning behaviors among a group of users.
- A user networking model with a set of measures to graph and quantify users' relationships in accordance with their direct and indirect interactions in a collaborative learning process.
- An integrated mechanism for the goal-driven learning recommendation based on learning behavior and user correlation analysis.

The rest of this paper is organized as follows. We give a brief overview on the related issues and works in Section 2. The behavioral modeling, including the learning action pattern and Goal-driven Learning Group, is introduced and defined in Section 3. The structural analysis for user correlations according to the user networking model is described in Section 4. The integrated mechanism for individualized recommendation of learning actions is described in Section 5. The experiment analysis and evaluation on this work are discussed in Section 6. We conclude this study and give some promising perspectives on future works in Section 7.

2 Related Work

Three main issues related to this study are walked through in this section. That is, issues of user behavior modeling and pattern, analyses for user communities and relationships, and studies on collaborative leaning and task-oriented research are addressed respectively.

2.1 User Behavior Modeling and Pattern Analysis

The user behavior modeling, as well as the behavior pattern analysis, has been developed by more and more researchers these years [5], [6], [7], [8], [9], [10], [11], [12], [13]. Razmerita [5] has proposed a generic ontology-based user modeling framework (OntobUMf) to model the user behaviors, and used it for the user classification, in order to enhance the personal knowledge management. Stolfo et al. [6] have used the Email Mining Toolkit (EMT) empowered with behavior-modeling techniques to compute behavior profiles of user email accounts, which can help detect the viral propagation problem. Chen et al. [7] proposed a data mining method to extract the user movement behavior patterns, in order to predict and recommend suitable services for users in a mobile service environment. Considering both user behaviors and collaborative filtering, Liu et al. [8] have proposed a semantic relatedness measure between words to retrieve related words and detect new word tasks, which can help enrich user experience and discover hidden information. Lee et al. [9] developed a non-supervised learning framework to discover the behavior patterns, in which a new cluster validity index was proposed for agglomerative iterative Bayesian fuzzy clustering, and the fuzzy-state Q-learning was proposed to learn the sequential actions. Yun and Chen [10] have proposed a model to mine mobile sequential patterns, in which both user moving patterns and purchase patterns have been taken into account, and they have further devised three algorithms (TJLS, TJPT, and TJPF) to determine the frequent sequential patterns in the mobile commerce environment. Munoz-Organero et al. [11] have utilized the behavior patterns to predict student motivation, which can further be used to predict the successful completion of an e-learning course, based on the analysis of relationships between the motivation and performance of 180 students who took an e-learning course deployed on a Moodle e-learning platform. Plantevit et al. [12] have presented a method to mine sequential patterns from multidimensional and multilevel databases, which took account of different dimensions and levels of granularity, in order to discover the regular specific patterns. Zhao and Ooi [13] proposed an access-pattern-driven distributed caching middleware named APRICOD, which gave more consideration on user interactions for media streaming applications.

2.2 User Community and Relationship Analysis

User relationship modeling and analysis, as well as their applications, have drawn a large body of researches in social media environment [14], [15], [16], [17], [18], [19], [20]. Wilson et al. [14] have proposed the interaction graphs to quantify user interactions in Facebook, in order to answer the question: "Are social links valid indicators of real user interaction?" Lin et al. [15] proposed the MetaFac framework which utilized various social contexts and interactions for community structure extraction, to support the community discovery process. Based on the analysis of the correlation between social and topical features in online social networks such as Flickr, Last.fm, and aNobii, Aiello et al. [16] built a user similarity network to perform the prediction of friendships. Leskovec and Horvitz [17] have constructed a communication graph to examine characteristics and patterns based on the collective dynamics of 240 million users rather than the individual actions or characteristics. Aiello et al. [18] analyzed the social network in a social bookmarking system named Nobii, in order to investigate the interplay of profile similarity and link creation according to interest-based factors. In order to find the strong relationships automatically in social networks, Xiang et al. [19] have developed a latent variable model to infer the relationship strength, in which both profile similarity and interaction activity are taken into account to improve the strength estimation. Leroy et al. [20] proposed a so-called cold start link prediction method which could detect potential social graph by using group membership information obtained from Flickr.

2.3 Collaborative Learning and Task-Oriented Study

As for the collaborative learning aspect [21], [22], [23], [24], [25], [26], Hamada [21] proposed an integrated web-based environment as a learning tool to support the active and collaborative learning, which can improve learners' motivation and performance for more knowledge and information seeking. Evidences in [22] revealed that the wiki-based collaboration system could prompt learners to share information and raise their interests in examining the discussion series. For the agent architecture design, Kumar and Rose [23] presented the Basilica, which adopted an object-oriented method to provide a capability for rich behavior expressions. Perera et al. [24] have utilized the clustering and sequential pattern mining methods to analyze the data generated from group working students, in order to support learning group skills in the context of a standard state-of-the-art tool. Boticki et al. [25] designed an architecture in which different contents or materials could be assigned to students to build the group, to support the content-independent collaborative mobile learning in a classroom. Papadimitriou et al. [26] developed an adaptive educational hypermedia system called MATHEMA, which provided users with the navigation support, in order to improve their performances through an interactive and constructivist environment.

Researchers have also considered more about the concept of task in their studies [27], [28], [29], [30], [31], [32], [33]. Benbunan-Fich et al. [27] developed a set of activity-based metrics for computer-based multitasking by measuring multitasking behaviors, which can be used to establish a conceptual and methodological foundation for further multitasking studies. Fetaji and Fetaji [28] used the task based learning model to develop and analyze a mobile software solution in order to enhance learning in the university environments. Ozawa et al. [29] have proposed an online learning algorithm, which can automatically detect task changes and transfer knowledge of previous tasks from one task to another, in order to solve the multitask pattern recognition problems. Benbunan-Fich et al. [30] have conceptualized the multitasking and developed a series of different metrics to investigate computer-based multitasking behavior, which can provide a conceptual and methodological way to better understand users' multitasking behaviors. Mairal et al. [31] proposed a task-driven framework for dictionary learning formulation, which can adapt to various tasks rather than only adapting to data reconstruction, in order to solve a few of issues, such as regression and classification. In order to facilitate the identification and detection of users' off-task behaviors, Cetintas et al. [32] proposed a machine learning model to analyze users' available actions in an intelligent tutoring system with multi-features. Based on the data-driven techniques, Bangalore et al. [33] have built the task structures for task-oriented individual dialogs, which can help classifications of dialog act and task/sub-task, as well as the further predictions of them.

2.4 Section Summary

Modeling for individual users or user communities, as well as their analyses, can provide comprehensive support not only in improvements of system performance, but also in individualized recommendations and predictions. Research works also pointed out the trend of task-based studies and applications, which can help better understand user behaviors in collaborative works, especially when facing multi-task issues. In this study, we try to find a way to describe and analyze users' learning behaviors and their interactive relationships in a task-oriented collaborative learning process, in order to support information and knowledge sharing in the technology-enhanced social learning environment.

3 Behavioral Modeling of Learning Activities

In this section, after introducing the modeling of a task-oriented learning process, we describe the concepts of learning action pattern and Goal-driven Learning Group to discover and analyze users' learning behaviors expressed by learning actions, the generation algorithms of them are also discussed respectively.

3.1 Definitions of Learning Actions

A series of definitions shall be introduced and defined in order to describe user information behaviors in a task-oriented learning process.

Learning action: A learning action is composed of the minimum unit of learning operations. A learning action may consist of a series of learning operations. For example, learning new foreign language words is regarded as a learning action, in which two operations, reciting and handwriting new language words, can be employed for two minimum units of it.

Learning activity: A learning activity is a set of learning actions, which constitute a purposeful learning process with a certain learning action sequence and time span. It is an educational process or procedure intending to motivate learning through actual experience. For example, in a foreign language lesson, a learning activity consists of learning actions—learning new words, new grammars and texts, doing exercises and quizzes. In this situation, the goal of this action sequence is to finish the final quiz. Besides, the sequence of learning actions for a learning activity can be optimized so as to satisfy different target students.

Learning sub-task: A learning sub-task is a set of learning activities toward a certain learning purpose in a specific learning stage. For example, each lesson could be viewed as a learning sub-task in the whole foreign language learning course.

Learning task: A learning task is a set of learning sub-tasks. It contains all learning action sequences that represent a complete learning process.

3.2 Hierarchical Model for Task-Oriented Learning Process

Following the definitions above, a hierarchical model is addressed to interpret the structure and relations in task-oriented learning processes.

Generally, as shown in Fig. 1, a learning course may be divided into several learning tasks with a certain sequence to achieve the final learning purpose. As for each learning task, it shall contain several sub-tasks in different learning stage with a specific sub-purpose. Continuously, in each learning sub-task, a series of learning objectives shall be established in different learning periods for different users, followed by a series of learning activities which are assigned to realize them. Likewise, the learning activity can further be divided into a sequence of learning actions which are composed of the minimum units of the learning operations. In details, they can be recorded as what has been done at

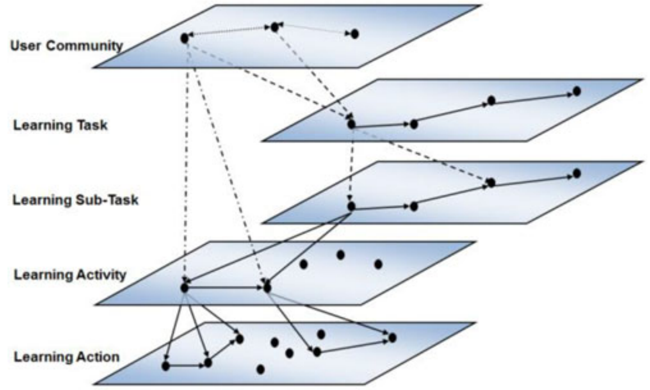


Fig. 1. Model of task-oriented learning process.

what time with the necessary materials for a specific learner. Take a specific English learner for instance, a learning action sequence can be recorded as memorizing new words in early morning, learning English grammar in the morning, and doing exercise in the afternoon.

As for users in a learning course, one learning task may be assigned to a group of users. To pursue the similar learning purpose, users may have some discussions through a web-based communication system, which may generate a number of direct interactions (such as reply, mention) and a variety of indirect interactions (e.g., posts containing similar topics). That is, all these direct and indirect interactions, which may contain users' hidden intentions, can be utilized to analyze user correlations within a learning task during a specific period. Moreover, based on these dynamical relationships calculated among users, the similar learning actions, especially those successful sequences of learning actions, can be shared along with successful learning experience and user-generated social knowledge.

3.3 Generation of Learning Action Patterns

The learning action pattern is defined to discover and describe individual users' information behavior patterns within a task-oriented learning process. Therefore, in order to calculate these patterns, the task-oriented learning shall be formalized as follows:

$act = \{U, O, Lr\}$: a non-empty set to describe the learning action which is the minimum unit in learning behavior modeling, where U indicates the user whom a specific learning action belongs to, O indicates the concrete operation of this learning action (e.g., submitting a report), and Lr indicates the learning resources that the user U has used to finish this learning action.

$Act = < act_1, act_2, \ldots, act_n, G >$: a non-empty set to describe the learning activity, represented as a sequence of learning actions, where act_i indicates the learning action that belongs to this learning activity, and G is a special learning action that indicates a goal of this learning action sequence. For example, if G is "start taking a quiz", the $< act_1, act_2, \ldots, act_n >$ indicates a series of learning actions that are performed to prepare for a better outcome of the quiz.

$S\text{-}Task = < Act_1, Act_2, \ldots, Act_n, T >$: a non-empty set to describe the learning sub-task, represented as a sequence of learning activities, where Act_i indicates the learning activity that belongs to this learning sub-task. T indicates a learning stage within the whole learning task, which can also be viewed as an end of time interval. For example, if the whole learning task is an English language learning course, each lesson can be viewed as a *S-Task*. Thus T indicates the learning period during this lesson, while $< Act_1, Act_2, \ldots, Act_n >$ indicates the learning activities that are taken for this lesson.

Based on these, the LA-Pattern, which is a sub-sequence of learning actions, can be defined as:

$$< act_i >^w_u \rightarrow G, \tag{1}$$

where $< act_i >$ denotes the learning action sequence which tends to constitute a purposeful learning process. w denotes the weight of this LA-Pattern, and specifically, the value indicates the frequency that this segment occurs in a whole learning action sequence (e.g., a learning action sequence generated from a learning sub-task). u denotes the user whom this learning action sequence belongs to. G denotes the learning goal of this sequence. In this way, the LA-Patterns can be viewed as a set of sequential learning behaviors frequently occurring in a specific user's learning action sequence, which intends to complete a certain learning purpose.

The *trie* [34], an ordered tree-based structure which can be used to store a dynamic string-like data set, has been well developed and applied in information storing and retrieving. For instance, Iglesias et al. [35] have applied the *trie* data structure in behavior profile creation and recognition for a computer user. In this study, we employ this tree-based data structure to find all the related sub-sequences with their frequency in a given learning action sequence, in order to calculate the weight w of each LA-Pattern. In particular, a certain learning action sequence with its subsequence suffixes which extend to the end of this sequence shall be all inserted into a *trie*, in order to calculate the frequency of each sub-sequence during the tree building process. For example, if the whole sequence is *<A, B, C, D>*, three sub-sequences *<B, C, D>*, *<C, D>* and *<D>* shall also be inserted. In this study, for a specific user, the input set of the learning action sequences shall consist of his/her learning sub-tasks.

Based on these discussed above, two criteria are given to generate the LA-Patterns.

Criteria 1—Basic criteria: Given a pre-defined learning goal set $G = \{G_1, G_2, \ldots, G_m\}$, and a sub-sequence q described as $< Act_1, Act_2, \ldots, Act_j, Act_n >^w_{u_i}$, if it satisfies that $n >= 2, w >= 2$, and $Act_n \in G$, then q is a LA-Pattern for user u_i, which can be described as $< act_1, act_2, \ldots, act_j >^w_{u_i} \rightarrow G_k$.

Criteria 2—Incorporation criteria: Given two sequences $q_1 : < act_1, act_2, \ldots, act_n >^{w_x}_{u_i}$ and $q_2 : < act_1, act_2, \ldots, act_m >^{w_y}_{u_i}$ for user u_i, if they satisfy that $w_x = w_y$, and $< act_1, act_2, \ldots, act_n > \subset < act_1, act_2, \ldots, act_m >$, then q_1 can be incorporated into q_2.

The algorithm for LA-Pattern generation based on these two criteria is described in Fig. 2.

3.4 Goal-Driven Grouping Based on LA-Patterns

Following the formalization and generation of LA-Patterns, in order to analyze the similarities in a user community in

Input: The Learning action sequence set $\{<act_1, act_2, \ldots, act_n>\}$
Output: The LA-Pattern set $\{< act_i >_u^w \rightarrow G\}$

Step 1: For each user u_i, divide the whole learning action sequence $< act_1, act_2, \ldots, act_n>$ into several sub-sequences in each corresponding pre-defined learning period T, which can be recorded as q_i: $\{< act_1, act_2, \ldots, act_j >_{T_1}, \ldots, < act_1, act_2, \ldots, act_k >_{T_n}\}_{u_i}$
Step 2: For each q_i, insert each sequence element with their corresponding subsequence suffixes into a *trie*, in order to build a tree-based data structure for each user u_i
Step 3: For each user u_i, traverse the *trie* from the root node to each leaf node in order to acquire all the sub-sequences with their corresponding frequency, which can be recorded as *sq*: $\{\{< act_k >^w\}_{u_i}\}$
Step 4: Filter each element in *sq* according to *Criteria 1*, if $k >= 2$, $w>=2$ and $act_k \in G$, remain the satisfied sequences, which can be recorded as *rsq*: $\{\{< act_j >^w \rightarrow G_n\}_{u_i}\}$
Step 5: For $\forall rsq_i, rsq_j \in rsq$, according to *Criteria 2*, if $w_x = w_y$, and $< act_i > \subset < act_j >$, let $rsq_j = rsq_i \cup rsq_j$
Step 6: Return the *LA-Pattern*: $\{< act_i >_u^w \rightarrow G\}$ for each user u_i

Fig. 2. Algorithm for LA-pattern generation.

accordance with each individual user's learning behaviors, we introduce the concept of Goal-driven Learning Group.

Goal-driven Learning Group: a Goal-driven Learning Group is given to model learning behaviors of a certain group of users, who have the same learning goal and similar learning actions according to the LA-Patterns in a specific learning period, which can be formalized as:

$$\left\{ \left[< act_k >^{w_1, w_2, \ldots, w_i}_{u_1, u_2, \ldots, u_i} \right] [G_j, T] \right\}, \tag{2}$$

where, $< act_k >$ denotes the learning action sequence in a specific LA-Pattern, u_i denotes the owner of this LA-Pattern, while w_i denotes the corresponding weight. G_j denotes goal actions representing the learning purpose for this learning group, and T denotes the specific learning period.

The Goal-driven Learning Group primarily refers to two portions: users' learning behaviors and the corresponding learning purpose. The former one represents the similarity of learning behaviors among users based on the LA-Patterns, while the latter one indicates the same learning goal that these users try to achieve within a specific learning period.

The learning period can be defined according to different factors and granularities. For instance, according to the curriculum, the learning period can be defined as one course or one lesson, while according to the time, the learning period can be defined as one month, one week, or one day. In each period, a group of users' learning patterns will be first grouped by their different learning purposes which indicate as goal actions, and then be further grouped by calculating the similarity of the corresponding learning action sequence $< act_k >$. That is, users will be assigned into groups according to their same learning purpose and similar learning behaviors. Note that, this process is not a partition, which means one user can be assigned into several Goal-driven Learning Groups. The algorithm to compute the Goal-driven Learning Group is described in Fig. 3.

Based on these discussed above, it can be viewed as that users in the same learning group may have the most similar learning behaviors when they pursue the same learning

Input: The LA-Pattern set $\{< Act_i >_u^w \rightarrow G\}$
The learning purpose set $\{G_n\}$
Output: The Goal-driven Learning Group set $\{\{[< Act_k >^{w_1, w_2, \ldots, w_i}_{u_1, u_2, \ldots, u_i}][G_j, T]\}\}$

Step 1: For each user u_i, according to the learning purpose set $\{G_n\}$, divide the LA-Pattern set $\{< act_i >_u^w \rightarrow G\}$ into several sub-sets, which can be recorded as $\{[< act_j >_u^w, \ldots, < act_k >_u^w] \rightarrow G_n\}_{u_i}$
Step 2: group the learning action pattern set $\{[< act_j >_u^w, \ldots, < act_k >_u^w] \rightarrow G_n\}_{u_i}$ by selecting the learning action sequence $[< act_j >_u^w, \ldots, < act_k >_u^w]$ with the same learning purpose G in a specific learning period T
Step 3: In each group, incorporate each $< Act_j >_u^w$ which may have the same learning action sequence, and record as $[< Act_k >^{w_1, w_2, \ldots, w_i}_{u_1, u_2, \ldots, u_i}]$
Step 4: Return the $\{\{[< Act_k >^{w_1, w_2, \ldots, w_i}_{u_1, u_2, \ldots, u_i}][G_j, T]\}\}$ as the final the Goal-driven Learning Group set

Fig. 3. Algorithm for Goal-driven Learning Group generation.

goal within a task-oriented learning process. Thus, in order to facilitate their collaboration works to reach better learning efficiency, in our system, we will try to share the information and knowledge among them by recommending a target user with the selected suitable learning actions as the possible further learning guide.

3.5 An Example

We consider three users with their learning action sequences to form an input user group as a simple example to demonstrate the generation process of LA-Patterns, as well as Goal-driven Learning Groups based on our algorithms.

Given users u_1, u_2, and u_3, with their recorded learning action sequences, $s_1 = < egiuvgiuvgi >$, $s_2 = < uvuvgiuvgi >$, and $s_3 = < uvgiuvgiuvgigi >$ in the same period $T = one\ week$.

As for user u_1, using the *trie*-based algorithm, the subsequences of learning actions with their corresponding frequencies can be calculated as $sq_{u_1} = \{< gi >^3, < egi >^1, < vgi >^2, \ldots\}$.

Assume the learning purpose is denoted by learning action i, based on *Criteria 1*, the preliminary learning action pattern set can be calculated as $rsq_{u_1} = \{< g >^3 \rightarrow i, < vg >^2 \rightarrow i, < uvg >^2 \rightarrow i\}$. According to *Criteria 2*, the sequences $< vg >^2$ and $< uvg >^2$ satisfy $< vg > \subset < uvg >$, and $w_{vg} = w_{uvg} = 2$. Thus we combine these two sequences together, and the LA-Pattern set for user u_1 can be calculated as $\{< g >^3_{u_1} \rightarrow i, < uvg >^2_{u_1} \rightarrow i\}$. Similarly, the patterns for other two users can be calculated as $\{< g >^2_{u_2} \rightarrow i, < uvg >^2_{u_2} \rightarrow i\}$, and $\{< g >^4_{u_3} \rightarrow i, < uvg >^3_{u_3} \rightarrow i\}$. Furthermore, the Goal-driven Learning Group for these three users during *one week* can be calculated as $\{[< g >^{3,\ 2,\ 4}_{u_1, u_2, u_3}, [< uvg >^{2,\ 2,\ 3}_{u_1, u_2, u_3}][i, one\ week]\}$.

4 Structural Analysis for User Correlations

In this section, we construct a user networking model in order to represent and analyze the user correlations based on users' direct and indirect interactions, which can further facilitate information and knowledge sharing in a task-oriented learning process.

4.1 Building User Networking Model

In our previous study, the so-called dynamically socialized user networking ($DSUN$) model has been proposed and constructed to discover and represent users' potential profiling and dynamical correlation based on the analysis of social streams [36], which can be defined and expressed as:

$$G(U, E, W), \tag{3}$$

where

$U = \{u_1, u_2, u_3, \ldots, u_n\}$: a non-empty set of vertexes in the network, each of which indicates a user in the system.

$E = \{e_{ij} : < u_i, u_j > \mid \mathit{if\ correlation\ exists\ between\ } u_i \mathit{\ and\ } u_j\}$: a collection of edges that connect the vertexes in U, which represent all the relationships among every vertex in the network.

$W = \{w_{ij} \mid \mathit{if} \exists e_{ij} \in \mathrm{E}\}$: the weight w_{ij}, appending on the corresponding edge, is developed to identify the strength of specific relation between two users. This value is employed to dynamically construct the model.

In this study, we apply and refine this model to build a user networking in a learning process, in order to support the calculation of similarity among users. In details, two important factors, users' communication actions and similarities of their posted contents, which indicate direct interactions and indirect interactions among users, are taken into account to build the user networking within a learning process. Thus, Eq. (4) is developed to quantify the value of w_{ij} between two vertexes u_i and u_j, which denote two related users:

$$w_{ij} = \alpha * w_{ar} + \beta * w_{cr}, \tag{4}$$

where w_{ar} denotes the relationships based on users' direct communication actions, and w_{cr} denotes the relationships based on similarities of users' posted contents. α and β are the importance coefficients which satisfy $\alpha + \beta = 1$.

The user relationship based on direct interactions, which can be viewed as direct relationship among users, is calculated in accordance with users' direct communications. That is, the value of w_{ar}, ranging from 0 to 1, indicates the interested degree from user u_i to user u_j based on the direct interaction times during a specific learning period (e.g., reply, mention), which can be quantified as:

$$w_{ar} = \frac{IR(u_i, u_j)}{IR(u_i)}, \tag{5}$$

where $IR(u_i, u_j)$ denotes the interaction number from user u_i to user $u_j, IR(u_i)$ denotes the total interaction number of user u_i.

The user relationship based on indirect interactions, which can be viewed as indirect relationship among users, is calculated in accordance with the similarities of users' posted contents. That is, the intentions hidden in users' posted contents will be taken into consideration, in order to analyze the potential user relationships. Based on our previous study [37], the term frequency-inverse document frequency (TFIDF)-based method is employed to extract users' recent interests hidden in their posted contents. Since users may not always discuss topics about their learning courses in the system, to restrict user relationships within the learning domain in user networking, we provide a set of keywords for each lesson, in order to extract more keywords related to users' learning intentions. Eq. (6) is used to extract keywords that can represent users' intentions more related to a specific lesson in a selected duration d, which can be expressed as:

$$K = \frac{TF(t_i, d_j)}{\sum_{k=1}^{n} TF(t_k, d_j)} * \frac{|D|}{|\{j : t_i \in d_j\}|} + \varphi C_{kos}, \tag{6}$$

where

$$C_{kos} = \begin{cases} 1, & \mathit{if\ this\ keyword\ exsists\ in\ the\ learning\ set}, \\ 0, & \mathit{otherwise}. \end{cases}$$

In Eq. (6), $TF(t_i, d_j)$ denotes the frequency of a word t_i in a selected duration d_j. $|D|$ is the number of all the durations divided from the whole period. φ is the equilibrium coefficient which ranges from 0 to 1.

For each keyword which can represent a specific user's potential intention, we connect those users who post contents related to this keyword with him/her in the user networking according to the weight w_{cr}, which can be quantified as:

$$w_{cr} = \frac{n_K}{M_j}, \tag{7}$$

where n_K denotes the frequency of the keyword that indicates the target user u_i's intention occurring in user u_j's posted contents. M_j denotes the total number of user u_j's posted words.

Based on these calculations, each extracted keyword calculated by Eq. (6) will represent one of a user's current intentions with its importance indicated by its weight within the current learning period. In this paper, we select one keyword with high importance (weight) as a user's intention and go further to calculate the correlations of contents posted by other users according to this intention using Eq. (7), in order to find those users who can provide information more related to it.

Based on these discussed above, finally, the value of w_{ij} between u_i and u_j with a specific threshold will determine whether the relation between these two users should be added into the use networking. Furthermore, in this user networking, the edge $\overrightarrow{u_j u_i}$, existing from u_j to u_i, indicates the direction of the relationship between these two users. That is, in this situation, user u_j can be viewed as the benefactor, which means he/she can provide helpful information to user u_i in order to support his/her learning purpose. On the contrary, user u_i can be viewed as the beneficiary, which means the information from user u_j could be useful to user u_i in accordance with his/her current intention. The algorithm to build the user networking is given in Fig. 4.

4.2 Measures for User Correlation Analysis

To analyze the correlations according to a specific user's intention in details, we introduce and define a set of measures as follows.

Input: The user set U: $\{u_1, u_2, \ldots, u_n\}$
Output: The user networking G in a specific period T

Step 1: For user u_i, calculate the weight $K = \frac{TF(t_i,d_j)}{\sum_{k=1}^{n} TF(t_k,d_j)} * \frac{|D|}{|\{j: t_i \in d_j\}|} + \varphi C_{kos}$ for each keyword in the specific time T, extract the keyword K with the highest weight to represent user u_i's intention
Step 2: Based on the keyword K, calculate the weight $w_{cr} = \frac{n_K}{M_j}$ for other users in the set $\{u_j \mid u_j \in U \ \&\& \ u_j \mathrel{!=} u_i\}$
Step 3: For each user u_i, calculate the weight $w_{ar} = \frac{IR(u_i,\ u_j)}{IR(u_i)}$
Step 4: According to ***Step 2*** and ***Step 3***, calculate the weight $w_{ij} = \alpha * w_{ar} + \beta * w_{cr}$ for each user u_i, if $w_{ij} > threshold\ \delta$, add user u_j as user u_i's neighbor node, which can be recorded as $\overrightarrow{u_j u_i}$
Step 5: Repeat ***Step 1***and ***Step 4***, until all the users in U find their neighborhood, return the user networking G

Fig. 4. Algorithm for building the user networking.

Interest degree: Interest degree is the in-degree of each vertex. That is, the interest degree of each vertex is the number of the arcs that take this vertex as the head, which can be expressed as $InD(u)$.

The interest degree of a specific vertex indicates the number of other users from whom this user may get helpful information related to his/her current intentions. The higher interest degree means the more related information and knowledge this user may obtain.

Popularity degree: Popularity degree is the out-degree of each vertex. That is, the popularity degree of each vertex is the number of the arcs that take this vertex as the tail, which can be expressed as $PoD(u)$.

The popularity degree of a specific vertex indicates the number of other users who may get suitable information related to his/her intentions from this user. The higher popularity degree means the more contribution this user may have.

Following these basic definitions, the following measure is introduced to analyze the correlations among users:

Contribution degree: Contribution degree is the value that reflects the contribution or importance of a vertex u_i to other vertexes that it points to, which can be expressed as $CoD(u_i) = \sum w_{ij}$.

Based on these, specifically, to calculate the better benefactor u_i among users linked to a target user u_j, the contribution degree between a pair of users, exactly, from u_i to u_j, can be quantified as:

$$CoD(u_i, u_j) = w_{ij} * \left(\frac{1}{\sum_{l=1}^{n} w_{il}} + \frac{1}{\sum_{k=1}^{n} w_{kj}} \right). \tag{8}$$

Specifically, for a pair of connected vertexes, $< u_j, u_i >$, the weight w_{ij} appending on this edge, as well as the contribution degree of u_i are taken account. That is, the value indicates in what degree the user u_i can support the user u_j in accordance with one of the current intentions. Thus, the correlation calculated based on $CoD(u_i, u_j)$ can be useful for the user u_j to better cope with his/her requirement in this situation, which means it can be applied to identify the most possible user u_i who can best support the target user u_j.

Input: An learning action sequence $s = < act_1, act_2, \ldots, act_n >$
A given Goal-driven Learning Group $LG = \{[< Act_k >_{u_1,u_2,\ldots,u_i}^{w_1,w_2,\ldots,w_i}][G_j, T]\}$
Output: The goal-driven learning action pattern set $DP = \{< act_i >^{w}\}$

Step 1: For all the learning action sequence $< Act_k >$ in the given Goal-driven Learning Group $LG = \{[< Act_k >_{u_1,u_2,\ldots,u_i}^{w_1,w_2,\ldots,w_i}][G_j, T]\}$, build a trie-based structure LG'
Step 2: For the inputting learning action sequence from a specific user u, find all the sub-sequences $< act_i >$ in accordance with the trie-based structure LG'
Step 3: For each $< act_i >$, calculate the frequency as the weight w, which can be recorded as $< act_i >^{w}$
Step 4: Return the set $\{< act_i >^{w}\}$ as the goal-driven learning action pattern set DP for user u

Fig. 5. Algorithm for goal-driven learning action pattern detection.

5 Learning Action Recommendation

In this section, an integrated method is proposed and developed to support a target user with the individualized learning action recommendation based on the analysis of both learning action patterns and user correlations within a task-oriented learning process.

5.1 Detection of Goal-Driven Learning Action Patterns

In Section 4, we have demonstrated how to extract the LA-Patterns from an individual, which can represent his/her personalized learning behaviors, and how to generate the Goal-driven Learning Groups in a user community, which can describe the similarities among a group of users. Based on these, in this section, we go further to discuss how to utilize other users' learning action patterns to support the detection of a specific user's learning behavior patterns, in order to facilitate the learning action recommendation process.

Specifically, to detect a specific user's learning behavior patterns toward a specific learning goal, the Goal-driven Learning Group is employed as a given learning action pattern set, which can help model an inputting target user's learning action sequence. That is, for a specific learning goal, we try to find all the matched sub-sequences in a target user's given learning action sequence in accordance with the learning action patterns selected in a Goal-driven Learning Group. For a higher efficiency in this process, the Aho-Corasick algorithm [38], which is one of the famous multi-string search algorithms in the pattern matching field, is employed. The algorithm to detect the goal-driven learning action pattern is expressed in Fig. 5.

5.2 Goal-Driven Learning Recommendation Mechanism

As shown in Fig. 6, both learning behavior pattern and user correlation are considered for the recommendation of learning action in a specific learning period. That is, the learning behavior pattern portion is taken into account of the information behavior factor which has been recorded as learning actions in the log data, while the user correlation portion indicates the interaction factor in both the direct and indirect way among a group of users within a learning course during a specific learning period. In details, the learning

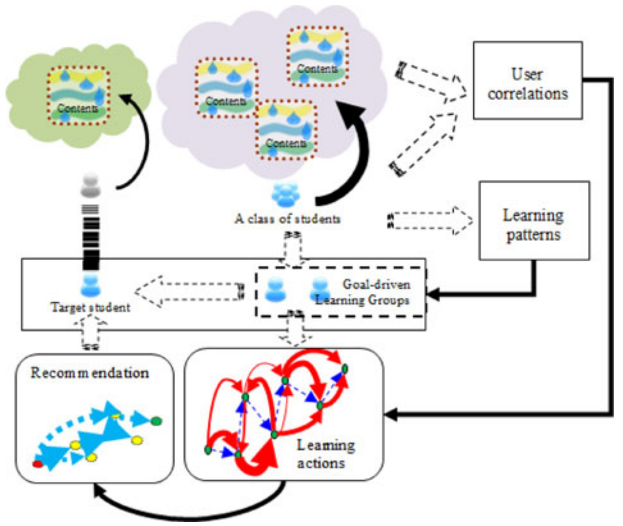


Fig. 6. Conceptual process of learning action recommendation.

action patterns, which are used to describe the users' learning behaviors in both the individual and collective way, can be extracted in accordance with different learning goals. Meanwhile, the user networking model can be constructed to represent users' potential and dynamical relationships based on the communication actions and posted contents, in which the specific correlations between the target user and other users can be further analyzed and extracted. Moreover, considering the detected goal-driven learning action patterns of a target user, three important weights: the weight which indicates the frequency of a LA-Pattern in a Goal-driven Learning Group, the weight which describes the users' relationships in the user networking model, and the weight which indicates the frequency of a detected goal-driven learning action pattern from the target user, can be used together to figure out the most suitable learning action from those similar users, which can be provided as the next learning step to serve the target user's specific learning purpose. Note that each recommended learning action refers to a target user for a specific learning goal within a selected learning period (e.g., one week for a lesson), which means the learning action we recommended is a specific action (e.g., view learning content regarding to the current lesson) following the current instructions, but not the generic action which will be suitable for the whole semester.

The formula to calculate the similar users is expressed as follows:

$$W_U = \gamma * \frac{\sum w_{DP_i} * w_{GP}(DP_i, u_j)}{\sum w_{DP}} + (1-\gamma) * \frac{CoD(u_j, u_i)}{\omega}, \tag{9}$$

where, $w_{GP}(DP_i, u_j) = \frac{w_p}{\sum w_p}$.

In Eq. (9), w_{DP} denotes the frequency-based weight for a detected goal-driven learning action pattern of the target user, $w_{GP}(DP_i, u_j)$ denotes the frequency-based weight for a LA-Pattern of user u_j in a Goal-driven Learning Group, $CoD(u_j, u_i)$ denotes the contribution degree calculated from the user networking model, and ω in the denominator is used for the normalization with a default value of 2.

Input: The target user u_i's current learning action act_{u_i}
A specific learning purpose G_i
Output: The recommended next learning action $actnext_{u_i}$

Step 1: For the whole user group, calculate the LA-Patterns $\{< act_i >_u^w \rightarrow G\}$ for each user u
Step 2: For the specific learning purpose G_i, generate the Goal-driven Learning Group LG: $\{[< Act_k >_{u_1,u_2,\dots,u_n}^{w_1,w_2,\dots,w_n}][G_i, T]\}$
Step 3: For the whole user group in a learning course, build the user networking model based on the weight $w_{ij} = \alpha * w_{ar} + \beta * w_{cr}$ in a selected timescale T.
Step 4: For the target user u_i, calculate the contribution degree $CoD(u_j, u_i) = w_{ji} * (\frac{1}{\sum_{l=1}^{n} w_{jl}} + \frac{1}{\sum_{k=1}^{n} w_{ki}})$
Step 5: For the target user u_i with his/her learning action sequence $s = < act_1, act_2, \dots, act_{u_i} >$, generate the goal-driven learning action pattern set $DP = \{<act_i>^w\}$
Step 6: For user u_i's current learning action act_{u_i}, find the following learning actions from users in the Goal-driven Learning Group LG and record as $actnext = \{actnext_{u_a}, actnext_{u_b}, \dots, actnext_{u_n}\}$
Step 7: Calculate the weight $W_U = \gamma * \frac{\sum w_{DP_i} * w_{GP}(DP_i, u_j)}{\sum w_{DP}} + (1-\gamma) * \frac{CoD(u_j, u_i)}{\omega}$ for each user in the Goal-driven Learning Group LG, in order to find the similar users
Step 8: For each element $actnext_{u_n}$ in list $actnext$, calculate the weight $W_{act} = \frac{\sum W_{U_j} * w_{na_i}}{\sum W_U}$
Step 9: Return the $actnext_{u_j}$ with the $Max(W_{act})$ to be the recommended next learning action $actnext_{u_i}$

Fig. 7. Algorithm for learning action recommendation.

For a target user u_i with the final learning action act_{u_i} in a given learning action sequence, Eq. (10) is used to calculate the weights of a set of learning actions that are selected from those similar users, which may further be inferred as the possible next learning actions:

$$W_{act} = \frac{\sum W_{U_j} * w_{na_i}}{\sum W_U}, \tag{10}$$

where, w_{na_i} denotes the frequency-based weight of a learning action generated by those similar users, following the learning action act_{u_i}. That is, the learning actions with a higher weight will be recommended to the target user as the next learning action. The recommendation algorithm is described in Fig. 7.

6 Implementation and Evaluation

In this section, after introducing the architecture of the implemented system, we show and discuss the experiment results in accordance with the analysis of learning behavior and user networking model respectively, in order to explain their practicability and usefulness. Based on these, the evaluation results are given to demonstrate the applicability and effectiveness of our proposed method.

6.1 System Architecture

The architecture for the individualized learning action recommendation is shown in Fig. 8.

This system consists of seven major components: Learning Action Pattern Analyzer, Goal-driven Learning Group Generator, User Interaction Analyzer, User Intention Analyzer & extractor, User Networking Builder, User Correlation Analyzer, and Individualized Action Recommender.

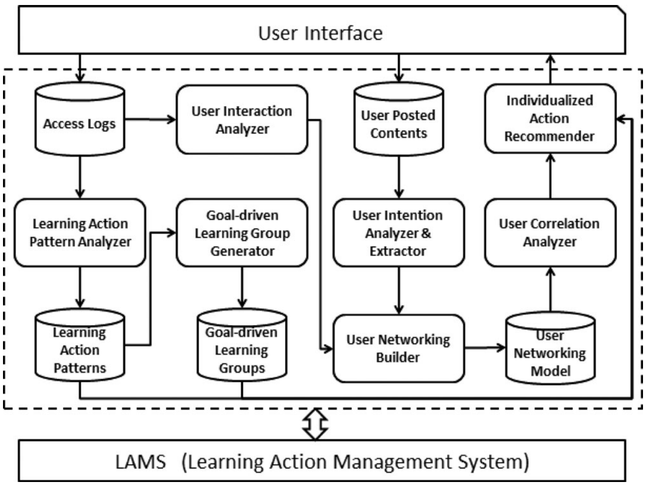


Fig. 8. Architecture of the task-oriented recommendation system.

As for the user information behavior analysis, the learning action pattern analyzer is used to calculate the LA-Patterns according to the analysis of learning behaviors described by learning actions which are generated in a specific learning course and stored in the Access Logs. The Goal-driven Learning Group Generator is employed to divide the users into several Goal-driven Learning Groups based on the similarity analysis of LA-Patterns. On the other hand, to calculate user correlations, the User Interaction Analyzer is integrated to compute users' direct communication behaviors, while the User Intention Analyzer & Extractor is responsible for the analysis and extraction of users' potential intentions. Based on these two modules, User Networking Builder can construct the user networking model, which can be further applied to figure out the correlations among the target user and other users. By the Individualized Action Recommender, after comparing a list of potential learning actions, the most possible next learning actions can be calculated as a feedback sent to the target user, which can benefit the action recommendation process among similar users. Finally, the Learning Action Management System (LAMS) manages all the learning actions and controls the whole recommendation process in this system.

We assume that there are no deviant users in our system, which means all users have used our system normally and generated no spam or irrelevant learning actions. Given that the assumption holds, we can ensure that all patterns contributed by learning action sequences are practicable and usable.

6.2 Experimental Analyses and Discussions

6.2.1 LA-Pattern Analysis Result

In this study, as shown in Table 1, we define totally 18 kinds of learning actions in our Moodle-based learning system. In order to facilitate the description in the following sections, we use 18 letters to represent them, which are also shown in Table 1.

Totally, 57 users have used this learning system, and 12,066 learning actions have been generated, in which, 13 kinds of learning actions have been generated among 18 kinds in total. In details, the learning action type u, view an assignment, containing 4,465 learning action records, is the most frequently used learning action, while the learning action type k, delete report in forum, containing only six learning action records, seems rarely to be used by users.

TABLE 1
Learning Actions and Their Notations

Learning Actions	Notation
view course introduction	a
view own learning history	c
lesson overview	d
view learning content	e
view posts in forum	g
discuss in forum	h
upload report in forum	i
update report in forum	j
delete report in forum	k
search in forum	l
View a quiz	o
start taking a quiz	p
refresh a quiz result	q
submit quiz result	r
review a quiz	s
view an assignment	u
upload a finished assignment	v
view user's profile	y

In our system, one course consists of 15 lessons. As for requirement, each student should conduct at least four actions for each lesson. Considering the absence, usually, a student should generate more than 60 actions to finish one course if not absent from a lesson. On the contrary, a student who generates too many actions (e.g., 600 actions) can be considered as a special case. In average, students generate 211 actions, and nearly four users generated less than 50 actions, which are 7, 35, 41 and 48 actions, while three users generated over 400 actions, which are 465, 668, and 732 actions.

As discussed in Section 3, LA-Patterns are a series of subsequences of the whole learning action sequence, which end with some special learning actions as the learning goals. That is, the learning action with the characteristics that can be viewed as a specific learning purpose will be selected as the goal action manually. Thus, according to the classification of learning actions (see Table 1), in this study, we predefine the following learning actions: i (upload report in forum), j (update report in forum), p (start taking a quiz), q (refresh a quiz result), r (submit quiz result), v (upload a finished assignment), to compose the learning goal set $G = \{i, j, p, q, r, v\}$, which leads to six Goal-driven Learning Groups based on the similarity of LA-Patterns. That is, we consider the learning action sequences ending with these learning actions in set G to generate LA-Patterns and Goal-driven Learning Groups as well.

For each user in one lesson, the learning actions sequence generated following the timeline composes an *S-Task*. All these *S-Tasks* will be used to extract the learning action subsequences which may become the LA-Patterns. Accordingly,

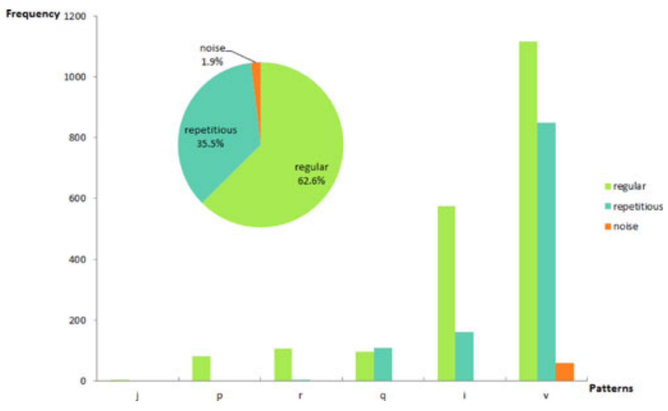


Fig. 9. Statistics and analysis for LA-patterns in each group.

according to *Criteria 1—basic criteria*, 838 learning action subsequences have been extracted, which can be categorized into six major types in details: i contains 215, j contains 1, p contains 29, q contains 78, r contains 49, and v contains 466, respectively. Consequentially, according to *Criteria 2—incorporation criteria*, after incorporating these learning action sub-sequences containing each user, the six types of LA-Patterns, have been refined to 183, 1, 24, 51, 37, and 359, respectively. Finally, we have obtained totally 655 LA-Patterns from more than 10,000 learning actions, which can be further assigned into six Goal-driven Learning Groups.

We further analyzed the features of each pattern according to two important factors: the frequency a certain pattern occurs in all users' learning action sequences, and the number of users who have conducted a certain pattern. Based on the statistics results, in each learning group, the patterns can be divided into three categories: *regular*, *repetitious* and *noise*, which are shown in Fig. 9. More detailed descriptions are given as follows:

- Regular patterns indicate these patterns that have no repeated action element in the sequence, which means each learning action in this pattern is unique.
- Repetitious patterns indicate these patterns that have repeated action elements in the sequence, which means there is at least one learning action that has been done at least twice in this pattern.
- Noise patterns indicate these patterns that are abnormal, which may be not correct or should not be recommended.

Note that only patterns in group-v and patterns in group-p contain some noise patterns, which is less than 2 percent among all patterns.

Based on these analyses, we can induce some useful insights as follows.

Generally, the shorter patterns may contain more users with higher frequency. For example, the *gi* in group-i, *op* in group-p and *uv* in group-v, which occupy nearly half in each group. These patterns can be viewed as a kind of common-use patterns or shortest patterns to complete a certain learning purpose. However, on the other hand, it does not mean only those patterns used by more users with higher frequency are useful. Moreover, some potential information (e.g., similarities among a small group of users) can be discovered and utilized from those patterns in spite of lower frequency. For instance, according to the patterns:

$$<uvuvuvvvuvuvuv>,$$
$$<uvuvuvvvuvuvuvuv>,$$
$$<uvuvuvvvuvuvuvuvuv>,$$

there are always three users: User u_{15}, User u_{25} and User u_{27}, in these patterns. It indicates that these three users may have a sort of behavior similarities to achieve the same learning goal: upload a finished assignment. Thus, more related information should be shared within them to pursue higher learning efficiency.

Furthermore, as for the categories, *regular*, *repetitious* and *noise*, in each Goal-driven Learning Group, holding the assumption in Section 6.1, the regular patterns can be viewed as basic patterns, which provide users with some basic steps as references to complete a certain learning goal, such as *ghi* in group-i, *opqr* in group-r and *euv* in group-v. The noise patterns refer to those patterns with none-recommended or incorrect sequence, such as *vueuv* in group-v and *qrop* in group-p. In these two groups respectively, we assume that in a well-defined LA-Pattern, learning action v cannot occur before u, and learning action q cannot occur before p, which means users should not upload a finished assignment before viewing it and users cannot redo a quiz before starting it. The repetitious patterns can be viewed as a positive means to better complete a certain learning purpose. Moreover, the so-called repeated factor can be extracted to facilitate the further learning action recommendation process. For instance, the sub-sequence *eoq*, which occurs multi-times in group-q, can become the repeated factor for this group. Then in the following recommendation process, when the learning action e is recommended to a specific user to refresh a quiz result, learning actions o and q can also be recommended to him/her.

Basically, our proposed methods mainly concentrate on the calculation of frequency of learning action sequences, rather than the meaning of each sequence. That is why some noise patterns have been extracted. However, the results discussed above certify that most of the LA-Patterns we extracted are meaningful and useful. We mainly employed the frequency factor in the following recommendation process, and considered the meaning of each pattern as the secondary factor to calculate the weight of each learning action.

6.2.2 User Correlation Analysis Result

Our learning system provides a discussion forum to allow both the students and the instructors to interact with each other asynchronously. That is, users can ask questions, discuss with the instructor and/or other students at any time. In this situation, we collect users' posts and record their direct interactions in our system to construct the user networking model, and further conduct the user correlation analysis, in order to improve the recommendation results.

The Yahoo API[1] (Japanese language morphological analysis), which can analyze the property of each keyword, as well as their frequencies, has been employed for the Japanese word segmentation process. To facilitate the following

1. http://developer.yahoo.co.jp/webapi/jlp/ma/v1/parse.html.

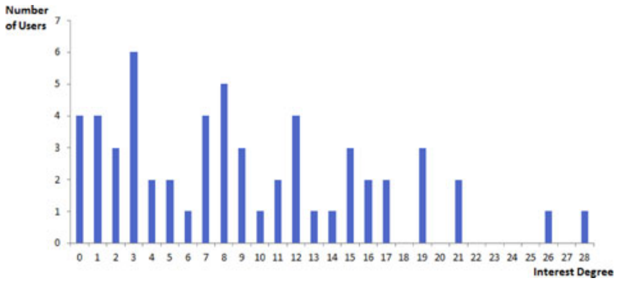


Fig. 10. Statistics data based on Interest degree.

analysis, we only select the nouns from users' posts as keywords. Finally, 294 nouns have been extracted to represent in part the users' intentions.

As discussed in Section 4, we set one week as the learning period T in terms of one lesson for the course to build the user networking. According to the calculation for the weight K for each keyword, in this study, we select the keyword with the highest weight to represent each user's intention, for instance, the keyword to represent User u_{16}'s intention is "OS". Then following Eq. (4), Eq. (5) and Eq. (7), where $\alpha = \beta = 0.5$, which means we assume that both of the two kinds of relationships considered in this study have the same importance, we obtain the weight w_{ij} for a pair of users. The average weight, $\overline{w_{ij}} = 0.21$, has been used as the $threshold\ \delta$.

Totally, 518 connections have been constructed among 57 users to represent their relationships in this lesson based on their direct and indirect interactions. Generally, in this user networking model, the more connections means the better interactions among users, which may lead to better collaboration works and better learning efficiency.

In average, each user has nearly nine relationships with others. We further give some statistics data in accordance with the interest degree and popularity degree. As shown in Figs. 10 and 11, the deviations of each degree are $Dev(InD) = 1.68$, and $Dev(PoD) = 1.66$. Moreover, in details, the top three users with their interest degrees are: $InD(u_{39}) = 28, InD(u_{69}) = 26$, and $InD(u_{65}) = 21$, which means these three users may get more helps from other users. The top three users with their popularity degrees are: $PoD(u_{36}) = 28, PoD(u_{50}) = 22$, and $PoD(u_{11}) = 21$, which means these three users are the more active users in this lesson. The top three users with their contribution degrees are: $CoD(u_{36}) = 12.3661, CoD(u_{50}) = 9.3553$, and

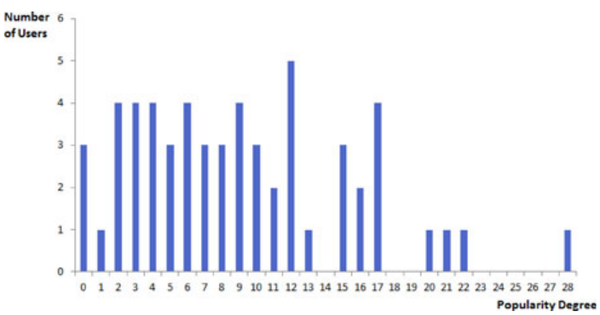


Fig. 11. Statistics data based on Popularity degree.

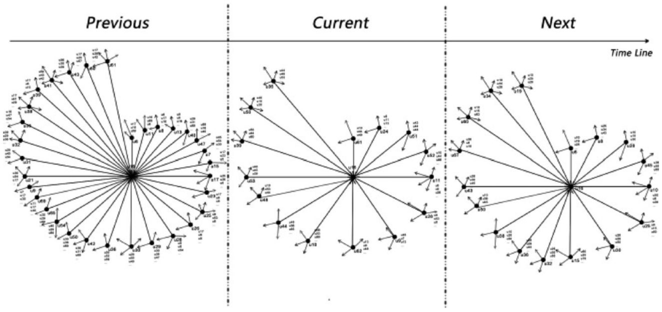


Fig. 12. An image of a specific user's time-varying correlations with others.

$CoD(u_{28}) = 7.1296$, which means these three users contribute to others more during this week and can be viewed as the foundational vertexes to provide information in this user networking model. On the other hand, the minimum value of each degree is 0 from User u_{27}, which means under this threshold, this vertex can get no relations with others.

We utilize these basic measures to serve the correlation analysis between two users. Fig. 12 shows an image of users' dynamical correlations. That is, take a specific User u_{16}, for example, in a selected learning period, say one week for a lesson, comparing with other two leaning periods ("Previous" and "Next"), in the current stage, 15 users have been involved from the user networking to construct the correlations with him/her, where the length of each edge indicates the weight of correlations between them. The detail analysis results for the current stage are shown in Table 2.

6.3 Evaluation

Based on these discussed above, the results of both LA-Patterns and user correlations are utilized for the final learning action recommendation.

TABLE 2
Results of Correlation Analysis for a Specific User

User ID	InD(v_j)	PoD(v_j)	CoD(v_j)	CoD(v_i,v_j)
9	8	12	3.9226	0.2376
10	15	12	4.0475	0.2350
11	19	21	6.8103	0.2546
24	0	4	1.2341	0.4228
28	1	20	7.1296	0.2512
36	14	28	12.3661	0.1796
39	28	17	5.9891	0.2082
44	19	13	4.9036	0.2206
48	17	16	5.4998	0.2132
50	9	22	9.3553	0.1882
51	0	10	2.9526	0.2656
52	11	12	3.1272	0.2592
53	11	15	5.7955	0.2102
61	15	9	2.7183	0.4564
62	4	11	3.9369	0.2372

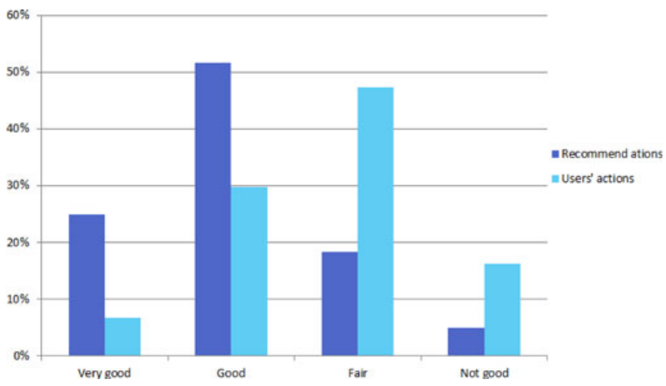


Fig. 13. Evaluation results in comparison.

For instance, in the situation discussed in Section 5.1, given User u_{16} with his/her recorded learning action sequence,

$$s = < egigeuvuvugigueuguvgigeugueuvugigeuvugiguvuv uvuvuvueuvgiguvugig >$$

assume the learning goal is i (upload report in forum), according to the algorithm shown in Fig. 5, the calculated goal-driven learning action pattern set is

$$P = \{ < ugi >^4, < uvugi >^4, < vugi >^4, < uvgi >^2, < uvuvugi >^1, < gi >^7, < vgi >^2, < euvuvugi >^1, < ueuvugi >^1, < vuvugi >^1, < euvugi >^2, < egi >^1 \}.$$

For other users, based on the analysis in the Goal-driven Learning Group-i, their frequency-based weights of each LA-Pattern, for example, $< uvgi >^{2,\ 2,\ 2,\ 3}_{u_{32}, u_{41}, u_{48}, u_{67}}$, together with the frequency-based pattern weight of User u_{16}, can be used to figure out the similarity between two users according to their learning behaviors.

Furthermore, according to the user correlation analysis results which have been shown in Table 2 (the weights of users who are not listed in the table will be set to 0), and considering that the user behavior patterns and user correlations have the same influence in the learning action recommendation process, we set γ to be 0.5 in Eq. (9). Finally, following Eq. (9) and Eq. (10), the learning actions with the top three weights will become the recommendation results for User u_{16} as the possible next learning step after learning action g (view posts in forum) in this learning period, which are learning action u (view an assignment) with 0.36, h (discuss in forum) with 0.31 and e (view learning content) with 0.12.

In order to evaluate the usefulness of recommendation results, other 19 users have been selected to conduct the recommendation process mentioned above. To each user, we recommend them with three learning actions sequentially in accordance with their weight-based rankings. Finally, totally 60 learning actions are selected and provided to these 20 users respectively as their recommended next learning steps. Besides, in order to evaluate the recommendation results, the likert scale-based questionnaire, which is the most widely used approach to scaling responses in the survey research [39], is given to classify and demonstrate the usefulness of the recommendations, where the subject will be asked to rate the recommendation results in the likert scales varying from "*Very good*", "*Good*", "*Fair*", and "*Not good*". For the fairness, the instructor of this course was asked to conduct the evaluations and grade the recommended learning actions according to the four scales. The evaluation results are shown in Fig. 13. In addition to the recommendation results, these 20 users' original decisions are graded to make the comparisons.

Basically, 95 percent of the recommended actions are considered as effective recommendations (among "*Very good*", "*Good*" and "*Fair*"). Moreover, comparing with users' original actions, over half of the recommended actions are "*Good*" which would lead to a high learning efficiency, while nearly half of users' original actions are "*Fair*" which would result in a low learning efficiency. In details, 25 percent recommended actions are "Very good" and 51.7 percent "Good", which can be considered to be efficient and useful, while only 6.7 percent users' actions are "Very good" and 29.7 percent "Good". On the contrary, less than 5 percent recommended actions are considered as unfavorable results, while users have made nearly 17.6 percent unfavorable decisions without the recommendations.

In addition to comparing the number of recommended actions with the number of users' original actions, the rankings of the recommendation results and the original actions of 20 users, are taken into account for the further comparisons. That is, the user's own decisions, which are sorted by their frequencies and ranked in a sequence, are employed to compare with the recommendation results sequence sorted by their weights. The comparison results are categorized into "*Higher*", "*Same*", "*Lower*", and "*Extra*", which indicates whether the ranking of recommended actions is higher, same, lower, comparing with the rankings of users' original decisions, or the recommendation results are even new actions which have not occurred in users' decision sequences.

Based on these, in each category, we employ the four scales mentioned above to give the assessment for each learning action. That is, when a recommended learning action is assigned to "*Very good*", and has a "*Higher*" ranking than user's original decision, it means this user's choice is not good enough. And if this user follows this suggestion, he/she may achieve a better score with a higher efficiency. Thus, following this comparison, we can further demonstrate the effectiveness of our recommendation results which can better assist users' decision-making to pursue a better learning efficiency.

Finally, as shown in Fig. 14, 42 percent recommended actions achieve a higher ranking, while 25 percent hold their rankings as same as users' original ones. Besides, 18 percent actions' ranking decline, and 15 percent are new actions.

We give the following discussions as further analysis,

1. Among the "*Higher*" recommended actions, the efficient actions reach to nearly 96 percent (both "*Very good*" and "*Good*"). Moreover, 36 percent of them are "*Very good*" which means by following these steps, the users can achieve their learning purpose in a more efficient way.
2. Among the "*Lower*" recommended actions, the most actions are the "*Fair*" and "*Not good*" actions (totally

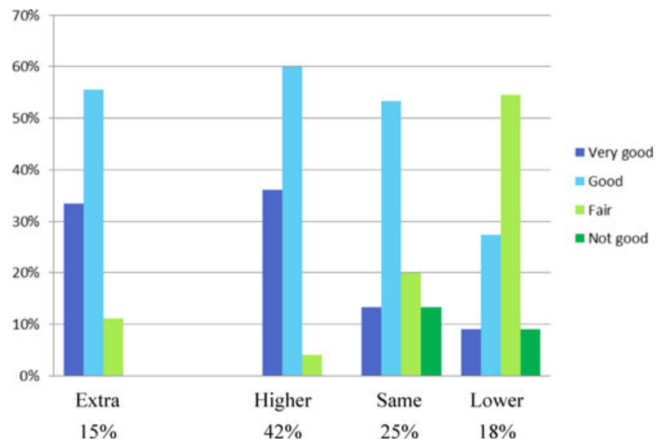


Fig. 14. Usefulness evaluation for recommendation results.

nearly 64 percent), which means according to these suggestions, the users can rearrange or drop those unfavorable actions which may not be suitable for their current learning goal, in order to avoid an inefficient learning style.

3. As for the "*Extra*" recommended actions, which are new actions not in users' original decision sequence, over 88 percent of them are efficient actions (both "*Very good*" and "*Good*"). These actions can be viewed as new perspectives to provide the users with more adaptive learning actions, in order to better pursue the learning purpose in a positive way.

According to the comparisons discussed above, in order to illuminate the value of our recommendation more clearly, a weighted matrix is designed for the score-based assessment, which is shown in Table 3. That is, each recommended action will be assigned an additional weight (score) in accordance with the usefulness ("*Very good*", "*Good*", "*Fair*" and "*Not good*") and effectiveness ("*Higher*", "*Same*", "*Lower*" and "*Extra*") analysis.

Finally, for each category of recommended actions, "*Very good*" scored totally 22 points, while "*Good*" scored 17, "*Fair*" scored 4, and "*Not good*" scored 2. For a clear illustration, the assessment results are shown in percentagewise in Fig. 15.

As a summary, all these results discussed above show that our recommendation method are practicable, and can effectively guide users to pursue their learning purposes in the task-oriented collaborative learning process.

7 Conclusion

In this study, we have presented an integrated method to model, discover and analyze users' learning behavior patterns and interaction-based correlations for the goal-driven learning action recommendations in a task-oriented learning process, in order to support the social knowledge delivery and task-oriented learning collaboration in the web-based learning environment. The main contributions of this paper can be summarized as follows.

TABLE 3
Values for Assessment of Recommended Actions

	Very good	Good	Fair	Not good
Higher	2	1	-1	-2
Same	0	0	0	0
Lower	-2	-1	1	2
Extra	2	1	-1	-2

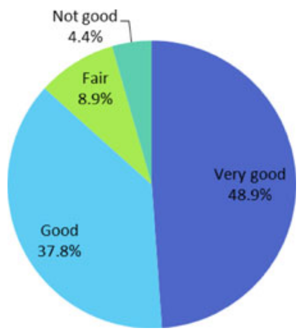


Fig. 15. Assessment results based on Table 3.

First, we have introduced a hierarchical model for the task-oriented collaborative learning process, in which the relationships among learning actions, learning activities, learning sub-tasks and learning tasks are described within a user community in an abstract level. Based on these, the LA-Pattern was proposed to discover and represent an individual user's learning behavior patterns which are described by the sequences of learning actions with their frequencies. Furthermore, the Goal-driven Learning Group was proposed to model a group of users' learning behaviors categorized by different learning goals, in order to analyze the similarities among users in terms of learning behavior patterns. Two algorithms were developed to generate the LA-Pattern and Goal-driven Learning Group respectively. The experimental results have demonstrated that the LA-Patterns extracted by our algorithm can correctly describe the users' learning behaviors, and the further analysis based on the Goal-driven Learning Groups also showed that our method could be applied to frequency-based learning pattern recognition and categorization according to different learning goals, which can benefit the task-oriented learning behavior analysis with high usability and practicability.

Second, following our previous study, we have developed and refined a user networking model to describe users' time-varying relationships in accordance with their interactions in the collaborative learning process. A TFIDF-based method was proposed to extracted keywords from the users' posted contents in order to capture their intentions and requirements within the learning domain. An algorithm was developed to construct the user networking model, which has taken account of both direct and indirect interactions among users. Based on these, a set of measures were introduced and defined to calculate and analyze users' correlations according to their learning requirements. The structural analysis based on the experiments showed that our user networking model could well represent users' relationships within an appointed learning period. And for a specific user, the proposed measures can effectively figure out all the users related to him/her, as well as their ranked correlations.

Finally, a goal-driven learning recommendation mechanism has been developed to utilize the LA-Patterns and user correlations, in order to recommend the suitable learning actions to users as their next adaptive learning steps, which are expected to help users to complete a specific learning goal in a more efficient way. The evaluations based on the empirical results demonstrated the usefulness and effectiveness of our proposed recommendation method, which can support and facilitate the task-oriented collaborative learning process.

However, there are some limitations. First, the learning actions are required to be pre-defined by the instructor in a learning system, which also limit the scale of the discovery of learning patterns. Second, since the indirect interactions to user correlations rely heavily on the keyword extraction process, the different precisions may lead to different results.

As for the future work, in addition to overcome the limitations mentioned above, the recommendation mechanism will be improved in order to fit the uncertain learning purpose and situation and automatically identify a user's learning goal. More experiments will be conducted to optimize the coefficients in the equations we proposed, in order to better adjust the weights of each factor to find the more suitable recommendations. We will also try to refine the learning action patterns and improve the algorithms to continuously deal with the emerging challenges in the enhanced social learning environments.

Acknowledgments

Qun Jin was the corresponding author.

References

[1] Social Learning Theory, http://teachnet.edb.utexas.edu/~Lynda_abbot/Social.html, 2014.

[2] J.E. Ormrod, *Human Learning,* third ed. Prentice-Hall, 1999.

[3] J. Vassileva, "Toward Social Learning Environment," *IEEE Trans. Learning Technologies*, vol. 1, no. 4, pp. 199-214, Oct.–Dec. 2008.

[4] B. Rienties, D. Tempelaar, P.V. Bossche, W. Gijselaers, and M. Segers, "The Role of Academic Motivation in Computer-Supported Collaborative Learning," *Computers in Human Behavior*, vol. 25, no. 6, pp. 1195-1206, Nov. 2009.

[5] L. Razmerita, "An Ontology-Based Framework for Modeling User Behavior—A Case Study in Knowledge Management," *IEEE Trans. Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 41, no. 4, pp. 772-783, July 2011.

[6] S.J. Stolfo, S. Hershkop, C.W. Hu, W.J. Li, O. Nimeskern, and K. Wang, "Behavior-Based Modeling and Its Application to Email Analysis," *ACM Trans. Internet Technology*, vol. 6, no. 2, pp. 187-221, May 2006.

[7] T.S. Chen, Y.S. Chou, and T.C. Chen, "Mining User Movement Behavior Patterns in a Mobile Service Environment," *IEEE Trans. Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 42, no. 1, pp. 87-101, Jan. 2012.

[8] Z.Y. Liu, Y.B. Zheng, L.X. Xie, M.S. Sun, L.Y. Ru, and Y. Zhang, "User Behaviors in Related Word Retrieval and New Word Detection: A Collaborative Perspective," *ACM Trans. Asian Language Information Processing*, vol. 10, no. 4, article 20, Dec. 2011.

[9] S.W. Lee, Y.S. Kim, and Z. Bien, "A Nonsupervised Learning Framework of Human Behavior Patterns Based on Sequential Actions," *IEEE Trans. Knowledge and Data Eng.*, vol. 22, no. 4, pp. 479-492, Apr. 2010.

[10] C.-H. Yun and M.-S. Chen, "Mining Mobile Sequential Patterns in a Mobile Commerce Environment," *IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Rev.*, vol. 37, no. 2, pp. 278-295, Mar. 2007.

[11] M. Munoz-Organero, P.J. Munoz-Merino, and C.D. Kloos, "Student Behavior and Interaction Patterns with an LMS as Motivation Predictors in E-Learning Settings," *IEEE Trans. Education*, vol. 53, no. 3, pp. 463-470, Aug. 2010.

[12] M. Plantevit, A. Laurent, D. Laurent, M. Teisseire, and Y.W. Choong, "Mining Multidimensional and Multilevel Sequential Patterns," *ACM Trans. Knowledge Discovery from Data*, vol. 4, no. 1, article 4, Jan. 2010.

[13] Z.W. Zhao and W.T. Ooi, "APRICOD: An Access-Pattern-Driven Distributed Caching Middleware for Fast Content Discovery of Noncontinuous Media Access," *ACM Trans. Multimedia Computing Comm. and Applications*, vol. 9, no. 2, article 15, May 2013.

[14] C. Wilson, A. Sala, K.P.N. Puttaswamy, and B.Y. Zhao, "Beyond Social Graphs: User Interactions in Online Social Networks and their Implications," *ACM Trans. Web*, vol. 6, no. 4, article 17, Nov. 2012.

[15] Y.R. Lin, J. Sun, H. Sundaram, A. Kelliher, P.l Castro, and R. Konuru, "Community Discovery via Metagraph Factorization," *ACM Trans. Knowledge Discovery from Data*, vol. 5, no. 3, article 17, Aug. 2011.

[16] L.M. Aiello, A. Barrat, R. Schifanella, C. Cattuto, B. Markines, and F. Menczer, "Friendship Prediction and Homophily in Social Media," *ACM Trans. Web*, vol. 6, no. 2, article 9, June 2012.

[17] J. Leskovec and E. Horvitz, "Planetary-Scale Views on a Large Instant-Messaging Network," *Proc. Int'l Conf. World Wide Web (WWW '08)*, pp. 915-924, Apr. 2008.

[18] L.M. Aiello, A. Barrat, C. Cattuto, G. Ruffo, and R. Schifanella, "Link Creation and Profile Alignment in the aNobii Social Network," *Proc. IEEE Second Int'l Conf. Social Computing (SocialCom)*, pp. 249-256, Aug. 2010.

[19] R. Xiang, J. Neville, and M. Rogati, "Modeling Relationship Strength in Online Social Networks," *Proc. 19th Int'l Conf. World Wide Web (WWW '10)*, pp. 981-990, Apr. 2010.

[20] V. Leroy, B.B. Cambazoglu, and F. Bonchi, "Cold Start Link Prediction," *Proc. 16th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '10)*, pp. 393-402, July 2010.

[21] M. Hamada, "An Integrated Virtual Environment for Active and Collaborative e-Learning in Theory of Computation," *IEEE Trans . Learning Technologies*, vol. 1, no. 2, pp. 117-130, Apr.–June 2008.

[22] W.T. Tsai, W. Li, J. Elston, and Y. Chen, "Collaborative Learning Using Wiki Web Sites for Computer Science Undergraduate Education: A Case Study," *IEEE Trans. Education*, vol. 54, no. 1, pp. 114-124, Feb. 2011.

[23] R. Kumar and C.P. Rose, "Architecture for Building Conversational Agents that Support Collaborative Learning," *IEEE Trans. Learning Technologies*, vol. 4, no. 1, pp. 21-34, Jan.–Mar. 2011.

[24] D. Perera, J. Kay, I. Koprinska, K. Yacef, and O.R. Zaiane, "Clustering and Sequential Pattern Mining of Online Collaborative Learning Data," *IEEE Trans. Knowledge and Data Eng.*, vol. 21, no. 6, pp. 759-772, June 2009.

[25] I. Boticki, H.W. Lung, and L. Chee-Kit, "Designing Technology for Content-Independent Collaborative Mobile Learning," *IEEE Trans. Learning Technologies*, vol. 6, no. 1, pp. 14-24, First Quarter 2013.

[26] A. Papadimitriou, M. Grigoriadou, and G. Gyftodimos, "Interactive Problem Solving Support in the Adaptive Educational Hypermedia System MATHEMA," *IEEE Trans. Learning Technologies*, vol. 2, no. 2, pp. 93-106, Apr.–Jun. 2009.

[27] R. Benbunan-Fich, R.F. Adler, and T. Mavlanova, "Measuring Multitasking Behavior with Activity-Based Metrics," *ACM Trans. Computer-Human Interaction*, vol. 18, no. 2, article 7, 2011.

[28] M. Fetaji and B. Fetaji, "Analyses of Task Based Learning in Developing 'M-Learn' Mobile Learning Software Solution: Case Study," *WSEAS Trans. Information Science and Applications*, vol. 6, no. 4, pp. 555-566, Apr. 2009.

[29] S. Ozawa, A. Roy, and D. Roussinov, "A Multitask Learning Model for Online Pattern Recognition," *IEEE Trans. Neural Networks*, vol. 20, no. 3, pp. 430-445, Mar. 2009.

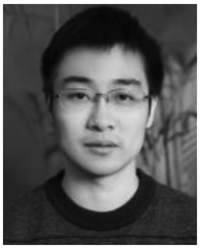
[30] R. Benbunan-Fich, R.F. Adler, and T. Mavlanova, "Measuring Multitasking Behavior with Activity-Based Metrics," *ACM Trans. Computer-Human Interaction*, vol. 18, no. 2, article 7, July 2011.

[31] J. Mairal, F. Bach, and J. Ponce, "Task-Driven Dictionary Learning," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 791-804, Apr. 2012.

[32] S. Cetintas, S. Luo, P.X. Yan, and C. Hord, "Automatic Detection of Off-Task Behaviors in Intelligent Tutoring Systems with Machine Learning Techniques," *IEEE Trans. Learning Technologies*, vol. 3, no. 3, pp. 228-236, July–Sept. 2010.

[33] S. Bangalore, G. Di Fabbrizio, and A. Stent, "Learning the Structure of Task-Driven Human–Human Dialogs," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 16, no. 7, pp. 1249-1259, Sept. 2008.

[34] E. Fredkin, “Trie Memory,” *Comm. ACM*, vol. 3, no. 9, pp. 490-499, Sept. 1960.
[35] J.A. Iglesias, P. Angelov, A. Ledezma, and A. Sanchis, “Creating Evolving User Behavior Profiles Automatically,” *IEEE Trans. Knowledge and Data Eng.*, vol. 24, no. 5, pp. 854-867, May 2012.
[36] X.K. Zhou and Q. Jin, “User Correlation Discovery and Dynamical Profiling Based on Social Streams,” *Proc. Int’l Conf. Active Media Technology (AMT ’12)*, pp. 53-62, Dec. 2012.
[37] X.K. Zhou, N.Y. Yen, Q. Jin, and T.K. Shih, “Enriching User Search Experience by Mining Social Streams with Heuristic Stones and Associative Ripples,” *Multimedia Tools and Applications*, vol. 63, no. 1, pp. 129-144, Mar. 2013.
[38] A.V. Aho and M.J. Corasick, “Efficient String Matching: An Aid to Bibliographic Search,” *Comm. ACM*, vol. 18, no. 6, pp. 333-340, June 1975.
[39] R. Likert, “A Technique for the Measurement of Attitudes,” *Archives of Psychology*, vol. 22, no. 140, pp. 5-55, June 1932.



Xiaokang Zhou is working toward the PhD degree in the Graduate School of Human Sciences, Waseda University, Japan, while as a research associate in the Department of Human Informatics and Cognitive Sciences. He has been engaged in the interdisciplinary research works in the fields of computer science and engineering, information systems, and human informatics. His current research interests include information seeking and recommendation, user modeling, and social computing. He is a member of the IEEE.

Jian Chen received the BS degree from North University of China, China, the MS degree from the University of Aizu, Japan, and the PhD degree from Waseda University, Japan, in 1990, 2003, and 2012, respectively. In 2013, he joined Taiyuan University of Technology, where he is currently in the College of Computer Science and Technology. His research interests include recommender systems, big data, and process mining, Internet of things, and cloud computing. He is a member of the IEEE and IPSJ.

Bo Wu received the MS degree in information science from Jilin University, China, in 2011. He is currently working toward the PhD degree in the Graduate School of Human Sciences, Waseda University, Japan. He has been engaged in the research fields of information and computer science, and social and human informatics. His current research interests include information search and recommendation, social data mining, social network analysis, and application systems development.

Qun Jin is a professor in the Department of Human Informatics and Cognitive Sciences, Waseda University, Japan. He has been engaged extensively in research works in the fields of computer science, information systems, and social and human informatics. His recent research interests cover ubiquitous computing, human-computer interaction, behavior and cognitive informatics, user modeling, information search and recommendation, and computing for well-being. He is a member of the IEEE, ACM, IEICE, IPSJ, JSAI, and CCF.