Knowledge-Based Systems 111 (2016) 36–50

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A behavioral sequence analyzing framework for grouping students in an e-learning system☆

Tao Xie, Qinghua Zheng, Weizhan Zhang*

MOEKLINNS Lab, Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, 710049, China

ARTICLE INFO

Article history:
Received 15 December 2015
Revised 29 July 2016
Accepted 1 August 2016
Available online 6 August 2016

MSC:
71.35.-y
71.35.Lk
71.36.+C

Keywords:
Grouping students
Behavioral sequence
Behavioral trajectory
Similarity measure

ABSTRACT

Grouping of students benefits the formation of virtual learning communities, and contributes to collaborative learning space and recommendation. However, the existed grouping criteria are mainly limited in the learning portfolios, profiles, and social attributes etc. In this paper, we aim to build a unified framework for grouping students based on the behavioral sequences and further predicting which group a newcomer will be. The sequences are represented as a series of behavioral trajectories. We discuss a shape descriptor to approximately express the geometrical information of trajectories, and then capture the structural, micro, and hybrid similarities. A weighted undirected graph, using the sequence as a node, the relation as an edge, and the similarity as the weight, is constructed, on which we perform an extended spectral clustering algorithm to find fair groups. In the phase of prediction, an indexing and retrieval scheme is proposed to assign a newcomer to the corresponding group. We conduct some preliminary experiments on a real dataset to test the availability of the framework and to determine the parameterized conditions for an optimal grouping. Additionally, we also experiment on the grouping prediction with a synthetic data generator. Our proposed method outperforms the counterparts and makes grouping more meaningful.

1. Introduction

All the time, grouping of students is a popular topic not only in classroom but also in many e-learning environments. In the classical education, the educators often deliberately design some team projects to promote students' collaborative and inquiry skills [1] or to influence their affections and values [2]. To this end, grouping is treated as a crucial process since its quality has much effect on the pedagogical objectives. As the computer and internet are widely used in education, the significance of grouping doesn't decline with the increase of transactional distance [3] between the teachers and students. Instead, its status has become more prominent due to the larger demand of personalization and the higher targets of collaboration and communication. Generally, two reasons lead to grouping. On the one hand, grouping in e-learning benefits the formation of so-called virtual learning communities [4] where the students within a community share the common interest, competence, or goal. On the other hand, grouping acts as a compromise of pursuing effect and avoiding cost in the personalized procedure. Although one student one program pattern may be more effective in learning, it's usually too expensive to hand out one program to each student especially when the user community is very large. In this case, grouping is a smart choice considering that students in the same group may have similar characteristics, interests, preferences, or abilities, while the ones in different groups may not have much in common. Besides, grouping makes special sense to collaborative learning space [5], along with recommendations [6] for future research.

Although the topic of grouping has been discussed for several years, most of them are based on the learning portfolios, profiles, social attributes, and the rising behavior-driven approaches. Little work has focused on the behavioral sequences composed of students' operation series as well as the insightful information behind. The modern learning theory indicates that students are the central part of the cognitive activity and the active constructors of the learning environment. Their actively constructed processes, i.e. interactions, may vary with the individuals' states, experiences,

☆ This research was partially supported by the National Science Foundation of China under Grant Nos. 61472317, 61428206, 61221063, 61472315, 91218301, the MOE Innovation Research Team No. IRT13035, the Co-ordinator Innovation Project for the Key Lab of Shaanxi Province under Grant No.2013SZS05-Z01, the Online Education Research Foundation of MOE Research Center for Online Education under Grant Nos. 2016YB165, 2016YB169, the Natural Science Basic Research Plan in Shaanxi Province of China Nos. 2016JM6027, 2016JM6080, and the Project of China Knowledge Centre for Engineering Science and Technology.

* Corresponding author.
E-mail addresses: graber@mail.xjtu.edu.cn (T. Xie), qhzheng@mail.xjtu.edu.cn (Q. Zheng), zhangwzh@mail.xjtu.edu.cn (W. Zhang).

http://dx.doi.org/10.1016/j.knosys.2016.08.001
0950-7051/

affections, expectations, competence, and so on. A learning behavior is generated as the interaction occurs, which is often stamped by time and accompanied with an operation type. The behaviors in durations would constitute all kinds of behavioral sequences depending on different users. We are curious that whether we would have an opportunity to extract underlying information that characterizes learning and even process grouping for a more homogenous structure. As the vigorous development of information technology, such behavioral sequences are easy to be captured but insufficient to be reported.

Our assumption is similar to Hilliard et al. (1992) that two groups of students would develop habits and preferences that cause them to perform in somewhat different ways owing to the diversities of behavioral styles [7]. The behavioral style is defined as a habitual behavior pattern, where we could group students by a certain differential thresholds. Previous studies used to adopt the sequential mining technology to address this discriminating task [8], which attempted to find the frequent patterns based on their occurrences. Nevertheless, it is insufficient to follow the sequences with interest only because a student's behavior is motivated by the complex inner process, where we may not get pedagogically-sound results until we take the deeper meaning behind the data into consideration. Moreover, the frequency is not a critical factor comparing with the students' interests in some cases. For example, an event with high frequency but low interest is invalid, and an event with low frequency but high interest, on the contrary, often counts. This intuition conforms to the paradoxical experimental conclusion in literature [9] that the more frequent a pattern is the more trivial or valueless it tends to be. Therefore, the sequential mining methods couldn't help us to acquire the desired results largely due to different data structure and different professional implications.

This study presents a view to analyze the behavioral sequences for the goal of grouping, which does not exist in other's work as far as we know. We attempt to identify students of which behavioral styles should be clustered in the same group and further estimate which group a newcomer will be as he/she styles his/her sequences. The first issue is a problem of similarity measurement. Although it has well been studied in a lot of fields including e-learning, nearly no one has considered the characteristics of sequences and data interaction simultaneously. The second issue is in essence a problem of prediction depending on the similarity models. To address these two issues effectively, we primarily take four actions. Firstly, we convert the sequences to behavioral trajectories. Secondly, we convert the calculation problem of behavior similarity to the one of trajectory similarity. Thirdly, we convert the problem of clustering to that of graph partitioning. And lastly, we convert the problem of complex prediction to that of simple character matching and distance measurement. The energy of the matrix is more concentrated after using Principle Component Analysis (PCA). The most promising group of the query sequence is the one that has the smallest distance to it.

The main contributions of this paper are:

(i) We propose a framework for grouping students based on their behavioral sequences logged in system. The grouping has dual meanings. One is dividing the sequences into several groups, and the other is assigning a new one into the most matching group. The central part of the framework is exploring the similarities of sequences and designing the prediction algorithm in terms of the newly defined data models.
(ii) An entirely different approach is presented to compute the similarity of sequences. We for the first time propose the conception of micro similarity of learning behaviors as well as the corresponding calculating method.
(iii) We propose a novel on-line predicting mechanism, i.e. indexing and retrieval, to effectively assign a student to the most matching group on a basis of the newly logged data and the historical models. This mechanism dedicates to condensing space of the models and carefully selecting important points, thus makes the space out of sparseness and keeps most of the features intact.
(iv) We conduct a series of extensive experiments on the learning behavioral logs collected during a whole school term to verify the availability of the proposed framework. We compare the gaps among approaches, study how to choose parameters for a better grouping, and examine the ability of online prediction compared to two common baselines via a synthetic behavior generator.

The rest of the paper is structured as follows. We discuss related works in Section 2. Section 3 introduces the research problems. Our proposed method is detailed in Section 4. In Section 5, we demonstrate the preliminary experimental results. Section 6 draws a conclusion and outlooks the future research.

2. Related work

Grouping has extensive applications, one of which is recommender system. Users are first partitioned into groups of similar ones, then recommendations are produced based on the characteristics of group members without searching the whole group base. For example, to recommend learning activities, Wu et al. proposed to match similar users based on a tree matching method and develop a novel similarity measure to handle the category issue of grouping [10]; to recommend point-of-interest locations, Chen et al. made groups according to users' demographics and frequently visited positions [11]. Without been exclusively emphasized, one sees that grouping has been treated as an important component of recommender system. In this section, we briefly review the previous studies mainly from the e-learning fields with three categories: grouping strategies, clustering approaches, and similarity measures.

Grouping techniques have been extensively researched recently. The big significance of grouping in teaching and learning environment is beyond all question in that there are considerable literatures have studied the positive effect of grouping on learning [12,13]. In traditional classroom, grouping is more based on the teachers' knowledge about students or class performance [14]. In computer-supported learning, the criteria for grouping are unrestricted. Many people, on a basis of the online test or usage logs, focused on the students' leaning portfolios, profiles, and social attributes etc. In literature [15] for example, the authors did a work on the students' portfolios. The portfolios usually comprised rich information figures, where students with similar portfolios were assembled in the same group while different groups revealed diversified characteristics. Their strategy was to collect the students' assignment scores, exam scores, and learning records, and then they used data mining techniques to obtain the grouping results. The work mainly depended on the teachers' observation about the students' learning activities, and most data sources originated from the students' mastery of knowledge. Thus their approaches lack strength to offer proper solutions to the problems we are currently focusing on.

The research on grouping based on the students' profiles is appeared in literature [16]. The profiles were modeled as the students' features such as personal details, course details, interests, team roles, preferences, trust ranking, and so on. They tried to achieve optimal grouping by satisfying some task constraints. However, the work is unable to handle the behavior data concerned in this paper. In addition, grouping with social attributes is also a