Automatic Dance Lesson Generation

Yang Yang, Howard Leung, LiHua Yue, and LiQun Deng

Abstract—In this paper, an automatic lesson generation system is presented which is suitable in a learning-by-mimicking scenario where the learning objects can be represented as multiattribute time series data. The dance is used as an example in this paper to illustrate the idea. Given a dance motion sequence as the input, the proposed lesson generation system automatically generates the lesson plan for students. It first extracts patterns from the input dance sequence to form the learning objects. The prerequisite structure is then built by considering the relations between the learning objects. Afterward the knowledge structure is constructed from the prerequisite structure based on the knowledge space theory. Finally, the learning path is derived according to an easy-to-complex manner while respecting the prerequisite relations. A user study that involved 40 students was conducted to evaluate the proposed work. The average learning time required for the treatment group (learning with the proposed system) was found to be lower than that of the control group (learning by free browsing) thus demonstrating the learning efficiency of the proposed system. The feedback from the questionnaires indicated that a majority of the subjects showed positive response toward the usefulness and rationality of our proposed system.

Index Terms—Computer-aided instruction, prerequisite structure, learning path, knowledge space theory

1 INTRODUCTION

The research of e-learning could be traced back to 1993 when an online computer-delivered lecture, tutorial, and assessment project described in [1] used several software programs to create a virtual classroom environment for students and instructors. The advantages of e-learning over traditional learning lie in the following aspects: 1) Self-paced learning: the e-learning environment allows students to learn at their own pace and production of e-learning content can be customized for individuals [2]; 2) Convenience and flexibility: students are not bounded to a specific time and a specific place to attend classes [3]; 3) Reduced cost: with an e-learning environment, the cost associated with instructors’ salaries, student travel, classroom, etc., can be reduced.

Different learning activities may be adopted depending on the learning content. For example, rote learning may be used for memorizing some fundamental knowledge while problem solving is ideal for enhancing logical thinking. The proposed work in this paper is suitable for a class of learning activity which is focused on learning by mimicking. For example, in learning dance, a student needs to mimic the teacher’s movement. Other examples that require learning by mimicking include handwriting or hand gesture. For this kind of learning activity, the learning objects can be represented as multiattribute time series data, for example, the 3D coordinates of the body parts in dance, the 2D trajectory of the pen in handwriting, the positions and orientations of the hand and fingers in gesture. Pattern discovery techniques can be applied to this kind of learning objects in order to extract patterns automatically from the multiattribute time series data and analyze them to provide useful information for the learning activity. In this paper, the dance is used as an example to illustrate the idea. In traditional dance education, students can either go to a dance lesson to learn from the dance teachers, or learn by simply watching videos or animations [4], [5]. The traditional learning in the former case lacks the advantages of e-learning that have been mentioned previously. In the latter case, since the whole dance may contain a lot of information, it may lead to the cognitive overload problem when the amount of information exceeds the capacity of the working memory.

In this work, an automatic lesson generation system is presented to provide an efficient learning experience to the students. Given an input dance sequence, the proposed system first extracts the patterns and builds the prerequisite structure from the patterns. The knowledge structure is then constructed based on the prerequisite structure. Finally, a learning path is derived by sequencing the learning objects in an easy-to-complex manner while respecting the prerequisite relations. Students can follow the proposed dance lessons in an organized way in order to fully reproduce the dance sequence. The remainder of the paper is organized as follows: The background of the related work is introduced in Section 2. The detail about the proposed lesson generation system is covered in Section 3. The user study, results, and discussion are provided in Section 4. The conclusion and future work are presented in Section 5.
2 BACKGROUND

2.1 Dance Education

The most commonly used method for dance education is demonstration-performance method. Demonstration-performance method is a simple yet sound way for people to learn physical and mental skills. In this approach, a demonstration must be given by the instructor, which will then be imitated by the students under close supervision. The feedback about the correctness of the performed dance moves will be provided by the instructor to help the students further improve their performance. This scheme is proven to be effective; however, the dependence on human instructor makes the learning process inconvenient and inflexible.

With the development of technology, dance lessons can be provided to learners in the absence of human instructor. Chua et al. [5] presented a wireless virtual reality system for Tai Chi training in which a virtual avatar is used to demonstrate the moves for the students. In the dance education systems proposed in [4], [6], the students could learn dance by imitating the model dance demonstrated by a virtual teacher. A virtual reality-based motion training system Just Follow Me (JFM) was provided in [7], and the result showed that JFM can produce the same or even better training result than the real-world training. However, these systems can only give demonstration and cannot give any feedback to help students improve. To overcome the lack of feedback, Naemura and Suzuki [8] estimated a dancer’s performance level through analyzing the video-captured dance movements. In the work reported in [9], the system can play back the motion segments which the student did not perform well. In [10], the system could track and capture the student’s motion in real time while the student is imitating the virtual avatar. By comparing the motions between the student and the virtual teacher, the replay could offer information about which parts of the body did not perform well.

However, demonstrating and imitating the full dance at once is not a practical way for people to learn because the full dance may be too long for students to remember and learn. In a traditional dance class, teachers would break down the full dance into basic steps and conduct modular lesson plan. As reported in [11], it is natural for people to segment an event automatically into hierarchically organized parts and subparts. In this paper, we propose a demonstration-performance learning system with an automatic way of dividing a full dance into an appropriate lesson plan.

2.2 Learning Objects

The term learning object is frequently explored in content creation and aggregation. There are multiple definitions provided by different communities [12], [13], [14]. A more general definition of learning object is proposed in [15] which defines a learning object as an independent and self-standing unit of learning content that is predisposed to be reused in multiple instructional contexts. This paper considers two important factors in creating the learning objects: granularity and reusability. With the consideration of granularity, it is argued in [14] that learning objects should be made of much smaller units of learning (say 2 to 15 minutes) rather than a chunk of several hours. Breaking down the learning objects into small units is appropriate but it should be noted that the learning time may vary among different people [15]. On the other hand, the common patterns can be extracted from the learning content to enable reusability. In [16], the patterns are extracted from some motions and modeled by a graph such that new motions can be generated by resequencing these patterns.

In this work, the dance motion is divided into patterns which are treated as learning objects. By extracting the patterns, the dance is divided into small motion segments which are easier to master to achieve granularity. In addition, the same frequently appearing motion segments are treated as a single pattern such that students only need to learn it once. On the other hand, these patterns could act as the components of other dances thus achieving reusability.

2.3 Prerequisite Structure

To facilitate organized learning, the prerequisite structure is built by finding all the prerequisite relations between the learning objects. The prerequisite relation is commonly used in educational setting to describe the dependency between learning objects. The prerequisite structure is composed of all the prerequisite relations on the set of learning objects. In many e-learning systems, such as Moodle or Blackboard, the prerequisite relations between the learning objects are specified by experts implying that the prerequisite structure must be built manually. In 1984, Novak [17] proposed the term concept map to represent knowledge as a graph consisted of nodes as concepts and links as the relations between the concepts. In fact, prerequisite structure is a special kind of concept map where the learning objects are treated as concepts and the relations between concepts are prerequisite relations. Concept map is of great significance in education but the construction of concept map is a nontrivial task. In [18], a method for semiautomatic construction of concept map is proposed which can arrange concepts or words in a map space using Kohonen’s self-organizing map algorithm. The semiconcept map is generated in [19] by allowing teachers and students to edit a set of concepts and links. However, semiautomatic construction of concept map still requires the users’ assistance. On the other hand, automatic construction of concept maps such as [20], [21] requires a large amount of the students’ testing records. A two-phase concept map construction approach is proposed in [20] by using the learners’ historical testing records. It first applied fuzzy set theory to transform the numerical testing records of learner into symbolic ones, found the grade fuzzy association rules and then adopted a heuristic algorithm to construct the concept map. The research in [21] attempted to improve the work in [20] by constructing the concept map by using fuzzy reasoning based on fuzzy rules. A novel method to automatically construct concept map is proposed in [22] but it also requires the students’ feedbacks.

In this work, an automatic method to build the prerequisite structure is proposed. The learning objects are represented by the patterns extracted automatically from the learning content. The prerequisite relation is defined based on the inclusion relation between the learning objects which can be determined automatically.
On the other hand, there may be some redundant relations in the prerequisite structure, so the transitive reduction algorithm is applied to remove the redundant prerequisite relations in order to get a compact prerequisite structure. The transitive reduction of a graph is sometimes referred as its minimal representation, i.e., a minimum equivalent graph which maintains the reachability relation.

### 2.4 Knowledge Space Theory

Knowledge space theory proposed by Doignon and Falmagne [23] often serves as a theoretical framework for student modeling. It offers a way to formally describe the structure of a given knowledge domain based on the prerequisite relation. According to this theory, a knowledge domain is defined as a set of learning objects in a specific field. The knowledge state of a student is identified with the subset of learning objects the student is capable of solving. However, due to the dependency between the learning objects, not every subset of the knowledge domain is a plausible knowledge state. Besides, the knowledge state of a student changes when he/she masters new learning objects. The knowledge structure of a given knowledge domain is referred as the collection of all the knowledge states in that knowledge domain.

Knowledge space theory has been implemented in many applications, for instance, Assessment and LEarning in Knowledge Spaces (ALEKS) [24], and Relational Adaptive Tutoring Hypertext (RATH). In addition, many researchers explore the usage of knowledge space theory in different ways. In [26], the students’ understanding in upper level undergraduate immunology course is evaluated by using knowledge space theory to analyze the difference between the students’ concept maps with the instructor’s concept maps. In [27], beginning college chemistry students’ understanding of stoichiometry is assessed using the knowledge space theory, with the results suggesting that there is a need for teaching students how to integrate their knowledge. A survey is provided in [28] on how people use the knowledge space theory to assess students’ background knowledge in order to provide personalized learning. The students’ thinking patterns are identified by using the knowledge space theory in [29] so the teachers can effectively guide the students along the critical learning paths suggested by the students themselves. In this work, the knowledge space theory is applied to derive the possible learning sequences.

### 2.5 Learning Path Planning

In planning the learning path, many researchers consider the difficulty level of the learning object as the criteria to find the best learning path. For example, an algorithm is proposed in [30] to select the shortest learning path by defining the weight of the relation between two learning objects as the difficulty of accessing from one learning object to another. The optimal learning path is determined in [31] by considering the learning time of each learning object in the optimization criterion. In [32], [33], the difficulty level of the course material serves as the criterion to select the optimal learning path from several personalized learning paths. Recently, there is an increasing interest in the research about personalized learning path based on two sources of personalization information: background knowledge and learning style. The background knowledge is considered in [34] which took into account of the student’s ability as a factor to implement personalized learning based on Item Response Theory. The research in [35] tried to improve the work in [34] by proposing a modified Item Response Theory. Regarding the learning style, an innovative way is proposed in [36] to detect it based on Bayesian networks. A client-based system is described in [37] which recommends relevant webpages according to the inferred learning style. A personalized learning environment is presented in [38] where the learning styles are inferred by monitoring the students’ interaction with the computers. Some researchers considered both the learning style and knowledge state in the personalization process [39], [40].

This work considers the difficulty level of the learning objects as the criterion to derive the learning path. It aims to automatically generate an appropriate lesson plan which can improve the learning efficiency of the students. The appropriate lesson plan will be regarded as a solid foundation for further extending the proposed system to a personalized learning environment.

### 3 Automatic Dance Lesson Generation

Fig. 1 shows the overview of the proposed automatic dance lesson generation framework. It can be divided into six steps which are described in the following sections.

#### 3.1 Capturing Motion Data

The motion data are captured by a marker-based optical 3D motion capture system with several cameras. The actor shown in Fig. 2a wears a tight suit attached with 35 optical markers on different body parts. The motion capture technology is capable of accurately tracking and recording human motion. The 3D motion of the actor is captured as a
time series data where each frame contains the 3D marker coordinates specifying the corresponding posture. A mesh model shown in Fig. 2b can be fit on the skeleton to generate a 3D virtual avatar.

### 3.2 Discovering Patterns

The learning content is broken down into smaller units called patterns which form the learning objects in the proposed work. The patterns could be repetitive or nonrepetitive, where the repetitive patterns refer to the frequently appearing dance motion segments, while the nonrepetitive patterns refer to the motion segments which only appear once. The granularity factor described in Section 2.1 is considered by dividing the dance into patterns as it is easier to master small dance motion segments. In addition, the same frequently appearing motion segments are grouped as a single repetitive pattern such that students only need to learn it once. On the other hand, these patterns could act as the components of other dances thus achieving reusability.

We use the algorithm proposed in [41] to find the repetitive patterns in a dance motion sequence which are frequently appearing motion segments. First a self-similarity matrix shown in Fig. 3 is computed which indicates whether two frames in the dance motion sequence are similar. Similar postures are represented by black pixels and dissimilar postures are represented by white pixels. Since the matrix is symmetric, only the lower triangular part is considered. The self-similarity matrix is then turned into a binary representation by a classifier which is trained to determine whether a pair of input postures is similar. The self-similarity matrix is then thinned and the start point of a repetitive pattern is located as the bottom-leftmost point of a diagonal line. The pixels are traced in the diagonal direction to find the longest similar segment pairs.

Fig. 4 demonstrates the keyframes of a sample dance and the extracted patterns. Each rectangular block represents a motion segment. Those rectangular blocks with the same texture are similar motion segments belonging to the same pattern. The labels $L_i$ denote the repetitive patterns and the labels $N_i$ denote the nonrepetitive patterns which are motion segments that only appear once. The patterns $E_i$ will be explained in the next section.

### 3.3 Building Prerequisite Structure

In this step, the prerequisite relation on the set of all patterns is obtained. In this work, the prerequisite relation corresponds to the inclusion property, i.e., if pattern $P$ is included in another pattern $S$, then $P$ is a prerequisite of $S$. A pattern may consist of several motion segments that are similar to each other. Pattern $P$ is included in pattern $S$, if at least one of the segments of $P$ is a subset of one of the segments in $S$. Intuitively, if pattern $P$ is included in pattern $S$, one would expect to learn pattern $P$ first before learning pattern $S$ thus explaining the rationality for the prerequisite relation. It should be noted that when there are overlapped patterns, the overlapped part may be also considered as a single pattern.

After identifying all prerequisite relations, the prerequisite structure can be created as a directed acyclic graph (DAG). In the graph, the nodes denote the patterns and the edges denote the prerequisite relations between those patterns. After computing the prerequisite relation of the sample dance (Fig. 4), the corresponding prerequisite structure as shown in Fig. 5a is obtained. For more details about the prerequisite relation, please refer to [42].

### 3.4 Eliminating Redundant Relations

The prerequisite structure obtained in the previous stage contains some redundant relations, making it difficult to visualize the graph to obtain a clear idea of the overall structure. As a result, the redundant relations need to be removed to result in a more compact representation. By applying the transitive reduction algorithm [43] on the prerequisite structure to remove such redundant relations, the compact representation can be derived. Fig. 5b shows the result after applying the transitive reduction of the prerequisite structure shown in Fig. 5a. It can be observed that the redundant relation $E_1 \rightarrow L_1$ is eliminated. For more detail about the transitive reduction, please refer to [42].
that the student is in the knowledge state \{E1\} (meaning that the student has already learned E1), if L2 is less complex than L3, then the next object to learn after E1 is L2. In this case, the complexity of L2 represents how difficult L2 is when the student has already learned E1. It should be noted that the repetitive patterns and overlapped patterns are placed in front of the nonrepetitive patterns in the learning path, because the organized learning on the repetitive patterns could probably help the learning of the nonrepetitive patterns which are less organized.

In Fig. 8, the complexity of going from one knowledge state to another is represented by the value associated with each link. It can be observed that from the knowledge state \{E1\}, the complexity to learn L2 is equal to 0.57, which is smaller than the complexity to learn L3 (0.89). L2 will thus be chosen as the next object to learn. The planned learning path of the repetitive patterns is \(E1 \rightarrow L2 \rightarrow L3 \rightarrow L1\). As for the nonrepetitive patterns, since they do not have relation with any other patterns, they will be simply arranged in the planned learning path according to their complexity. For example, the complexity values of N1, N2, N3 are equal to 0.84, 0.48, 0.91, respectively; thus, the planned learning path of nonrepetitive patterns is \(N2 \rightarrow N1 \rightarrow N3\). The full planned learning path is obtained by merging the planned learning path of repetitive patterns in front of the planned learning path of nonrepetitive patterns. The full planned learning path is illustrated as the dashed arrows in Fig. 8 as \(E1 \rightarrow L2 \rightarrow L3 \rightarrow L1 \rightarrow N2 \rightarrow N1 \rightarrow N3\).

4 User Study
4.1 System Demonstration
The proposed dance lesson dance generation system has been implemented. The system interface is shown in Fig. 9.
It consists of three panels: “Prerequisite structure,” “Learning path,” “Time sequence.” The prerequisite structure of the dance is demonstrated in the “Prerequisite structure” panel, with the nonrepetitive patterns arranged on the right side of the panel (p15, p16, p17, p18, p19). The “Learning path” panel is used to guide students through their learning process. The “Time sequence” panel shows the corresponding segments of the selected patterns in the timeline. Whenever the student clicks on any of the patterns, the corresponding pattern ID in both “Prerequisite structure” and “Learning path” panels will be highlighted. In the meantime, the corresponding segments will be indicated in the “Time sequence” panel and the corresponding motion will be played as a 3D animation with a virtual avatar. The student can further select multiple patterns by clicking on the target patterns while holding the Ctrl key. The student can also click on the segment in the “Time sequence” panel, and then the corresponding motion segment will be played.

4.2 User Study Setting
A total of 40 college students, 27 female and 13 male, were involved in this user study. None of these students has ever taken any formal dance lessons before. They were divided into two groups. Each group was invited to learn two dances (Latin and Hip Hop) in two different modes: the treatment mode and the control mode. The treatment mode refers to the learning with this proposed work of dance lesson generation system, while the control mode refers to the learning by freely browsing the dance motions. Note that the group who was asked to learn Latin in control mode is the same group as the group who was asked to learn Hip Hop in treatment mode. Alternatively, the treatment group who was asked to learn Latin is the same group as the group who was asked to learn Hip Hop in control mode. In this way, all 40 students in these groups had a chance to get exposed to both learning modes with the two dances. The students kept learning the dances until they had mastered them and the learning time for each student was recorded. Three judges who had taken dance lessons were invited to assess the students’ dancing performance. The students were considered to be successful in mastering the dance when at least two affirmative votes were obtained from the three judges.

After the students finished learning, they were asked to complete a questionnaire with three designed questions. The questionnaire adopts a five-point Likert scale [45] to evaluate the usefulness and the rationality of the proposed system. The three questionnaires are:

1. What do you think of the proposed dance lesson generation system?
2. What do you think of the learning material?
3. What do you think of the learning process?

The choices for the first question are: very useful, useful, neutral, useless, very useless. The choices for the last two questions are: very rational, rational, neutral, irrational, very irrational.

4.3 Results and Discussion
Table 1 shows the average learning time of the students with control mode and with the treatment mode. It can be concluded that the average learning time for the treatment mode is less than that with the control mode for both Latin and Hip Hop dances. This shows that the proposed scheme could improve the learning efficiency. For the two dances involved in the user study, the average learning time of the treatment mode outperforms that of the control mode by about 12 minutes. By defining the learning efficiency $LE$ as the reciprocal of the learning time, the improvement of learning efficiency $ILE$ for a dance is given by (1). In this user study, the proposed system leads to an $ILE$ of 16.3 percent.

<table>
<thead>
<tr>
<th>Dance</th>
<th>Average learning time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin</td>
<td>97.4</td>
</tr>
<tr>
<td>Hip Hop</td>
<td>72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mode</th>
<th>Control</th>
<th>Treatment</th>
<th>Control</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin</td>
<td>97.4</td>
<td>84.7</td>
<td>72</td>
<td>61</td>
</tr>
<tr>
<td>Hip Hop</td>
<td>72</td>
<td>61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 9. Dance lesson generation system.
The students’ responses to the questionnaire are shown in Figs. 10, 11, and 12. From Fig. 10, it can be observed that a majority of students (75 percent) believed that the proposed dance lesson generation system is useful in learning dance while only 5 percent of the students showed negative opinions. The result shown in Fig. 11 is also encouraging, with 80 percent of the students considering the learning material to be very rational or rational. Furthermore, Fig. 12 indicates that 65 percent of the students confirmed or strongly confirmed the rationality of the learning process in the proposed system. The proposed work on automatic dance lesson generation is deemed useful and rational for most students who have not received any formal dance training.

5 CONCLUSION

In this paper, an automatic dance lesson generation is proposed which is suitable in a learning-by-mimicking scenario where the learning objects can be represented as multiattribute time series data. The proposed automatic lesson generation system identifies the compact prerequisite structure, determines the knowledge states and plans a learning path by considering the complexity of the learning objects. The final planned learning path facilitates learning in an easy-to-complex manner while respecting the prerequisite relations. A demonstration of the proposed dance lesson generation system has been implemented. The user study showed that the system provides a practical way for people to learn dance.

Fig. 10. Students’ responses to the question “What do you think of the proposed dance lesson generation system?”

\[
ILE = \frac{LE_{\text{TREATMENT}} - LE_{\text{CONTROL}}}{LE_{\text{CONTROL}}}.
\]

Fig. 11. Students’ responses to the question “What do you think of the learning material?”

Fig. 12. Students’ responses to the question “What do you think of the learning process?”

ACKNOWLEDGMENTS

The work described in this paper was fully supported by a grant from the Research Grants Councils of the Hong Kong Special Administration Region, China (Project No. CityU 1165/09E).

REFERENCES


The work described in this paper was fully supported by a grant from the Research Grants Councils of the Hong Kong Special Administration Region, China (Project No. CityU 1165/09E).

Fig. 11. Students’ responses to the question “What do you think of the learning material?”

\[
ILE = \frac{LE_{\text{TREATMENT}} - LE_{\text{CONTROL}}}{LE_{\text{CONTROL}}}.
\]

Fig. 12. Students’ responses to the question “What do you think of the learning process?”

ACKNOWLEDGMENTS

The work described in this paper was fully supported by a grant from the Research Grants Councils of the Hong Kong Special Administration Region, China (Project No. CityU 1165/09E).

REFERENCES


Yang Yang is currently working toward the PhD degree in the Department of Computer Science and Technology at the University of Science and Technology of China (USTC). He is also enrolled in a joint PhD program offered by the City University of Hong Kong and USTC. His research interests include human motion analysis, human computer interaction, and e-learning.

Howard Leung received the BEng degree in electrical engineering from McGill University, Canada, in 1996, and the MSc and PhD degrees in electrical and computer engineering from Carnegie Mellon University in 1999 and 2003, respectively. He is currently an assistant professor in the Department of Computer Science at the City University of Hong Kong. His current research projects include 3D human motion analysis, brain informatics, and intelligent tools for Chinese handwriting education. He is currently a member of the Multimedia Systems and Applications Technical Committee, which is a subsidiary organization of the IEEE Circuits and Systems Society.

Lihua Yue is currently a professor in the Department of Computer Science and Technology at the University of Science and Technology of China. Her main research projects include database system and information integration.

Liqun Deng is currently working toward the PhD degree in the Department of Computer Science and Technology at the University of Science and Technology of China (USTC). He is also enrolled in a joint PhD program offered by City University of Hong Kong and USTC. His research interests include human motion analysis, human computer interaction, and pattern recognition.