ISBN 978-981-14-4787-7
Proceedings of 2020 the 10th International Workshop on Computer Science and Engineering (WCSE 2020)
Shanghai, 19-21 June, 2020, pp. 25-29
doi: 10.18178/wcse.2020.06.005

Design and Applications of Learner Model Based on Behavior Analysis

Yifei Yao, Fanhua Yu + and Xu Zhang

College of Computer Science and Technology, Changchun Normal University, Changchun 130032, China

Abstract. As a deep integration of information science and education, learning behavior analysis is one of the hot topics to realize personalized study. A lightweight learner model is proposed to generate customized quiz for individuals, which is along with an explicit and an implicit update algorithm. While learning interest is a three-level queue which is inferred from learning behavior and learning level is a binary vector which is inferred from test result. Commentary is provided on two fronts, concerning the extreme and unexpected cases. Through the analysis of simulation results, it is confirmed that the model is helpful to enhance the outcomes of online learning.

Keywords: learner model, learning analytics, behavior analysis, adaptive educational system.

1. Introduction

Online learning is growing together with the development of MOOC and SPOC. Almost every university has its own online courses, while the demand for personalized learning is increasing sharply. Several countries have developed a few policies about learning analysis, which not only emphasized its importance but also raised the main criticisms of privacy issues and the quality of interactive data. At present, the research of learning recommendation mainly focuses on learning interest modelling and learning behaviour analysis, in which collaborative filtering based on learning analysis is the most successful technology. These technologies have been widely used in classroom teaching and mobile assisted language learning [1,2]. There are even many practical systems on use, including foreign language learning, art interactive learning and consumer rating sharing preferences [3]. Some researchers studied the open learning mode of university curriculum by observing and interviewing with different students. He said that the students were satisfied with the settings in which they had the opportunity to change the model data and give control.

In the past, the quizzes provided by online courses are mostly designed by teachers, while all the questions are fixed and present at the end of each chapter. It is regarded as a passive test for students, lacking personalization, and an opposite way from personalized learning for students. Due to this issue, adaptive testing technology took a booming explosion, which can provide customized exercises for students according to their needs and abilities. Meanwhile, feedback are collected in time, thus automatic intervention becomes a reality. On the Internet, each one learns independently, while he wants a test on his own demand rather than a one-size-fits-all. In order to realize the goal of personalized test, a system needs to evaluate students' learning level and understand their interests. In return, it can not only help learners improve the effect of independent learning, but also provide a good measurement for teachers to evaluate students' learning level.

This paper mainly focuses on the quantization-based learner model that uses explicit and implicit update to infer learning interest for each individual. Our contributions to this topic are summarized as follows:

1. The leaner model is designed based on the two main types of behaviours in the use of learning platform, and it is updated in two ways of explicit and implicit.

+ Corresponding author. Tel.: +86 0431-86168137; fax: +86 0431-86168137.
E-mail address: 46995842@qq.com.

2. The three-level queue is used to store the learning level, which can describe the three different states of knowledge learning.

3. Taking the course of Program Design Foundation as an example, the algorithm is simulated and analyzed compared with the traditional quiz system.

2. Related Work

Leaning Analytics collects data related to learning activities and interprets the data from all aspects by using a variety of methods and tools. It can record and analyze learning environment and discover learning rules even predict learning results. Robert and Katrien divided learning analytics systems into two categories: providing suggestions and providing data visualization, and proposed that cross-disciplinary cooperation should be considered. Mohammed took learning analytics into a blended medical course [1], got a high prediction accuracy of 42.3% for students with poor grades through database queries and Moodle plugins. This research can be used for early warning and timely intervention.

Learner Model is the key to realize a personalized teaching system. [4] compared 10 different adaptive learning systems, and proposed that a dynamic learner model relies on special behaviours. Ding suggested four basic features to describe a learner. She also gave the method of initialization and updating, which was practiced in the course of "software engineering". Bayesian networks and artificial intelligence are used to deal with the uncertainty in the dynamic leaner modelling. However, cold start is a big problem because it needs time to get enough data for confidence coefficient.

3. Design of Learner Model

There are mainly two attributes to describe the learner, which are learning interest and learning level. Learning interest is a multi-level queue, which is sorted according to the measurement learners mastering the knowledge. Learning level is a linear table, which reflects learners' ability by their performance in tests.

3.1. Learning Level

A list is used to describe the learning level of learner p. Learning level describes the ability of learner p to answer all the question sets, including number of correct answers $r_{p,k}$, number of wrong answers $w_{p,k}$.

$$level_p = \left\{ \left\{ T_1, \left(r_{p,1}, w_{p,1} \right) \right\}, \left\{ T_2, \left(r_{p,2}, w_{p,2} \right) \right\}, \cdots, \left\{ T_n, \left(r_{p,n}, w_{p,n} \right) \right\} \right\} \tag{1}$$

3.2. Learning Interest

Learning interest is expressed in the form of a multi-level queue. The learning interest of student p is formalized as follows.

$$interlist_p = \left\{ \left\{ K_p^1, accu_p^1 \right\}, \left\{ K_p^2, accu_p^2 \right\}, \cdots, \left\{ K_p^{m_p}, accu_p^{m_p} \right\} \right\} \tag{2}$$

There is m_p knowledge points in the learning interest $interlist_p$ of student p, while $accu_p^i$ is the accuracy of knowledge point K_p^i.

The reason why using multi-level queue rather than single-queue is that students are always more interested in what they have just learned. To measure the ability of a student mastering the knowledge points, the weighted average sum of question set accuracy should be calculated as below. Because this value never can be 100%, knowledge point is considered to be mastered if it is more than 80%. It can also be divided into several levels for need, for example, less than 50% is poor, 50% to 80% is medium, and above 80% is good.

$$accu_p^i = \frac{1}{n} \sum_{j=1}^{n} \frac{r_j}{r_j + w_j} \tag{3}$$

3.3. Representation of Learner Model

Vector Space Model (VSM) is a kind of representation for vector space widely used in computer science, and it refers to describe the characteristics of learner and learning resources in form of vectors. Each vector consists of one or more feature terms and their corresponding weights, so VSM is suitable to calculate in computer. Here is an example, which denotes the learning interest model of n-dimensional-vector.

$$Learner = \{l_1, l_2, \cdots, l_q\} = \{(k_1, \theta_1), (k_2, \theta_2), \cdots, (k_q, \theta_q)\} \tag{4}$$

Where l_q denotes the interest node of user q, which is composed of interest keyword k_q and its weight θ_q.

4. Updating Algorithms of Learner Model

Supposing that the learner has only two kinds of behaviours, taking quiz and watching learning videos, then the model is explicitly updated when the learner watches videos and implicitly updated after each test.

4.1. Initialization Algorithm

At the very beginning, the learner model should be initialized as below.

- Generate a two-value list for learning interest as $interlist_p$;
- For every knowledge point, set $accu_p^i = -1$;
- Initialize queue L_p^1, L_p^2, L_p^3 empty.

4.2. Explicit Update Algorithm of Learner Model

When the learner chooses to study by watching video or other learning materials, the three-level queue will updated as in Algorithm 1. It will take different actions in three cases, such as learning new in a new chapter, learning new in an old chapter and reviewing.

Algorithm 1: Explicit update algorithm of learner model

Input: k is the knowledge point which the learner studied;
$threshold$ is the value measure learning, which means the student i has mastered the knowledge point p if $accu_p^i > threshold$;
m_1, m_2 : the number of knowledge points in Queue L_p^1 and L_p^2;

Output: The updated learner model.

if $k \in$ new for a new chapter **then**
  for $i \leftarrow 1$ to m_2, m_1 **do**
    if $accu_p^i > threshold$ **then**
      move K_p^i to the head of Queue L_p^3
    end if
  end for
  for $i \leftarrow 1$ to m_1 **do**
    move K_p^i to the head of Queue L_p^2
  end for
  set k as the head of Queue L_p^1
else
  if $k \in$ new for an old chapter **then**
    set k as the head of Queue L_p^1
    set accuracy value of k ($accu_p^k$) as -1
  else
    if $k \in L_p^1$ or L_p^2 **then**
      set k as the head of L_p^1 with value -1
    else
      if $timedifference = time_{now} - timestamp > threshold$ **then**
        set k as the head of L_p^1 with value -1
      end if
    end if
  end if
end if

4.3. Implicit Update Algorithm of Learner Model

After the learner took a quiz, the model is updated implicitly as in Algorithm 2.

Algorithm 2: Implicit update algorithm of learner model

Input: k is the number of knowledge points in this test;
n is the number each knowledge point appears in total;

Output: The updated learner model.

1: **for** $i \leftarrow 1$ to k **do**

2: $accu_p^i = \frac{1}{n}\sum_{j=1}^{n}\frac{r_j}{r_j + w_j}$

3: **end for**

4: **for** $i \leftarrow 1$ to 3 **do**

5: sort the knowledge points according to $accu_p^i$ in L_p^i

6: **end for**

---

5. Simulation and Analysis

5.1. Personalized Self-test Generation System

When the user takes the initiative to start the test, the system analyzes the user's behaviour data and generates the user's current test paper adaptively. Firstly, the learner model is updated based on the user's behaviour from the time of last test. Then, the learning level matrix is renewed according to the result in last test. After that, candidate question sets are selected related to the knowledge points, followed by a normalization processing to get the issue's number of each question set. At last, the final test paper of n questions is recommended for the user. The recommendation process is shown in Fig. 1.

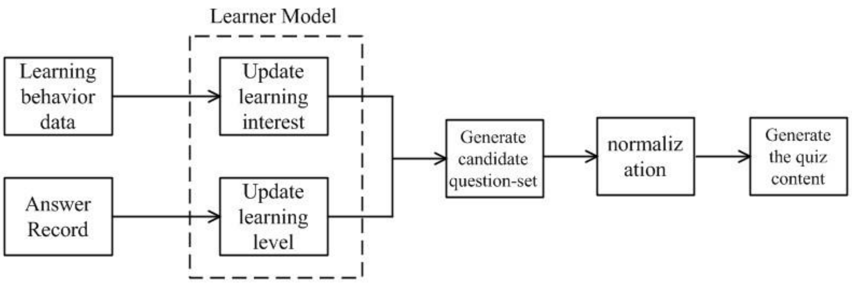

Fig. 1: Recommendation process

5.2. Simulation

Take the course of "C Language Program Design" as an example, this section describes the simulation of the algorithm. For the sake of simplification, it is assumed that the course can be divided into eight chapters, with four or eight knowledge points in each chapter showed in Fig. 2.

Suppose student p has learned chapter 3 and is starting for chapter 4. His learning behaviour sequence is e31-study, e32-study, e33-study, test 1, e34-study, e35-study, e36-study, test 2, e37-study, e38-study, test 3.

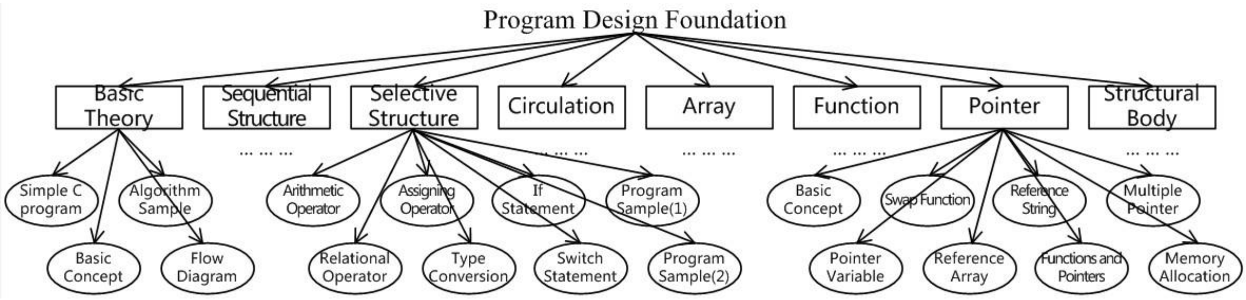

Fig. 2: Tree diagram of the knowledge points

In each test, there are 10 questions, and the results are as follows:

- Test 1: A0301 has 4 questions, 3 correct and 1 wrong; A0302 has 3 questions, 2 correct and 1 wrong; A0303 has 3 questions, 2 correct and 1 wrong.

- Test 2: A0304 has 6 questions, 4 correct and 2 wrong; A0303 has 3 questions, 2 correct and 1 wrong; A0302 has 1 question, and the answer is correct.
- Test 3: A0304 has 8 questions, 6 correct and 2 wrong; A0303 has 1 question, and the answer is wrong; A0302 has 1 question, and the answer is correct.

Then the parameters of learning interest are as below.

K34	K32	K33	K38	K37	K36	K35	K31
0.5	0.65	0.65	0.71	0.71	0.71	0.71	0.75

5.3. Analysis and Discussion

Extreme case: Suppose some extreme cases will occur as below.

- Always wrong: If the student makes mistakes frequently, $accu$ value turns smaller and smaller, so its probability of being tested increases, which is helpful for learners to take tests in time.
- Always right: This case indicates that the student has mastered the knowledge point. Then, the corresponding questions will only be in review test.
- Tests without learning new: At the end of the term, students may keep on testing by themselves. In this case, $accu$ values may be refreshed frequently. When all the knowledge points are learned, they will be reviewed from old to new according to the time stamp.

Discussion: $accu$ value will change in the following cases.

- $accu$ value of corresponding knowledge point will be reset to -1 if the learner study the video repeatedly. In this case, the knowledge point should be treated as a new one no matter what the $accu$ value is, which makes sure that corresponding point can appear in time at the next quiz.
- $accu$ values will be updated when a test problem is related to other knowledge points. Considering the relevance of knowledge learning, learning some knowledge may lead to insight or confusion of other knowledge, so it is reasonable to change the $accu$ vale of related knowledge points together.

6. Conclusions

Since the booming of online learning, learner modelling becomes a major issue of adaptive educational system. In order to provide students with better outcome, their study interest is identified by behaviour analysis while their learning level is defined according to their correct answer rate. Together with the explicit and implicit algorithm, learner model will be updated dynamically, and it can be used to generate self-test questions suitable for each student’s ability. Simulations and discussions on different situations show that the proposed model is able to find a better solution t for adaptive learning compared to traditional ways.

7. Acknowledgements

The authors would like to acknowledge the financial support from the Jilin Provincial Education Department Science and Technology Research Program (Grant No. JJKH20181179KJ) and the China Scholarship Council (No. 201908220058).

8. References

[1] S. Mohammed, F. Uno, and T. Matti, "How learning analytics can early predict under-achieving students in a blended medical education course," Medical Teacher vol. 39, no. 7, pp 757-767, 2017.

[2] Vo NgocHoi, "'Understanding higher education learners' acceptance and use of mobile devices for language learning: A Rasch-based path modeling approach," Computers & Education vol. 146, article 103761, 2020.

[3] K. Maimaiti, "Design of cloud computing-based foreign language teaching management system based on parallel computing," International Journal of Information and Communication Technology vol. 16, no. 1, pp. 17-29, 2020.

[4] MA. Tadlaoui, RN. Carvalho, and M. Khaldi, "A learner model based on multi-entity Bayesian networks and artificial intelligence in adaptive hypermedia educational systems," International Journal of Advanced Computer Research vol. 8, no. 37, pp. 148-160, 2018.