Personalized Content Sequencing Based on Choquet Fuzzy Integral and Item Response Theory

Ahmad Kardan, Roya Hosseini
Department of Computer Engineering and IT
Amirkabir University of Technology
Tehran, Iran
e-mail: {aakardan, hosseini.ec}@aut.ac.ir

Abstract—**Personalization of learning path is an important research issue in current E-learning systems because learners differ from various aspects such as knowledge level, experience, and ability. Therefore, most personalized systems focus on learner preferences, and browsing behavior for providing adaptive learning path guidance. However, these systems usually neglect to consider the dependence among the learning concept difficulty and the learner model. Generally, a learning concept has varied difficulty for learners with different levels of knowledge. Considered the importance of learning path with learning concepts difficulty that are highly matched to the learner's knowledge and ability, this paper proposes a system based on Choquet Fuzzy Integral and Item Response Theory. This system recommends appropriate learning contents to learner during the learning process.**

Keywords-E-learning; difficulty level; Choquet fuzzy integral; learning path; Item response theory

I. INTRODUCTION

Exponential growth of Internet as well as recent advances of Information Technology in educational context, have strong impact on educational process. One of the aspects of education that has been greatly influenced is learning. In this regard, Information and Communication Technologies (ICTs) provide opportunities to use new and efficient educational tools and methods that support learners during the learning process.

Computer-Based Training (CBT) is known as a practical tool in the learning process. E-Learning is the advanced form of CBT and is capable of enhancing both the effectiveness and efficiency of learning [1]. However, the development of Web-Based tutoring systems has created a great deal of hypermedia in learning sessions. Consequentially, learners are unable to learn very efficiently due to information overload and disorientations [2].

On the other hand, as long as the users differ on their goals, knowledge level, and interest, it is not effective to provide a fixed sequence of learning contents for all learners. Hence, personalized curriculum sequencing is necessary for adaptive learner support in Web-Based tutoring systems [3].

According to [4] curriculum sequencing is an important element of personalized learning; it provides each learner the most suitable individually planned sequence of concepts to learn and contents to work with.

So far, different systems were proposed for supporting learners through their learning process. Examples of such systems include systems that support learner by providing personalized sequence of exercises, or problems.

Most of the personalized tutoring systems consider learner knowledge, ability, and preferences for providing adaptive and personalized content sequencing. But they still ignore the dependence among the difficulty level of learning concepts and the learner's knowledge. To illustrate it more, a learning concept may have different difficulty for learners with different knowledge level and ability.

Dynamic problem difficulty is obviously better matched to the leaner's knowledge level, and hence can ease the comprehension of learning concepts. The related work in [5] calculates problem difficulty with regards to the learner model and recommends suitable problems to the learner. Experimental results of [5] confirm that dynamic problem difficulty performs well for a wide range of learners, whereas static problem complexity performs well for students of intermediate ability, but rather badly for beginners and advanced learners. The main drawback of the proposed method in [5] is the use of constraint based modeling for calculation of problem difficulty. According to [6] this model requires a huge set of constraints to be defined and this adds to its complexity. Moreover, it is not always possible to define constraint for some knowledge domains.

To construct a learning path with learning concepts difficulties that are highly matched to the learner's knowledge and ability, this research presents a Web-Based tutoring systems based on Choquet Fuzzy Integral and Item Response Theory. Concept difficulty is calculated by using the information provided in the learner model and the knowledge domain. Then, appropriate contents matched with the learner's ability are recommended to him/her by using Item Response Theory.

So, the organization of this paper is as follows: The following section presents the architecture of the proposed systems. Section 3 proposes a methodology for content sequencing. Finally, section 4 provides a conclusion for this work.

II. SYSTEM ARCHITECTURE

This section presents the architecture of the system for learning content sequencing based on Choquet Fuzzy Integral and Item Response Theory. As it is shown in Fig. 1,

978-981-08-9195-4/11 ©2011 IACSIT

this system consists of five major units as well as four major databases. The five major units are: Learner Interface Unit, Dynamic Concept Difficulty Calculation (DCDC) Unit, Feedback Unit, Content Recommendation Unit, and Learning Content Repository. The four major databases are: Knowledge Domain Database, Learner Account Database, Concept Difficulty Database, and Learner Model Database.

The Learner Interface Unit provides a learning interface for learners to interact with the Feedback Unit and the Content Recommendation Unit. DCDC Unit dynamically calculates concept difficulty based on the information in the learner model and the knowledge domain. The Feedback Unit aims to collect learner explicit feedback information from the Learner Interface Unit and stores it in the Learner Model Database. Content Recommendation Unit is in charge of recommending suitable contents to learner based on the learner's ability and difficulty of learning concepts by using Item Response Theory. To this end, this unit uses the values of difficulty level of concepts that are calculated by the DCDC Unit. Learning Content Repository consists of learning contents which are presented to the learner during the learning process. Each content is designed to convey a certain concept of the knowledge domain. Also, this repository contains test items. A test is designed for each content to evaluate the understanding degree of a learner. Having finished reading a concept, the learner must answer the corresponding test item. The numbers on the arrows in Fig. 1 show the steps related to data flow within the system. These steps are as follows:

Step 1. The learner logs in the system by providing his/her user account information.

Step 2. The user login information is compared to the information in the Learner Account Database.

Step 3. If the information is valid, then the user selects the Sub-domain to be learned from the system home page and step 4 is followed. Otherwise, the learner should try logging the system until he/she is authenticated.

Step 4. The system searches and selects the set of concepts that are related to the Learner-Selected Sub-domain. The concepts that the learner has already learned are removed from this set.

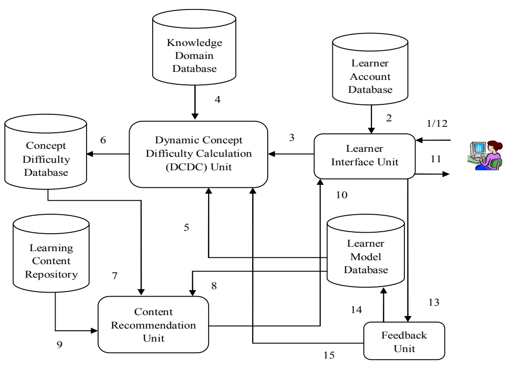

Figure 1. The system architecture.

The learner is considered to learn a concept if he/she answers its corresponding test item correctly. The final set servers as one of the inputs of the DCDC Unit.

Step 5. The learner information is the other input that is used by the DCDC Unit.

DCDC Process: Using the inputs provided by Step 4 and Step 5, this process calculates the concept difficulty by using Choquet Fuzzy Integral which is discussed in detail in section 3.2.

Step 6. After the DCDC process is finished, the difficulty level of each concept is temporarily stored in the Concept Difficulty Database.

Step 7. The information of the Concept Difficulty Database serves as one of the inputs of the Content Recommendation Unit.

Step 8. The information in the learner model is the other input that is used by the Content Recommendation Unit.

Step 9. The information in the Learning Content Repository serves as the third input of the Content Recommendation Unit.

Content Recommendation Process: Using the information provided by Step 7, Step 8, and Step 9, this unit selects and ranks suitable learning contents for learner based on Item Response Theory which is discussed in detail in section 3.3.

Step 10. The list of recommended contents is transferred to the Learner Interface Unit.

Step 11. The Learner Interface Unit presents the recommended contents to learner.

Step 12. After the learner finishes reading certain content, its corresponding test item is presented to learner to evaluate his/her underrating degree. This unit receives the learner's responses to the tests.

Step 13. The Leaner Interface Unit sends the received Feedback to the Feedback Unit for updating the information in the learner model and recommending suitable contents to learner.

Step 14. The Feedback Unit updates the information in the learner model based on the received feedback.

Step 15. To recommend learning content based on the learner feedback, concept difficulty level is calculated again for the concepts that have not been learned by the learner yet. Steps 4 to Step 15 are repeated until the learner learns all the concepts in the selected Sub-domain.

III. METHODOLOGY

This section presents the proposed method for learning path construction. The following subsections first introduce the influential parameters on calculating learning concept difficulty. Then Choquet Fuzzy Integral is used to calculate the difficulty of learning concepts. Finally, regarding the difficulty of concepts, a method based on Item Response Theory is presented to recommend suitable contents to learner.

A. Parameters

Dynamic calculation of difficulty level for a concept relies on the information in the learner model and the knowledge domain. Therefore, both the learner and the domain model are two influential elements in determining the learning concept difficulty. The learner model consists of two parts, namely the learner's knowledge model and the learner's behavioral model. The former models the learner's knowledge in concepts of the knowledge domain and the latter models the learner's behavior and more specifically the number of time the learner studies each concept of the knowledge domain. The details of these parameters are as follows:

Domain Model: The domain model consists of the concepts as well as the relations between them. In this paper, the concept that the difficulty level is calculated for it is called the Main Concept. Semantic relationship between the Main concept and other concepts of the certain knowledge domain can be described in two forms: 1- Prerequisite Concepts that are necessary to perceive the Main Concept, and 2- Related Concepts that are related to the main concept and are part of the same Sub-domain.

Learner Behavioral Model: This model stores the information about the learner's activity, i.e. the number of times the certain concept is studied by the learner. Hence, the learner activity in the Main Concept, its Prerequisite Concepts, and its Related Concepts is important in determining the difficulty level of the Main Concept.

Leaner Knowledge Model: The learner's wrong answer to the tests related to the Main concept, its Prerequisite Concepts, and its Related Concepts implies the learner Knowledge in the concepts of the knowledge domain.

B. Concept Difficulty Calculation Using Choquet Fuzzy Integral

The Choquet Fuzzy Integral is an operator that is used for the aggregation of the interdependent parameters based on the fuzzy measure. According to [7] suitability of this integral is proved for the Real-Time applications. Therefore, Choquet Fuzzy Integral can improve the response time of an E-learning system. In this paper, Choquet Fuzzy Integral is used to aggregate the influential parameters on concept difficulty. The definition of Choquet Fuzzy Integral is as follows:

Definition: Choquet Fuzzy Integral is an integral that uses fuzzy measure to aggregate the set of input parameters. According to [8] it is defined as (1):

$$E_g(h) = \int_X h(.) \circ g(.) = \sum_{i=1}^{n} [h(x_i) - h(x_{i-1})] g(A_i) \qquad (1)$$

where $h(x_1) \le h(x_2) \le \dots\dots\dots\dots \le h(x_n)$ and $h(x_0) = 0$. The variables that are used in (1) are as follows:

n Number of the input parameters. In this paper, $n = 3$.

X The set of input parameters of the Choquet Fuzzy Integral. This set is shown as $X = \{x_1, x_2, \dots, x_n\}$.

x_1 Represents the number of the relations between the Main Concept, its Prerequisite Concepts, and its Related Concepts.

x_2 Represents learner's wrong answers to the test items related to the Main Concept, its Prerequisite Concepts, and its Related Concepts.

x_3 Represents the number of the learner's study activity in the Main Concept, its Prerequisite Concepts, and its Related Concepts.

h The function that determines the value of the input variables. For example, $h(x_i)$ is the value of x_i.

g The λ-fuzzy measure which is defined as $g : P(X) \to [0,1]$ such that:

$$g(\phi) = 0, g(X) = 1 \qquad (2)$$

If $A, B \in P(X)$ and $A \subset B$, then $g(A) \le g(B)$ (3)

If $A, B \subset X$, $A \cap B = \phi$, then

$$g(A \cup B) = g(A) + g(B) + \lambda g(A) g(B) \qquad (4)$$

for some fixed $\lambda > -1$.

The value of λ is found from the equation $g(X) = 1$ that is equivalent to solve (5):

$$g_\lambda(X) = \frac{1}{\lambda}\left(\prod_{i=1}^{n}(1 + \lambda g_i) - 1\right), \lambda \ne 0 \qquad (5)$$

A_i This variable represents the set of input parameters in the form of $A_i = \{x_i, x_{i+1}, \dots, x_n\}$.

$g(A_i)$ This value is recursively calculated by (6) and (7):

$$g(A_n) = g(\{x_n\}) = g_n \qquad (6)$$

$$g(A_i) = g_i + g(A_{i+1}) + \lambda g_i g(A_{i+1}), 1 \le i < n \qquad (7)$$

To calculate the difficulty level of the concepts in the knowledge domain the following steps should be followed:

Step 1. The concept difficulty is calculated by using (1). $h(x_1)$ is the sum of Prerequisite and Related Concepts of the Main Concept; $h(x_2)$ is the learner's wrong answers to test items related to Main Concept, its Prerequisite Concepts, and its Related Concepts; $h(x_3)$ is the learner's study activity in the Main Concept, its Prerequisite Concepts, and its Related Concepts.

Step 2. According to the assumption of (1), the values of $h(x_i)$ are sorted ascendingly such that $h(x_1) \le h(x_2) \le \ldots \le h(x_n)$.

Step 3. The value of the fuzzy measure is calculated for each of the three input parameters. This research uses the function introduced in [9] which calculates the fuzzy measure values as (8):

$$g_i = \frac{1}{1 + d(h_i, h_0)} \qquad i = 1,2,3 \tag{8}$$

g_i represents the fuzzy measure value of x_i. $d(h_i, h_0)$ is the Euclidean distance between $h(x_i)$ and h_0. h_0 is an optional value from which the distance of all $h(x_i)$ is calculated. In this research, it is assumed that $h_0 = 0$.

Step 4. The value of λ is calculated by using (5).

Step 5. The value of $g(A_1), g(A_2)$, and $g(A_3)$ is calculated according to the value of λ using (6) and (7).

Step 6. To calculate the difficulty of the Main Concept, the values of $h(x_i)$ and g are used in (1).

C. Content Recommendation

The proposed system recommends appropriate contents to the learner by estimating learner's ability and ranking the contents accordingly. Item Response Theory (IRT) evaluates the ability of leaner and recommends suitable content to him/her. The goal of IRT is to estimate the learner's ability with regards to his/her dichotomous answers to test items. The IRT model illustrates the relationship between learner's answer and a test item through the Item Characteristic Curve (ICC). As shown in Fig. 2 this curve is S-shaped. The horizontal axis represents the ability of learner in a limited range and the vertical axis represents the probability that the learner with certain ability can answer the test correctly [10].

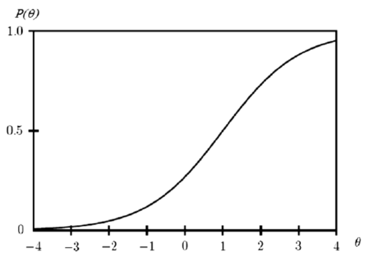

Figure 2. A typical Item Characteristic Curve.

The Item Characteristic Curve is recognized with the cumulative form of the logistic function. This paper uses the one-parameter logistic model (1PL) or Rasch model to estimate the ability of the learner. According to [10] the formula of 1PL is defined as (9):

$$P_i(\theta) = \frac{1}{1 + \exp(-D(\theta - b_i))} \tag{9}$$

where $P_i(\theta)$ represents the probability of a correct answer to a test item i, b_i is the difficulty of the $i-$th concept which, in this paper, is calculated by the DCDC Unit, and D is a constant 1.702. Since each content conveys certain concept, both the concept and its corresponding content will have the same difficulty level.

This paper estimate learner's ability by using Bayesian estimation. The Hermite-Gauss quadrature [11] is used in to approximate the ability of the learner as (10):

$$\theta = \frac{\sum_k^q \theta_k L(u_1, u_2, \ldots, u_n \mid \theta_k) A(\theta_k)}{\sum_k^q L(u_1, u_2, \ldots, u_n \mid \theta_k) A(\theta_k)} \tag{10}$$

where θ_k is one of q equidistant quadrature points comprised between the limited range of ability. $A(\theta_k)$ is the weight associated to each of the quadrature points according to a normal probability distribution with mean zero and variance one. $L(u_1, u_2, \ldots, u_n \mid \theta_k)$ is the likelihood of the answers pattern after the administration of i items. In (10), the likelihood function $L(u_1, u_2, \ldots, u_n \mid \theta_k)$ is calculated as (11):

$$L(u_1, u_2, \ldots u_n \mid \theta_k) = \prod_{i=1}^{n} P_i(\theta_k)^{u_i} Q_i(\theta_k)^{1-u_i} \tag{11}$$

As in (9), $P_i(\theta_k)$ represents the probability that learner answers the $i-$th test item correctly. $Q_i(\theta_k)$ represents the probability that learner cannot give a correct answer to the $i-$th test item and is calculated as $Q_i(\theta_k) = 1 - P_i(\theta_k)$. u_i is the result of the learner answer to the test item i and is 1 for correct answer and 0 for the incorrect answer to item i.

The proposed system uses the feedbacks from the leaner for estimation of his/her ability. If learner gives correct answers to the test items, then his/her ability will be increased using the formula mentioned in (10); otherwise, learner's ability will be decreased. Having estimated the learner's ability, the Content Recommendation Unit selects and ranks a series of appropriate learning contents in the Learning Content Repository based on the value of the Item Information Function. This value depends on the matching degree between the difficulty of the item and the learner's ability. Contents that their test item has higher information value are more appropriate for the learner. The Item information function is defined as (12):

$$I_i(\theta) = P_i(\theta)Q_i(\theta) \tag{12}$$

where $P_i(\theta_k)$ and b_i are the information value and the difficulty parameter of the $i-$th test item respectively. As mentioned previously, the value of b_i is calculated by the DCDC Unit. Hence, the Content Recommendation Unit ranks the contents according to the information values of their corresponding test items. Finally, this unit recommends the ranked contents to the learner.

IV. CONCLUSION

This research proposes a system to provide personalized learning path for the learner. Choquet Fuzzy Integral is used to dynamically calculate the difficulty level of learning concepts. Then the content sequencing is adapted to the learner's ability by using Item Response Theory. Choquet Fuzzy Integral considers both the information in the learner model and the relationship between the concepts of the knowledge domain, and hence is more precise than the static methods which consider equal concept difficulty for all the learners with different levels of knowledge. Also, Choquet Fuzzy Integral has less complexity compared to the work done in [5] due to the use of Constraint-Independent learner model. Moreover, it can be easily applied to different knowledge domains. Using both the Choquet Fuzzy Integral and Item Response Theory, the presented system in this paper provides a personalized learning path for the learner such that the difficulties of the contents are highly matched to his/her knowledge and ability. This, in turn, leads to a more efficient learning process. Future works aims to provide a prototype system and evaluate its efficiency through experimental analysis.

ACKNOWLEDGEMENTS

This paper is provided as a part of the research performed in the Advanced E-learning Technologies Laboratory in Amirkabir University of Technology and is financially supported by Iran Telecommunication Research Center (ITRC). Therefore, the authors would like to thank this center for its significant help.

REFERENCES

[1] S. Alexander, "E-Learning developments and experiences," Education+Training, vol. 43, pp. 240–248, 2001.

[2] C.-M. Chen, C.-Y., Liu and M.-H., Chang, "Personalized curriculum sequencing utilizing modified item response theory for web-based instruction," Expert Systems with Applications, vol. 30, pp. 378–396, 2006.

[3] P. Brusilovsky, "Adaptive and intelligent technologies for web-based education," Künstliche Intelligenz, vol. 4, pp. 19-25, 1999.

[4] F. Zhu and N. Yao, "Ontology-based learning activity sequencing in personalized education system," Proc. 2009 International Conference on Information Technology and Computer Science, itcs, vol. 1, pp.285-288, doi:10.1109/ITCS.2009.64.

[5] A. Mitrovic and B. Martin, "Evaluating adaptive problem selection," Lecture Notes in Computer Science, vol. 3137, pp. 185–194, 2004.

[6] T. WEI, "An intelligent tutoring system for Thai writing using constraint based modeling," Master of Science, School of Computing, National University of Singapore, 2005.

[7] J.-I. Shieh, H.-H., Wu and H.-C., Liu, "Applying a complexity-based choquet integral to evaluate students' performance," Expert Systems with Applications, vol. 36, pp. 5100–5106, 2009.

[8] J.-H. Chiang, "Aggregating membership values by a choquet-fuzzy-integral based operator," Fuzzy Sets and Systems, vol. 114, pp. 367-375, 2000.

[9] S. Medasani, J., Kim and R., Krishnapuram, "An overview of membership function generation techniques for pattern recognition," International Journal of Approximate Reasoning, vol. 19, pp. 391–417, 1999.

[10] F. Baker, "The basics of item response theory," ERIC clearing house on assessment and evaluation, 2001.

[11] J.-G. Blais and G. Raîche, "Some features of the sampling distribution of the ability estimate in computerized adaptive testing according to two stopping rules," presented at the 11th International Objective Measurement Workshop, New Orleans, 2002.