Personalized Adaptive Learning using Neural Networks

Devendra Singh Chaplot
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
chaplot@cmu.edu

Eunhee Rhim
Samsung Electronics Co., Ltd.
Suwon, South Korea
eunhee.rhim@samsung.com

Jihie Kim
Samsung Electronics Co., Ltd.
Suwon, South Korea
jihie.kim@samsung.com

L@S 2016, April 25-26, 2016, Edinburgh, Scotland Uk
ACM 978-1-4503-3726-7/16/04.
http://dx.doi.org/10.1145/2876034.2893397

Abstract
Adaptive learning is the core technology behind intelligent tutoring systems, which are responsible for estimating student knowledge and providing personalized instruction to students based on their skill level. In this paper, we present a new adaptive learning system architecture, which uses Artificial Neural Network to construct the Learner Model, which automatically models relationship between different concepts in the curriculum and beats Knowledge Tracing in predicting student performance. We also propose a novel method for selecting items of optimal difficulty, personalized to student's skill level and learning rate, which decreases their learning time by 26.5% as compared to standard pre-defined curriculum sequence item selection policy.

Author Keywords
Adaptive Learning; Neural Networks; Learner Model; Instructional Model; Student Model; Personalized Item Selection.

Introduction
Adaptive learning refers broadly to a learning process where the content taught or the way such content is presented changes, or "adapts," based on the responses of the individual student [5]. It is the core technology for intelligent tutoring systems having 3

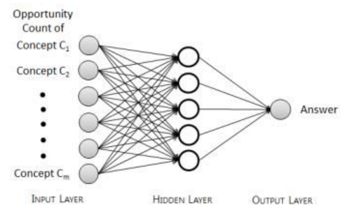

Figure 2: Artificial Neural Network used for Learner Model

Refined NN	Equation
NN-I	$f_{NN}(\boldsymbol{C}) + b_j$
NN-S	$\alpha_k f_{NN}(\boldsymbol{C})$
NN-SI	$\alpha_k f_{NN}(\boldsymbol{C}) + b_j$

Table 1: Equations for refined Neural Networks

major components: model of content to be learned (Content Model), model to estimate student proficiency (Learner Model) and a model to present content to the student in a personalized fashion based on his proficiency (Instructional Model).

The proposed adaptive learning system overcomes two important shortcomings of existing adaptive learning systems: (1) inability of Learner Model to handle multi-concept problems and (2) inability of Instructional Model to systematically select problems of appropriate difficulty for the student to maximize learning gain. We propose a new adaptive learning system architecture as shown in Figure 1, based on Artificial Neural Networks, which overcomes these shortcomings.

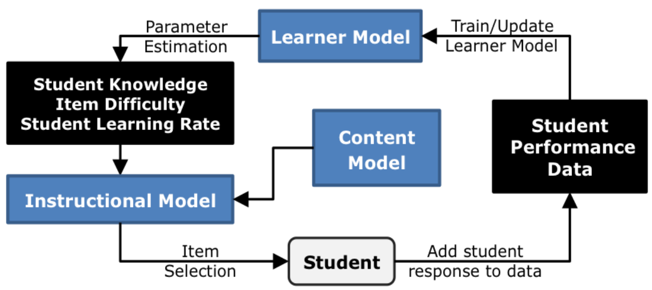

Figure 1: Adaptive Learning System Architecture

Content Model

The Content Model consists of a set of concepts, which are represented in a pre-requisite graph [1]. The pre-requisite graph is a directed acyclic graph with edges denoting the order in which concepts need to be mastered. The content model also consists of items. An 'item' refers to any entity like problem, question, step, quiz, etc. which is not broken down into smaller entities and involves certain concept(s).

Learner Model

Classical learner models in both Logistic Regression and Bayesian Knowledge Tracing (BKT) [2] families are unable to handle multi-concept items. Conjunctive BKT, Additive Factor Model and Conjunctive Factor Model were proposed to handle this problem, but they are limited by mathematical assumptions in the underlying cognitive model [3]. In order to handle multi-concept problems, we propose a new learner model using an artificial neural network, which does not assume any relationship between the inputs (concepts in Content Model) contrary to previous methods and can leverage huge amount of student performance data available for educational data mining to identify complex non-linear relationships between the concepts. Student performance data contains student-item transactions, each containing Student ID $T = T_k$, Item ID $M = M_j$, set of concepts involved in item M_j (denoted by S_j), Current Opportunity Count(s) (OC) [2] of concept(s) in set S_j and student response X_t (1 for correct, 0 for incorrect). The OC(s) of concept(s) involved in the item and corresponding student response are used as input and output, respectively, for training the Neural Network as shown in Figure 2.

Estimating Item Difficulty

Let's denote the output of the neural network for an item M_j, by $f_{NN}(OC_i : C_i \subseteq S_j)$. The output of the neural network is the mean of predicted performance over all items involving input concepts. We estimate difficulty b_j of item M_j, by calculating the average difference between predicted and real values of student performance on that item.

$$b_j = \frac{1}{n_j} \sum_{M=M_j} \left(f_{NN}(OC_i : C_i \subseteq S_j) - X_t \right)$$

Here, n_j is the number of transaction for item M_j.

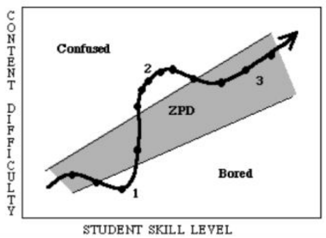

Figure 3: Zone of Proximal Development (ZPD) illustration [4]

ESTIMATING STUDENT LEARNING RATE

We create an individualized neural network for each student T_k, which is trained only on transactions by that student. Then we take the ratio of sum of outputs from individualized NN to the sum of outputs from general NN to estimate student learning rate α_k:

$$\alpha_k = \frac{\sum_{T=T_k} f_{NN_k}(OC_i : C_i \subseteq S_j)}{\sum_{T=T_k} f_{NN}(OC_i : C_i \subseteq S_j)}$$

Here, the individualized neural network for student k is denoted by f_{NN_k}. The original Neural Network can be refined using these estimates of item difficulty (NN-I), student learning rate (NN-S) or both (NN-SI) as described in Table 1.

Instructional Model

Instructional Model is responsible for selecting practice items of optimal difficulty, which maximize `Learning Gain' or the increase in student skill level. Theories of zone of proximal development (ZPD) [4] (See Figure 3) and item information function [1] indicate that an item of appropriate difficulty matches the current student skill level. Probability of solving an item correctly is analogous to the current student skill level on the concepts involved in the item (when difficulty is constant). As the item difficulty increases, the probability of correctly solving the item decreases, and vice versa. Consequently, we formulate `Learning Gain' of an item M_j as the geometric mean of both the quantities so that it is maximized when chances of correctness and difficulty are balanced:

$$LG\ (M_j) = \rho \sqrt{sigmoid(b_j) * P(X = 1)}$$

where ρ is a constant
b_j is the difficulty of item M_j, $b_j \in (-\infty, \infty)$
$P(\mathrm{X}{=}1)$ is the prob. of solving item correctly

Intuitively, a student with higher learning rate should be given more challenging items and should have higher learning gain than another student having the same skill level but a lower learning rate. Thus, we define 'Personalized Learning Gain' of an item M_j for student T_k, having learning rate $\alpha_k \in (0, \infty)$, as

$$PLG\ (M_j, T_k) = \rho \sqrt{sigmoid(b_j) * P(X = 1)^{\frac{1}{\alpha_k}}}$$

The learning gain for each item can be calculated using estimates of b_j, α_k and $P(X = 1)$ from the Learner Model.

We propose two item selection policies:

1. Max Learning Gain (NN): Selecting the item with highest LG.
2. Max Personalized Learning Gain (NN): Selecting the item with highest PLG.

In both the policies, if all concepts involved in an item are already mastered, then that item is discarded. Student response to the selected item is used to update the learner model, which is used to select the next item. This process is repeated until all concepts are mastered i.e. when probability of knowing the concept is greater than a particular threshold, typically 0.95.

Experiments & Results

PREDICTING STUDENT PERFORMANCE

Results on Algebra 2008-2009 Course Data from KDD Cup 2010 show that Neural Networks outperform standard and individualized BKT [6] on the task of predicting student performance, as shown in Table 2.

Method	Accuracy	RMSE	AUC
Standard BKT	82.7	0.363	-
Individualized BKT	82.8	0.361	-
Neural Network	88.3	0.320	0.687
NN-S	88.4	0.318	0.693
NN-I	88.8	0.308	0.708
NN-SI	**88.9**	**0.307**	**0.713**

Table 2: Comparison of Learner Models

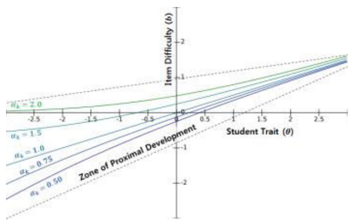

Figure 4: Optimal Item difficulty as a function of student trait for students with different pace of learning

Comparing Item Selection Policies

We generated synthetic data simulating 3000 students and 183 items over 11 concepts in Grade 6 Expressions and Equations for validating item selection policies due to unavailability of real data. The most generic 4-parameter logistic model of IRT was modified to generate the data. We defined θ_i as the latent trait over each concept C_i rather than having a single latent trait and modeled learning gain over time as defined earlier. The prerequisite graph in the content model was incorporated by constraining θ_is of post-requisites to be always less than that of prerequisites.

We evaluate item selection policy by calculating the average number of items required by students to master all concepts. Pre-defined curriculum sequence is the most common item selection policy, which selects concepts in the order pre-defined in the curriculum. To evaluate the formulation of LG and PLG, we use Maximizing Learning Gain and Maximizing Personalized Learning Gain in idealized setting using real values of parameters (item difficulty and student learning rate), which were used to generate data. As the real values will not be available in practice, we also maximize LG and PLG using parameter estimates from Neural Networks. Results in Table 3 show that maximizing PLG (ideal) reduces items required to achieve mastery by 26.5% over pre-defined curriculum sequence policy. Max PLG (NN) is able to achieve learning efficiency comparable to the ideal scenario. The plot of optimal item difficulty (b), which maximizes personalized learning gain, as a function of Student Skill level (θ) for different values of α_k is shown in Figure 4. It shows that optimal item difficulty is not only proportional to the current student skill level but also consistent with the student pace of learning.

Item Selection Policy	Avg. #items to reach mastery
Pre-defined curriculum sequence	85.29
Max Learning Gain (ideal)	70.88
Max Personalized Learning Gain (ideal)	62.66
Max Learning Gain (NN)	69.96
Max Personalized Learning Gain (NN)	64.51

Table 3: Comparison of Item Selection Policies

Conclusion

We have proposed an adaptive learning system architecture based on Artificial Neural Networks which handles multi-concept items for effective prediction of student performance and selects practice items of optimal difficulty personalized to student's skill level.

References

1. Barla, M., Bieliková, M., Ezzeddinne, A. B., Kramár, T., Šimko, M., and Vozár, O. 2010. On the impact of adaptive test question selection for learning efficiency.*Computers & Education*, *55*(2), 846-857.
2. Corbett, A. T., and Anderson, J. R. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, *4*(4), 253-278.
3. Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A. and Rosé, C. P. 2015. Data mining and education. WIREs Cogn Sci, 6: 333–353
4. Murray, T., and Arroyo, I. 2002. Toward measuring and maintaining the zone of proximal development in adaptive instructional systems. In *Intelligent Tutoring Systems* (pp. 749-758)
5. Oxman, S., and Wong, W. 2014. White Paper: Adaptive Learning Systems. Snapwiz. *Integrated Education Solutions*, 1.
6. Yudelson, M. V., Koedinger, K. R., and Gordon, G. J. 2013. Individualized bayesian knowledge tracing models. In *Artificial Intelligence in Education* (pp. 171-180).