A Neural Network Approach to Estimate Student Skill Mastery in Cognitive Diagnostic Assessments

Qi Guo, Maria Cutumisu, Ying Cui
Department of Educational Psychology, University of Alberta
{qig, cutumisu, yc}@ualberta.ca

ABSTRACT

In computer-based tutoring systems, it is important to assess students' mastery of different skills and provide remediation. In this study, we propose a novel neural network approach to estimate students' skill mastery patterns. We conducted a simulation to evaluate the proposed neural network approach and we compared the neural network approach with one of the most widely used cognitive diagnostic algorithm, the DINA model, in terms of skill estimation accuracy and the ability to recover skill prerequisite relations. Results suggest that, while the neural network method is comparable in skill estimation accuracy to the DINA model, the former can recover skill prerequisite relations more accurately than the DINA model.

Keywords

prerequisite discovery, skills, neural network, student modeling, cognitive diagnosis model

1. INTRODUCTION

In intelligent tutoring systems, assessing students' skill mastery patterns and determining skill prerequisite relationship are two important areas of research. Various approaches are proposed to solve these two problems, including Educational Data Mining (EDM) approaches, such as Bayesian Knowledge Tracing, Learning and Performance Factor Analysis (for a comparison see [5]), and psychometric approaches, such as Cognitive Diagnostic Models (CDMs) [2, 6]. Compared to CDMs, which assess student skill mastery based on their responses to a test administered at one time point (i.e., no learning occurs during the test), the EDM approaches have the advantage of assessing student learning dynamically. However, unlike CDMs, which estimate every item's psychometric properties, the EDM approaches often assume all test items that measure the same set of skills have the same psychometric properties (e.g., same guessing and slipping parameters). This assumption is unlikely to be tenable in practice, and it may lead to less accurate skill estimation and less efficient item selection. While both approaches have their strengths and weaknesses, this study will focus on developing a new CDM approach using the neural networks, and evaluate the proposed approach by comparing it with the current most popular CDM method, the DINA (deterministic inputs, noisy "and" gate) model [2] using simulated data.

2. A BRIEF INTRODUCTION TO NEURAL NETWORKS

A neural network is a supervised classification algorithm that consists of several layers of neurons (i.e., processing units) [4]. Each neuron linearly combines information from previous layers and applies a non-linear *activation* function. The most commonly-used activation function is the logistic/sigmoid function. A typical feedforward neural network consists of a layer of hidden units and a layer of output units. Mathematically, it can be represented as:

$$Y_{n,q} = sigmoid\big(\vec{1}_{n,1}\vec{b}'_{1,q} + sigmoid(\vec{1}_{n,1}\vec{b}'_{1,k} + X_{n,p}W_{p,k})W_{k,q}\big),$$

where $Y_{n,q}$ is the output matrix consisting of *n* subjects' values on *q* output variables, $X_{n,p}$ is the input matrix consisting of *n* subjects' values on *p* input variables, $\vec{b}'_{1,k}$ is a vector of intercept values for *k* hidden units, $W_{p,k}$ is the weight matrix between *p* input variables and *k* hidden units, $\vec{b}'_{1,q}$ is a vector of intercept values for *q* output units, and $W_{k,q}$ is the weight matrix between *k* hidden units and *q* output units.

One challenge in applying neural networks to estimate students' skill mastery patterns is that students' skill mastery patterns are unobserved. Thus, we only have observed values for the input variables (students' item response patterns) but not for the output variables (students' skill mastery pattern).

3. METHODOLOGY: THE PROPOSED NEURAL NETWORK APPROACH

To overcome the problem mentioned above, we propose a novel neural network model that has the same input and output (i.e., students' item response patterns). The core idea underlying our approach is to first reduce the input (student item response patterns) to a smaller number of hidden units representing students' latent skills and then use these hidden units to best reproduce student item response vectors (i.e., output) with the restriction of the Q-matrix, a matrix that specifies the set of skills measured by each item. A conceptual diagram of the proposed network is shown in **Figure 1**.

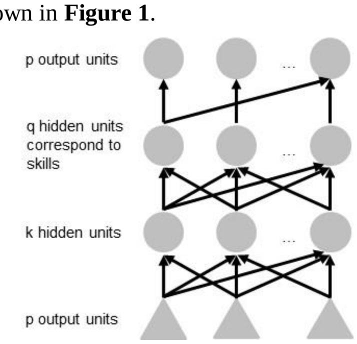

Figure 1. A diagram of the proposed neural network. Relations between skill hidden units and output units are specified based on the Q-matrix.

It is important to note that the relation between the second layer of hidden units and output units is specified based on the Q-matrix, which specifies which skills are required by each item. Intuitively, the network first extracts features from student item response patterns and then it dictates the relations between features and student item response patterns based on the Q-matrix. Mathematically, the model can be represented as follows:

$$Y_{n,p} = sigmoid\big(\vec{1}_{n,1}\vec{b}'_{1,p} + sigmoid\big(\vec{1}_{n,1}\vec{b}'_{1,q} + sigmoid(\vec{1}_{n,1}\vec{b}'_{1,k} + X_{n,p}W_{p,k})W_{k,q}\big)W_{q,p} \odot Q'_{q,p}\big),$$

where $\odot$ represents elementwise multiplication, and $Q'_{q,p}$ is the Q-matrix.

Similar to a regular neural network, the proposed model uses maximum likelihood to define the cost function and it can be optimized using some variants of gradient descent (e.g., rprop [4]). To speed up the optimization, it is important to choose meaningful starting values for the weight matrices. To initialize $W_{q,p}$, we can first train a multivariate logistic regression with all the theoretically possible skill patterns (i.e., expected theoretical plausible skill patterns) as input, and their corresponding expected item response patterns (i.e., item response pattern assuming no slips and guesses) as output, assuming slipping and guessing parameters are 0. Then, we use the weight matrix from this multivariate logistic regression as the starting values of the proposed neural network.

4. EVALUATION

In order to demonstrate the accuracy of the proposed neural network, we conducted a preliminary simulation study. Five thousand students' responses (correct/incorrect) to 28 test items were generated based on a skill prerequisite model shown in Figure 2. Skill prerequisite relations, true model used in the simulation (left); recovered using DINA skill estimates (middle) and neural network skill estimates (right)

To evaluate the recovered prerequisite relationship, we counted the number of estimated causal links that were not in the true model, and the number of missing causal links that were in the true model.

and a Q-matrix (available upon request). The guessing and slipping parameters for all items were set to 0.1. We compared the proposed method with the DINA model in terms of accuracy of 1) student skill pattern estimates and 2) skill prerequisite relation recovery. Accuracy of skill pattern estimates is defined as:

$$accuracy = 1 - \frac{|estimated\ skill\ pattern\ matrix - true\ skill\ pattern\ matrix|}{n * q},$$

where n is the sample size, and q is the number of skills in the Q-matrix. The skill prerequisite relations were recovered by using a Bayesian network to model the relations among estimated student skills. The causal direction in the Bayesian network is determined by the following heuristic [1]:

If P(skill1=0)<P(skill2=0), then skill1 is the prerequisite of skill2.

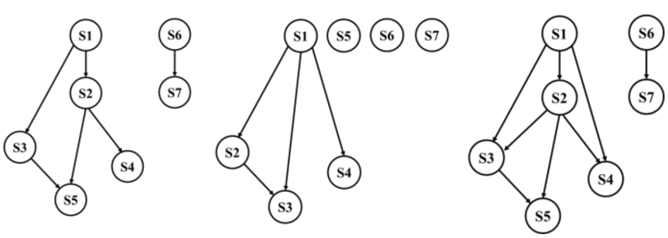

Figure 2. Skill prerequisite relations, true model used in the simulation (left); recovered using DINA skill estimates (middle) and neural network skill estimates (right)

To evaluate the recovered prerequisite relationship, we counted the number of estimated causal links that were not in the true model, and the number of missing causal links that were in the true model.

We programed our proposed neural network using Python. The number of hidden units in the first layer was set to 56. The number of hidden units in the second layer was set to seven, corresponding to seven skills in the Q-matrix. The Rprop algorithm was used to optimize the neural network. For the DINA analysis, we used the *CDM* R package [6]. For the Bayesian network analysis, we used the *bnlearn* R package's mmhc algorithm [7] and *Rgraphviz* R package [3].

The results suggested that the proposed method had similar or slightly better accuracy (89.2%) at estimating skill patterns than the DINA model (87.9%). Moreover, the proposed method was better at recovering the skill prerequisite relations. The recovered skill prerequisite relations by the DINA model and the proposed method are shown in Figure . The prerequisite relations recovered based on the DINA skill estimates only contained two arcs from the true model (i.e., S1 to S2, S1 to S3), and they contained two arcs that were not in the true model (S1 to S4, S2 to S3). The prerequisite relations recovered based on the neural network skill estimates contained all the arcs from the original model, as well as two arcs that were not in the true model (S1 to S4, S2 to S3). Overall, the results suggested that the proposed network had slightly better skill estimation accuracy than the DINA model and it was more accurate at recovering skill prerequisite relations than the DINA model.

5. CONCLUSIONS AND DISCUSSION

This study proposed a novel neural network approach to estimate student skill mastery patterns in CDM. Traditionally, parameter estimation of models with latent variables usually depends on Expectation Maximization or Markov Chain Monte Carlo methods. The proposed neural network approach frames the latent variable model problem as a supervised problem and it solves it using the gradient descent method. Initial evidence suggests that the proposed method has comparable skill estimation accuracy as the DINA model, but it can recover skill prerequisite relations better than the DINA model. Further research is needed to rigorously evaluate this method.

6. REFERENCES

[1] Chen, Y., Gonzlez-Brenes, J. and Tian, J. 2016. Joint Discovery of Skill Prerequisite Graphs and Student Models. In *International Conference on Educational Data Mining.* Raleigh, NC, IEDMS, 46-53.

[2] de La Torre, J. 2009. DINA Model and Parameter Estimation: A Didactic. *Journal of Educational and Behavioral Statistics, 34*, 1, 115-130.

[3] Gentry, J., Long, L., Gentleman, R., Falcon, S., Hahne, F., Sarkar, D. and Rgraphviz, K. H. 2009. Provides plotting capabilities for R graph objects. *R package version, 2*, 0.

[4] Goodfellow, I., Bengio, Y., Courville, A. 2016. *Deep Learning*. MIT Press.

[5] Pavlik, P. I., Cen, H., Koedinger, K.R. 2009. Performance factors analysis - A new alternative to knowledge tracing. *Proceedings of the 2009 conference on artificial intelligence in education: building learning systems that care: from knowledge representation to affective modelling* (Brighton, 2009), 531-538.

[6] Robitzsch, A., Kiefer, T., George, A. and Uenlue, A. 2014. *CDM: Cognitive diagnosis modeling.R package version 3.1-14,* .

[7] Scutari, M. bnlearn: Bayesian Network Structure Learning, R package version 2.7 (2011). *URL* http://www.bnlearn.com.