AMCS Comprehensive Examination Proposal

Attn: Dr. Rodica Curtu, AMCS Interim Director

September 11, 2019

Student

Geoffrey Converse

Committee Members

Dr. Suely Oliveira[†] (Chair)	Computer Science
Dr. David Stewart[†]	Mathematics
Dr. Xueyu Zhu[†]	Mathematics
Dr. Bruce Ayati[†]	Mathematics
Dr. Jonathan Templin	Psychological and Quantitative Foundations

Date and Time

Tuesday, October 1
3:30 pm to 5:30 pm
Location TBD

[†] AMCS Affliated

Abstract

Item Response Theory (IRT) is an area of Educational Assessment that studies how a subject's responses to an assessment can serve as indicators of that subject's underlying skills, or latent traits. For example, a student may answer 88% of questions correctly on a math exam. But this accuracy score isn't a great measure of skill - other students could have received the same score, yet it is unlikely that they have the exact same skill level. This latent trait of "math skill" is not directly observable and can't be discerned from a test score accuracy. In order to quantify a latent trait, it is assumed that each subject's value is drawn from a standard normal distribution, $\mathcal{N}(0, 1)$.

This talk introduces basic IRT models, particularly the 2-Parameter Logistic Model (2PL), which gives the probability of a student answering a particular assessment item correctly. Formally, the probability of student j answering item i correctly (event $X_{ji} = 1$) is given by

$$P(X_{ji} = 1|\theta_j) = \frac{1}{1 + \exp[a_i\theta_j - b_i]}, \tag{1}$$

where θ_j is subject j's skill value, a_i is a discrimination parameter which quantifies the amount of skill that item i requires, and b_i is the difficulty parameter of item i. Note that this can be extended to higher dimensional latent traits, so that a subject has a set of J independent, normally distributed latent traits $\Theta_j = (\theta_1, ..., \theta_J)^T$; this model is then called the Multidimensional Logistic 2-Parameter (ML2P) model, and is the focus of our work.

While Equation 1 gives the probability of a correct answer given certain parameters, this equation is not often directly applicable. The skill θ is not observable, and it can be difficult to accurately assign the values a_i and b_i. Given results to particular assessment, there are various

methods of estimating the parameters a_i and b_i, in addition to predicting students' latent trait values. These typically include Maximum Likelihood Estimation (MLE) or the Expectation-Maximization (EM) algorithm. However, some of these methods can only estimate either the model parameters **or** skill parameters, not both. This is an issue since there are often times where none of these values are known. Though there exists methods to estimate both at the same time, such as Joint Maximum Likelihood Estimation (JMLE), there are many limitations in implementation and practical applications.

In recent work, we have developed new methods of simultaneous parameter recovery and latent trait prediction by using a Variational Autoencoder (VAE), a special class of neural network. A VAE aims to break down (encode) data to a smaller dimension and fit it to a specified distribution, then reconstruct (decode) the original input. Neural networks provide an interesting relationship to the ML2P model in that Equation 1 is very similar to the common sigmoidal activation function. A modification to the architecure of a VAE, using a binary matrix relating skills and items, allows interpretation of the weight and bias parameters of the decoder to serve as discrimination and difficulty parameter estimates. This also allows interpretation of the learned distribution hidden layer of the VAE.

This talk summarizes the new model, ML2P-VAE, along with a comparison to a similar method. In addition, we provide an additional application of the same model architecture to the field of baseball analytics. Currently, we are working on developing a software package for easy implementation of ML2P-VAE, and are exploring using latent traits that are not independent of one another. Specifically, we are interested in directly comparing the ML2P-VAE method to more traditional estimation methods when latent traits are correlated with one another.