Skills Diagnosis with Latent Variable Models

Abstract

The primary aim of skills diagnosis is to develop and analyze tests in ways that reveal information with more diagnostic value, when compared with traditional approaches. In the methods for skills diagnosis that we consider mastery of particular skills or states of knowledge can be represented by a list of binary latent variables, indicating mastery of each of a finite set of skills under diagnosis. The main objective of skills diagnosis is to classify examinees according to this list of skills. In this training session, several popular modeling and classification approaches will be discussed. Three conjunctive latent class models known as the DINA, NIDA, and Fusion models will be introduced, and software for fitting these models with Mplus will be demonstrated. Because of the multidimensional nature of these models, estimation benefits greatly if it can adapt to previous responses. To address this, computerized adpative testing (CAT) is considered. Because Fisher information does not apply to discrete latent variables, alternative and computationally simple item selection rules are introduced. For CAT settings in which both traditional and diagnostic models are being used, CAT algorithms are introduced for ensuring reliable information for these dual objectives. In addition to sequential methods of test construction, indices for use in fixed-length test construction are also given. The training session is meant to provide practical guidelines for implementing skills diagnosis, and considers the essential topics of identifying the attributes measured by items as well as test equating.

Topics

Multiple Classification Latent Class Models (DINA, NIDA, Fusion): Models for item responses in skills diagnosis often arise from constructing a sequence of unobservable responses to subtasks that must all be correct in order to correctly answer the item (Embretson 1984, 1997; Maris, 1999). By recognizing that the performances of examinees cannot be precisely predicted from their list of mastered skills, stochastic models allow for the possibility of "slips" and "guesses". The models we consider are largely distinguished by whether slips and guesses occur at the subtask level (NIDA model) or at the item level (DINA model). Dr. Jeff Douglas will describe the DINA and NIDA models with extensions and appl ications.

Q-matrix identification: Dr. Jonathon Templin will discuss how to determine the matrix that relates items to attributes. Methods based on expert opinion as well as exploratory

methods will be introduced.

Software and Model Estimation: Dr. Jonathon Templin will demonstrate software for fitting skills diagnosis models that works in conjunction with the software Mplus. Mplus can fit latent class models with general constraints, and Dr. Templin will show how these constraints may be specified using his software.

Equating: Equating is a common concern in any testing operation using item response theory or classical methods. Dr. de la Torre will discuss the role that equating plays in skills diagnosis and gives some results from his studies on equating in this setting.

Goodness-of-Fit: Assessing the quality of a model and choosing among competing models requires techniques for assessing model fit. Dr. Jimmy de la Torre will describe the use of posterior predictive checks and related residual analyses that can be used for diagnosing the appropriateness of a model. These methods utilize many of the familiar ideas in confirmatory factor analysis. In addition, he will discuss Bayes factors as a way of choosing a model when Bayesian estimation techniques are used.

CAT: Classification of examinees according to a list of binary skill indicators can take place more accurately and more efficiently if items are selected in an adaptive manner, depending on an examinees previous responses. Dr. Hua Chang has done extensive research on the theory and methods for doing this as efficiently as possible. He will describe the Shannon entropy criterion as well as other criteria for item selection in cognitive diagnosis. These techniques are designed to quickly distinguish true from false attribute patterns, and identify an examinee's true knowledge state. As a separate topic, many testing programs are interested in estimating a broadly defined latent trait to summarize scores, while also gleaning some diagnostic information from an exam. Dr. Hua-Hua Chang will also present procedures using ideas of item bank stratification that will carefully balance the ability of a test to estimate a single summary while also yielding reliable information for cognitive diagnosis in the context of a computerized adaptive test.

Test Construction: Little has been developed to specify how one might construct a good test using aspects of skills diagnosis models. To address this, Dr. Henson will discuss the derivation of three indices based on Kullback-Leibler information, which serves as a measure of how informative an item is for the classification of examinees. The effectiveness of these indices are briefly examined for items calibrated using the Deterministic Input; Noisy "and" Gate model (DINA) and the Reparameterized Unified Model (RUM) by implementing a simple heuristic to construct a test from an item bank. Real world appli-

cations of these indices will then be discussed in both test construction and instrument refinement.

Schedule

A tentative schedule for the training session is given below. It will begin at 9:00 am and will end at 3:45 pm, and will include several breaks to interact with participants as well as a 75 minute lunch break.

9:00-9:15	Introduction	Jeff Douglas
9:15-10:00	Latent Variable Models	Jeff Douglas
10:00-10:15	Break	
10:15-11:00	Q-Matrix Construction	Jon Templin
11:00-11:45	Test Construction	Robert Henson
11:45-1:00	Lunch	
1:00-2:00	Software Demonstration	Jon Templin
2:00-2:30	Equating	Jimmy de la Torre
2:30-2:45	Break	
2:45-3:15	Goodness of Fit	Jimmy de la Torre
3:15-3:45	CAT	Hua-Hua Chang