A Longitudinal Diagnostic Facet Status Model: Tracking the Learning Trajectory

He Ren and Chun Wang
College of Education, University of Washington
Supported by NSF #2300382

NCME 2025, 4/26, Denver

UNIVERSITY *of* WASHINGTON | COLLEGE OF EDUCATION

U.S. National Science Foundation

2025 ANNUAL MEETING ncme DENVER

Outline

- Background
- Methods
- Simulation Study
- Discussions

Background: Cognitive Diagnostic Models (CDMs)

- Assess students' fine-grained learning status

	A1	A2	A3
Item 1	1	0	0
Item 2	1	0	1
⋮			

- Each item measures certain knowledge or skills (i.e., attributes)
- Each student has a latent attribute profile

	A1	A2	A3
STU 1	1	1	1
STU 2	1	0	0
⋮			

- Predict students' response by linking their profiles with the attributes targeted by the item

Background: Diagnostic Facet Status Model (DFSM)

- A CDM for multiple-choice items (Wang, 2024)
  - Build on the concept of facets rather than attributes
  - A facet is one piece or a set of several pieces of learner's understanding on certain knowledge
  - Each response option targets at certain facets and all response options are included

Background: Diagnostic Facet Status Model (DFSM)

- Frank was thinking about the forces acting on a box at rest on his desk. Which statement do you think best represents the forces acting on the box?
  A) No forces act on the box at rest.
  B) The force of gravity is the only force acting on the box.
  C) Earth pulls down and the table pushes up equally on the box.
  D) The downward pull of gravity is larger than the upward push of the desk.

Option-facet mapping for the example item

OPTION	α1	α2	α3	β4	β5	β6
A					1	1
B					1	
C	1	1	1			
D	1	1		1		

α1 Students can identify all the relevant sources of forces on an object.

α2 Students can correctly identify all the relevant direction a force is acting.

α3 Students can compare the relative sizes of forces on static objects.

β4 For an object at rest or moving horizontally, the downward force is greater than the upward force.

β5 Passive objects cannot exert a force even though they touch another object.

β6 Motion determines the existence of a force or forces.

Methods: DFSM

- Probability of choosing option *k*

$$\theta_{ijk} \equiv P(X_{ij} = k \mid \boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, \boldsymbol{\lambda}_j)$$
$$= \frac{\exp(\lambda_{j,0} + \boldsymbol{\lambda}_{j,1}^T[\boldsymbol{h}(\boldsymbol{\alpha}_i, \boldsymbol{q}_{j,k}^g) - \boldsymbol{h}(\boldsymbol{\alpha}_i, \boldsymbol{q}_{j,K}^g)] + \boldsymbol{\lambda}_{j,2}^T \boldsymbol{h}(\boldsymbol{\beta}_i, \boldsymbol{q}_{j,k}^p))}{1 + \sum_{k=1}^{K-1} \exp(\lambda_{j,0} + \boldsymbol{\lambda}_{j,1}^T[\boldsymbol{h}(\boldsymbol{\alpha}_i, \boldsymbol{q}_{j,k}^g) - \boldsymbol{h}(\boldsymbol{\alpha}_i, \boldsymbol{q}_{j,K}^g)] + \boldsymbol{\lambda}_{j,2}^T \boldsymbol{h}(\boldsymbol{\beta}_i, \boldsymbol{q}_{j,k}^p))},$$

  - $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ refer to the goal and intermediate facets, respectively
  - $\lambda_{j,0}$ is the intercept
  - $\boldsymbol{\lambda}_{j,1}^T$ and $\boldsymbol{\lambda}_{j,2}^T$ are slopes for goal and intermediate facets
  - $\boldsymbol{\lambda}_{j,1}^T \boldsymbol{h}(\boldsymbol{\alpha}_i, \boldsymbol{q}_{j,k}^g)$ and $\boldsymbol{\lambda}_{j,2}^T \boldsymbol{h}(\boldsymbol{\beta}_i, \boldsymbol{q}_{j,k}^p)$ are linear combinations of main effects and interactions

Methods: Regularized EM for DFSM

- The marginal likelihood is

π_l: mixing proportion of latent profile l

$$L(\boldsymbol{\lambda}, \boldsymbol{\pi} \mid \mathbf{X}) = \prod_{i=1}^{N} \left\{ \boxed{\sum_{l=1}^{M} \pi_l} \prod_{j=1}^{J} \prod_{k=1}^{K} \theta_{ijk}^{I(x_{ij}=k)} \right\},$$

  - Lots of latent profiles are **impermissible**

- Permissible and impermissible profile
  - With D (e.g., $D = 4$) facets, there are 2^D ($2^4 = 16$) possible facet profile
  - Profile: $\alpha1\ \alpha2\ \alpha3\ \beta4$

	0000	1000	1100	1101	1111	0100	0110	…
π_l	0.1	0.2	0.4	0.2	0.1	0	0	0

$$\alpha1 \rightarrow \alpha2 \rightarrow \beta4 \rightarrow \alpha3$$

Methods: Regularized EM for DFSM

- To shrink the $\pi_l s$ corresponding to impermissible patterns to 0, a regularized expectation maximization (REM) algorithm has been proposed

- The penalized log likelihood is

$$\ell(\boldsymbol{\lambda}, \boldsymbol{\pi} \mid \mathbf{X}) = \underbrace{\boxed{\sum_{i=1}^{N} \log \left\{ \sum_{l=1}^{M} \pi_l \prod_{j=1}^{J} \prod_{k=1}^{K} \theta_{ijk}^{x_{ij}=k} \right\}}}_{\text{The log-likelihood}} + \underbrace{\boxed{\gamma \sum_{l=1}^{M} \log_{\rho_N}(\pi_l),}}_{\text{penalty term}}$$

penalty term
note $\pi_l \in (0,1)$

Methods: Regularized EM for DFSM

- EM algorithm is a popular method for model estimation with missing variable

- E-step: Expectation of the complete data log likelihood

log likelihood of complete data (both $\mathbf{X}$ and $\boldsymbol{\theta}$)

$$
\begin{aligned}
\boxed{\ell(\boldsymbol{\lambda}, \boldsymbol{\pi} \mid \mathrm{X}, \boldsymbol{\theta})} &= \log L(\boldsymbol{\lambda}, \boldsymbol{\pi} \mid \mathrm{X}, \boldsymbol{\theta}) \\
&= \log \left[L(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\pi} \mid \mathrm{X})\, P(\boldsymbol{\theta} \mid \boldsymbol{\pi}, \boldsymbol{\lambda}) \right] \\
&= \boxed{\log L(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \mathrm{X})} + \boxed{\log P(\boldsymbol{\theta} \mid \boldsymbol{\pi})}
\end{aligned}
$$

log likelihood of observed data

log probability of the latent profile

- The bold $\boldsymbol{\theta}$ refers to the latent profile $(\boldsymbol{\alpha}, \boldsymbol{\beta})$

Methods: Regularized EM for DFSM

- E-step: Expectation with respect to the posterior of latent profile

$$\ell(\boldsymbol{\lambda}, \boldsymbol{\pi} \mid \mathbf{X}, \boldsymbol{\theta}) = \underbrace{\boxed{\sum_{i=1}^{N} \left[\sum_{j=1}^{J} \sum_{k=1}^{K} \mathbb{I}(x_{ij} = k) \log \theta_{ijk}}}_{\text{log likelihood of observed data}} + \boxed{\log P(\boldsymbol{\theta}_i \mid \boldsymbol{\pi}) \Big]}$$

log probability of the latent profile

log likelihood of observed data

$$E[l(\boldsymbol{\lambda}, \boldsymbol{\pi} \mid \mathbf{X}, \boldsymbol{\theta})] = \sum_{j=1}^{J} \left\{ \sum_{i=1}^{N} \sum_{l=1}^{M} \hat{\eta}_{il} \sum_{k=1}^{K} I(x_{ij} = k) \log \theta_{ijk} \right\} + \sum_{i=1}^{N} \sum_{l=1}^{M} \hat{\eta}_{il} \log \pi_l \quad + \gamma \sum_{l=1}^{M} \log_{\rho_N}(\pi_l)$$

- where $\hat{\eta}_{il}$ is the posterior of the latent profile

$$\begin{aligned} \hat{\eta}_{il} &\equiv P(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_l, \boldsymbol{\beta}_i = \boldsymbol{\beta}_l \mid \boldsymbol{\lambda}, \boldsymbol{\pi}^{(r)}) \\ &= \frac{\pi_l^{(r)} \prod_{j=1}^{J} \prod_{k=1}^{K} \theta_{ijk}^{I(x_{ij}=k)}}{\sum_{l=1}^{M} \pi_l^{(r)} \prod_{j=1}^{J} \prod_{k=1}^{K} \theta_{ijk}^{I(x_{ij}=k)}}, \end{aligned}$$

Methods: Regularized EM for DFSM

- M-step: Maximize the expectation with respect to our parameters

$$E[l(\boldsymbol{\lambda}, \boldsymbol{\pi} \mid \mathbf{X}, \boldsymbol{\theta})] = \boxed{\sum_{j=1}^{J} \left\{ \sum_{i=1}^{N} \sum_{l=1}^{M} \hat{\eta}_{il} \sum_{k=1}^{K} I(x_{ij} = k) \log \theta_{ijk} \right\}} + \sum_{i=1}^{N} \sum_{l=1}^{M} \hat{\eta}_{il} \log \pi_l + \gamma \sum_{l=1}^{M} \log_{\rho_N}(\pi_l)$$

π_l is not involved

- Due to the constrain $\sum_{l=1}^{M} \pi_l = 1$, a Lagrange multiply s was employed

$$E[l(\boldsymbol{\lambda}, \boldsymbol{\pi} \mid \mathbf{X}, \boldsymbol{\theta})]^{\boldsymbol{\pi}} = \sum_{i=1}^{N} \sum_{l=1}^{M} \hat{\eta}_{il} \log \pi_l + \gamma \sum_{l=1}^{M} \log_{\rho_N}(\pi_l) + \boxed{s(1 - \sum_{l=1}^{M} \pi_l),}$$

Lagrange multiply

- Take the derivative with respect to π_l and set it to 0

$$\pi_l = \frac{\gamma + \sum_{i=1}^{N} \hat{\eta}_{il}}{\sum_{l=1}^{M} (\gamma + \sum_{i=1}^{N} \hat{\eta}_{il})}$$

Methods: Longitudinal DFSM

- The marginal likelihood for longitudinal DFSM

transition parameter

$$L(\boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{\tau} \mid \mathbf{X}) = \prod_{i=1}^{N} \left\{ \boxed{\sum_{l_1=1}^{M} \sum_{l_2=1}^{M}} \pi_{l_1} \boxed{\tau_{l_2|l_1}} \prod_{t=1}^{2} \prod_{j=1}^{J_t} \prod_{k=1}^{K} \log \theta_{ijkt}^{I(x_{ijt}=k)} \right\}$$

l_1 and l_2 are profiles at two time points

	0000	1000	1100	1101	1111	0100	…
π_l	0.1	0.2	0.4	0.2	0.1	0	0

$\tau_{l_2\|l_1}$	0000	1000	1100	1101	1111	0100	…
0000						0	
1000						0	
1100						0	
1101						0	
1111						0	
0100	0	0	0	0	0	0	0
…						0	

Methods: Regularized EM for Longitudinal DFSM

- REM algorithm 1 (One-stage method)

penalty term

$$\ell(\boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{\tau} \mid \mathbf{X}, \boldsymbol{\theta}) = \sum_{i=1}^{N} \log \left\{ \sum_{l_1=1}^{M} \sum_{l_2=1}^{M} \pi_{l_1} \tau_{l_2|l_1} \prod_{t=1}^{2} \prod_{j=1}^{J_t} \prod_{k=1}^{K} \log \theta_{ijkt}^{I(x_{ijt}=k)} \right\} + \boxed{\gamma \sum_{l=1}^{M} \log_{\rho_N}(\pi_{l1})}$$

- In the r-th iteration in EM

when $\pi_{l_1}^{(r)} = 0$ and set the corresponding **rows and columns** in $\tau_{l_2|l_1}^{(r)} = 0$

Methods: Regularized EM for Longitudinal DFSM

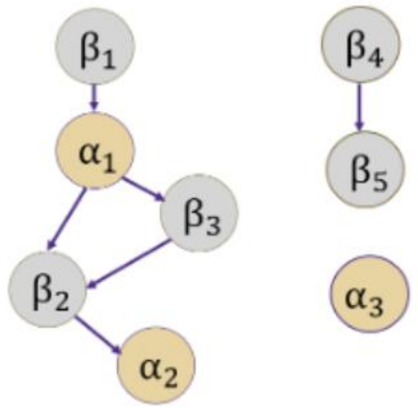
- REM algorithm 2 (Two-stage method)
  - Estimate DFSM with time point 1 responses to obtain the impermissible $\pi_l s$
  - Fix the corresponding rows in $\tau_{l_2|l_1}$ at 0
  - Estimate the longitudinal DFSM with group-wise penalty on the column of $\tau_{l_2|l_1}$

penalty term

$$\ell(\boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{\tau} \mid \mathbf{X}, \boldsymbol{\theta}) = \sum_{i=1}^{N} \log \left\{ \sum_{l_1=1}^{M} \sum_{l_2=1}^{M} \pi_{l_1} \tau_{l_2|l_1} \prod_{t=1}^{2} \prod_{j=1}^{J_t} \prod_{k=1}^{K} \log \theta_{ijkt}^{I(x_{ijt}=k)} \right\} + \boxed{\gamma \sum_{l_2} \log_{\rho_N} \Big(\sum_{l_1} \tau_{l_2|l_1} \Big)}$$

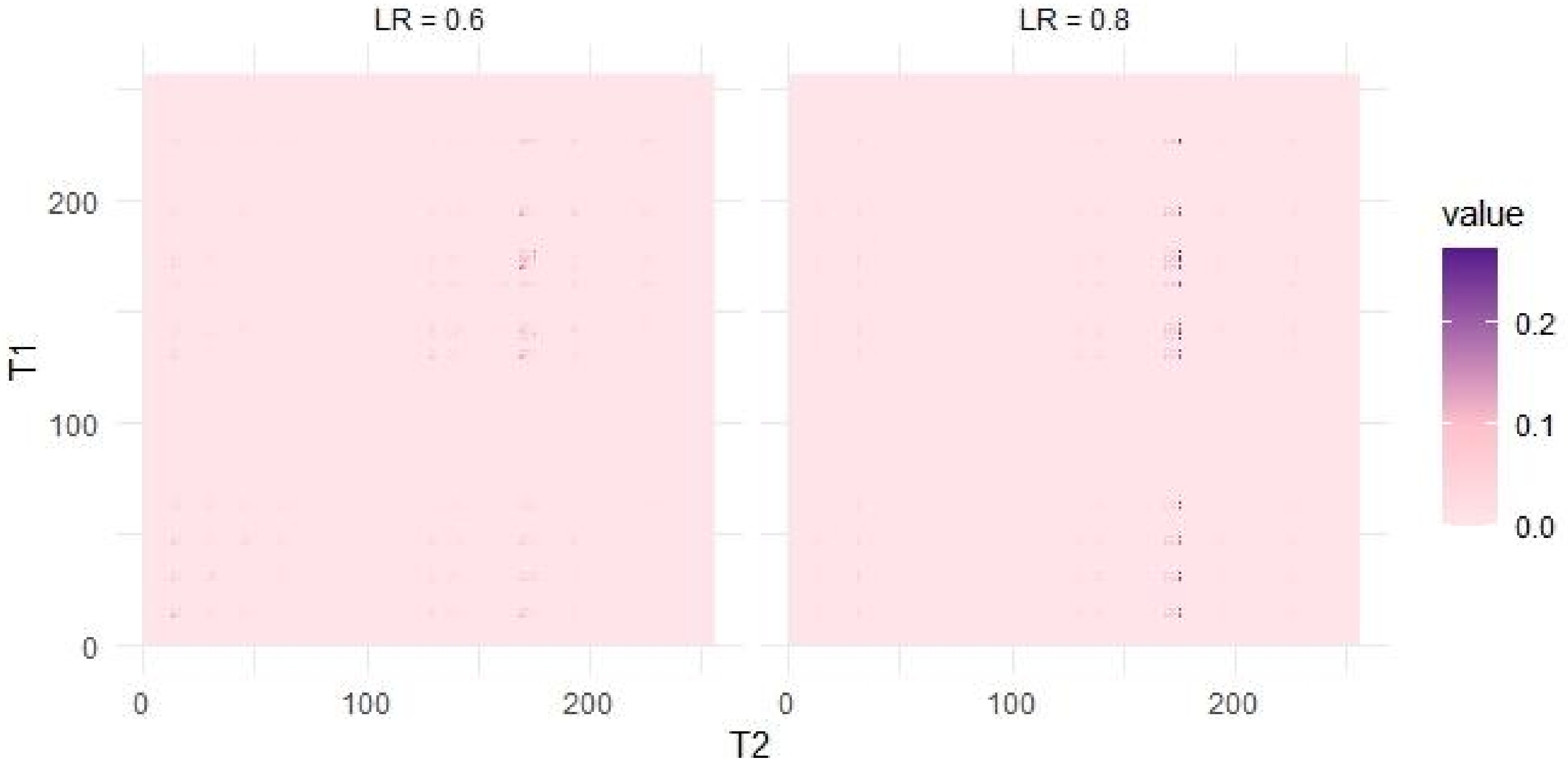
Simulation Study: Design

- 27 items, each with 4 options
- 8 facets (3 goal facets and 5 intermediate facets)
- Facets relationship results in 36 permissible patterns



- Sample size (500 vs 2,000)
- Learning and forgetting rate (LR = 0.8 and FR = 0.15 vs LR = 0.6 and FR = 0.2)

Simulation Study: Design



Simulation Study: Results

Table 1

Estimation results of $\boldsymbol{\tau}$.

Method	Size	Learning	Sum of abs. bias	Recovery error of sparsity	Recovery of true zeros	Recovery of true non-zeros
Algorithm 1	500	High	62.010	0.010	0.999	0.542
		Low	58.575	0.010	0.999	0.543
	2000	High	60.948	0.005	0.997	0.864
		Low	52.904	0.005	0.997	0.884
Algorithm 2	500	High	64.181	0.036	0.968	0.803
		Low	60.772	0.041	0.960	0.923
	2000	High	62.261	0.068	0.932	0.977
		Low	55.768	0.082	0.916	1.000

Discussion

- The proposed REM algorithms can effectively shrink proper rows and columns of the transition matrix to 0
- Provide a valuable psychometric model that can infer how students transit across different facet profiles

- DFSM item parameters are assumed to be known
- A uniform distribution over the permissible facet profiles is assumed
- Exploring different hierarchical structures, and thus different sparsity structures

Thanks!
heren@uw.edu

He Ren and Chun Wang
College of Education, University of Washington

This study is supported by **NSF #2300382**

UNIVERSITY *of* WASHINGTON | COLLEGE OF EDUCATION

Personal Homepage

U.S. National Science Foundation

DENVER 2025 ANNUAL MEETING ncme