Learning or Forgetting? A Dynamic Approach for Tracking the Knowledge Proficiency of Students

ZHENYA HUANG, QI LIU, and YUYING CHEN, University of Science and Technology of China, China
LE WU, Hefei University of Technology, China and iFLYTEK Co., Ltd, China
KELI XIAO, Stony Brook University, USA
ENHONG CHEN, University of Science and Technology of China, China
HAIPING MA, Anhui University, China
GUOPING HU, iFLYTEK Research, China

Reporter: Zhenya Huang

Outline

1. Background
2. Problem & Overview
3. Model
4. Experiment
5. Conclusion & Future work
Background

- Cognitive diagnosis for knowledge proficiency
  - Domain: **Education**, Recruitment, Sports, Game, etc
  - Goal: Evaluating how much students learn about different knowledge concepts
    - Math subject: Function, Set, Inequality, etc
  - Fundamental task
    - Evaluation, Testing, Recommendation, etc
Background

- **Learning activities**
  - Taking courses, Practicing exercises, Taking Tests, etc
  - Classroom-based
    - Rely on expertise of teachers
    - Hard to record data
  - Online learning
    - Open environment with computer-aided technology
    - Learning data of students can be recorded
    - KhanAcademy, MOOC, etc
Background

➢ Cognitive diagnosis Problem

<table>
<thead>
<tr>
<th>March</th>
<th>April</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>(e_1)</td>
<td>(e_5)</td>
<td>(e_9)</td>
</tr>
<tr>
<td>(e_2)</td>
<td>(e_6)</td>
<td>(e_{10})</td>
</tr>
<tr>
<td>(e_3)</td>
<td>(e_7)</td>
<td>(e_{11})</td>
</tr>
<tr>
<td>(e_4)</td>
<td>(e_8)</td>
<td>(e_{12})</td>
</tr>
</tbody>
</table>

\(u_1, u_2\): Student \(e_1 \sim e_{12}\): Exercise \(k_1, k_2\): Knowledge concept \(\checkmark\): Right answer \(\times\): Wrong answer

<table>
<thead>
<tr>
<th>Q-matrix</th>
<th>(k_1)(Function)</th>
<th>(k_2)(Inequality)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(e_1)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(e_2)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(e_3)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(e_4)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(e_5)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(e_6)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(e_7)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(e_8)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(e_9)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(e_{10})</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(e_{11})</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(e_{12})</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Related work

- **Static modeling**
  - IRT: Item Response Theory

  \[
P(X_{ij} = 1 | \theta_j) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta_j - b_i)]}
  \]

- **DINA:**

  \[
P_j(\alpha_i) = P(X_{ij} = 1 | \alpha_i) = g_j^{1 - a_{ij}}(1 - s_j)^a_{ij}.
  \]

- **PMF: Probabilistic Matrix Factorization**

  \[
p(R|U, V, \sigma^2) = \prod_{i=1}^{N} \prod_{j=1}^{M} \left[ \mathcal{N}(R_{ij}|U_i^TV_j, \sigma^2) \right]^{I_{ij}}
  \]
Related work

- Dynamic modeling
  - BKT: Bayesian Knowledge Tracing
  - Hidden Markov Model
  - Tracing for single concept
  - Discrete results (mastered or non-mastered)

![Dynamic Modeling Diagram]

- Student’s knowledge state
- Probability transition matrix

\[ P(L_0) \xrightarrow{P(T)} P(L_{t-1}) \xrightarrow{P(T)} P(L_t) \xrightarrow{P(T)} P(L_{t+1}) \]

\[ P(G), P(S) \]

\[ X_{t-1}, X_t, X_{t+1} \]

Observation, 0 for wrong, 1 for correct
Related work

- Dynamic modeling
  - DKT: Deep Knowledge Tracing
    - Apply RNNs (LSTM) to model student knowledge over time
    - Tracing all concepts together
    - Hidden states can represent the latent knowledge states
Background

- Limitation
  - Ignoring the dynamic memory factors
    - How can we learn and remember knowledge?
    - Why do we forget what we have learned?
  - Lack of interpretability
    - Don’t know the meaning of latent vectors/hidden states
  - Learning records are sparse
    - Students practice very few exercises
### Problem & Overview

**Given**
- Exercising logs as a score tensor: \( R \in \mathbb{R}^{N \times M \times T} \)
- Q-matrix representing exercise-knowledge relation: \( Q \in \mathbb{R}^{M \times K} \)

**Goal**
- Tracking the change of knowledge proficiency of students from time 1 to T
- Predicting her proficiency on K concepts and performance scores on specific exercises at time T + 1

#### (a) Exercising log example

<table>
<thead>
<tr>
<th>Student</th>
<th>Exercise</th>
<th>Time</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u_1 )</td>
<td>( e_1 )</td>
<td>( t_1 )</td>
<td>0</td>
</tr>
<tr>
<td>( u_1 )</td>
<td>( e_5 )</td>
<td>( t_2 )</td>
<td>0.25</td>
</tr>
<tr>
<td>( u_2 )</td>
<td>( e_2 )</td>
<td>( t_1 )</td>
<td>0</td>
</tr>
<tr>
<td>( u_2 )</td>
<td>( e_3 )</td>
<td>( t_3 )</td>
<td>1</td>
</tr>
<tr>
<td>( u_2 )</td>
<td>( e_1 )</td>
<td>( t_3 )</td>
<td>0.75</td>
</tr>
<tr>
<td>( u_3 )</td>
<td>( e_4 )</td>
<td>( t_4 )</td>
<td>1</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
</tbody>
</table>

#### (b) Q-matrix example

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Knowledge concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( k_1 )</td>
</tr>
<tr>
<td>( e_1 )</td>
<td>1</td>
</tr>
<tr>
<td>( e_2 )</td>
<td>0</td>
</tr>
<tr>
<td>( e_3 )</td>
<td>0</td>
</tr>
<tr>
<td>( e_4 )</td>
<td>0</td>
</tr>
<tr>
<td>( e_5 )</td>
<td>1</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
</tbody>
</table>
Problem & Overview

- **Model overview**
  - KPT: Knowledge Proficiency Tracing model
  - EKPT: Exercise-correlated Knowledge Proficiency model

![Diagram showing model overview]

- **Input**
  - Exercising logs
    - $u_1$, $u_2$, ..., $u_T$

- **Educational factors**
  - Student-related
    - Forgetting
    - Learning
  - Exercise-related
    - Knowledge order
    - Connectivity

- **Modeling**
  - KPT: Knowledge Proficiency Tracing
  - EKPT: Exercise-correlated Knowledge Proficiency Tracing

- **Output**
  - Student proficiency vectors
    - $e_1$, $e_2$, ..., $e_T$
    - Knowledge order
    - Connectivity
  - Exercise knowledge vectors
    - Connectivity

- **Applications**
  - Knowledge estimation
    - $k_1$, $k_2$, $k_3$
    - $0.8$, $0.6$, $0.2$
    - $0.5$, $0.3$, $0.1$
  - Score prediction
    - $e_1$, $e_2$, ..., $e_T$
    - Connectivity
  - Visualization
Outline

1. Background
2. Problem & Overview
3. Model
4. Experiment
5. Conclusion & Future work
KPT model

- Probabilistic modeling
  - For each student and exercise, modeling the responses as:

  \[ p(R|U, V, b) = \prod_{t=1}^{T} \prod_{i=1}^{N} \prod_{j=1}^{M} \mathcal{N}(R_{ij}^t|\langle U_i^t, V_j \rangle - b_j, \sigma_R^2)^{I_{ij}}, \]

  - \( U_i^t \in \mathbb{R}^{K \times 1} \): proficiency vector of student i, representing how much students learn on K concepts at time t
  - \( V_j \in \mathbb{R}^{K \times 1} \): knowledge vector of exercise j, denoting the latent correlation between exercise j and K concepts

- How to establish the corresponding relationship among students, exercises and knowledge concepts?
KPT model

- Modeling V with Q-matrix prior
  - Goal: project exercise into knowledge space, enhancing interpretability
  - Traditional Q-matrix
    - Denoting exercise-knowledge correlation
    - Binary entries: do not fit for probabilistic modeling
  - Our work assumption
    - If $Q_{jq} = 1$, then this concept $q$ is more relevant to exercise $j$ than all other concepts with mark 0

\[
\forall p, q \in K, p \neq q, \text{if } Q_{jq} = 1 \text{ and } Q_{jp} = 0 \Rightarrow q \succ j p,
\]
\[
\forall p, q \in K, p \neq q, \text{if } Q_{jq} = 1 \text{ and } Q_{jp} = 1 \Rightarrow q \prec j p,
\]
\[
\forall p, q \in K, p \neq q, \text{if } Q_{jq} = 0 \text{ and } Q_{jp} = 0 \Rightarrow q \preceq j p.
\]

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Knowledge concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k_1$</td>
</tr>
<tr>
<td>$e_1$</td>
<td>1</td>
</tr>
<tr>
<td>$e_2$</td>
<td>0</td>
</tr>
<tr>
<td>$e_3$</td>
<td>0</td>
</tr>
<tr>
<td>$e_4$</td>
<td>0</td>
</tr>
<tr>
<td>$e_5$</td>
<td>1</td>
</tr>
</tbody>
</table>

\{ e_1, k_1, k_2 \}
\{ e_1, k_1, k_3 \}
\{ e_2, k_3, k_1 \}
\{ e_5, k_1, k_2 \}
\{ e_5, k_1, k_4 \}
\ldots
KPT model

- Modeling U with learning theories
  - Goal: explain the dynamic factors in the learning process

\[
p(U_i^t) = \mathcal{N}(U_i^t | \bar{U}_i^t, \sigma_U^2 I), \text{ where } \bar{U}_i^t = \{\bar{U}_{i1}^t, \bar{U}_{i2}^t, \ldots, \bar{U}_{iK}^t\},
\]

\[
\bar{U}_{ik}^t = \alpha_i L_{ik}^t(\ast) + (1 - \alpha_i) F_{ik}^t(\ast), \text{ s.t. } 0 \leq \alpha_i \leq 1,
\]

- Two learning theories
  - **Learning curve**: The more exercises she does, the higher level of proficiency on the related knowledge she will get

\[
L_{ik}^t(\ast) = U_{ik}^{t-1} \frac{D f_{ik}^t}{f_{ik}^t + r},
\]

  - Number of practice times

  - **Forgetting curve**: The longer the time passes, the more knowledge she will forget

\[
F_{ik}^t(\ast) = U_{ik}^{t-1} e^{-t/s},
\]

  - Time interval
EKPT model

- Sparsity problem
  - Students practice very few exercises compared with the huge exercise space
  - Inaccurate if students just practices few exercises at each time

- EKPT model
  - Exercise connectivity assumption
    - Students may get consistent scores on these knowledge-based exercises
    - Learning each exercise vector with its similar ones

![Diagram of EKPT model with learning process and Q-matrix]

<table>
<thead>
<tr>
<th>Learning Process</th>
<th>March</th>
<th>April</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u₁, u₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>e₁, e₂, e₃, e₄</td>
<td>e₅, e₆, e₇, e₈</td>
<td>e₉, e₁₀, e₁₁, e₁₂</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q-matrix</th>
<th>k₁(Function)</th>
<th>k₂(Inequality)</th>
</tr>
</thead>
<tbody>
<tr>
<td>e₁</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>e₂</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>e₃</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>e₄</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>e₅</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>e₆</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>e₇</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>e₈</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>e₉</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>e₁₀</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>e₁₁</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>e₁₂</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

u₁, u₂: Student  e₁~e₁₂: Exercise  k₁, k₂: Knowledge concept  ✔: Right answer  ✗: Wrong answer
EKPT model

- Modeling V with exercise connectivity
  - For exercise j, we define a neighbor set
    \[ N_{V_j} = \{ l | k \in j \cap l, l \in \hat{V}, k \in K \} \]
  - The knowledge vector of exercise j is influenced by the set:
    \[ V_j = \sum_{l \in N_{V_j}} w(j, l) \times V_l + \theta_V, \theta_V \sim \mathcal{N}(0, \sigma^2_V). \]
  - \( w(j, l) \) is the weight influence, which can be any weight function, like
    \[ V_j = \frac{1}{|N_{V_j}|} \sum_{l \in N_{V_j}} V_l + \theta_V, \theta_V \sim \mathcal{N}(0, \sigma^2_V). \]

Equal contribution for all neighbor exercises
Model Comparasion

KPT

\[ \lambda_p > \lambda_V \]

\[ b_j \]

\[ \lambda_{U1}, \lambda_U \]

EKPT

\[ \lambda_p > \lambda_3 \]

\[ b_j \]

\[ \lambda_{U1}, \lambda_U \]
Model Learning

KPT

\[
\min_{\Phi} \mathcal{E}(\Phi) = \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{M} I_t \left( \hat{R}_{ij} - R_{ij} \right)^2 \\
- \lambda_p \sum_{j=1}^{K} \sum_{p=1}^{K} I \left( q > p \right) \ln \left( \frac{1}{1 + e^{-(V_{jq} - V_{jp})}} \right) + \frac{\lambda_V}{2} \sum_{j=1}^{M} \|V_j\|_F^2 \\
+ \frac{\lambda_U}{2} \sum_{t=2}^{T} \sum_{i=1}^{N} \|U_t^j - U_{t-1}^j\|_F^2 + \frac{\lambda_{U(t)}}{2} \sum_{i=1}^{N} \|U_i^j\|_F^2.
\]

The algorithm for parameter learning of the KPT model:

**ALGORITHM 1:** Parameter Learning of the KPT Model

Initialize $U$, $V$, $\alpha$ and $b$;

while not converged do

for $i = 1, 2, \ldots, N$ do

for $t = 1, 2, \ldots, T$ do

for $k = 1, 2, \ldots, K$ do

Fix $V$, $\alpha$, $b$, update $U_{ik}$ by Equation (15) using SGD;

end for

Fix $U$, $V$, $\alpha$, $b$, update $\alpha_i$ by Equation (17) and Equation (19) using PG;

end for

end for

Return $U$, $V$, $\alpha$ and $b$;

end while

EKPT

\[
\min_{\Phi} \mathcal{E}(\Phi) = \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{M} I_t \left( \hat{R}_{ij} - R_{ij} \right)^2 \\
- \lambda_p \sum_{j=1}^{K} \sum_{p=1}^{K} I \left( q > p \right) \ln \left( \frac{1}{1 + e^{-(V_{jq} - V_{jp})}} \right) + \frac{\lambda_S}{2} \sum_{j=1}^{M} \|V_j\|_F^2 \\
+ \frac{\lambda_{U(t)}}{2} \sum_{i=1}^{N} \|U_t^j - U_{t-1}^j\|_F^2 + \frac{\lambda_{U(t)}}{2} \sum_{i=1}^{N} \|U_i^j\|_F^2.
\]

The algorithm for parameter learning of the EKPT model:

**ALGORITHM 2:** Parameter Learning of the EKPT Model

Initialize $U$, $V$, $\alpha$ and $b$;

while not converged do

for $i = 1, 2, \ldots, N$ do

for $t = 1, 2, \ldots, T$ do

for $k = 1, 2, \ldots, K$ do

Fix $V$, $\alpha$, $b$, update $U_{ik}$ by Equation (15) using SGD;

end for

Fix $U$, $V$, $\alpha$, $b$, update $\alpha_i$ by Equation (17) and Equation (19) using PG;

end for

end for

for $j = 1, 2, \ldots, M$ do

for $k = 1, 2, \ldots, K$ do

Fix $U$, $\alpha$, $b$, update $V_{jk}$ by Equation (25) using SGD;

end for

Fix $U$, $V$, $\alpha$, update $b$ by Equation (18) using SGD;

end for

Return $U$, $V$, $\alpha$ and $b$;

end while
Model

- **Application**
  - Knowledge Proficiency Estimation
    \[
    \hat{U}_i^{(T+1)} = \{\hat{U}_{i1}^{(T+1)}, \hat{U}_{i2}^{(T+1)}, \ldots, \hat{U}_{iK}^{(T+1)}\},
    \]
    \[
    \hat{U}_{ik}^{(T+1)} \approx \alpha_i U_{ik}^T \frac{Df_{ik}^{T+1}}{f_{ik}^{T+1} + r} + (1 - \alpha_i) U_{ik}^T e^{-\frac{\Delta(T+1)}{S}},
    \]

- **Student Performance Prediction**
  \[
  \hat{R}_{ij}^{(T+1)} \approx \langle U_i^{(T+1)}, V_j \rangle - b_j.
  \]
  \[
  \hat{R}_{ij}^{(T+1)} = \begin{cases} 
  \hat{R}_{ij}^{(T+1)} & \text{if } 0 \leq \hat{R}_{ij}^{(T+1)} \leq 1, \\
  0 & \text{if } \hat{R}_{ij}^{(T+1)} < 0, \\
  1 & \text{if } \hat{R}_{ij}^{(T+1)} > 1.
  \end{cases}
  \]

- **Diagnosis results explanation and visualization**
Outline

1. Background
2. Problem & Overview
3. Model
4. Experiment
5. Conclusion & Future work
## Experiment

### Dataset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Math1</th>
<th>Math2</th>
<th>Assist</th>
<th>Adaptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training logs</td>
<td>521,248</td>
<td>347,424</td>
<td>263,327</td>
<td>229,848</td>
</tr>
<tr>
<td>Testing logs</td>
<td>74,464</td>
<td>18,312</td>
<td>43,888</td>
<td>38,308</td>
</tr>
<tr>
<td># of students</td>
<td>9,308</td>
<td>1,306</td>
<td>7197</td>
<td>3,217</td>
</tr>
<tr>
<td># of exercises</td>
<td>64</td>
<td>280</td>
<td>3211</td>
<td>411</td>
</tr>
<tr>
<td># of time windows</td>
<td>4</td>
<td>10</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td># of knowledge concepts</td>
<td>12</td>
<td>13</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Avg. knowledge concepts per exercise</td>
<td>1.15</td>
<td>1.3215</td>
<td>1.5073</td>
<td>1.06</td>
</tr>
</tbody>
</table>

### Baseline

<table>
<thead>
<tr>
<th>Model</th>
<th>Data Source</th>
<th>Application</th>
<th>Dynamic Explanation?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q-matrix</td>
<td>Multi-Skill</td>
<td>Repeating</td>
</tr>
<tr>
<td>IRT [17]</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>DINA [15]</td>
<td>√</td>
<td>√</td>
<td>×</td>
</tr>
<tr>
<td>PMF [63]</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>BKT [31]</td>
<td>√</td>
<td>×</td>
<td>√</td>
</tr>
<tr>
<td>LFA [9]</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>DKT [52]</td>
<td>×</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>QMIRT</td>
<td>√</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>QPMF</td>
<td>√</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>KPT</td>
<td>√</td>
<td>√</td>
<td>×</td>
</tr>
<tr>
<td>EKPT</td>
<td>√</td>
<td>√</td>
<td>×</td>
</tr>
</tbody>
</table>
Knowledge Proficiency Estimation

**DOA**: if $a$ masters better than $b$ on a concept $k$ at time $T$, then $a$ will have a higher probability to get correct answers to the exercises related to concept $k$ than $b$ at time $T$.

$$DOA(k) = \sum_{j=1}^{M} \sum_{a=1}^{N} \sum_{b=1}^{N} \delta \left( U_{ak}^{T+1}, U_{bk}^{T+1} \right) \cap \delta \left( R_{aj}^{T+1}, R_{bj}^{T+1} \right),$$

Our models perform better than baselines

EKPT is better than KPT on sparse dataset
Experiment

- Student Performance Prediction
  - MAE, RMSE

- Dynamic models are better than static ones
- Deep learning based models (DKT) perform not very good
  - Possible: Time is not longer enough, Data volume may not support

- Diagnosis results visualization

- The student practices many times on K3, knowledge proficiency increases
- The student practices very few exercises on K4, she may forget what she have learned
Experiment

- Model Analysis
- Computational Performance
  - Though our model needs more time for training, they are competitive compared with DKT (deep learning based ones)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Time</th>
<th>Stastic Models</th>
<th>Dynamic Models</th>
<th>Variants</th>
<th>Our Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IRT</td>
<td>DINA</td>
<td>PMF</td>
<td>BKT</td>
</tr>
<tr>
<td>Math1</td>
<td>Each</td>
<td>0.022</td>
<td>0.316</td>
<td>0.023</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1.960</td>
<td>18.05</td>
<td>1.833</td>
<td>1.516</td>
</tr>
<tr>
<td>Math2</td>
<td>Each</td>
<td>0.011</td>
<td>0.616</td>
<td>0.021</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1.051</td>
<td>57.28</td>
<td>1.283</td>
<td>0.581</td>
</tr>
<tr>
<td>Assist</td>
<td>Each</td>
<td>0.015</td>
<td>/</td>
<td>0.033</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.320</td>
<td>/</td>
<td>4.951</td>
<td>1.275</td>
</tr>
<tr>
<td>Adaptive</td>
<td>Each</td>
<td>0.013</td>
<td>/</td>
<td>0.029</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2.154</td>
<td>/</td>
<td>3.466</td>
<td>1.017</td>
</tr>
</tbody>
</table>

- Parameter sensitivity

![Graphs showing the impact of λ_U on four datasets.](image-url)
Experiment

- Model Analysis
- Exercise relationship
- Exercise with same concepts are grouped together

(a) Math1

(b) Math2

(c) Assist

(d) Adaptvie
<table>
<thead>
<tr>
<th>1</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Problem &amp; Overview</td>
</tr>
<tr>
<td>3</td>
<td>Model</td>
</tr>
<tr>
<td>4</td>
<td>Experiment</td>
</tr>
<tr>
<td>5</td>
<td>Conclusion &amp; Future work</td>
</tr>
</tbody>
</table>
Conclusion & Future work

**Conclusion**

- A focused study on tracking the knowledge proficiency of students
- Two explanatory probabilistic models considering different educational factors
  - Incorporating learning theories for explaining the knowledge change
  - Incorporating Q-matrix for improving the interpretability
  - Incorporating exercise connectivity property to address sparsity problem
- Experiments on different datasets show the both effectiveness and explanatory power of our models

**Future work**

- Consider different specific modeling for learning and forgetting factors
- Consider student behaviors and social connections for more precise diagnosis
- Consider different learning scenarios
  - Game
  - Multiple-attempt response
  - Repeated learning
Thanks for your listening!

huangzhy@mail.ustc.edu.cn