Adaptive Learning with Cognitive Contexts Modeling

(基于认知情境建模的自适应学习方法)

Qi Liu (刘 淇)
qiliuql@ustc.edu.cn
University of Science and Technology of China
http://staff.ustc.edu.cn/~qiliuql/
# Outline

1. Background of Adaptive Learning
2. Cognitive Diagnosis Methods
3. Adaptive Testing Frameworks
5. Discussion and Conclusion
• Adaptive is important for “learning”

➢ **Adaptive learning** is an educational method which uses computer algorithms to orchestrate the interaction with the learner and deliver customized resources and learning activities to address the unique needs of each learner [1].

Some representative “adaptive cases” in machine learning

- Adaptive Sampling [AdaBoost’ EuroCOLT’95]
- Re-weighted cost [BAN’ICML’18]
- Adaptive learning rate [Adam, ICLR’15]
- Parameter initialization [MAML, ICML’17]
- Dropout [AE, NIPS’13]
- Model layer selection [AdaShare, ArXiv’19]

Background

Data Selection
- Some representative "adaptive cases" in machine learning

Parameter Training

Architecture Optimization

- Adaptive Sampling [AdaBoost’ EuroCOLT’95]
- Re-weighted cost [BAN’ICML’18]
- Adaptive learning rate [Adam, ICLR’15]
- Parameter initialization [MAML, ICML’17]
- Dropout [AE, NIPS’13]
- Model layer selection [AdaShare, ArXiv’19]
Background

- Adaptive in human learning

- “Adapts” based on the responses of the individual student, which dynamically adjusts the level or types of instruction based on individual student abilities or preferences. [Wikipedia]
Background

**Experienced teacher**
- Evidence: Limited learning records
- Evaluation: via score
- Solution: Offline teaching

**Adaptive Learning Algorithms**
- Large number of learning records
- Evaluation: via Cognitive diagnosis
- Solution: Adaptive teaching
How to **accurately** evaluate the cognitive level of each student?

- **Learning records**
- **Traditional Model**
- **Embedding**
- **Cognitive modeling**
- **Cognitive structure**
- **Prediction**

![Diagram showing the evaluation process](image)
How to make fast & comprehensive testing to evaluate the cognitive level of student?

Testing machines → Massive training data

Testing human

Traditional testing

Adaptive testing + Cognitive modeling → fast & comprehensive

Cognitive structure
Given the cognitive level of each student, how to **adaptively recommend** personalized educational resources?

- Implicit context
- Explicit cognitive context
- Cognitive context-aware
- User interest-aware

**Adaptive in machine learning**

**Adaptive in human learning**

**Traditional personalized recommendation**
**Background**

**Challenges**

- How to **accurately** evaluate the cognitive level of each student?
- How to make **fast & comprehensive** testing to evaluate the cognitive level of student?
- How to **adaptively recommend** personalized educational resources?

**Solutions**

- Cognitive Diagnosis
- Adaptive Testing
- Cognitive Context-aware Recommendation

**Response matrix R**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

**Q-matrix**

<table>
<thead>
<tr>
<th>一次函数 二次函数 时代性观念</th>
</tr>
</thead>
<tbody>
<tr>
<td>试题1</td>
</tr>
<tr>
<td>试题2</td>
</tr>
<tr>
<td>试题3</td>
</tr>
<tr>
<td>试题4</td>
</tr>
</tbody>
</table>

**Exercise Text**

We must keep in mind that the population change of the environment (such as the climate, social, and cultural environment) is a very important factor that affects the development of scientific and technological innovation. In this regard, we need to understand the development of the scientific and technological environment.
Outline

1. Background of Adaptive Learning
2. Cognitive Diagnosis Methods
3. Adaptive Testing Frameworks
5. Discussion and Conclusion
Cognitive Diagnosis

- Cognitive diagnosis (认知能力诊断)
  - Definition: diagnose the proficiency of participants on specific skills/concepts
  - Necessary: adaptive learning is based on cognitive diagnosis
  - Real-world scenarios: games, sports, recruitment, education, etc.
Cognitive Diagnosis

- **Problem Definition** of Cognitive diagnosis

- **Given** (Input):
  - Student-Exercise matrix $R$
    - $R_{ij}$ denotes the response score of student $i$ on exercise $j$
  - Exercise-Concept Q-matrix
    - If exercise $j$ relates to concept $k$, $Q_{jk} = 1$

- **Goal** (Output):
  - Student’s proficiency on each concept/skill
    - E.g. in the range of [0,1]
Cognitive Diagnosis

- Technological development

  - early 20th century~1960s:
    - Ability level paradigm
    - Classical Measurement Theory (CTT)
    - Item Response Theory (IRT)
    - Describe the relationship among true score, errors and observation score

  - 1960s~2010s:
    - Cognitive level paradigm
    - Multidimensional IRT (MIRT)
    - Describe the relationship between latent trait and item characteristics

  - Now~:
    - Deep cognitive diagnosis paradigm
    - Deep learning-based cognitive diagnosis
    - Describe complex relationships

References:
- Harold Gulliksen. 1950. Theory of mental tests
Cognitive Diagnosis

- Item Response Theory (IRT)
  \[
  P(R_{uv} = 1 | \theta_u, a_v, b_v, c_v) = c_v + \frac{1 - c_v}{1 + \exp(-1.7a_v(\theta_u - b_v))}
  \]

- DINA
  \[
  P(Y_{ij} = 1 | \alpha_i) = (1 - S_j)^{\eta_{ij}} (g_j)^{1 - \eta_{ij}}
  \]

References:
Cognitive Diagnosis

- **Summary of existing methods**
  - **Problems** in the interaction functions of traditional methods:
    - manually designed → labor intensive
    - mostly linear function → limited approximation ability
    - just numerical data → cannot mine heterogeneous big data

- **Deep cognitive diagnosis paradigm**
  - Learn interaction function automatically from data with deep learning
    - manually designed, limited ability → automatically learned, high ability
Neural Cognitive Diagnosis

Model Design:

- **Student Factors**: knowledge proficiency vector $F^s$
- **Exercise Factors**: knowledge relevancy vector $F^{kn}$
  - other exercise factors $F^{other}$ (optional): e.g., difficulty, discrimination
- **Interaction Function**: interactive multi-layers
- **Output**: The probability that the student would correctly answer the exercise

Fei Wang, Qi Liu, Enhong Chen et al. Neural Cognitive Diagnosis for Intelligent Education Systems. *AAAI2020, Accepted.*
Neural Cognitive Diagnosis

- **Explainable**
  - Explainably model student knowledge states
  - $F^s \circ F^{kn}$: attach each entry of $F^s$ to a specific knowledge concept

- **Monotonicity Assumption**: The probability of correct response to the exercise is monotonically increasing at any dimension of the student’s knowledge proficiency. (widely applicable)
**NeuralCDM:** One basic implementation of NeuralCD with Q-matrix

- **Training:**
  - **Input Layer:**
    - Directly from Q-matrix
  - **Interaction Layer:**
    - Full connection
    - Positive weights
    - Monotonicity Assumption
  - Training with cross entropy loss

**Diagram:**

- **Input Layer:**
  - $Q_e \cdot (h^s - h^{diff}) \times h^{disc}$
  - $F^{kn} \cdot F^s$

- **Output Layer:**
  - $y$

- **Interaction Layer:**
  - Positive Full Connection
  - Positive Full Connection
  - Positive Full Connection

- **Monotonicity Assumption:**
  - Knowledge Relevancy
  - Knowledge Proficiency
  - Knowledge Difficulty
  - Exercise Discrimination
Neural Cognitive Diagnosis

**General:**
- NeuralCD framework is general and can cover some traditional models
- e.g., MF, IRT, MIRT

![Diagram of NeuralCDM and IRT models](image)
Neural Cognitive Diagnosis

- **Extendible**
  - refine Q-matrix with exercise texts
  - pre-train a CNN to predict knowledge concepts of the input exercise
  - combine with Q-matrix through a partial order probabilistic scheme:

  knowledge relevancy: $Q$-matrix $\geq$ predicted $>$ other = 0

---

NeuralCDM+
Neural Cognitive Diagnosis

**Experiment**

**Datasets**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Math</th>
<th>ASSIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Students</td>
<td>10,268</td>
<td>4,163</td>
</tr>
<tr>
<td>#Exercises</td>
<td>917,495</td>
<td>17,746</td>
</tr>
<tr>
<td>#Knowledge concepts</td>
<td>1,488</td>
<td>123</td>
</tr>
<tr>
<td>#Response logs</td>
<td>864,722</td>
<td>324,572</td>
</tr>
<tr>
<td>#Knowledge concepts per exercise</td>
<td>1.53</td>
<td>1.19</td>
</tr>
<tr>
<td>AVG #log</td>
<td>2.28</td>
<td>8.05</td>
</tr>
<tr>
<td>STD #log&gt;1</td>
<td>0.305</td>
<td>0.316</td>
</tr>
</tbody>
</table>

**Student performance prediction**

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>RMSE</th>
<th>AUC</th>
<th>Accuracy</th>
<th>RMSE</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DINA</td>
<td>0.593±.001</td>
<td>0.487±.001</td>
<td>0.686±.001</td>
<td>0.650±.001</td>
<td>0.467±.001</td>
<td>0.676±.002</td>
</tr>
<tr>
<td>IRT</td>
<td>0.782±.002</td>
<td>0.387±.001</td>
<td>0.795±.001</td>
<td>0.674±.002</td>
<td>0.464±.002</td>
<td>0.685±.001</td>
</tr>
<tr>
<td>MIRT</td>
<td>0.793±.001</td>
<td>0.378±.002</td>
<td>0.813±.002</td>
<td>0.701±.002</td>
<td>0.461±.001</td>
<td>0.719±.001</td>
</tr>
<tr>
<td>PME</td>
<td>0.763±.001</td>
<td>0.407±.001</td>
<td>0.792±.002</td>
<td>0.661±.002</td>
<td>0.476±.001</td>
<td>0.732±.001</td>
</tr>
<tr>
<td>NeuralCDM</td>
<td>0.792±.002</td>
<td>0.378±.001</td>
<td>0.820±.001</td>
<td>0.719±.008</td>
<td>0.439±.002</td>
<td>0.749±.001</td>
</tr>
<tr>
<td>NeuralCDM+</td>
<td>0.804±.001</td>
<td>0.371±.002</td>
<td>0.835±.002</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- **Math**: private, mathematical exercises *(with texts)* and logs
- **ASSIST**: public, mathematical exercises *(without texts)* and logs
Adaptive Diagnosis

- a student’s performance on 3 exercises in ASSIST
- and his/her diagnosed result

Q-Matrix

<table>
<thead>
<tr>
<th>Exercise 1</th>
<th>Exercise 2</th>
<th>Exercise 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Line</td>
<td>Solving Inequalities</td>
<td>Add Whole Numbers</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Student Response

- Exercise 1: Cross
- Exercise 2: Checkmark
- Exercise 3: Checkmark

The student is more likely to answer correctly when his/her knowledge proficiency satisfies the requirement of the exercise.

Code for NeuralCDM is available at https://github.com/bigdata-ustc/NeuralCD
Outline

1. Background of Adaptive Learning
2. Cognitive Diagnosis Methods
3. Adaptive Testing Frameworks
5. Discussion and Conclusion
What is Adaptive Testing

- **Adaptive Testing**: a promising form of computer-administrated examination
  - different from traditional “Paper and Pencil testing”
  - interactively select questions for students, based on his/her current performance

![Diagram of Adaptive Testing]
Application of Adaptive Testing

- More and more adaptive testing applications
  - Improve the experience of both examiners and examinees
- Not only in educational exams
  - Will be applied more in the future

- GRE
- GMAT
- TOEFL

- Knewton iFlyTek
- … …

- PROMs VR-12
- … …

- HSK
- ISPNN
- … …

Educational Examination
Online Education
Sports & Health
Qualification Examination
Formalization of Adaptive Testing

- **Adaptive Testing Framework**
  - **Cognitive Diagnosis Model (CDM)**
    - **Input**: examinee records
    - **Output**: diagnosis for examinee ability
    - **Example**: IRT, NeuralCD
  - **Selection Strategy**
    - **Input**: examinee ability
    - **Output**: the next question

- **Adaptive Testing Procedure**
  - **Input**: new examinee $e_1$
  - **Procedure**: step $t \rightarrow$ select $\rightarrow$ answer $\rightarrow$ diagnose $\rightarrow$ step $t + 1$
  - **Output**: diagnosis for $e_1$
Related Work in Adaptive Testing

- **Methodology:**
  - Heuristic: $h(q) = \text{informativeness of question}$
  - Greedy: $q_t^* = \text{argmax}_{q \in Q} h(q)$

- **By informativeness we mean**
  - Generally: information a question offers while answered by the examinee
  - Specifically: accuracy of parameter estimation of the CDM

- **Example: a simple IRT-based strategy**
  - Intuition: matching difficulty with ability
  - Procedure: (bOpt)

$$p(a, b | \theta) = \frac{1}{1 + e^{-a(\theta - b)}}$$
Following the same idea, there are a number of adaptive testing strategies:

- Categorized by their underlying CDMs.
Challenge in Adaptive Testing

- **Model-specific** methodology leads to inflexible framework
  - IRT model → **IRT-based** strategy → only dedicated to IRT
  - MIRT → **redesign MIRT-based** strategy → only dedicated to MIRT
  - NeuralCDM → **hard** to redesign → **no** suitable strategy currently

- We need a **model-agnostic** framework
Inspiration from Machine Learning

- **Adaptive Testing**: from a data scientist’s perspective
  - Selects valuable data samples for models
- **Active Learning**: Inspire us with a model-agnostic solution
  - Applies data querying strategies to a wide range of models and tasks
  - ① Abstract the underlying models without specific assumptions
  - ② Uplift the objectives to a high level

<table>
<thead>
<tr>
<th>Active Learning Concepts</th>
<th>Adaptive Learning Counterparts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeled dataset</td>
<td>Tested question set</td>
</tr>
<tr>
<td>Unlabeled dataset</td>
<td>Untested question set</td>
</tr>
<tr>
<td>Learning model</td>
<td>Diagnosis model</td>
</tr>
<tr>
<td>Active query selection</td>
<td>Question selection</td>
</tr>
<tr>
<td>Person annotator</td>
<td>Student</td>
</tr>
</tbody>
</table>
How to Achieve Model-Agnostic

- **CDM abstraction** from specific models
  - Ability parameters *without* interpretation assumption

\[
\begin{align*}
\text{IRT} & \quad p(\theta) = \frac{1}{1 + e^{-a(\theta - b)}} \\
\text{MIRT} & \quad p(\hat{\theta}) = \frac{1}{1 + e^{-\hat{a}r(\hat{\theta} - \hat{b})}} \quad \Rightarrow \quad M(\theta)
\end{align*}
\]

- **Optimization objective** at a high level
  - Original objective: make parameters estimated accurately
  - From an educational perspective, is that **enough**?
Quality and Diversity: Proper Objectives

What should appropriate questions look like?

- **High-quality**: contains enough information
  - Understand examinees **accurately**
  - The more information, the more uncertainty in the change of model parameters
- **Diverse**: covers enough knowledge points
  - Understand examinees **comprehensively**

Nothing useful if the student will **definitely** answer the question right/wrong

Biased results if the knowledge coverage is **imbalanced**
Our Work: Framework Overview

- **Model-Agnostic Adaptive Testing (MAAT)**
  - Work with an abstract model to achieve model-agnostic
  - Aim at the two high-level objective in a two-stage method

- **Quality Module**
  - Expected Model Change

- **Diversity Module**
  - Importance Weighted Knowledge Coverage

- **Importance Module**
  - Test-Effect Embedding

- **Step t**
  - Abstract CDM with parameters

- **$Q_U$**
  - Untested Question Set

- **$Q_C$**
  - Top $K_C$ Candidate Question Set

- **$Q_T$**
  - Tested Question Set

- **$q_i$**
  - Selected Question
Quality Module: Select High-quality Questions

- **Quantify** the quality of a question
  - How we measure the information the abstract CDM obtains

- **Expected Model Change (EMC)**
  - Idea: the more information of question, the more uncertainty in the change of model parameters
  - **Challenge**: change is unknown before the student answer the question
    - **Solution**: estimate with expectation
      
      $\text{EMC}(q_j) = E_{a_{ij}\sim p} \Delta M(<e_i, q_j, a_{ij}>)$

      \[
      \begin{align*}
      \text{EMC} & = 0.4 \times 10 + 0.6 \times 15 = 13 \\
      \text{EMC} & = 0.9 \times 1 + 0.1 \times 100 = 10.9
      \end{align*}
      \]

      - $p = 0.4, \Delta \theta = 10$
      - $p = 0.6, \Delta \theta = 15$
      - $p = 0.9, \Delta \theta = 1$
      - $p = 0.1, \Delta \theta = 100$
Diversity Module: Select Diverse Questions

- **Quantify** the diversity of the candidate questions
  - Intuitively done with knowledge concepts related to questions
- **Importance Weighted Knowledge Coverage (IWKC)**
  - Idea: the more knowledge concepts covered, the more diverse of the questions

\[
IWKC(Q_T) = \frac{\sum_{k \in K} w_k \cdot IncCov(k, Q_T)}{\sum_{k \in K} w_k},
\]

\[
IncCov(k, Q_T) = \frac{cnt(k, Q_T)}{cnt(k, Q_T) + 1},
\]

- **Challenge:**
  - **Difficulty**: subset selection with maximizing IWKC is **NP-hard**
  - **Solution**: fortunately with the **submodular property** of IWKC, the simple greedy algorithm can achieve a suboptimal solution with optimal ratio \(1 - \frac{1}{e}\)
Experiments

- Two real **datasets** on education
- **Quality** comparison
  - AUC on performance prediction
- **Diversity** comparison
  - Coverage on knowledge

<table>
<thead>
<tr>
<th>Methods</th>
<th>IRT @25</th>
<th>IRT @50</th>
<th>MIRT @25</th>
<th>MIRT @50</th>
<th>NCDM @25</th>
<th>NCDM @50</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAND</td>
<td>0.6435</td>
<td>0.7076</td>
<td>0.7426</td>
<td>0.7767</td>
<td>0.7081</td>
<td>0.7566</td>
</tr>
<tr>
<td>MFI</td>
<td>0.7092</td>
<td>0.7207</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>KLI</td>
<td>0.7081</td>
<td>0.7257</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Opt</td>
<td>-</td>
<td>-</td>
<td>0.7515</td>
<td>0.7710</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MKLI</td>
<td>-</td>
<td>-</td>
<td>0.7502</td>
<td>0.7747</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MAAT</td>
<td><strong>0.7192</strong></td>
<td><strong>0.7319</strong></td>
<td><strong>0.7600</strong></td>
<td><strong>0.7861</strong></td>
<td><strong>0.7614</strong></td>
<td><strong>0.7868</strong></td>
</tr>
</tbody>
</table>

![Graph](image)

(a) IRT on EXAM

<table>
<thead>
<tr>
<th>MAAT</th>
<th>D-Opt</th>
<th>MKLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td>0.6666</td>
<td>Function</td>
</tr>
<tr>
<td>Set</td>
<td>0.6710</td>
<td>Equation</td>
</tr>
<tr>
<td>Equation</td>
<td>0.6763</td>
<td>Equation</td>
</tr>
<tr>
<td>Triangle</td>
<td>0.6841</td>
<td>Triangle</td>
</tr>
<tr>
<td>Algebra</td>
<td>0.6905</td>
<td>Geometry</td>
</tr>
<tr>
<td>Triangle</td>
<td>0.6961</td>
<td>Function</td>
</tr>
<tr>
<td>Coordinates</td>
<td>0.7022</td>
<td>Geometry</td>
</tr>
<tr>
<td>Geometry</td>
<td>0.7087</td>
<td>Triangle</td>
</tr>
<tr>
<td>Real Number</td>
<td>0.7136</td>
<td>Algebra</td>
</tr>
<tr>
<td>Equation</td>
<td>0.7188</td>
<td>Function</td>
</tr>
</tbody>
</table>
Outline

1. Background of Adaptive Learning
2. Cognitive Diagnosis Methods
3. Adaptive Testing Frameworks
5. Discussion and Conclusion
Traditional recommendation

Collaborative filtering on student response matrix and recommend the exercise more likely to be correctly answered.

Based on learning trajectory and knowledge structure, adopt expert system to recommend learning path.

Assuming student state is static makes it hard to capture the dynamic cognitive context, which impairs the model ability to provide suitable recommendation.

Thai-Nghe et al. 2010

Zhu et al. 2018
Cognitive Context-aware Recommendation

- Adapt to the implicit **evolving cognitive context** of the learner
- Keep the learning path be **in accordance with the logicality** determined by the knowledge structure (e.g. prerequisites)

**knowledge structure of items**

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>one digit addition</td>
</tr>
<tr>
<td>B</td>
<td>two digit addition</td>
</tr>
<tr>
<td>C</td>
<td>count number within 100</td>
</tr>
<tr>
<td>D</td>
<td>multiplication</td>
</tr>
<tr>
<td>E</td>
<td>common table form</td>
</tr>
<tr>
<td>F</td>
<td>reading tables</td>
</tr>
</tbody>
</table>

**cognitive context of learner**
Cognitive Context-aware Recommendation

Traditional static adaptive recommendation
- Recommend all items at one time
- Without Interaction
- Context-blind

Interactive dynamic adaptive recommendation
- Multi-round interaction
- Adaptively adjust recommendation strategy according to cognitive context

Limitation
Without precisely modeling cognitive context, it is hard to recommend suitable items.

Challenge
How to model the evolving cognitive context?
How to stay with the logicality of knowledge structure?
How to maximize the overall gain along the learning path?
Evolving Cognitive Context Modeling

- Cognitive Diagnosis
  - IRT, MIRT, …
  - DIRT, NeuralCD
  - …
  - Problem: only model the static cognitive context

- Knowledge Tracing (Dynamic Cognitive Diagnosis)
  - BKT (Bayesian knowledge tracing)
  - DKT
  - EKT
  - …

- How to model the dynamic evolving cognitive context along time?

---

Evolving Cognitive Context Modeling

- Deep Knowledge Tracing (DKT)
  - Use RNN to model the dynamic evolving cognitive context according to the learning records
  - The hidden state vector in RNN: dynamic cognitive context

Learning Path Recommendation

- The learning path should be **in accordance with the logicality** determined by the **knowledge structure**
- Avoid exploring the effect of recommending *calculus* to *junior students*
- Learning path recommendation should **maximize the overall gain** along the whole learning path instead

Cognitive Navigation algorithm

Actor-critic recommender
Cognitive Context-aware Recommendation

- CSEAL: Cognitive Structure Enhanced framework for Adaptive Learning
**Cognitive Context-aware Recommendation**

- **CSEAL**: Cognitive Structure Enhanced framework for Adaptive Learning

### Table 3: Overall results of $E_p$.  

<table>
<thead>
<tr>
<th></th>
<th>KSS</th>
<th>KES</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN</td>
<td>0.000700</td>
<td>0.257919</td>
</tr>
<tr>
<td>GRU4Rec</td>
<td>0.007727</td>
<td>0.201219</td>
</tr>
<tr>
<td>MC-10</td>
<td>0.110577</td>
<td>0.002236</td>
</tr>
<tr>
<td>MC-50</td>
<td>0.108636</td>
<td>-0.005103</td>
</tr>
<tr>
<td>DQN</td>
<td>0.100610</td>
<td>0.002688</td>
</tr>
<tr>
<td>CSEAL-NCN</td>
<td>0.222363</td>
<td>0.003354</td>
</tr>
<tr>
<td>CN-Random</td>
<td>0.272784</td>
<td>0.138526</td>
</tr>
<tr>
<td>Cog</td>
<td>0.164128</td>
<td>0.166560</td>
</tr>
<tr>
<td>CSEAL</td>
<td>0.346883</td>
<td>0.405823</td>
</tr>
</tbody>
</table>

$$E_p = \frac{E_e - E_s}{E_{sup} - E_s},$$

**Student performance improves in simulated environments**

**Better evaluation results given by experts**
CSEAL: Cognitive Structure Enhanced framework for Adaptive Learning

More efficient and logical learning path

Figure 9: Visualization of different recommended learning paths for the learning item 642, i.e., completing_the_square_1.
How to **accurately** evaluate the cognitive level of each student?

How to make **fast & comprehensive** testing to evaluate the cognitive level of student?

How to **adaptively recommend** personalized educational resources?
Discussion-1

- Understanding students better
  - Learning behaviors and feedbacks
  - Learning psychology
  - … …

- Understanding resources better
  - Multimodal learning
  - Knowledge structure
  - … …
We test students with exams and diagnose them with cognitive diagnosis.

Now we test machines with plenty of data but do not diagnose them.

How to diagnose machines?
Discussion-3

- Adaptive learning enhanced with cognitive diagnosis
  - Modeling human ability and learning resources simultaneously
  - Adaptively matching resources with ability → promoting learning efficiency

- How about adaptive machine learning?
Reference

- Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, Hui Xiong, Yu Su and Guoping Hu. EKT: Exercise-aware Knowledge Tracing for Student Performance Prediction. IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE), accepted.
- Yu Su, Qingwen Liu, Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, Chris Ding, Si Wei, Guoping Hu, Exercise-Enhanced Sequential Modeling for Student Performance Prediction, AAAI2018, 2435-2443.
http://base.ustc.edu.cn/

https://github.com/bigdata-ustc

Thanks!