Structure-based Knowledge Tracing: An Influence Propagation View

Shiwei Tong\textsuperscript{1}, Qi Liu\textsuperscript{1,\#}, Wei Huang\textsuperscript{1}, Zhenya Huang\textsuperscript{1}, Enhong Chen\textsuperscript{1}, Chuanren Liu\textsuperscript{2}, Haiping Ma\textsuperscript{3}, Shijin Wang\textsuperscript{4,5}

\textsuperscript{1}Anhui Province Key Laboratory of Big Data Analysis and Application, School of Data Science & School of Computer Science and Technology, University of Science and Technology of China
\textsuperscript{2}Business Analytics and Statistics, University of Tennessee, \textsuperscript{3}Anhui University,
\textsuperscript{4}iFLYTEK Research, iFLYTEK CO., LTD., \textsuperscript{5}State Key Laboratory of Cognitive Intelligence,
{tongsw, uste0411}@mail.ustc.edu.cn, {qiliuql, huangzhy, cheneh}@ustc.edu.cn
chuanren@xminer.org, hpma@ahu.edu.cn, sjwang3@iflytek.com

Reporter : Shiwei Tong
Date : 2/14/21
# Outline

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Problem Definition</td>
</tr>
<tr>
<td>3</td>
<td>SKT</td>
</tr>
<tr>
<td>4</td>
<td>Experiments</td>
</tr>
<tr>
<td>5</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>
Background

- Traditional Learning
  - Classroom & Homework & Examination

- Limitations
  - Resources
  - Share
  - Personalized

customized learning tempo
Background

- Online Education Systems
  - MOOC, ITS, OJ
A fundamental problem

- Predict student performance in the future
Background

- A fundamental problem
  - Predict student performance in the future
- Knowledge Tracing
  - Trace the evolving knowledge states of learners on the concepts
Related Work

- Bayesian Knowledge Tracing
  1. Single skill tracing
  2. Mastered or non-mastered results

- Deep Knowledge Tracing
  State in a high-dimensional and continuous representation
Although with significant improvement by utilizing knowledge structure, previous works ignore the propagated influence among concepts.
Background

- Knowledge Tracing
  - Trace the evolving knowledge states of learners on the concepts

Legend
- ✔ Correct Response
- ✗ Wrong Response
- ➔ Prerequisite
- ➔ Similarity
- ○ Learning
- ● Influenced

<table>
<thead>
<tr>
<th>ID</th>
<th>Concept 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>one digit multiplication</td>
</tr>
<tr>
<td>E</td>
<td>two digit multiplication</td>
</tr>
<tr>
<td>F</td>
<td>decimal multiplication</td>
</tr>
</tbody>
</table>
Background

- Knowledge Tracing
  - Trace the evolving knowledge states of learners on the concepts

Legend
- ✓ Correct Response
- ✗ Wrong Response
- → Prerequisite
- ➔ Similarity
- ○ Learning
- ◆ Influenced

<table>
<thead>
<tr>
<th>ID</th>
<th>Concept 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>one digit multiplication</td>
</tr>
<tr>
<td>E</td>
<td>two digit multiplication</td>
</tr>
<tr>
<td>F</td>
<td>decimal multiplication</td>
</tr>
</tbody>
</table>
Background

- Knowledge Tracing
  - Trace the evolving knowledge states of learners on the concepts

Legend
- ✓ Correct Response
- ✗ Wrong Response
- → Prerequisite
- 💼 Similarity
  - Green: Learning
  - Pink: Influenced

<table>
<thead>
<tr>
<th>ID</th>
<th>Concept 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>one digit multiplication</td>
</tr>
<tr>
<td>E</td>
<td>two digit multiplication</td>
</tr>
<tr>
<td>F</td>
<td>decimal multiplication</td>
</tr>
</tbody>
</table>
Knowledge Tracing

- Trace the evolving knowledge states of learners on the concepts
Knowledge Tracing

- Trace the evolving knowledge states of learners on the concepts
Background

- Knowledge Tracing
  - Trace the evolving knowledge states of learners on the concepts

Knowledge can be transferred among concepts
Background

- **Knowledge Tracing**
  - Trace the evolving knowledge states of learners on the concepts

- **Key issue**
  - The knowledge can be transferred among concepts
    - The learning influence can be propagated along the multiple relations in the knowledge structure
  - It is essential to consider the influence propagation when utilizing the knowledge structure for knowledge tracing
Two key challenges

- Two types of learning effects
  - The temporal effect from the exercise sequence
  - The spatial effect from the knowledge structure
Two key challenges

- Two types of learning effects
  - The temporal effect from the exercise sequence
  - The spatial effect from the knowledge structure
Two key challenges

Two types of learning effects
- The temporal effect from the exercise sequence
- The spatial effect from the knowledge structure
Challenge

- Two key challenges
  - Two types of learning effects
    - The temporal effect from the exercise sequence
    - The spatial effect from the knowledge structure
  - Knowledge structure with multiple relations
Outline

1 Introduction

2 Problem Definition

3 SKT

4 Experiments

5 Conclusion
Problem Definition

- Knowledge Tracing
  - Past Exercise Sequence
  - Knowledge Structure

- Knowledge Structure

Knowledge Tracing
- Past Exercise Sequence
- Knowledge Structure

Knowledge Structure
Knowledge Structure

Knowledge Graph
- Node → entity/concept
- Edge → relation

Educational knowledge graph
- Entity
  - Concept, exercise, course, ...
- Relation
  - Prerequisite, similarity, collaboration, Hierarchy/Taxonomy

CONCEPT MAP
- Prerequisite Relationship
- Collaboration Relationship
Knowledge Structure

- **Multiple relations**
  - **Prerequisite**
    - Directed
    - The **hierarchical structure** existing among the learning items.
  - **Similarity**
    - Undirected
    - Linked vertexes are involved in the same topic or area and may overlap in some knowledge

- **Definition**
  - \( G(V, E) \)
    - \( V = \{v_1, v_2, ..., v_N\} \) and each vertex \( v \in V \) corresponds to one concept
    - \( E = \{E^r, r = 1, ..., R\} \), where \( r \) stands for a certain type of relations, \( E^r \) represents of the type \( r \). \( R \) is the number of relation types

![Diagram](image-url)
Problem Definition

- **Given:**
  - Historical learning records \( X = \{x_t, 1, ..., T\} \)
    - \( x_t = (e_t, p_t), p_t \in \{0, 1\} \)
  - Knowledge Structure
    - \( G(V, E) \)

- **Goal:**
  - Modeling a learner’s knowledge state through their performance sequence
  - Predicting how a learner will perform on future exercises
Outline

1. Introduction
2. Problem Definition
3. SKT
4. Experiments
5. Conclusion
Two key challenges

- Two types of learning effects
  - temporal and spatial effect
- Knowledge structure with multiple relations
Overview

- Two key challenges
  - Two types of learning effects
    - **temporal** and **spatial** effect
  - Knowledge structure with **multiple relations**

- **SKT**
  - At each step, model the knowledge state **temporally** and **spatially**
  - Two influence propagation methods for **different types** of relations
The $i$-th row corresponds to the hidden state on concept $i$.
Cascade Influence Propagation (CIP): *jointly* model the **temporal and spatial effects** on concepts.
Cascade Influence Propagation

- Respectively model the temporal and spatial effect
Respectively model the **temporal** and spatial effect
The temporal effect from the exercise sequence $X = \{x_t, 1, \ldots, T\}$
The temporal effect from the exercise sequence $X = \{x_t, 1, \ldots, T\}$

$$x^t_j = \begin{cases} 
1 & \text{if } j = 2 \cdot e_t + p_t, \\
0 & \text{otherwise.} 
\end{cases}$$
Temporal Effect

- The temporal effect from the exercise sequence $X = \{x_t, 1, \ldots, T\}$
- Convert the performance vector into temporal effect vector

$$x_j^t = \begin{cases} 
1 & \text{if } j = 2 \cdot e_t + p_t, \\
0 & \text{otherwise.}
\end{cases}$$

$$\mathcal{E}_T^t = x^t \mathbf{E}_r,$$
Temporal Effect

- The temporal effect from the exercise sequence $X = \{x_t, 1, \ldots, T\}$
  - Convert the performance vector into temporal effect vector
  - Use a gate function to get the temporal effected hidden state

$$x_j^t = \begin{cases} 1 & \text{if } j = 2 \cdot e_l + p_i, \\ 0 & \text{otherwise}. \end{cases}$$

$$\mathcal{E}_T = x^t E_T,$$

$$h_{i,T}^t = G(\mathcal{E}_T^t, h_i^t),$$
Cascade Influence Propagation

- Respectively model the temporal and **spatial** effect
Spatial Effect

- Once the state of a certain concept is changed, **the influence will be propagated** to the related concepts along the multiple relations.

- **Challenge**: it is not easy to model the spatial effect on a knowledge structure with **multiple relations**
Two influence propagation methods

- Partial propagation for directed relations
- Synchronization propagation for undirected relations
Spatial Effect

- Two influence propagation methods
  - Partial propagation for directed relations
  - Synchronization propagation for undirected relations
Partial Propagation

- Partial propagation for directed relations
  - The influence is unidirectionally propagated from only predecessors to successors

\[ P_{ij}^r = (h_i^{t,T} - h_i^t) \oplus E_c(j). \]
Partial Propagation for directed relations

- The influence is unidirectionally propagated from only predecessors to successors

\[ P_{ij}^r = (h_i^{t,T} - h_i^t) \oplus E_c(j). \]

\[ f_{part}(h_i^{t,T}, h_i^t, E_c(j)) = \text{relu}(W_p^r P_{ij}^r + b_p^r), \]

Concept hidden state

Temporal effected hidden state

Concept feature vector

Spatial effect vector
Spatial Effect

- Two influence propagation methods
  - Partial propagation for directed relations
  - Synchronization propagation for undirected relations
- Synchronization propagation for undirected relations
  - The influence is bidirectionally propagated between neighbor concepts
    - To neighbors
    - From neighbors

![Diagram of synchronization propagation](image.png)
Synchronization Propagation

- Synchronization propagation for undirected relations
  - The influence is bidirectionally propagated between neighbor concepts
    - To neighbors
    - From neighbors

\[
S_{ij}^{r} = h_{i}^{t,T} \oplus h_{j}^{t} \oplus E_{c}(j). \quad f_{sync}(h_{i}^{t,T}, h_{j}^{t}, E_{c}(j)) = relu(W_{s}S_{ij}^{r} + b_{s})
\]
Synchronization Propagation

- Synchronization propagation for undirected relations
  - The influence is bidirectionally propagated between neighbor concepts
    - To neighbors
    - From neighbors

\[ R_i^r = (h_i^{t,T} + \sum_{j \in \mathcal{N}^r(i)} h_j^t) \oplus E_c(i), \quad sync_i^r = \text{relu}(W_{ss}^r R_i^r + b_{ss}^r), \]

Concept hidden state  Temporal effected hidden state  Concept feature vector
State update

- **Aggregate influence**
  - **Learning**
    - \( A_i = \sum_r sync_i^r \)
  - **Influenced**
    - Neighbors & successors
    - \( A_j = (1 - \alpha) \cdot \sum_r part_{ij}^r + \alpha \sum_r sync_{ij}^r \)
    - \( \alpha \) is a hyper-parameter to balance the influence from two relations

- **Aggregated Influence**
  - \( I_j = relu(W_j A_j + b_i) \)

- **State Update**
  - \( h_{j}^{t+1} = g(I_j, h_j^t) \)
Prediction and Training

- **Prediction**
  - $\hat{p}_i^t = f_{out}(h_i^t)$
  - $f_{out}(h_i^t) = \sigma(W_o h_i^t + b_o)$
  - $P(p_t = 1|e_t, x_1,...,t-1, G) = \hat{p}_{e_t}^t$
**Prediction and Training**

- **Prediction**
  - $\hat{p}_i^t = f_{out}(h_i^t)$
  - $f_{out}(h_i^t) = \sigma(W_o h_i^t + b_o)$
  - $P(p_t = 1|e_t, x_1, ..., x_{t-1}, G) = \hat{p}_e^t$

- **Loss Function**
  - $\mathcal{L} = - \sum_t (p_t \log \hat{p}_t + (1 - p_t) \log (1 - \hat{p}_t))$

**Diagram:**
- Concept hidden state $\mathcal{H}$
- Knowledge State $y_t = \{\hat{p}_1^t, ..., \hat{p}_N^t\}$

**True label**

**Prediction**

---

Anhui Province Key Laboratory Of Big Data Analysis and Application
Outline

1. Introduction
2. Problem Definition
3. SKT
4. Experiments
5. Conclusion
Experiments

- **Dataset**
  - ASSISTments
    - Knowledge structure is constructed from learning records
      - Code is available in https://github.com/bigdata-ustc/EduData
  - Junyi
    - Knowledge structure is provided

- **Data Analysis**
  - Conditional correctness $P_{ij} = \frac{n_c(j|i)}{n(j|i)}$
    - $n_c(j|i)$ is the times that concept $j$ is correctly answered in the first time step when its neighbor or predecessor $i$ has been correctly answered
    - $n(j|i)$ is the times that concept $j$ is answered no matter when its neighbor or predecessor $i$ has been correctly answered.
  - Non-conditional correctness $P^n_j = \frac{n_c(j)}{n(j)}$

<table>
<thead>
<tr>
<th>Statistics</th>
<th>ASSISTments</th>
<th>Junyi</th>
</tr>
</thead>
<tbody>
<tr>
<td># learners</td>
<td>19,840</td>
<td>1,000</td>
</tr>
<tr>
<td># sequence</td>
<td>19,840</td>
<td>59,792</td>
</tr>
<tr>
<td># exercise-performance pair</td>
<td>683,801</td>
<td>4,049,359</td>
</tr>
<tr>
<td># vertexes</td>
<td>100</td>
<td>835</td>
</tr>
<tr>
<td># prerequisite relations</td>
<td>1,112</td>
<td>978</td>
</tr>
<tr>
<td># similarity relations</td>
<td>1,512</td>
<td>1,040</td>
</tr>
</tbody>
</table>

![Correct Rate](image)

**TABLE I**: The statistics of the dataset.

55.13% average $P^n$  
69.55% average $P^p$  
73.25% average $P^s$

Fig. 3: Correct rate comparison.
Experiments

- **Dataset**
  - ASSISTments
    - Knowledge structure is constructed from learning records
    - Code is available in https://github.com/b giant-data-u stc/EduData
  - Junyi
    - Knowledge structure is provided

- **Data Analysis**
  - Conditional correctness $P_{ij} = \frac{n_c(j|i)}{n(j|i)}$
    - $n_c(j|i)$ is the times that concept $j$ is correctly answered in the first time step when its neighbor or predecessor $i$ has been correctly answered
    - $n(j|i)$ is the times that concept $j$ is answered no matter when its neighbor or predecessor $i$ has been correctly answered.
  - Non-conditional correctness $P_j^n = \frac{n_c(j)}{n(j)}$

**TABLE I: The statistics of the dataset.**

<table>
<thead>
<tr>
<th>Statistics</th>
<th>ASSISTments</th>
<th>Junyi</th>
</tr>
</thead>
<tbody>
<tr>
<td># learners</td>
<td>19,840</td>
<td>1,000</td>
</tr>
<tr>
<td># sequence</td>
<td>19,840</td>
<td>59,792</td>
</tr>
<tr>
<td># exercise-performance pair</td>
<td>683,801</td>
<td>4,049,359</td>
</tr>
<tr>
<td># vertexes</td>
<td>100</td>
<td>835</td>
</tr>
<tr>
<td># prerequisite relations</td>
<td>1,112</td>
<td>978</td>
</tr>
<tr>
<td># similarity relations</td>
<td>1,512</td>
<td>1,040</td>
</tr>
</tbody>
</table>

**Fig. 3: Correct rate comparison.**

55.13% average $P^n$

69.55% average $P^p$

73.25% average $P^s$
Experiments

- Dataset
  - ASSISTments
    - Knowledge structure is constructed from learning records
    - Code is available in https://github.com/bigdata-ustc/EduData
  - Junyi
    - Knowledge structure is provided

- Data Analysis
  - Conditional correctness $P_{ij} = \frac{n_c(j|i)}{n(j|i)}$
    - $n_c(j|i)$ is the times that concept $j$ is correctly answered in the first time step when its neighbor or predecessor $i$ has been correctly answered
    - $n(j|i)$ is the times that concept $j$ is answered no matter when its neighbor or predecessor $i$ has been correctly answered
  - Non-conditional correctness $P_j^n = \frac{n_c(j)}{n(j)}$

Table 1: The statistics of the dataset.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>ASSISTments</th>
<th>Junyi</th>
</tr>
</thead>
<tbody>
<tr>
<td># learners</td>
<td>19,840</td>
<td>1,000</td>
</tr>
<tr>
<td># sequence</td>
<td>19,840</td>
<td>59,792</td>
</tr>
<tr>
<td># exercise-performance pair</td>
<td>683,801</td>
<td>4,049,359</td>
</tr>
<tr>
<td># vertexes</td>
<td>100</td>
<td>835</td>
</tr>
<tr>
<td># prerequisite relations</td>
<td>1,112</td>
<td>978</td>
</tr>
<tr>
<td># similarity relations</td>
<td>1,512</td>
<td>1,040</td>
</tr>
</tbody>
</table>

When the neighbors and predecessors have been learned, the correctness probability is promoted.

Fig. 3: Correct rate comparison.
Experiments

- Characteristics of the comparison methods

<table>
<thead>
<tr>
<th>Experimental Settings</th>
<th>Modeling Concept Relations</th>
<th>Directed</th>
<th>Undirected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BKT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DKT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DKT+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DKVMN</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GKT</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Variants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKT_TE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKT_Part</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>SKT_Sync</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Our Model</td>
<td>SKT</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

- Metrics: AUC and F1
Performance Comparison

TABLE III: Performance comparison on the KT task.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Eval</th>
<th>BKT</th>
<th>DKT</th>
<th>DKT+</th>
<th>DKVMN</th>
<th>GKT</th>
<th>SKT (ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSISTments</td>
<td>AUC</td>
<td>0.678</td>
<td>0.727</td>
<td>0.728</td>
<td>0.730</td>
<td>0.735</td>
<td>0.746</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>0.554</td>
<td>0.541</td>
<td>0.572</td>
<td>0.575</td>
<td>0.577</td>
<td>0.607</td>
</tr>
<tr>
<td>Junyi</td>
<td>AUC</td>
<td>0.831</td>
<td>0.847</td>
<td>0.889</td>
<td>0.890</td>
<td>0.893</td>
<td>0.908</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>0.760</td>
<td>0.779</td>
<td>0.819</td>
<td>0.817</td>
<td>0.825</td>
<td>0.835</td>
</tr>
</tbody>
</table>

- Our proposed SKT achieves a better performance than any other baseline.
Our proposed SKT achieves a better performance than any other baseline.

Among baselines, DKVMN and GKT are the best two models, which either model the relations of concepts or explicitly utilize the existing knowledge structures.
Performance Comparison

TABLE III: Performance comparison on the KT task.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Eval</th>
<th>BKT</th>
<th>DKT</th>
<th>DKT+</th>
<th>DKVMN</th>
<th>GKT</th>
<th>SKT (ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSISTments</td>
<td>AUC</td>
<td>0.678</td>
<td>0.727</td>
<td>0.728</td>
<td>0.730</td>
<td>0.735</td>
<td>0.746</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>0.554</td>
<td>0.541</td>
<td>0.572</td>
<td>0.575</td>
<td>0.577</td>
<td>0.607</td>
</tr>
<tr>
<td>Junyi</td>
<td>AUC</td>
<td>0.831</td>
<td>0.847</td>
<td>0.889</td>
<td>0.890</td>
<td>0.893</td>
<td>0.908</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>0.760</td>
<td>0.779</td>
<td>0.819</td>
<td>0.817</td>
<td>0.825</td>
<td>0.835</td>
</tr>
</tbody>
</table>

Utilizing the concept relations (i.e., knowledge structure), no matter explicitly or implicitly, does provide additional useful information for estimating learners’ knowledge states.
Performance Comparison

TABLE III: Performance comparison on the KT task.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Eval</th>
<th>BKT</th>
<th>DKT</th>
<th>DKT+</th>
<th>DKVMN</th>
<th>GKT</th>
<th>SKT (ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSISTments</td>
<td>AUC</td>
<td>0.678</td>
<td>0.727</td>
<td>0.728</td>
<td>0.730</td>
<td>0.735</td>
<td>0.746</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>0.554</td>
<td>0.541</td>
<td>0.572</td>
<td>0.575</td>
<td>0.577</td>
<td>0.607</td>
</tr>
<tr>
<td>Junyi</td>
<td>AUC</td>
<td>0.831</td>
<td>0.847</td>
<td>0.889</td>
<td>0.890</td>
<td>0.893</td>
<td>0.908</td>
</tr>
<tr>
<td></td>
<td>F1</td>
<td>0.760</td>
<td>0.779</td>
<td>0.819</td>
<td>0.817</td>
<td>0.825</td>
<td>0.835</td>
</tr>
</tbody>
</table>

- Modeling the **temporal and spatial effect** based on influence propagation
- Respectively modeling the propagation ways along different relations

Simultaneously combining **temporal** and **spatial** information and considering the multiple relations among the **knowledge structure**.
Ablation Study

- Variants
  - SKT_TE
    - only models the temporal effect
  - SKT_Part
    - Only models partial propagation
  - SKT_Sync
    - Only models synchronization propagation

<table>
<thead>
<tr>
<th>Model</th>
<th>ASSISTments</th>
<th>Junyi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>F1</td>
</tr>
<tr>
<td>SKT_TE</td>
<td>0.710</td>
<td>0.533</td>
</tr>
<tr>
<td>SKT_Part</td>
<td>0.711</td>
<td>0.548</td>
</tr>
<tr>
<td>SKT_Sync</td>
<td>0.736</td>
<td>0.579</td>
</tr>
<tr>
<td>SKT</td>
<td>0.746</td>
<td>0.607</td>
</tr>
</tbody>
</table>

TABLE IV: Performance comparison of SKT and its variants.
Ablation Study

- **Variants**
  - **SKT_TE**
    - only models the temporal effect
  - **SKT_Part**
    - Only models partial propagation
  - **SKT_Sync**
    - Only models synchronization propagation

- **SKT_Part** and **SKT_Sync** have a better performance than **SKT_TE**.
  - It is important to model the influence propagated in the knowledge structure.

### TABLE IV: Performance comparison of SKT and its variants.

<table>
<thead>
<tr>
<th>Model</th>
<th>ASSISTments</th>
<th>Junyi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>F1</td>
</tr>
<tr>
<td>SKT_TE</td>
<td>0.710</td>
<td>0.533</td>
</tr>
<tr>
<td>SKT_Part</td>
<td>0.711</td>
<td>0.548</td>
</tr>
<tr>
<td>SKT.Sync</td>
<td>0.736</td>
<td>0.579</td>
</tr>
<tr>
<td>SKT</td>
<td><strong>0.746</strong></td>
<td><strong>0.607</strong></td>
</tr>
</tbody>
</table>
Ablation Study

- Variants
  - SKT_TE
    - only models the temporal effect
  - SKT_Part
    - Only models partial propagation
  - SKT_Sync
    - Only models synchronization propagation

- SKT_Part and SKT_Sync have a better performance than SKT_TE.
- It is important to model the influence propagated in the knowledge structure.
- SKT has a significant promotion
  - It is critical to consider the different ways of the propagation along different relations.

<table>
<thead>
<tr>
<th>Model</th>
<th>ASSISTments AUC</th>
<th>ASSISTments F1</th>
<th>Junyi AUC</th>
<th>Junyi F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SKT_TE</td>
<td>0.710</td>
<td>0.533</td>
<td>0.887</td>
<td>0.824</td>
</tr>
<tr>
<td>SKT_Part</td>
<td>0.711</td>
<td>0.548</td>
<td>0.898</td>
<td>0.829</td>
</tr>
<tr>
<td>SKT_Sync</td>
<td>0.736</td>
<td>0.579</td>
<td>0.899</td>
<td>0.828</td>
</tr>
<tr>
<td>SKT</td>
<td>0.746</td>
<td>0.607</td>
<td>0.908</td>
<td>0.835</td>
</tr>
</tbody>
</table>
\( \alpha \) plays a crucial role which balances the contribution from different influences.

Fig. 4: Influence of \( \alpha \).
Parameter Sensitivity

- $\alpha$ plays a crucial role which balances the contribution from different influences.
  - When $\alpha$ increases
    - the performance increases at the beginning.

Fig. 4: Influence of $\alpha$. 
Parameter Sensitivity

- $\alpha$ plays a crucial role which balances the contribution from different influences.
  - When $\alpha$ increases
    - the performance increases at the beginning.
    - the performance afterwards decreases

Fig. 4: Influence of $\alpha$. 
α plays a crucial role which balances the contribution from different influences.

- When α increases
  - the performance increases at the beginning.
  - the performance afterwards decreases

Properly **balancing** the influence from different relations is vital for achieving **more accurate** prediction performance.

Fig. 4: Influence of α.
Case Study

The proficiency of related concepts gets promoted.

1. Decimals on the number line 1
2. Decimals on the number line 1
3. Number line
4. Telling time
5. Subtraction 3

(a) Time series performance diagram.
(b) Subsequence diagram.
(c) Radar diagram.

The proficiency of the predecessor will not be influenced by the successor.
Generate the influence feature vector by $J_{ij} = \frac{y(j|i)}{\sum_k y(j|k)}$.

$y(j|i)$ is the average correctness probability assigned by SKT to exercise $j$ when exercise $i$ is answered correctly at the first time step.

![Concept Clustering Diagram]
Outline

1 Introduction
2 Problem Definition
3 SKT
4 Experiments
5 Conclusion
Conclusion

- Knowledge Tracing
  - Goal: precisely trace the evolving knowledge states of learners
  - Key issue: the knowledge can be transferred among concepts
    - Influence propagation in knowledge structure

- Two challenges
  - Two types of learning effects: temporal and spatial effect
  - Multiple relations in knowledge structure

- Structure-based Knowledge Tracing
  - Respectively model the temporal and spatial effect
  - Propose two influence propagation methods for directed relations and undirected relations

- Extensive experiments on real-world data demonstrate the effectiveness and interpretability of SKT

- Our codes are available in https://github.com/bigdata-ustc/XKT
Thank you

Q&A

wx13011065 tongsw@mail.ustc.edu.cn tswsxk.github.io

bigdata.ustc.edu.cn