# Federated Deep Knowledge Tracing

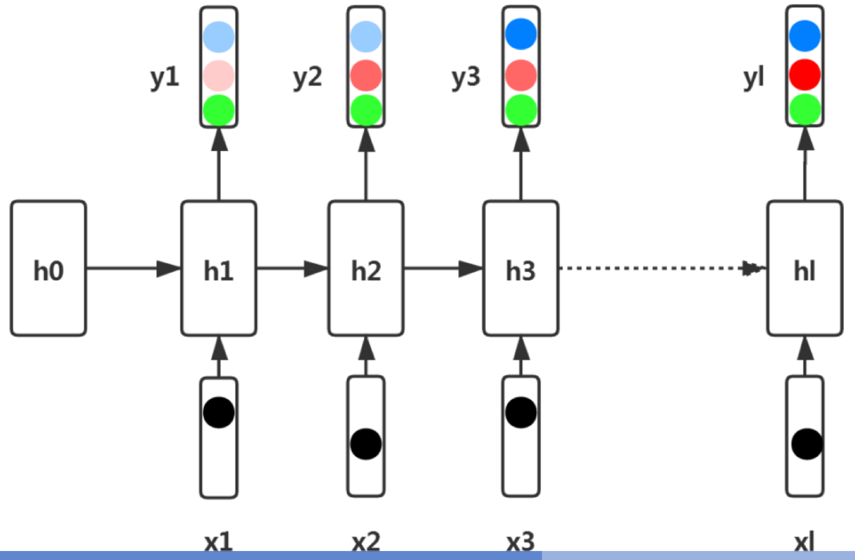
Jinze Wu[1], Zhenya Huang[1], Qi Liu[1,*], Defu Lian[1], Hao Wang[1], Enhong Chen[1]
Haiping Ma[2], Shijin Wang[3]
[1]Anhui Province Key Laboratory of Big Data Analysis and Application,
School of Computer Science and Technology, University of Science and Technology of China
[2]Anhui University
[3]iFLYTEK Research & State Key Laboratory of Cognitive Intelligence, iFLYTEK Co., Ltd,
{hxwjz,wanghao3}@mail.ustc.edu.cn,{huangzhy,qiliuql,liandefu,cheneh}@ustc.edu.cn,
hpma2020@163.com,sjwang3@iflytek.com

**Reporter: Jinze Wu**

# Outline

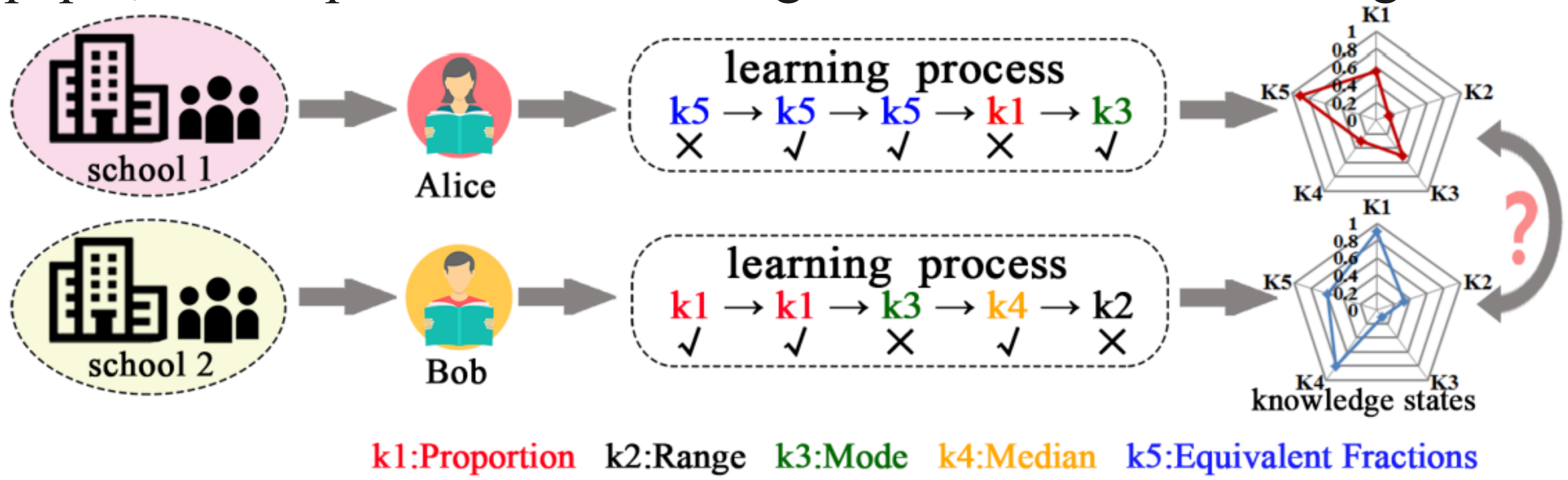
# Background

- Knowledge Tracing
  - A fundamental task in **intelligent education**.
  - Aim to trace knowledge states of students based on their historical learning trajectories
  - **Deep knowledge tracing** (DKT) leverage recurrent neural networks to capture the changes of students' knowledge states
  - Traditionally, researchers process a central training to learn DKT models.



# Background

- Problems and Challenges
  - learn high-quality DKT models for isolated silos (e.g., different schools) while avoiding **data isolation problem**
  - Data scarcity: distributed learning data is difficult to be collected
  - Different data quality: different teaching arrangements lead to different data quality
  - Data incomparability: the information mined in different scenes is incomparable
  - In this paper, we adopt federated learning to overcome the challenges



# Outline

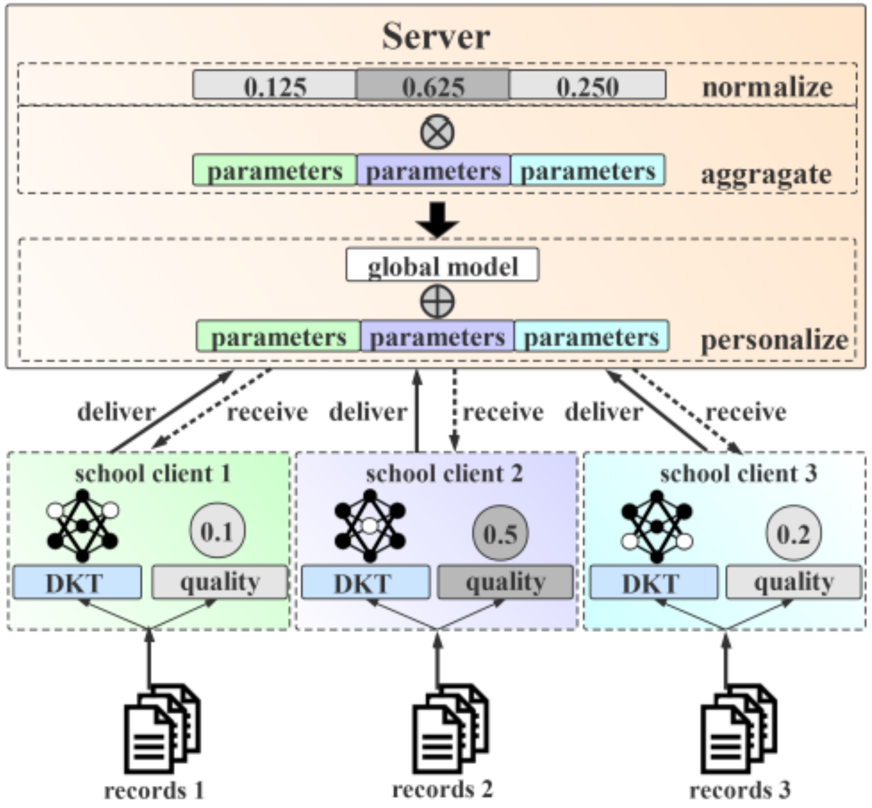
# Problem Definition

- Problem Definition
  - Given:
    - S : set of schools
    - $Q_s$: set of exercises
    - $N_s$: set of students
    - r = { $(q_0, g_0), (q_1, g_1), \ldots, (q_l, g_l)$ } : learning records of a certain student
      - $q_l \in Q_s, g_l \in \{0, 1\}$
  - Goal: train |S| local DKT models, where the s-th DKT model can trace the students in school s of knowledge states (represent the students' mastery of concepts).

# Outline

# FDKT Framework

- FDKT Framework
  - Server
    - receive and aggregate local DKTs delivered
    - compose and updating the models for local clients.
  - Client
    - train an independent DKT with the private data
    - evaluate the data quality with confidence measurements



# FDKT Framework

- Client
  - Local DKT modeling
    - given the learning records of a certain student, DKT uses a RNN to model her knowledge presentations $\{h_1, h_2, \ldots, h_l\}$ and output her knowledge states (mastery levels) $\{y_1, y_2, \ldots, y_l\}$ on multiple concepts

$$h_l = \tanh(\mathbf{W_{hx}} x_l + \mathbf{W_{hh}} h_{l-1} + \mathbf{b_h}),$$
$$y_l = \mathrm{sigmoid}(\mathbf{W_{yh}} h_l + \mathbf{b_y}),$$

  - Data quality evaluation
    - propose two data quality evaluation methods with confidence estimation for $Q_s$ items following educational measurement theories, i.e., Classical Test Theory and Item Response Theory, respectively.
    - CTT confidence: $\alpha_{CTT} = F\ (P(Q_s), D(Q_s), CR(Q_s))$
    - IRT confidence: $\alpha_{IRT} = \max(\sum_{i=1}^{|Q_s|} \beta_i\ * I_i(\theta))$

# FDKT Framework

- Server
  - Model aggregation
    - integrate all local DKT models to a global one based on the local confidences

$$\Theta^t = \sum_{s=1}^{S} \hat{\alpha}_s^t \times \Theta_s^t$$

  - Model update
    - hierarchical model interpolation: obtain the new personalized model from both the global model and the local model

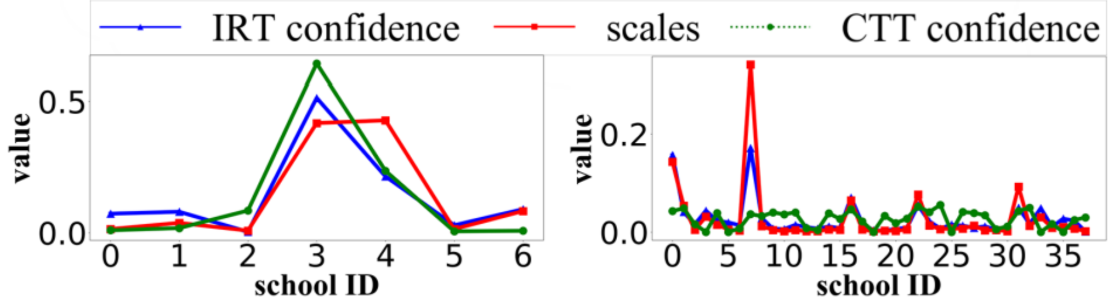
$$\Theta_s^t = \lambda \cdot \Theta_s^t + (1 - \lambda) \cdot \Theta^t.$$

# Outline

# Experiment

- Dataset
  - MATH dataset
  - ASSIST
- Data analysis
  - Inconsistency between data scales and data quality
- Baseline methods
  - centrally training methods:
    BKT, DKT
  - Traditional federated methods:
    FedSGD, FedAvg, FedAtt, LoAdaboost
  - Our methods:
    FedInter, FedCTT, FedIRT
- Evaluation metrics
  - ACC, AUC, RMSE, epoch, DOA

| Statistics | MATH | ASSIST |
|---|---|---|
| # of schools | 7 | 38 |
| # of records | 204,293 | 801,645 |
| # of students | 3,830 | 7,395 |
| # of exercises | 4,145 | 27,288 |
| # of knowledge concepts | 112 | 200 |
| Avg. records per student | 53.34 | 108.40 |
| Avg. exercises per concept | 37.01 | 136.44 |



(a) MATH (b) ASSIST

# Experiment

- Over performances
  - Task: student performance prediction
- Observation
  - Methods with federated learning settings perform better: harness more data
  - FDKTIRT has the best performances: quality-oriented aggregation is beneficial
  - FDKTCTT performs no most outstanding result: the limitations of CTT confidence

**Table 2: Results of student performance prediction under four metrics.**

(a) Results of student performance prediction on MATH

| method \ dataset | MATH | | | |
|---|---|---|---|---|
| | epoch | RMSE | AUC | ACC |
| BKT | - | 0.463 | 0.692 | 0.701 |
| DKT | - | 0.453 | 0.705 | 0.712 |
| FedSGD | - | 0.455 | 0.696 | 0.694 |
| FedAvg | 13 | 0.449 | 0.721 | 0.713 |
| FedAtt | 14 | 0.453 | 0.718 | 0.708 |
| LoAdaboost | 8 | 0.450 | 0.726 | 0.708 |
| FedInter | 8 | 0.449 | 0.733 | 0.719 |
| FDKTCTT | **4** | 0.448 | 0.735 | 0.717 |
| FDKTIRT | **4** | **0.446** | **0.739** | **0.721** |

(b) Results of student performance prediction on ASSIST

| method \ dataset | ASSIST | | | |
|---|---|---|---|---|
| | epoch | RMSE | AUC | ACC |
| BKT | - | 0.452 | 0.743 | 0.681 |
| DKT | - | 0.413 | 0.814 | 0.75 |
| FedSGD | - | 0.425 | 0.798 | 0.746 |
| FedAvg | 20 | 0.387 | 0.861 | 0.791 |
| FedAtt | 22 | 0.386 | 0.862 | 0.792 |
| LoAdaboost | 28 | 0.384 | 0.863 | 0.792 |
| FedInter | **11** | 0.376 | 0.875 | 0.796 |
| FDKTCTT | 17 | 0.379 | 0.872 | 0.795 |
| FDKTIRT | **11** | **0.375** | **0.877** | **0.802** |

# Experiment

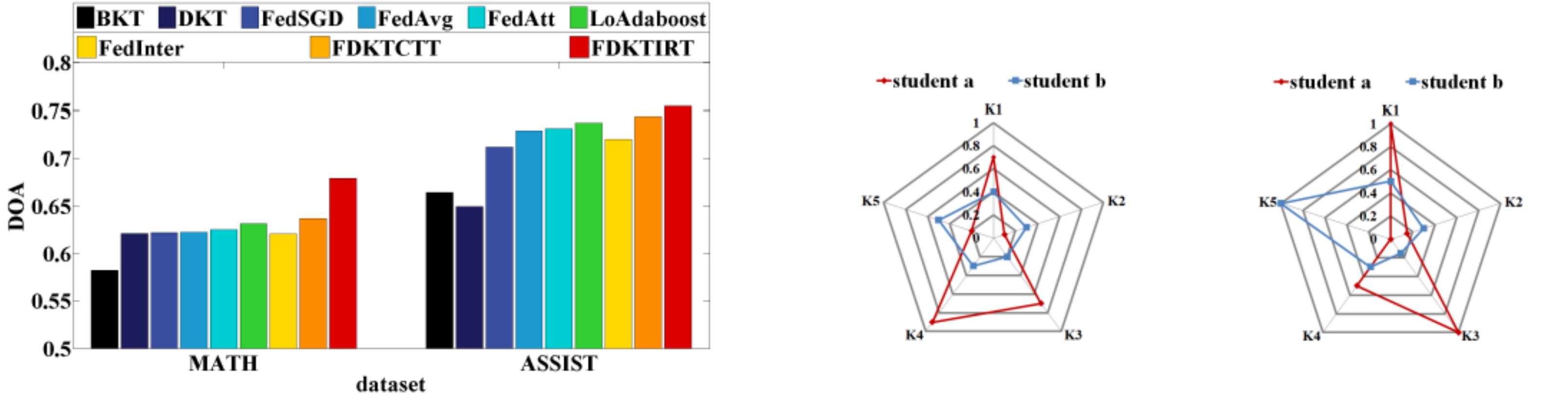
- Effectiveness of Data quality
  - Task: report the performances of different methods on isolated schools
- Observation
  - DKT performs significantly poorly: federated learning settings expand available data
  - FedInter achieves competitive results: effectiveness of model personalization
  - FDKTIRT and FDKTCTT perform better: data quality is a more important factor

**Table 3: Statistics of confidence, scale and results of student performance prediction with AUC of partial datasets.**

| category | name | school 1 | school 2 | school 3 | school 4 | school 5 | school 6 |
|---|---|---|---|---|---|---|---|
| statistic | scales | 114,627 | 42,899 | 10554 | 9,743 | 3103 | 1,163 |
| | CTT confidence | 0.043 | 0.048 | 0.041 | 0.049 | 0.034 | 0.012 |
| | IRT confidence | 0.157 | 0.041 | 0.022 | 0.016 | 0.011 | 0.010 |
| method | DKT | 0.871 | 0.830 | 0.838 | 0.809 | 0.798 | 0.535 |
| | LoAdaboost | 0.858(-1.3%) | 0.816(-1.4%) | 0.811(-2.7%) | 0.796(-1.3%) | 0.865(+6.7%) | 0.737(+20.2%) |
| | FedInter | 0.876(+0.5%) | 0.843(+1.3%) | 0.846(**+0.8%**) | 0.867(+5.8%) | 0.927(+12.9%) | 0.801(+26.6%) |
| | FDKTCTT | 0.871(+0.0%) | 0.844(+1.4%) | 0.842(+0.4%) | 0.871(+6.2%) | 0.934(+13.6%) | 0.801(+26.6%) |
| | FDKTIRT | 0.879(**+0.8%**) | 0.848(**+1.8%**) | 0.846(**+0.8%**) | 0.875(**+6.6%**) | 0.937(**+13.9%**) | 0.805(**+27.0%**) |

# Experiment

- Performance on Comparability
  - Task: compute the comparability of students
- Observation
  - BKT and DKT perform worst: independent training on isolated schools is incomparable
  - Federated learning methods perform better: effective for federated learning strategies
  - FDKTIRT performs best: quality-oriented aggregation strategies with personalization update strategy are more effective at achieving comparable results




(a) Overall performances on comparability with metric DOA. (b) Example: performances on states. (c) Example: performances on scores.

**Figure 4: Left bar chart is DOA results of methods. Right radars are comparable examples of two students' knowledge states and true scores from isolated schools. (K1: Scatter Plot; K2: Proportion; K3: Point Plotting; K4: Graph shape; K5: Congruence;)**

# Outline

# Conclusion & Future work

- Overall results
  - Design a novel client-server architecture framework to solve data isolation problem
  - Combine federated learning to train DKT models while alleviating data scarcity
  - In the client part, two implementations of quality-oriented aggregation strategies are provided; in the server part, hierarchical model interpolation is explored

- Future work
  - Explore more applications of federated learning in the educational field,
  - Extend FDKT to many other KT methods and develop a general platform
  - Explore ways to model item and user characteristics appropriately under federated settings

# Q & A

Thanks!

hxwjz@mail.ustc.edu.cn