IJCAI 2021 MONTREAL

Item Response Ranking for Cognitive Diagnosis

[1]Anhui Province Key Laboratory of Big Data Analysis and Application, School of Computer Science and Technology & School of Data Science, University of Science and Technology of China
[2]University of California, Berkeley
{tongsw, yrunl, ustc0411}@mail.ustc.edu.cn, {qiliuql, huangzhy}@ustc.edu.cn,
{pardos, jiangwj}@berkeley.edu

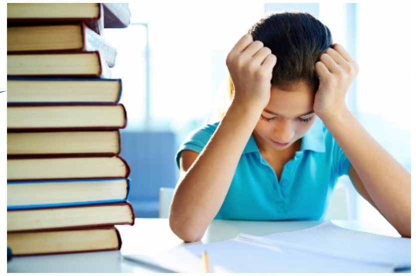
Reporter：Shiwei Tong
Date：6/25/2021

Outline

30th International Joint Conference on Artificial Intelligence

Background

- Cognitive diagnosis, a fundamental task in education area
  - reveal the proficiency level of students on knowledge concepts.
- Assist tutors to give proper instruction based on individual characteristics
- Help students be aware of their learning progress



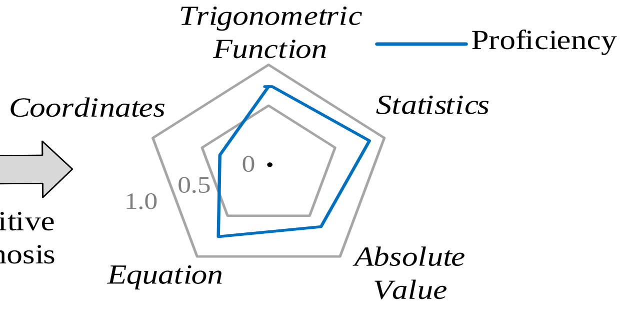
Exercise	Knowledge Concepts
e_1	*Trigonometric Function*
e_2	*Statistics*
e_3	*Coordinates, Equation*
e_4	*Absolute Value, Equation*

Practice

Exercise	e_1	e_2	e_3	e_4
Response	✓	✓	✗	✓

Response Logs

Cognitive Diagnosis



Diagnostic Report

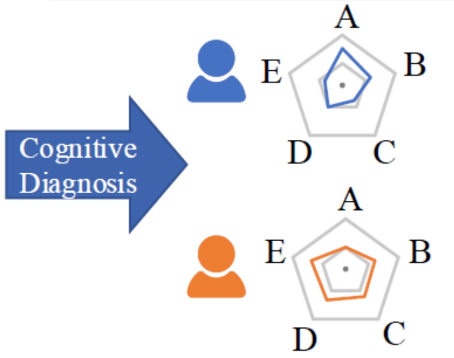
Cognitive Diagnosis Model (CDM)

IJCAI 2021 MONTREAL

- General form
  - $P(y_{ie}|\boldsymbol{u}_i, \boldsymbol{v}_e) = f(\boldsymbol{u}_i, \boldsymbol{v}_e)$
  - $\boldsymbol{u}_i$ is the latent traits of students
  - $\boldsymbol{v}_e$ is the latent traits of items
  - f is called the interaction function

Exercise	Knowledge Concepts	Ada	Bob
e_1	A	✓	✗
e_2	B	✓	✓
e_3	A, B	✗	✗
e_4	C, D	?	✓
e_5	C, E	✗	?

Legend
A: Function
B: Derivative
C: Geometry
D: Coordinates
E: Equation
✓ Correct
✗ Wrong



Monotonicity

Student’ s proficiency is monotonic with the probability of giving the right response to a test item.

Key Question

IJCAI 2021 MONTREAL

General Form

$$P(y_{ie}|\boldsymbol{u}_i, \boldsymbol{v}_e) = f(\boldsymbol{u}_i, \boldsymbol{v}_e)$$

Objective Function

$$r_{ie} \leftarrow P(y_{ie})$$

True response

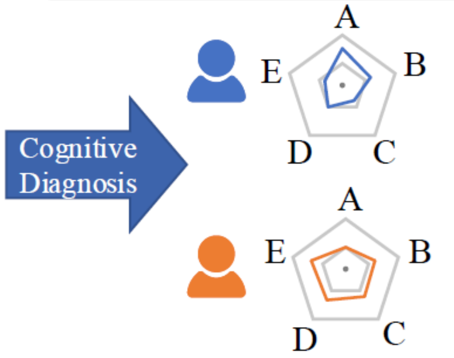
Predicted probability of giving the right response

Cognitive Diagnosis Model (CDM)

- General form
  - $P(y_{ie}|\boldsymbol{u}_i, \boldsymbol{v}_e) = f(\boldsymbol{u}_i, \boldsymbol{v}_e)$
  - $\boldsymbol{u}_i$ is the latent traits of students
  - $\boldsymbol{v}_e$ is the latent traits of items
  - f is called the interaction function
- Basic theories
  - Local dependence
  - **Monotonicity**

Exercise	Knowledge Concepts	Ada	Bob
e_1	A	✓	✗
e_2	B	✓	✓
e_3	A, B	✗	✗
e_4	C, D	?	✓
e_5	C, E	✗	?

Legend
A: Function
B: Derivative
C: Geometry
D: Coordinates
E: Equation
✓ Correct
✗ Wrong



Monotonicity

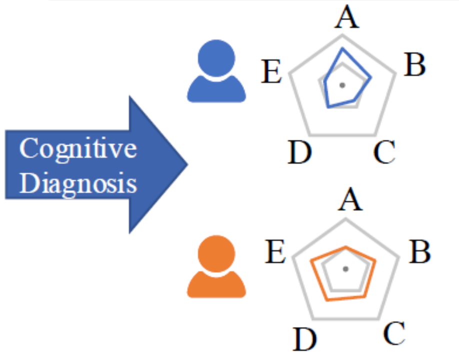
Student's proficiency is monotonic with the probability of giving the right response to a test item.

Cognitive Diagnosis Model (CDM)

- General form
  - $P(y_{ie}|\boldsymbol{u}_i, \boldsymbol{v}_e) = f(\boxed{\boldsymbol{u}_i, \boldsymbol{v}_e})$
  - $\boldsymbol{u}_i$ is the latent traits of students
  - $\boldsymbol{v}_e$ is the latent traits of items
  - f is called the interaction function
- Basic theories
  - Local dependence
  - **Monotonicity**

Exercise	Knowledge Concepts	Ada	Bob
e_1	A	✓	✗
e_2	B	✓	✓
e_3	A, B	✗	✗
e_4	C, D	?	✓
e_5	C, E	✗	?

Legend
A: Function
B: Derivative
C: Geometry
D: Coordinates
E: Equation
✓ Correct
✗ Wrong



Monotonicity

Student’ s proficiency is monotonic with the probability of giving the right response to a test item.

Related Work

IJCAI 2021 MONTREAL

- Traditional Methods
  - Item-level
    - IRT
      - $P(y_{ie}|\theta, a, b) = \frac{1}{1+e^{-1.7a(\theta-b)}}$
    - MIRT
      - $P(y_{ie}|\boldsymbol{\theta}, \boldsymbol{a}, b) = \frac{1}{1+e^{-\boldsymbol{a}\boldsymbol{\theta}+b}}$
  - Concept level
    - DINA
      - $P(y_{ie}|\theta, g, s, \beta) = g^{1-\eta}(1-s)^{\eta}$
      - $\eta = \prod_k \theta_k^{\beta_k}$
      - $\theta_k \in \{0,1\}, \beta_k \in \{0,1\}$
- Deep learning methods
  - NeuralCD

General Form

$$P(y_{ie}|\boldsymbol{u}_i, \boldsymbol{v}_e) = f(\boldsymbol{u}_i, \boldsymbol{v}_e)$$

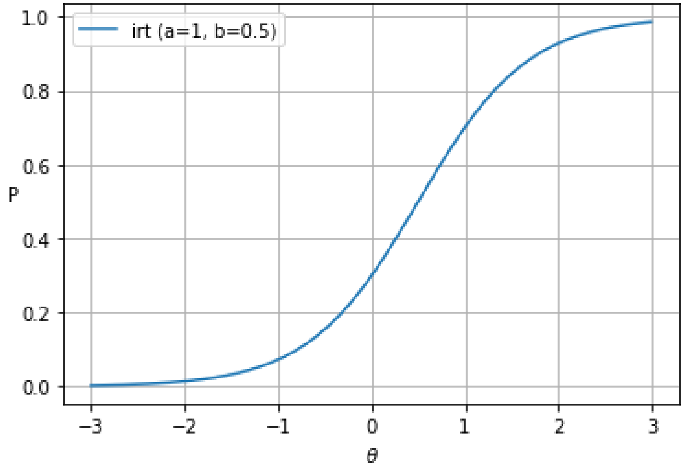


Figure 1: Example of IRT

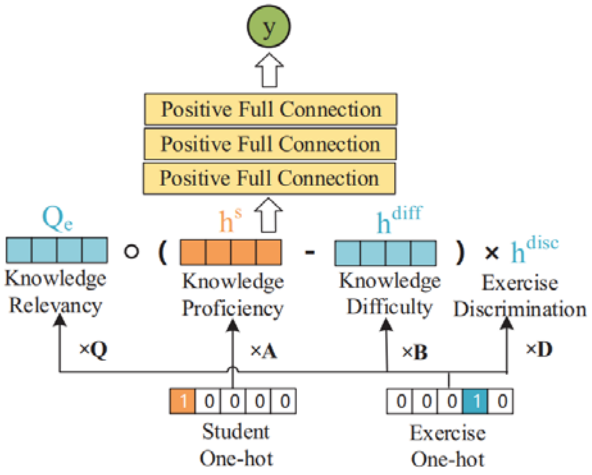


Figure 2: NeuralCD Framework

Related Work

- Traditional Methods
  - Item-level
    - IRT
      - $P(y_{ie}|\theta, a, b) = \frac{1}{1+e^{-1.7a(\theta-b)}}$
    - MIRT
      - $P(y_{ie}|\boldsymbol{\theta}, \boldsymbol{a}, b) = \frac{1}{1+e^{-\boldsymbol{a}\boldsymbol{\theta}+b}}$
  - Concept level
    - DINA
      - $P(y_{ie}|\theta, g, s, \beta) = g^{1-\eta}(1-s)^{\eta}$
      - $\eta = \prod_k \theta_k^{\beta_k}$
      - $\theta_k \in \{0,1\}, \beta_k \in \{0,1\}$
- Deep learning methods
  - NeuralCD

General Form

$$P(y_{ie}|\boldsymbol{u}_i, \boldsymbol{v}_e) = f(\boldsymbol{u}_i, \boldsymbol{v}_e)$$

Extend traits dimension

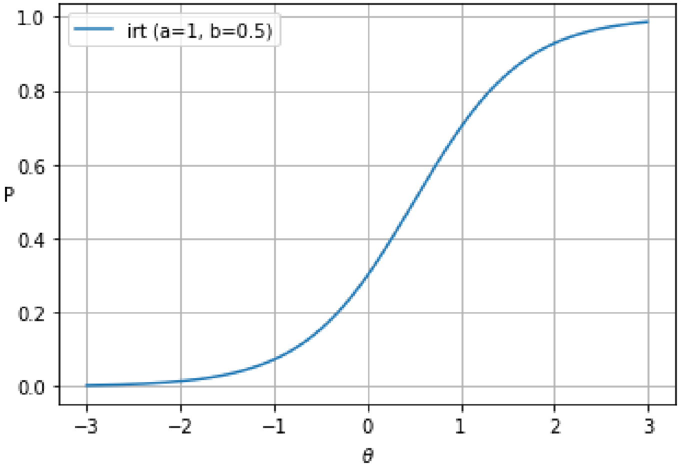


Figure 1: Example of IRT

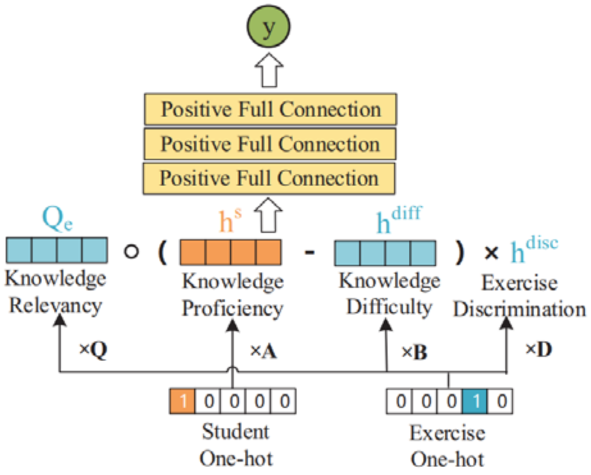


Figure 2: NeuralCD Framework

Related Work

IJCAI 2021 MONTREAL

- Traditional Methods
  - Item-level
    - IRT
      - $P(y_{ie}|\theta, a, b) = \frac{1}{1+e^{-1.7a(\theta-b)}}$
    - MIRT
      - $P(y_{ie}|\boldsymbol{\theta}, \boldsymbol{a}, b) = \frac{1}{1+e^{-\boldsymbol{a}\boldsymbol{\theta}+b}}$
  - Concept level
    - DINA
      - $P(y_{ie}|\theta, g, s, \beta) = g^{1-\eta}(1-s)^{\eta}$
      - $\eta = \prod_k \theta_k^{\beta_k}$
      - $\theta_k \in \{0,1\}, \beta_k \in \{0,1\}$
- Deep learning methods
  - NeuralCD

General Form

$$P(y_{ie}|\boldsymbol{u}_i, \boldsymbol{v}_e) = f(\boldsymbol{u}_i, \boldsymbol{v}_e)$$

f is human-designed

f is learned from data

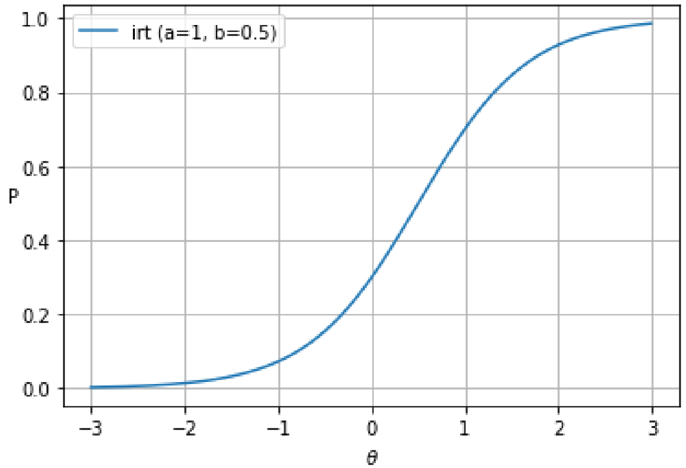


Figure 1: Example of IRT

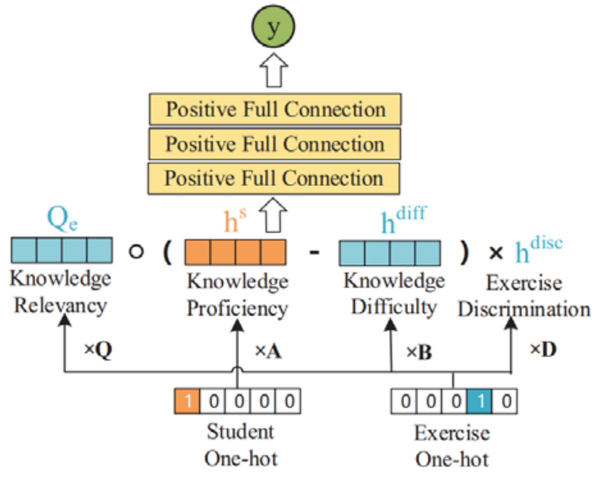


Figure 2: NeuralCD Framework

Key Question

IJCAI 2021 MONTREAL

General Form

$$P(y_{ie}|\boldsymbol{u}_i, \boldsymbol{v}_e) = f(\boldsymbol{u}_i, \boldsymbol{v}_e)$$

Objective Function

$$r_{ie} \leftarrow P(y_{ie})$$

Monotonicity

Previous methods apply a monotone function as the interaction function rather than considering it in the optimization criteria.

Question

Is it possible to create an optimization criteria to enhance the monotonicity?

Motivation

Exercise	Knowledge Concepts	Ada	Bob
e_1	A	✓	✗
e_2	B	✓	✓
e_3	A, B	✗	✗
e_4	C, D	?	✓
e_5	C, E	✗	?

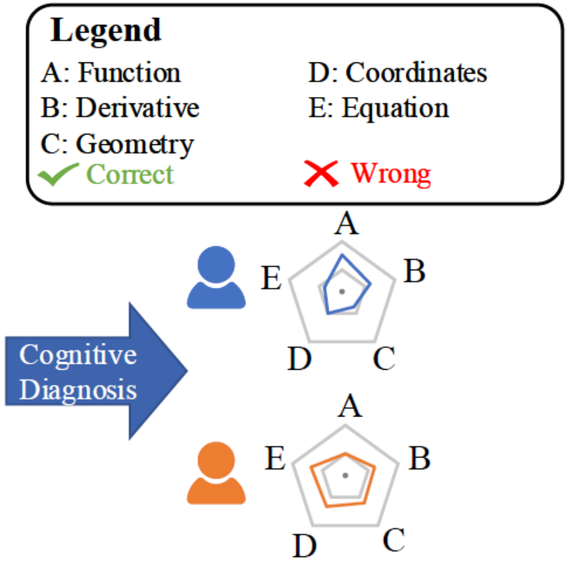


Figure 1: Illustration of cognitive diagnosis. The left part presents the response logs of students on test items, and the right-bottom part shows corresponding diagnostic results.

Motivation

Observation

Ada with a right response to item e_1 is considered as having a higher proficiency level on the related concept A (i.e., Function) than Bob with a wrong response.

Exercise	Knowledge Concepts	Ada	Bob
e_1	A	✓	✗
e_2	B	✓	✓
e_3	A, B	✗	✗
e_4	C, D	?	✓
e_5	C, E	✗	?

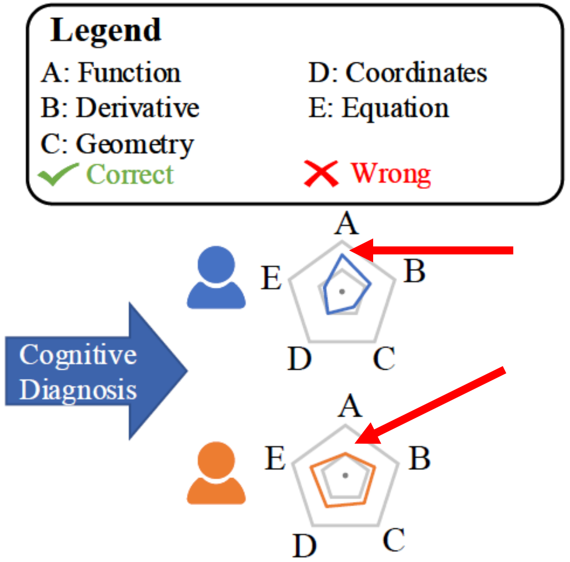


Figure 1: Illustration of cognitive diagnosis. The left part presents the response logs of students on test items, and the right-bottom part shows corresponding diagnostic results.

Motivation

IJCAI 2021 MONTREAL

Observation

Ada with a right response to item e_1 is considered as having a higher proficiency level on the related concept A (i.e., Function) than Bob with a wrong response.

Induction

Students with correct responses should be more proficient than students with wrong responses.

Exercise	Knowledge Concepts	Ada	Bob
e_1	A	✓	✗
e_2	B	✓	✓
e_3	A, B	✗	✗
e_4	C, D	?	✓
e_5	C, E	✗	?

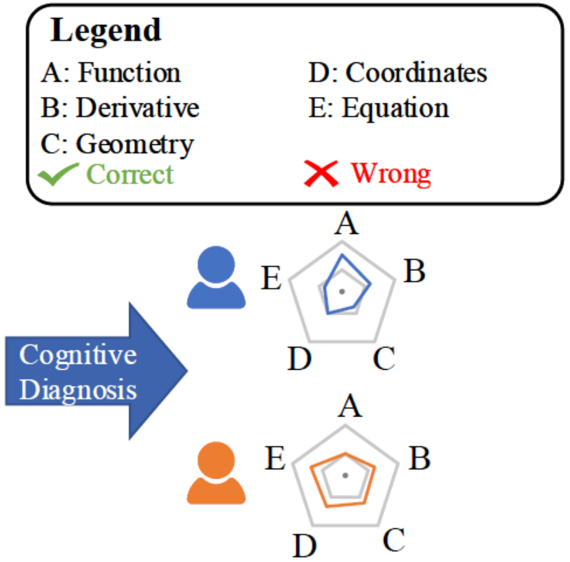


Figure 1: Illustration of cognitive diagnosis. The left part presents the response logs of students on test items, and the right-bottom part shows corresponding diagnostic results.

Motivation

IJCAI 2021 MONTREAL

Observation

Ada with a right response to item e_1 is considered as having a higher proficiency level on the related concept A (i.e., Function) than Bob with a wrong response.

Induction

Students with correct responses should be more proficient than students with wrong responses.

Motivation

Based on the partial order between item responses, we can create a new optimization criteria to exploit the response pairs to enhance the monotonicity.

Exercise	Knowledge Concepts	Ada	Bob
e_1	A	✓	✗
e_2	B	✓	✓
e_3	A, B	✗	✗
e_4	C, D	?	✓
e_5	C, E	✗	?

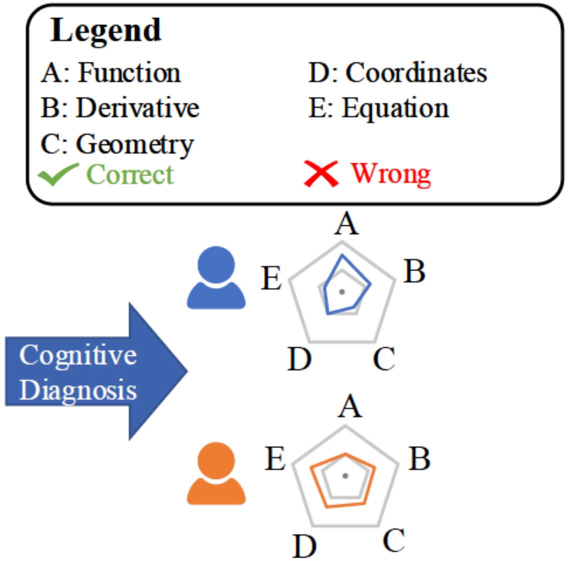


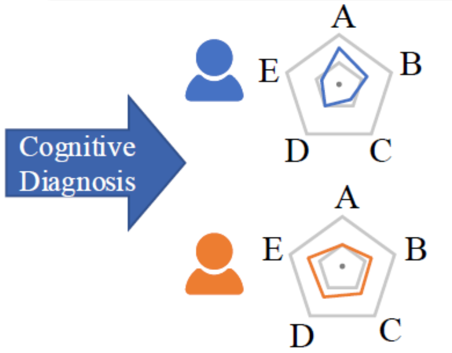
Figure 1: Illustration of cognitive diagnosis. The left part presents the response logs of students on test items, and the right-bottom part shows corresponding diagnostic results.

Challenge

- It is hard for us to compare item responses across different items related to non-overlapped concepts.

Exercise	Knowledge Concepts	Ada	Bob
e_1	A	✓	✗
e_2	B	✓	✓
e_3	A, B	✗	✗
e_4	C, D	?	✓
e_5	C, E	✗	?

Legend
A: Function
B: Derivative
C: Geometry
D: Coordinates
E: Equation
✓ Correct
✗ Wrong

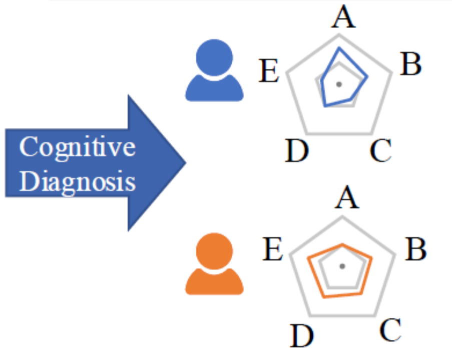


Challenge

- It is hard for us to compare item responses across different items related to non-overlapped concepts.

Exercise	Knowledge Concepts	Ada	Bob
e_1	A	✓	✗
e_2	B	✓	✓
e_3	A, B	✗	✗
e_4	C, D	?	✓
e_5	C, E	✗	?

Legend
A: Function D: Coordinates
B: Derivative E: Equation
C: Geometry
✓ Correct ✗ Wrong



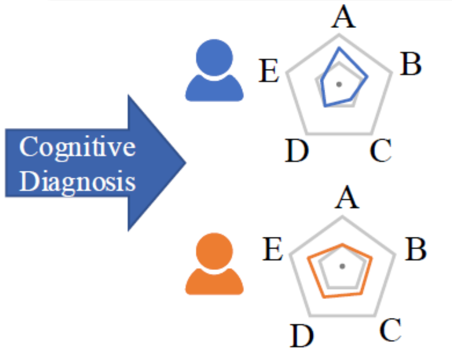
Challenge

- It is hard for us to compare item responses across different items related to non-overlapped concepts.

Exercise	Knowledge Concepts	Ada	Bob
e_1	A	✓	✗
e_2	B	✓	✓
e_3	A, B	✗	✗
e_4	C, D	?	✓
e_5	C, E	✗	?

Legend

A: Function D: Coordinates
B: Derivative E: Equation
C: Geometry
✓ Correct ✗ Wrong



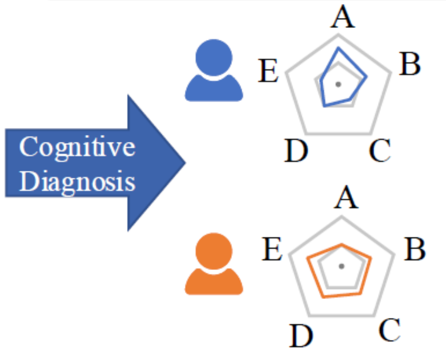
Challenge

IJCAI 2021 MONTREAL

- It is hard for us to compare item responses across different items related to non-overlapped concepts.
- There exist many unobserved responses.

Exercise	Knowledge Concepts	Ada	Bob
e_1	A	✓	✗
e_2	B	✓	✓
e_3	A, B	✗	✗
e_4	C, D	?	✓
e_5	C, E	✗	?

Legend
A: Function
B: Derivative
C: Geometry
D: Coordinates
E: Equation
✓ Correct
✗ Wrong

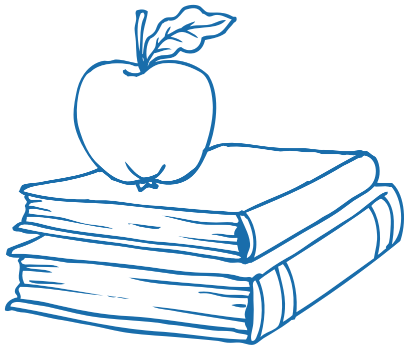


Challenge

- It is hard for us to compare item responses across different items related to non-overlapped concepts.
- There exist many unobserved responses.
- How to find an objective function so that the monotonicity can be directly optimized.

Outline

30th International Joint Conference on Artificial Intelligence



1. Introduction
2. **Problem Definition**
3. IRR Framework
4. Experiments
5. Conclusion

Problem Definition

IJCAI 2021 MONTREAL

Given

- N students: $S = \{s_1, s_2, \dots, s_N\}$
- M test items: $E = \{e_1, e_2, \dots e_M\}$
- L knowledge concepts: $K = \{k_1, k_2, \dots, k_L\}$
- Q-matrix $Q = \{Q_{ij}\}_{M \times L}$
- Response logs $\mathcal{R} = \{(s, e, r)\}$

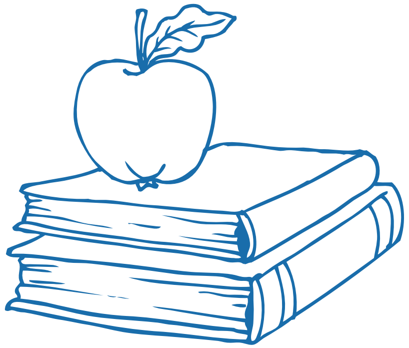
	k_1	k_2	k_3
e_1	1	0	0
e_2	0	1	1
e_3	1	1	1

An example of Q-matrix

Goal

Mine students' proficiency on knowledge concepts.

Outline



Monotonicity

IJCAI 2021 MONTREAL

Monotonicity

In the literature, the monotonicity theory assumes that student' s proficiency is monotonic with the probability of giving the right response to a test item.

$$r_{ie} \leftarrow P(y_{ie})$$

Monotonicity

IJCAI 2021 MONTREAL

Monotonicity

In the literature, the monotonicity theory assumes that student' s proficiency is monotonic with the probability of giving the right response to a test item.

$$r_{ie} \leftarrow P(y_{ie})$$

Pairwise Monotonicity

Given a specific test item, the students with right responses are more skilled than those with wrong responses.

Monotonicity

IJCAI 2021 MONTREAL

Monotonicity

In the literature, the monotonicity theory assumes that student' s proficiency is monotonic with the probability of giving the right response to a test item.

$$r_{ie} \leftarrow P(y_{ie})$$

Pairwise Monotonicity

Given a specific test item, the students with right responses are more skilled than those with wrong responses.

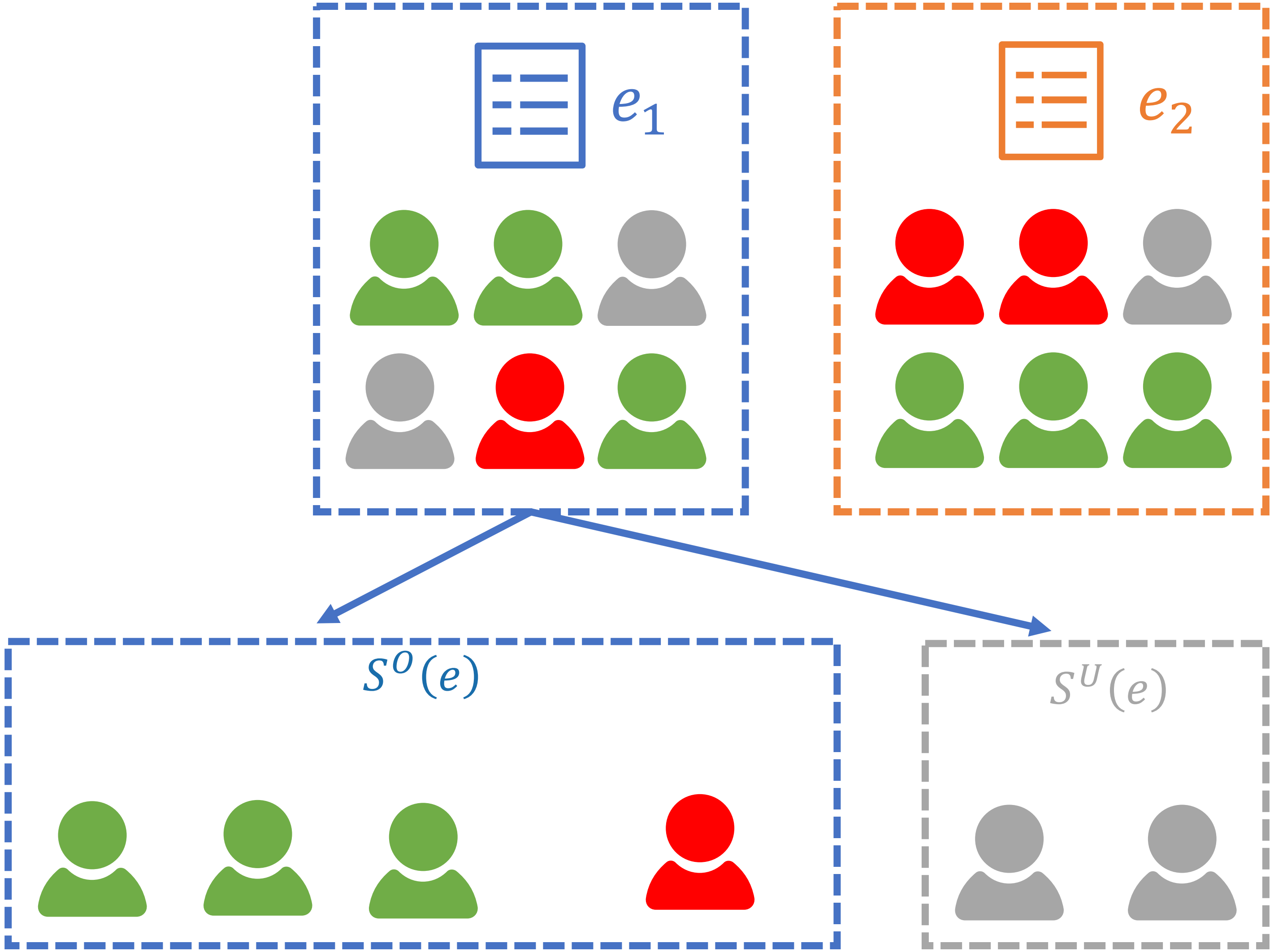
$$(r_{ie} - r_{je}) \leftarrow P(y_{ie} - y_{je})$$

Overview

- Objective
  - Traditional pointwise: $r_{ie} \leftarrow P(y_{ie})$
  - **Pairwise Monotonicity :** $\boldsymbol{(r_{ie} - r_{je}) \leftarrow P(y_{ie} - y_{je})}$
- Two issues
  - How to construct the response pairs
    - non-overlapped problem
    - unobserved responses
  - Objective function design

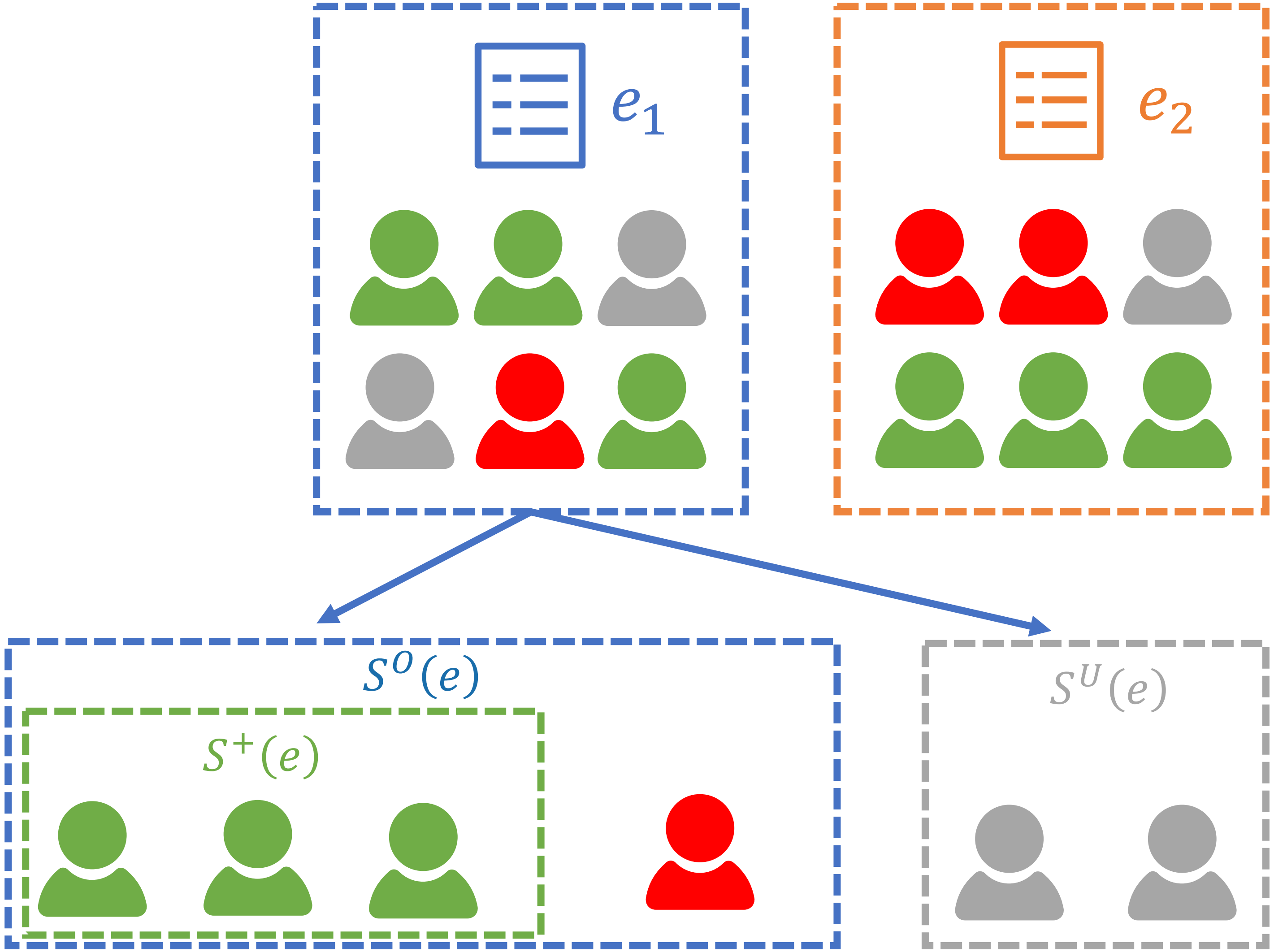
Pair Construction

- Item specific sampling for non-overlapped problem
  - Observed students $S^O(e)$
  - Unobserved students $S^U(e)$



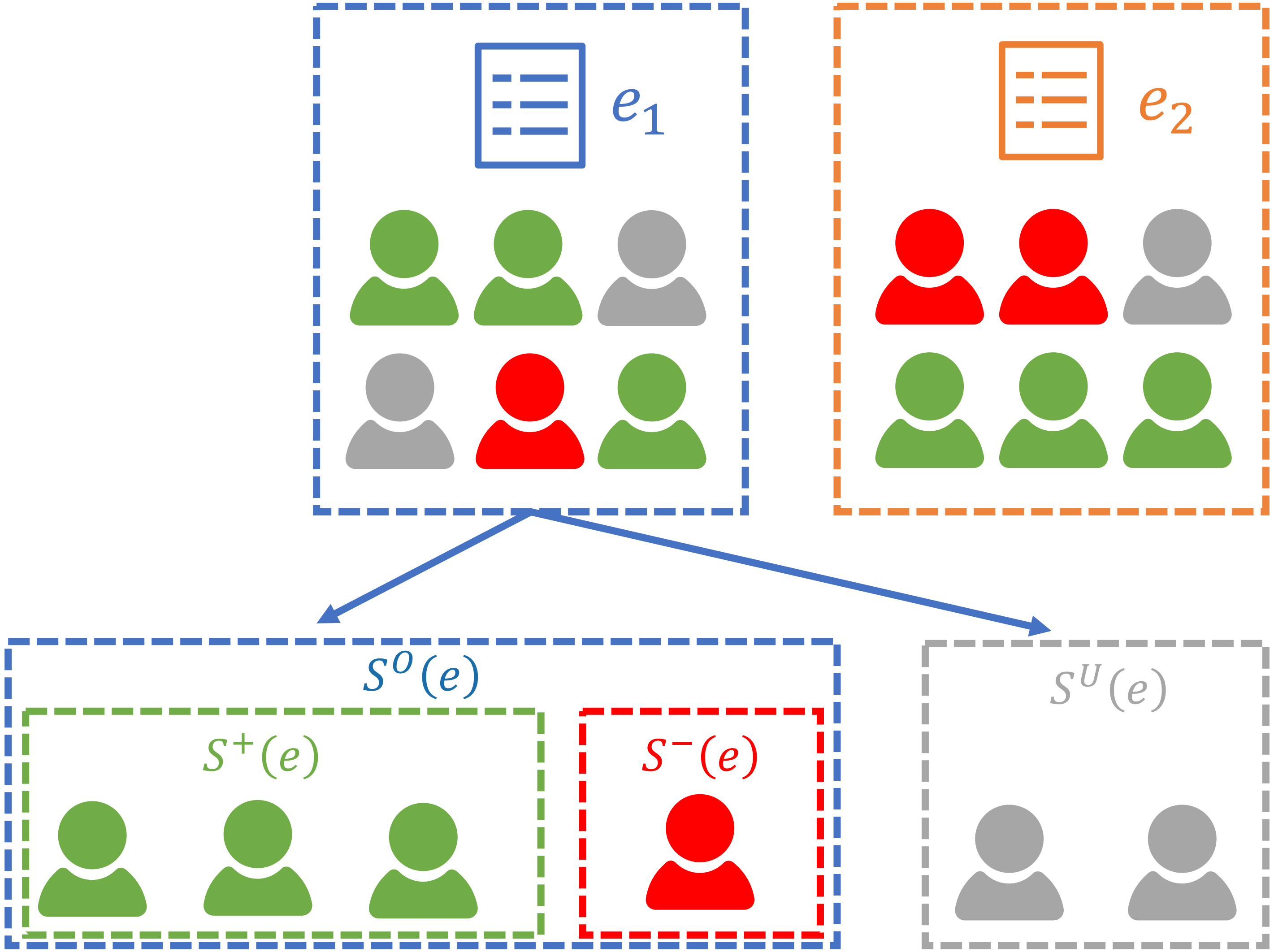
Pair Construction

- Item specific sampling for non-overlapped problem
  - Observed students $S^O(e)$
    - Positive students $S^+(e)$
    - Negative students $S^-(e)$
  - Unobserved students $S^U(e)$



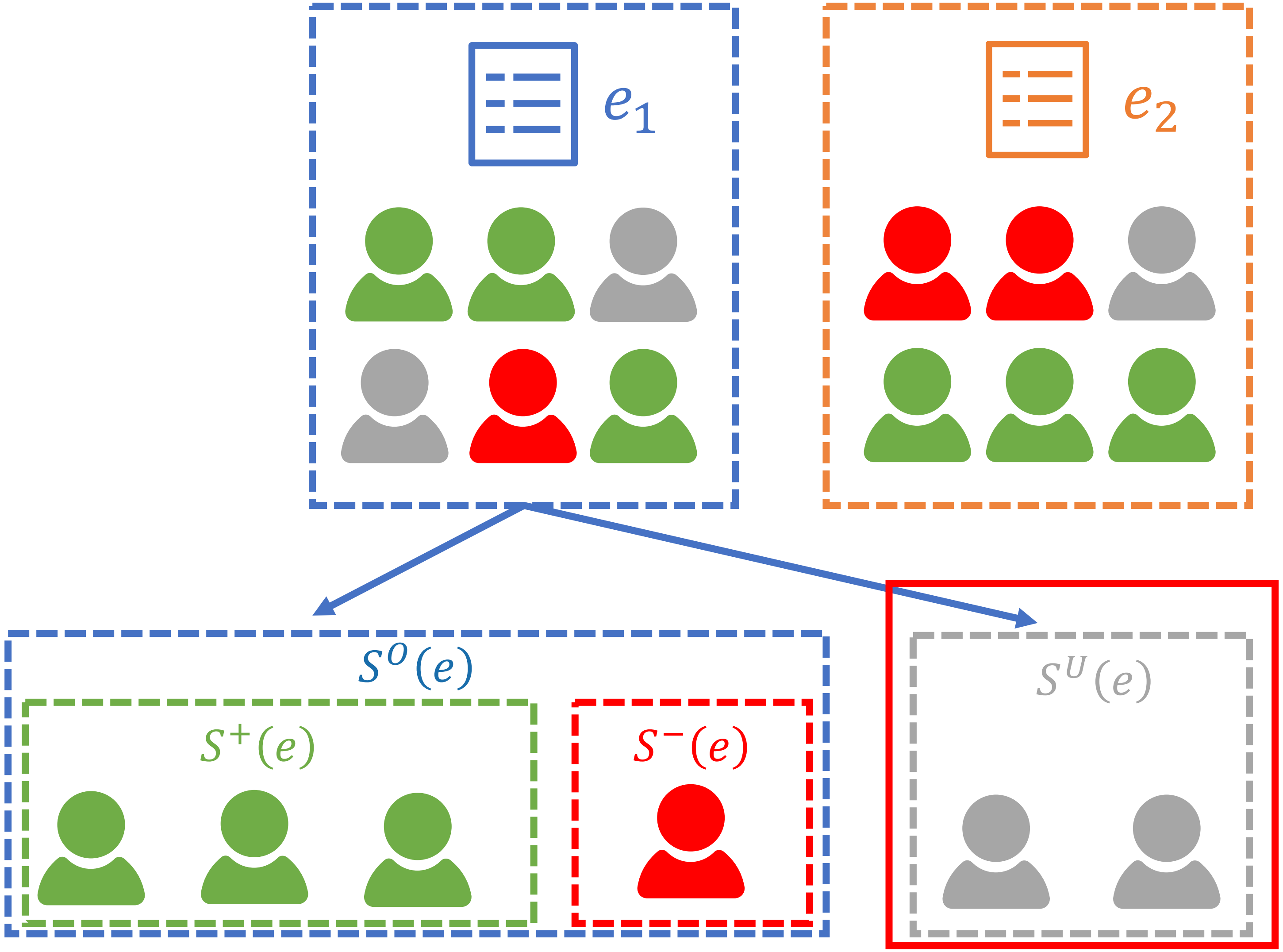
Pair Construction

- Item specific sampling for non-overlapped problem
  - Observed students $S^O(e)$
    - Positive students $S^+(e)$
    - Negative students $S^-(e)$
  - Unobserved students $S^U(e)$



Pair Construction

- Item specific sampling for non-overlapped problem
  - Observed students $S^O(e)$
    - Positive students $S^+(e)$
    - Negative students $S^-(e)$
  - Unobserved students $S^U(e)$



Pair Construction

- Item specific sampling for non-overlapped problem

- Two branch sampling for unobserved responses

	s_1	s_2	s_3	s_4
e_1	1	0	?	1
e_2	0	?	1	0
e_3	?	?	1	?
e_4	1	0	1	0
e_5	1	?	?	0

Student

Exercise

? as negative (0)

Pair Construction

- Item specific sampling for non-overlapped problem

- Two branch sampling for unobserved responses

Exercise \ Student	s_1	s_2	s_3	s_4
e_1	1	0	?	1
e_2	0	?	1	0
e_3	?	?	1	?
e_4	1	0	1	0
e_5	1	?	?	0

$$\rightarrow \quad 0 \le P(r_{s^-e}) \le P(r_{ue}) \le P(r_{s^+e}\}) \le 1$$

Pair Construction

- Item specific sampling for non-overlapped problem

- Two branch sampling for unobserved responses

	s_1	s_2	s_3	s_4
e_1	1	0	?	1
e_2	0	?	1	0
e_3	?	?	1	?
e_4	1	0	1	0
e_5	1	?	?	0

Student

Exercise

Positive sampling

$$P(r_{ue}) \geq P(r_{s^-e}) \geq 0,$$

$s^- \in S^-(e)$? → 1

$$0 \leq P(r_{s^-e}) \leq P(r_{ue}) \leq P(r_{s^+e}\}) \leq 1$$

Pair Construction

- Item specific sampling for non-overlapped problem
- Two branch sampling for unobserved responses

	s_1	s_2	s_3	s_4
e_1	1	0	?	1
e_2	0	?	1	0
e_3	?	?	1	?
e_4	1	0	1	0
e_5	1	?	?	0

Student

Exercise

Positive sampling

$$P(r_{ue}) \geq P(r_{s^-e}) \geq 0,$$

$s^- \in S^-(e)$ $\uparrow$? → 1

$$0 \leq P(r_{s^-e}) \leq P(r_{ue}) \leq P(r_{s^+e}\}) \leq 1$$

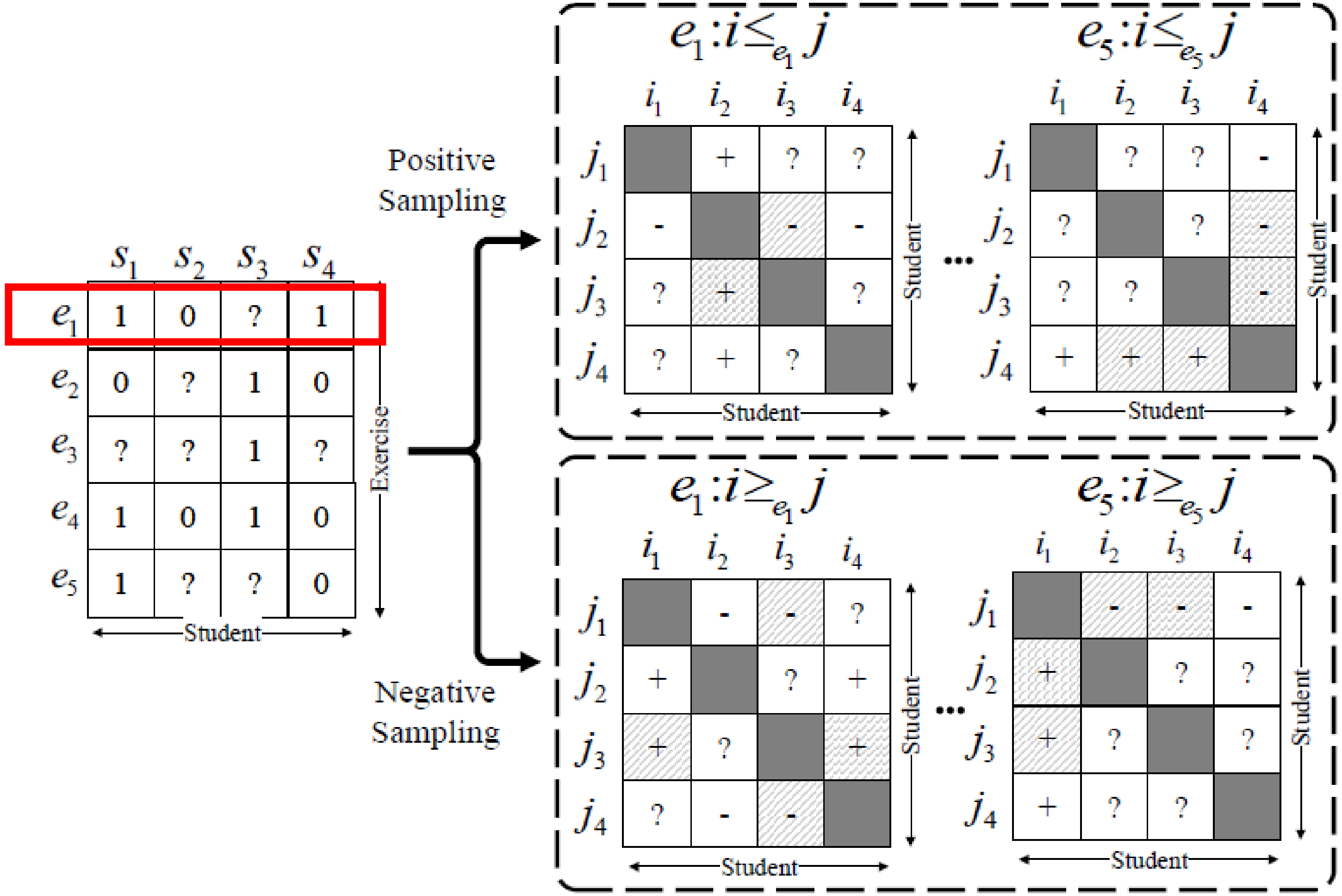
$s^+ \in S^+(e)$ $\downarrow$? → 0

$$P(r_{ue}) \leq P(r_{s^+e}) \leq 1$$

Negative sampling

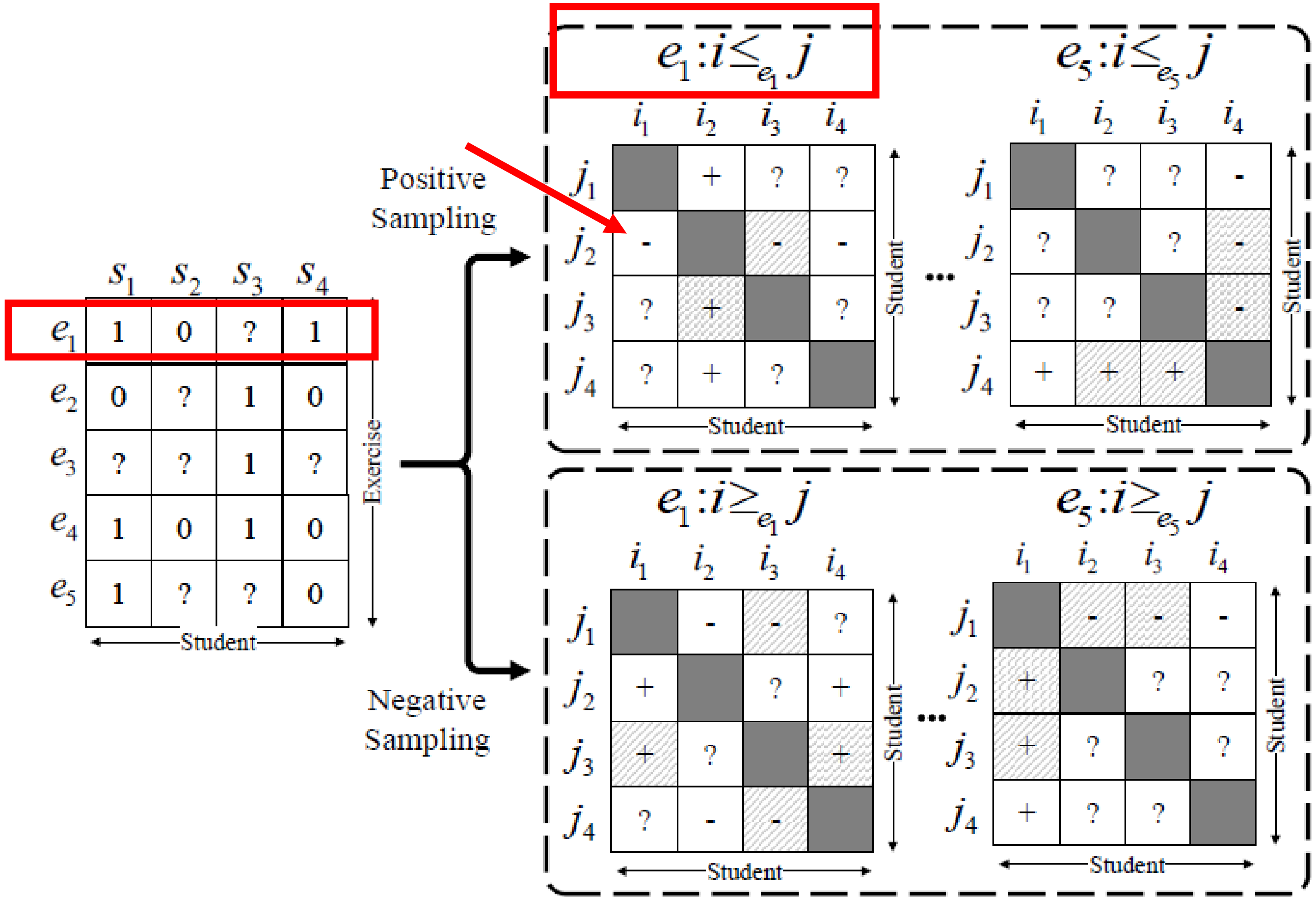
Pair Construction

- Item specific sampling for non-overlapped problem
- Two branch sampling for unobserved responses



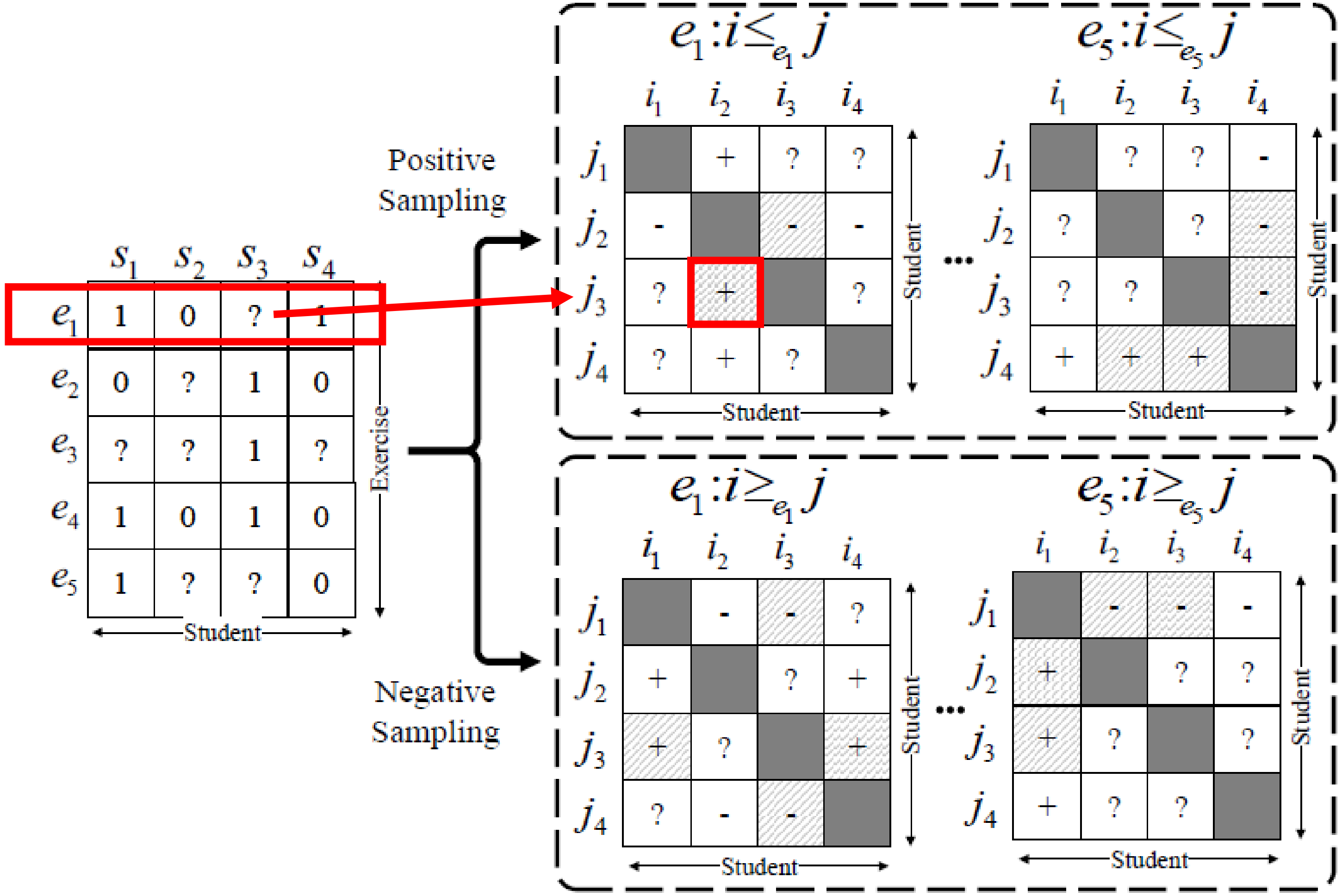
Pair Construction

- Item specific sampling for non-overlapped problem
- Two branch sampling for unobserved responses



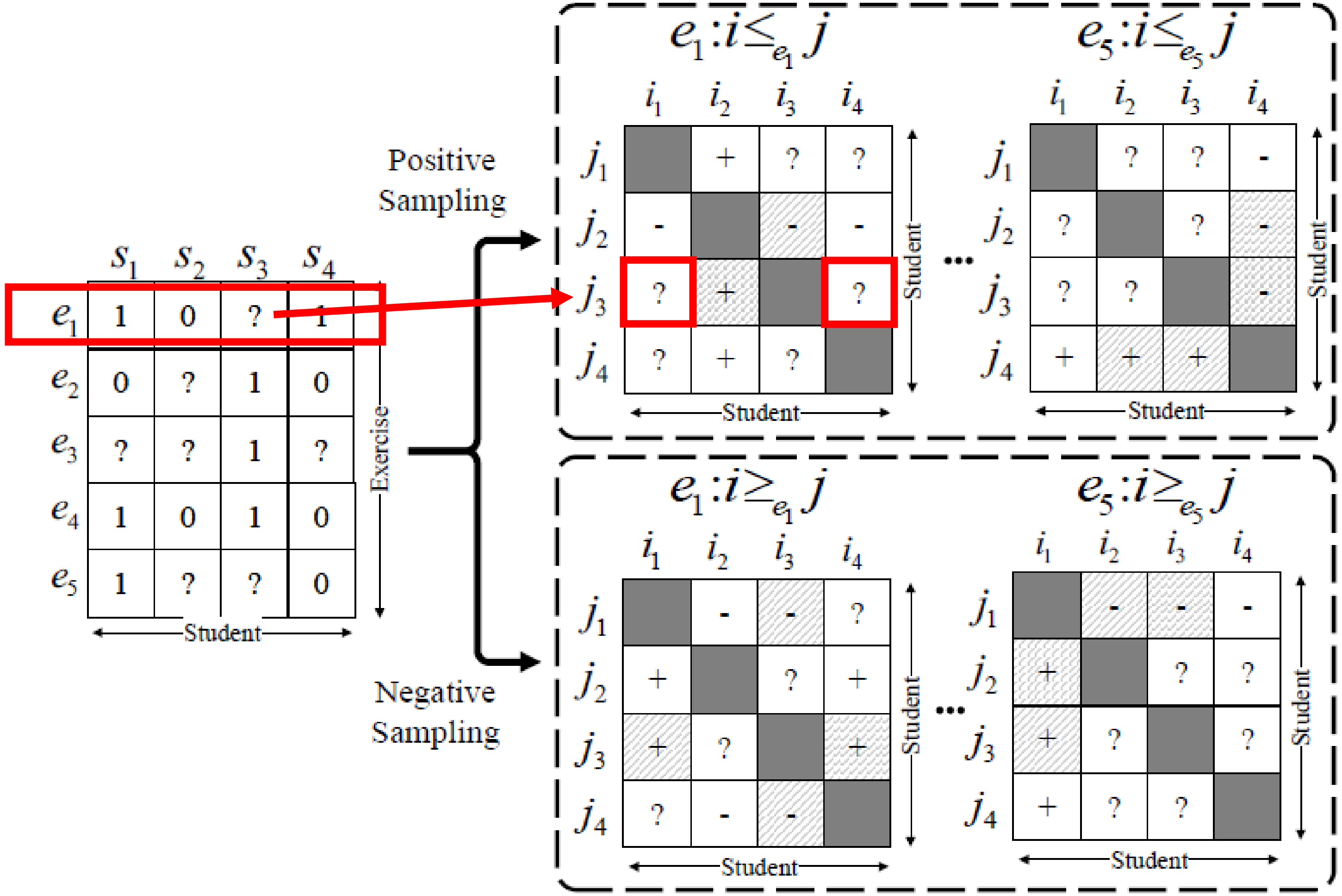
Pair Construction

- Item specific sampling for non-overlapped problem
- Two branch sampling for unobserved responses



Pair Construction

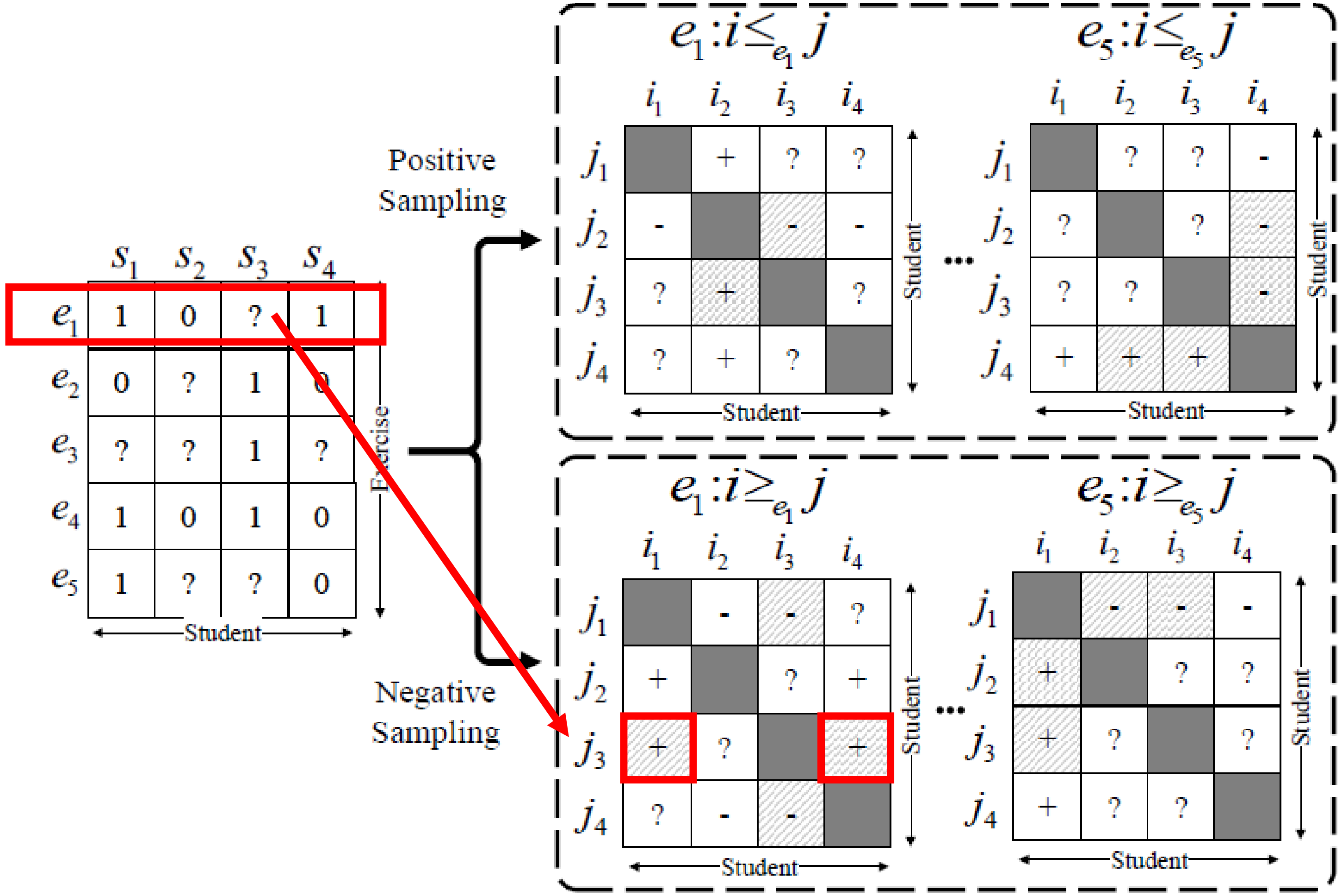
- Item specific sampling for non-overlapped problem
- Two branch sampling for unobserved responses



Pair Construction

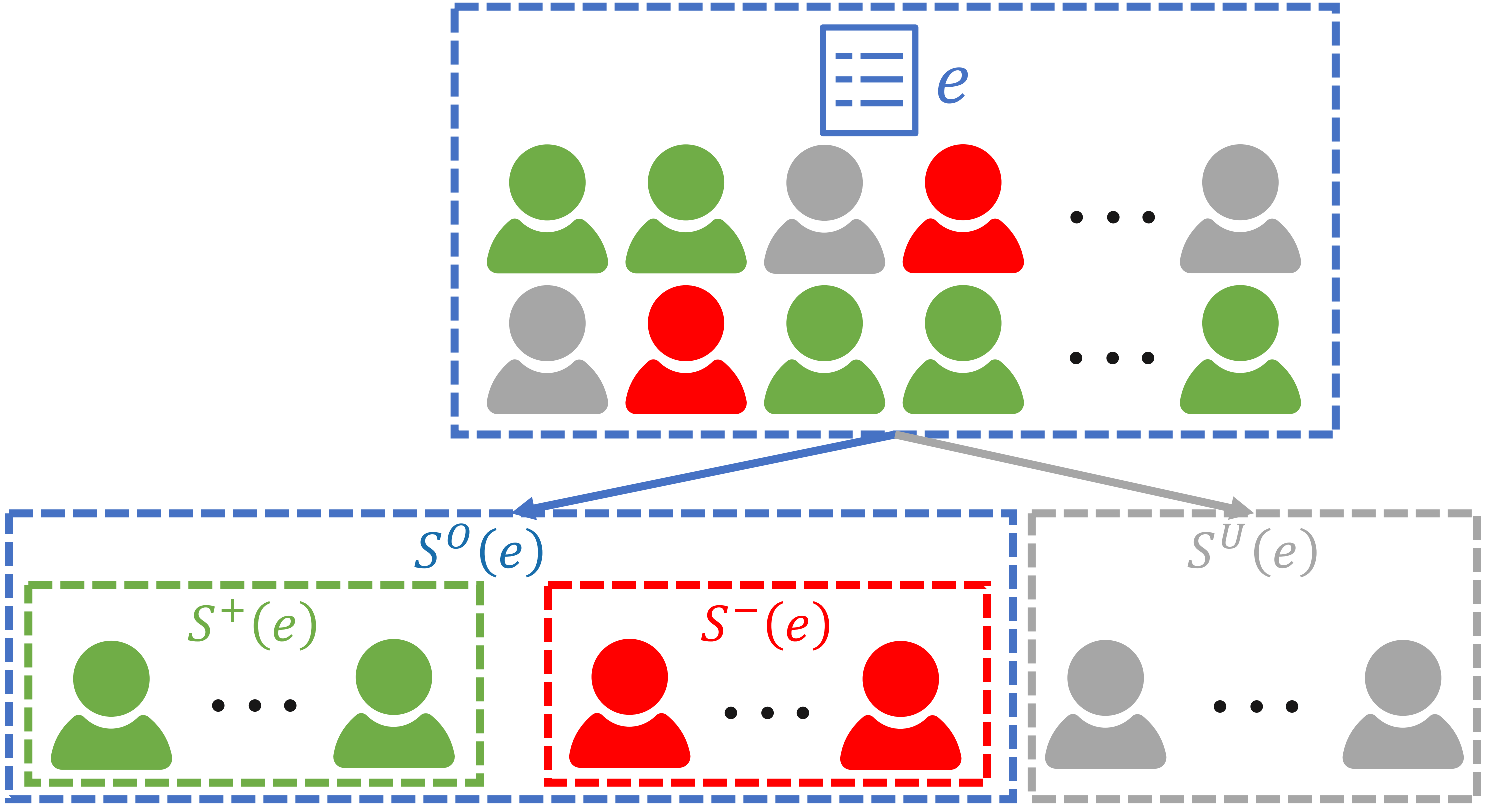
- Item specific sampling for non-overlapped problem

- Two branch sampling for unobserved responses



Pair Construction

- Item specific sampling for non-overlapped problem
- Two branch sampling for unobserved responses
- Training samples $T(R)$



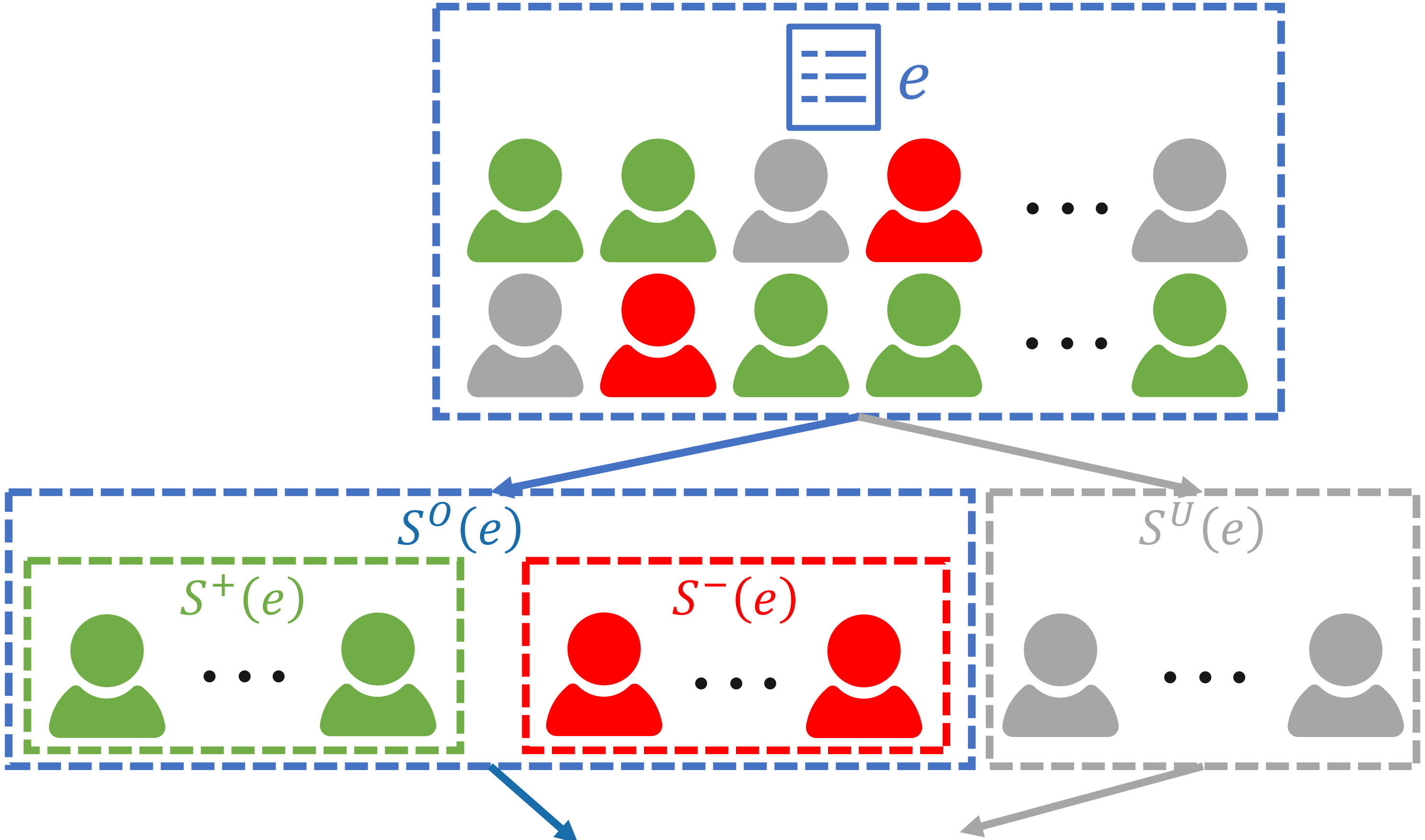
$$|S^+(e) \times (S - S^+(e))| + |S^-(e) \times (S - S^-(e))|$$

Pair Construction

- Item specific sampling for non-overlapped problem

- Two branch sampling for unobserved responses

- Training samples $T(R)$



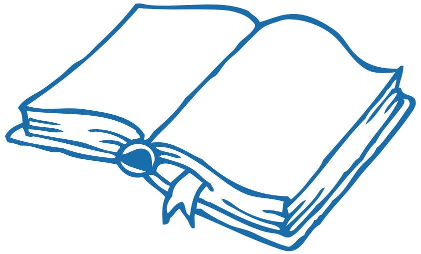
$$T(R) = \begin{cases} \{(s, s') | s' \in \underset{N^O}{\Lambda}(S^-(e)) \oplus \underset{N^U}{\Lambda}(S^U(e))\} & r_{se} = 1, \\ \{(s', s) | s' \in \underset{N^O}{\Lambda}(S^+(e)) \oplus \underset{N^U}{\Lambda}(S^U(e))\} & r_{se} = 0, \end{cases}$$

Objective Function

- Log-likelihood
  - $\ln IRR = \ln IRR^{+} + \ln IRR^{-}$
  - $IRR^{+} = \prod_{e \in E} \prod_{i \in S^{+}(e)} \prod_{j \in S - S^{+}(e)} P\left(r_{ie} \geq r_{je}\right) \times \left(1 - P\left(r_{je} \geq r_{ie}\right)\right)$
  - $IRR^{-} = \prod_{e \in E} \prod_{i \in S^{-}(e)} \prod_{j \in S - S^{-}(e)} P\left(r_{ie} \leq r_{je}\right) \times \left(1 - P\left(r_{je} \leq r_{ie}\right)\right)$
- Loss function
  - $\mathcal{L} = -\sum_{(i,j) \in T(R)} \log \frac{\exp(P(r_{ie}|\Theta))}{\exp(P(r_{ie}|\Theta)) + \exp\left(P(r_{je}|\Theta)\right)} + \lambda(\Theta)$
  - $\lambda(\Theta)$ is the regularization term

Outline

30th International Joint Conference on Artificial Intelligence



1 Introduction

2 Problem Definition

3 IRR Framework

4 **Experiments**

5 Conclusion

Experimental Setup

IJCAI 2021 MONTREAL

- Dataset
  - ASSISTments
  - MATH
- Divide the students on each test item into training : test = 8:2
  - 10% of training used for hyper-parameter adjustment
- Hyper-parameters
  - $\lambda \leftarrow \{0.1, 0.01, 0.001, 0.0001\}$
    - $\lambda = 0.0001$
  - $N^o, N^U \leftarrow \{1, 5, 10, 30\}$
    - $N^O = N^U = 10$
- Baselines
  - IRT, MIRT, DINA, NeuralCD

Statistics	ASSISTments	MATH
# users	4,163	10,268
# items	17,746	917,495
# knowledge concepts	123	1,488
# response logs	324,572	864,722

Table 1: The statistics of the dataset.

Parameter Setting

Dimension of latent traits equal to the number of concepts

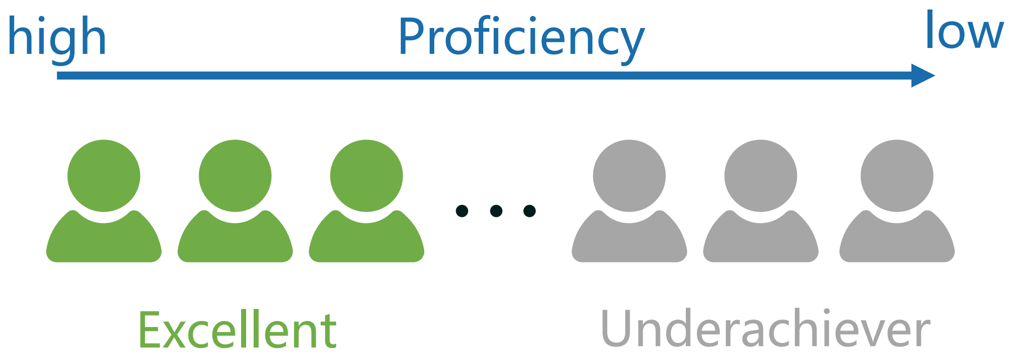
- 123 in ASSISTments
- 1488 in MATH

Evaluation Metrics

- Classification Metrics
  - AUC
  - Precision
  - Recall
  - F1

Evaluation Metrics

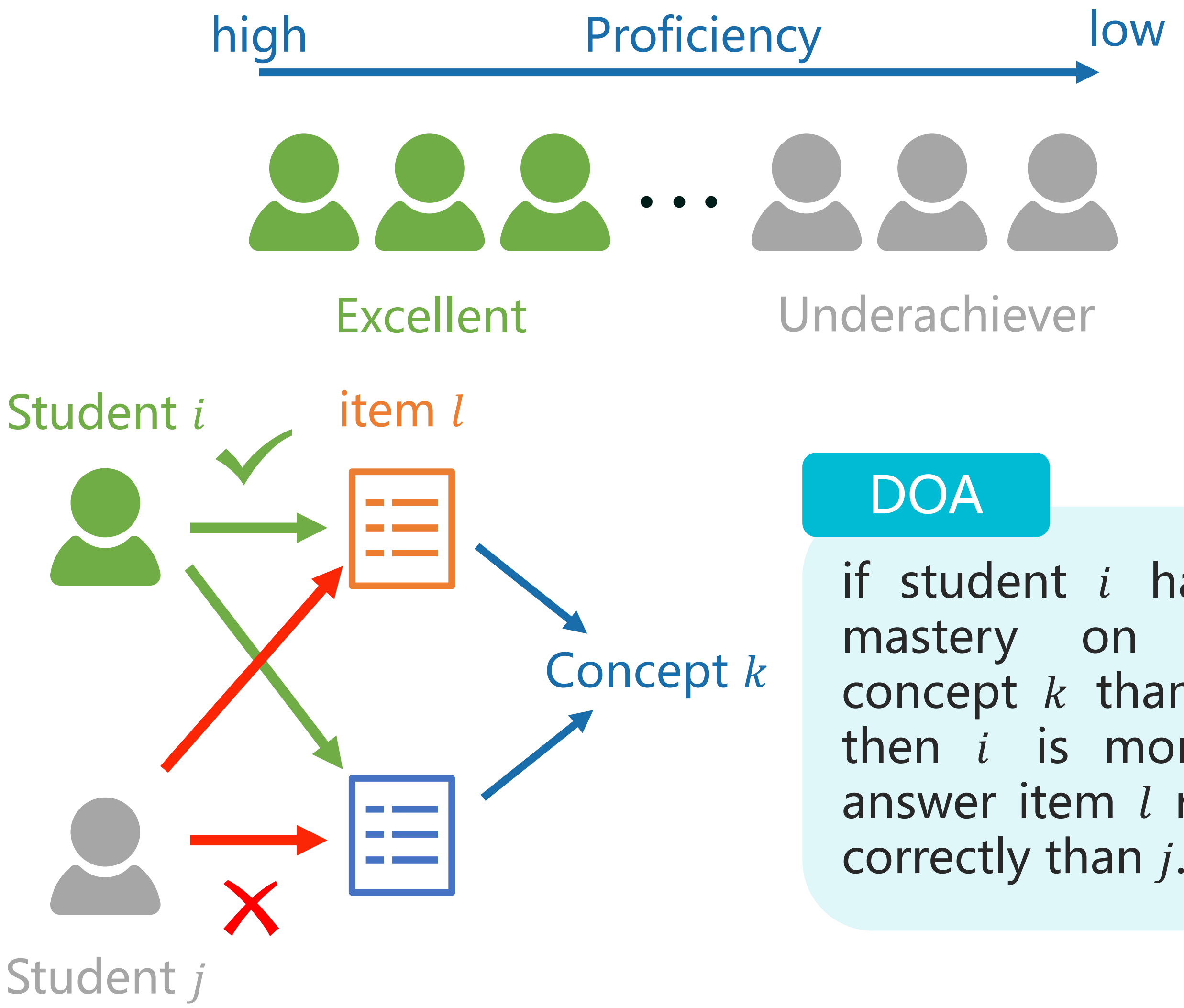
- Classification Metrics
  - AUC
  - Precision
  - Recall
  - F1
- Ranking Metrics
  - MAP(E), NDCG(E)
  - MAP(U), NDCG(U)



Evaluation Metrics

IJCAI 2021 MONTREAL

- Classification Metrics
  - AUC
  - Precision
  - Recall
  - F1
- Ranking Metrics
  - MAP(E), NDCG(E)
  - MAP(U), NDCG(U)
- Monotonicity Metrics
  - DOA



DOA

if student i has a better mastery on knowledge concept k than student j, then i is more likely to answer item l related to k correctly than j.

Student Performance Ranking

IJCAI 2021 MONTREAL

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.786	**0.815**	0.776	**0.848**	0.777	**0.889**	0.817	**0.839**
	Precision	0.689	**0.713**	0.665	**0.752**	0.716	**0.785**	0.706	**0.734**
	Recall	0.485	**0.553**	0.585	**0.607**	0.737	**0.645**	**0.662**	0.579
	F1	0.518	**0.598**	0.534	**0.654**	0.666	**0.697**	0.625	**0.626**
Ranking	MAP(E)	0.840	**0.853**	0.817	**0.913**	0.824	**0.937**	0.867	**0.907**
	NDCG@5(E)	0.871	**0.881**	0.867	**0.894**	0.868	**0.912**	0.881	**0.889**
	MAP(U)	0.780	**0.799**	0.754	**0.880**	0.761	**0.908**	0.826	**0.875**
	NDCG@5(U)	0.464	**0.481**	0.462	**0.503**	0.462	**0.524**	0.487	**0.500**

(a) ASSISTments

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.549	**0.567**	0.537	**0.666**	0.537	**0.686**	0.596	**0.608**
	Precision	0.717	**0.725**	0.637	**0.760**	0.583	**0.767**	0.686	**0.737**
	Recall	0.447	**0.707**	0.722	**0.743**	0.699	**0.750**	0.655	**0.719**
	F1	0.500	**0.711**	0.524	**0.747**	0.588	**0.754**	0.602	**0.724**
Ranking	MAP(E)	0.735	**0.740**	0.728	**0.808**	0.729	**0.814**	0.765	**0.772**
	NDCG@5(E)	0.852	**0.855**	0.845	**0.906**	0.844	**0.909**	0.853	**0.866**
	MAP(U)	0.356	**0.360**	0.345	**0.494**	0.347	**0.508**	0.431	**0.437**
	NDCG@5(U)	0.488	**0.498**	0.479	**0.654**	0.477	**0.677**	**0.587**	0.569

(b) MATH

Student Performance Ranking

IJCAI 2021 MONTREAL

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.786	**0.815**	0.776	**0.848**	0.777	**0.889**	0.817	**0.839**
	Precision	0.689	**0.713**	0.665	**0.752**	0.716	**0.785**	0.706	**0.734**
	Recall	0.485	**0.553**	0.585	**0.607**	0.737	**0.645**	**0.662**	0.579
	F1	0.518	**0.598**	0.534	**0.654**	0.666	**0.697**	0.625	**0.626**
Ranking	MAP(E)	0.840	**0.853**	0.817	**0.913**	0.824	**0.937**	0.867	**0.907**
	NDCG@5(E)	0.871	**0.881**	0.867	**0.894**	0.868	**0.912**	0.881	**0.889**
	MAP(U)	0.780	**0.799**	0.754	**0.880**	0.761	**0.908**	0.826	**0.875**
	NDCG@5(U)	0.464	**0.481**	0.462	**0.503**	0.462	**0.524**	0.487	**0.500**

(a) ASSISTments

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.549	**0.567**	0.537	**0.666**	0.537	**0.686**	0.596	**0.608**
	Precision	0.717	**0.725**	0.637	**0.760**	0.583	**0.767**	0.686	**0.737**
	Recall	0.447	**0.707**	0.722	**0.743**	0.699	**0.750**	0.655	**0.719**
	F1	0.500	**0.711**	0.524	**0.747**	0.588	**0.754**	0.602	**0.724**
Ranking	MAP(E)	0.735	**0.740**	0.728	**0.808**	0.729	**0.814**	0.765	**0.772**
	NDCG@5(E)	0.852	**0.855**	0.845	**0.906**	0.844	**0.909**	0.853	**0.866**
	MAP(U)	0.356	**0.360**	0.345	**0.494**	0.347	**0.508**	0.431	**0.437**
	NDCG@5(U)	0.488	**0.498**	0.479	**0.654**	0.477	**0.677**	**0.587**	0.569

(b) MATH

Student Performance Ranking

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.786	**0.815**	0.776	**0.848**	0.777	**0.889**	0.817	**0.839**
	Precision	0.689	**0.713**	0.665	**0.752**	0.716	**0.785**	0.706	**0.734**
	Recall	0.485	**0.553**	0.585	**0.607**	0.737	**0.645**	**0.662**	0.579
	F1	0.518	**0.598**	0.534	**0.654**	0.666	**0.697**	0.625	**0.626**
Ranking	MAP(E)	0.840	**0.853**	0.817	**0.913**	0.824	**0.937**	0.867	**0.907**
	NDCG@5(E)	0.871	**0.881**	0.867	**0.894**	0.868	**0.912**	0.881	**0.889**
	MAP(U)	0.780	**0.799**	0.754	**0.880**	0.761	**0.908**	0.826	**0.875**
	NDCG@5(U)	0.464	**0.481**	0.462	**0.503**	0.462	**0.524**	0.487	**0.500**

(a) ASSISTments

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.549	**0.567**	0.537	**0.666**	0.537	**0.686**	0.596	**0.608**
	Precision	0.717	**0.725**	0.637	**0.760**	0.583	**0.767**	0.686	**0.737**
	Recall	0.447	**0.707**	0.722	**0.743**	0.699	**0.750**	0.655	**0.719**
	F1	0.500	**0.711**	0.524	**0.747**	0.588	**0.754**	0.602	**0.724**
Ranking	MAP(E)	0.735	**0.740**	0.728	**0.808**	0.729	**0.814**	0.765	**0.772**
	NDCG@5(E)	0.852	**0.855**	0.845	**0.906**	0.844	**0.909**	0.853	**0.866**
	MAP(U)	0.356	**0.360**	0.345	**0.494**	0.347	**0.508**	0.431	**0.437**
	NDCG@5(U)	0.488	**0.498**	0.479	**0.654**	0.477	**0.677**	**0.587**	0.569

(b) MATH

Student Performance Ranking

- For every CDM, the IRR-CDM significantly outperform the pointwise-CDM on classification metrics on all datasets.

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.786	**0.815**	0.776	**0.848**	0.777	**0.889**	0.817	**0.839**
	Precision	0.689	**0.713**	0.665	**0.752**	0.716	**0.785**	0.706	**0.734**
	Recall	0.485	**0.553**	0.585	**0.607**	0.737	**0.645**	**0.662**	0.579
	F1	0.518	**0.598**	0.534	**0.654**	0.666	**0.697**	0.625	**0.626**
Ranking	MAP(E)	0.840	**0.853**	0.817	**0.913**	0.824	**0.937**	0.867	**0.907**
	NDCG@5(E)	0.871	**0.881**	0.867	**0.894**	0.868	**0.912**	0.881	**0.889**
	MAP(U)	0.780	**0.799**	0.754	**0.880**	0.761	**0.908**	0.826	**0.875**
	NDCG@5(U)	0.464	**0.481**	0.462	**0.503**	0.462	**0.524**	0.487	**0.500**

(a) ASSISTments

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.549	**0.567**	0.537	**0.666**	0.537	**0.686**	0.596	**0.608**
	Precision	0.717	**0.725**	0.637	**0.760**	0.583	**0.767**	0.686	**0.737**
	Recall	0.447	**0.707**	0.722	**0.743**	0.699	**0.750**	0.655	**0.719**
	F1	0.500	**0.711**	0.524	**0.747**	0.588	**0.754**	0.602	**0.724**
Ranking	MAP(E)	0.735	**0.740**	0.728	**0.808**	0.729	**0.814**	0.765	**0.772**
	NDCG@5(E)	0.852	**0.855**	0.845	**0.906**	0.844	**0.909**	0.853	**0.866**
	MAP(U)	0.356	**0.360**	0.345	**0.494**	0.347	**0.508**	0.431	**0.437**
	NDCG@5(U)	0.488	**0.498**	0.479	**0.654**	0.477	**0.677**	**0.587**	0.569

(b) MATH

Student Performance Ranking

- For every CDM, the IRR-CDM significantly outperform the pointwise-CDM on classification metrics on all datasets.
- IRR-CDM also achieve higher scores in ranking metrics.

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.786	**0.815**	0.776	**0.848**	0.777	**0.889**	0.817	**0.839**
	Precision	0.689	**0.713**	0.665	**0.752**	0.716	**0.785**	0.706	**0.734**
	Recall	0.485	**0.553**	0.585	**0.607**	0.737	**0.645**	**0.662**	0.579
	F1	0.518	**0.598**	0.534	**0.654**	0.666	**0.697**	0.625	**0.626**
Ranking	MAP(E)	0.840	**0.853**	0.817	**0.913**	0.824	**0.937**	0.867	**0.907**
	NDCG@5(E)	0.871	**0.881**	0.867	**0.894**	0.868	**0.912**	0.881	**0.889**
	MAP(U)	0.780	**0.799**	0.754	**0.880**	0.761	**0.908**	0.826	**0.875**
	NDCG@5(U)	0.464	**0.481**	0.462	**0.503**	0.462	**0.524**	0.487	**0.500**

(a) ASSISTments

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.549	**0.567**	0.537	**0.666**	0.537	**0.686**	0.596	**0.608**
	Precision	0.717	**0.725**	0.637	**0.760**	0.583	**0.767**	0.686	**0.737**
	Recall	0.447	**0.707**	0.722	**0.743**	0.699	**0.750**	0.655	**0.719**
	F1	0.500	**0.711**	0.524	**0.747**	0.588	**0.754**	0.602	**0.724**
Ranking	MAP(E)	0.735	**0.740**	0.728	**0.808**	0.729	**0.814**	0.765	**0.772**
	NDCG@5(E)	0.852	**0.855**	0.845	**0.906**	0.844	**0.909**	0.853	**0.866**
	MAP(U)	0.356	**0.360**	0.345	**0.494**	0.347	**0.508**	0.431	**0.437**
	NDCG@5(U)	0.488	**0.498**	0.479	**0.654**	0.477	**0.677**	**0.587**	0.569

(b) MATH

Student Performance Ranking

IJCAI 2021 MONTREAL

- For every CDM, the IRR-CDM significantly outperform the pointwise-CDM on classification metrics on all datasets.
- IRR-CDM also achieve higher scores in ranking metrics.

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.786	**0.815**	0.776	**0.848**	0.777	**0.889**	0.817	**0.839**
	Precision	0.689	**0.713**	0.665	**0.752**	0.716	**0.785**	0.706	**0.734**
	Recall	0.485	**0.553**	0.585	**0.607**	0.737	**0.645**	**0.662**	0.579
	F1	0.518	**0.598**	0.534	**0.654**	0.666	**0.697**	0.625	**0.626**
Ranking	MAP(E)	0.840	**0.853**	0.817	**0.913**	0.824	**0.937**	0.867	**0.907**
	NDCG@5(E)	0.871	**0.881**	0.867	**0.894**	0.868	**0.912**	0.881	**0.889**
	MAP(U)	0.780	**0.799**	0.754	**0.880**	0.761	**0.908**	0.826	**0.875**
	NDCG@5(U)	0.464	**0.481**	0.462	**0.503**	0.462	**0.524**	0.487	**0.500**

(a) ASSISTments

Metrics		DINA		IRT		MIRT		NeuralCD	
		Pointwise	IRR	Pointwise	IRR	Pointwise	IRR	Pointwise	IRR
Classification	AUC	0.549	**0.567**	0.537	**0.666**	0.537	**0.686**	0.596	**0.608**
	Precision	0.717	**0.725**	0.637	**0.760**	0.583	**0.767**	0.686	**0.737**
	Recall	0.447	**0.707**	0.722	**0.743**	0.699	**0.750**	0.655	**0.719**
	F1	0.500	**0.711**	0.524	**0.747**	0.588	**0.754**	0.602	**0.724**
Ranking	MAP(E)	0.735	**0.740**	0.728	**0.808**	0.729	**0.814**	0.765	**0.772**
	NDCG@5(E)	0.852	**0.855**	0.845	**0.906**	0.844	**0.909**	0.853	**0.866**
	MAP(U)	0.356	**0.360**	0.345	**0.494**	0.347	**0.508**	0.431	**0.437**
	NDCG@5(U)	0.488	**0.498**	0.479	**0.654**	0.477	**0.677**	**0.587**	0.569

(b) MATH

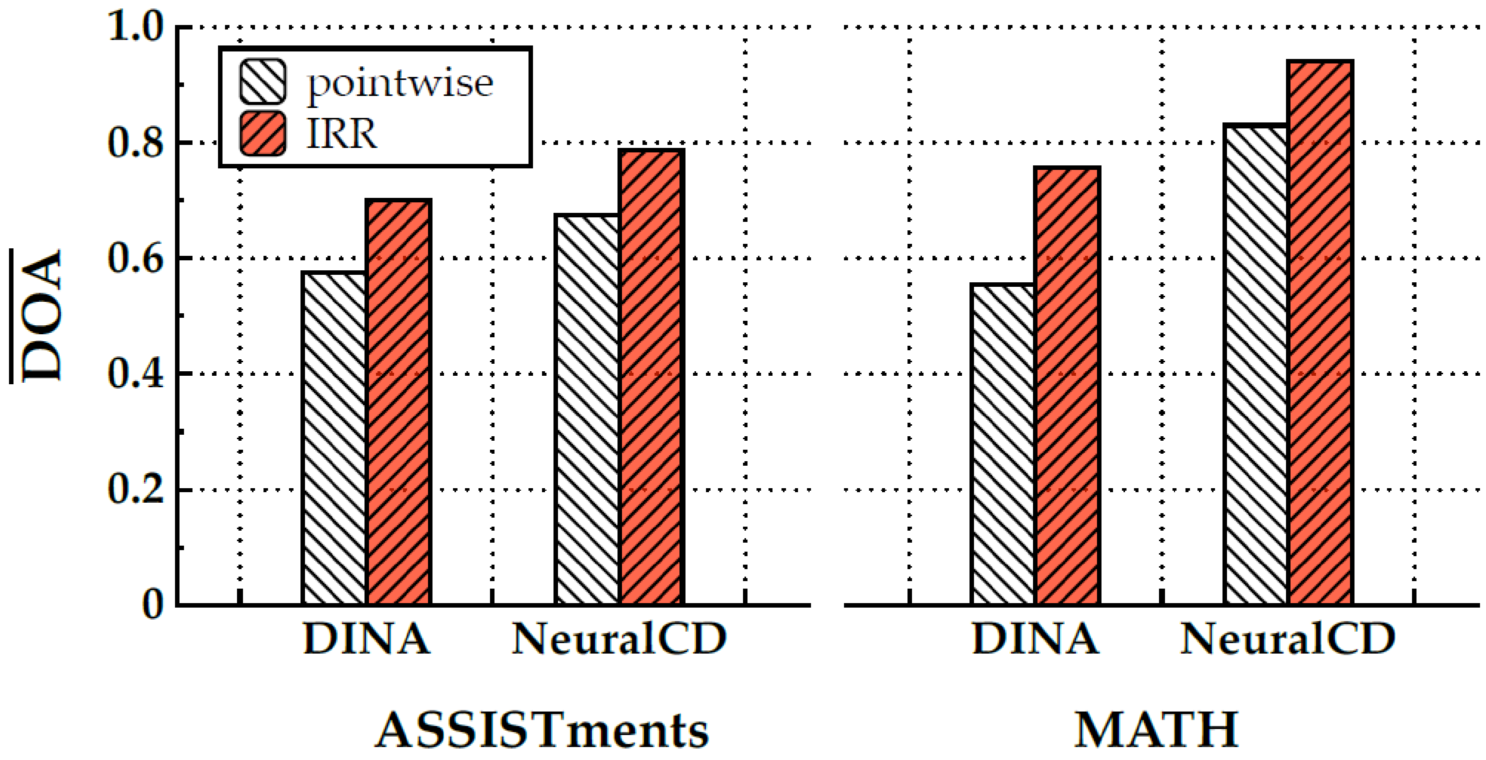
Conclusion

By exploiting the partial order between responses, IRR can not only help CDMs **promote the diagnosis precision** but also **better maintain the monotonicity**.

Knowledge Proficiency Monotonicity

IJCAI 2021 MONTREAL

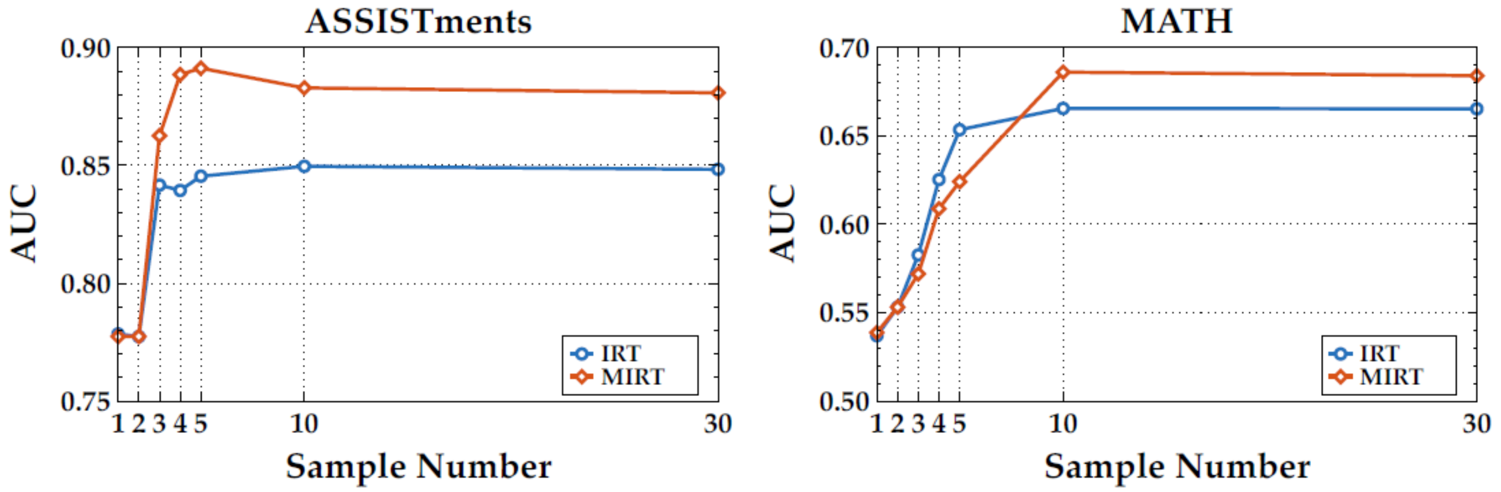
- DOAs of IRR-CDMs are higher than pointwise-CDMs



Conclusion

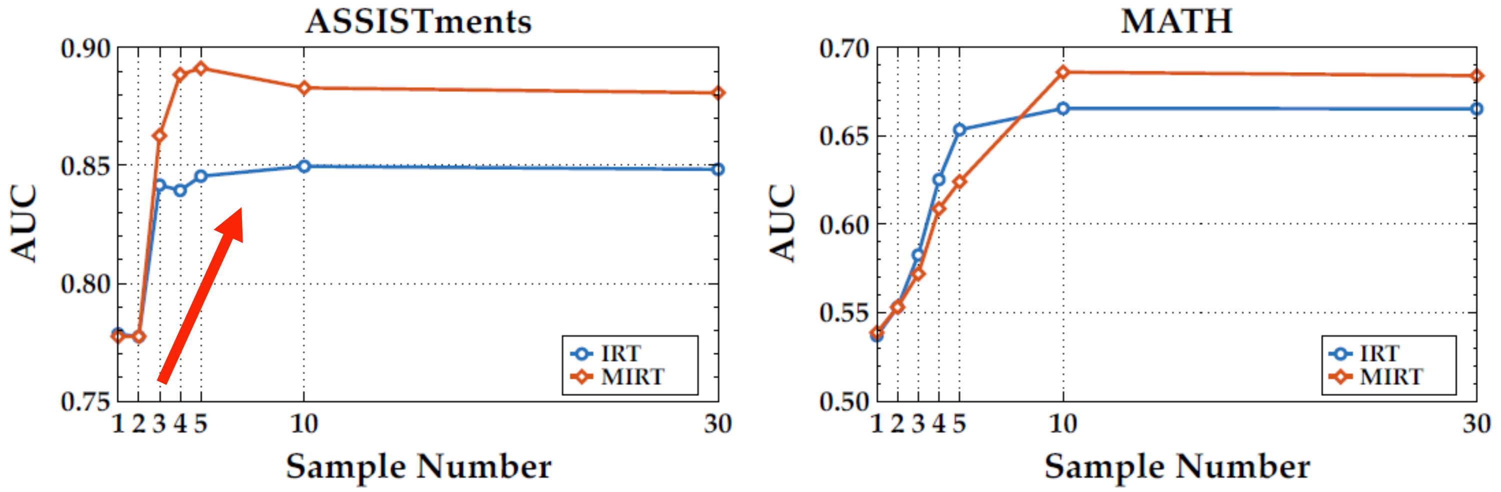
Knowledge proficiency diagnosed by IRR is **more monotonic**.

Different Sample Number



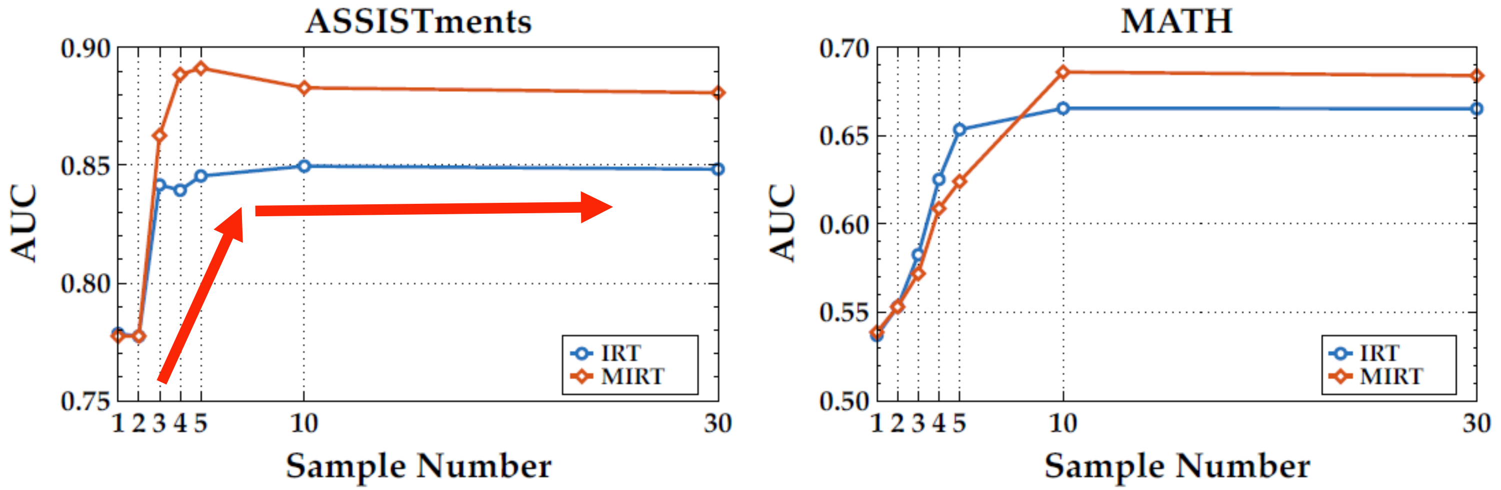
Different Sample Number

- As the sample number increases, the performance of IRR-CDMs
  - increases at the beginning



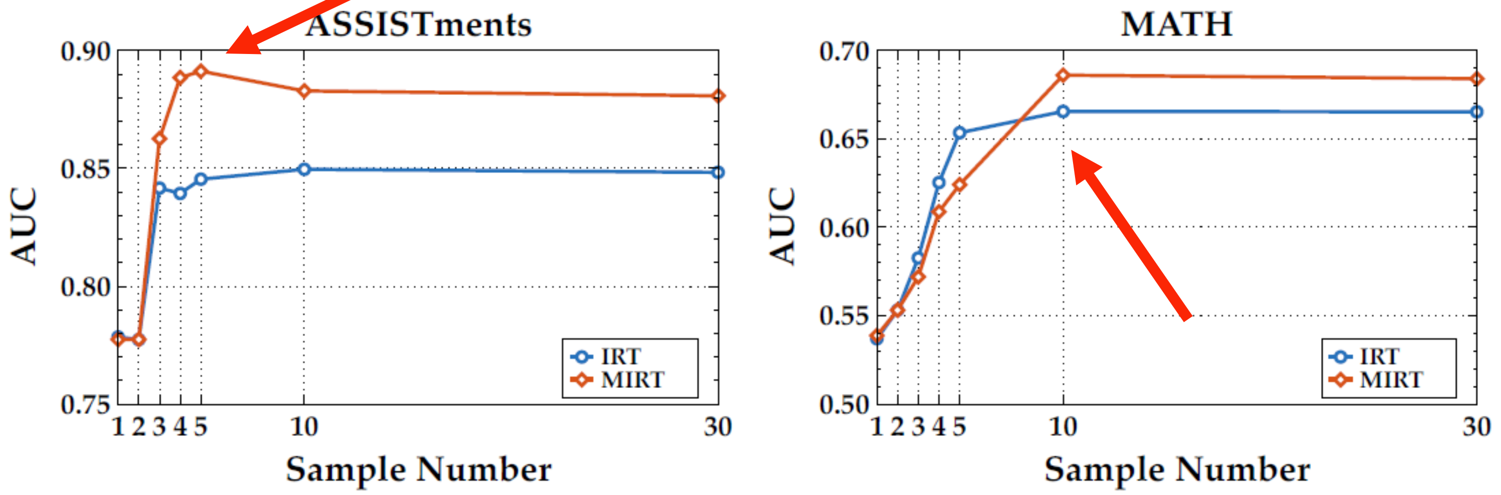
Different Sample Number

- As the sample number increases, the performance of IRR-CDMs
  - increases at the beginning
  - converges afterwards



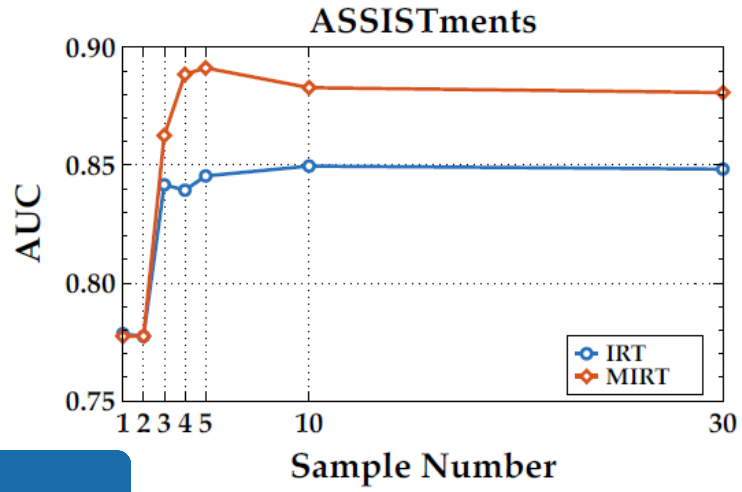
Different Sample Number

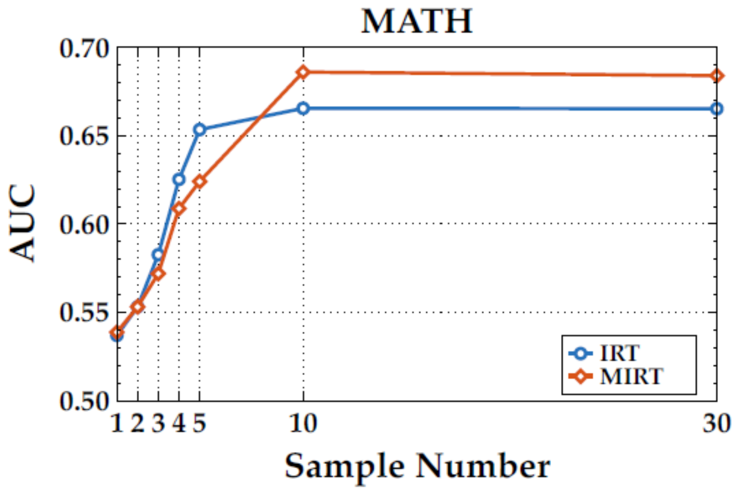
- As the sample number increases, the performance of IRR-CDMs
  - increases at the beginning
  - converges afterwards
- Compared with MATH, the performance of IRR-CDMs on ASSISTments converges earlier.



Different Sample Number

- As the sample number increases, the performance of IRR-CDMs
  - increases at the beginning
  - converges afterwards
- Compared with MATH, the performance of IRR-CDMs on ASSISTments converges earlier.





Conclusion

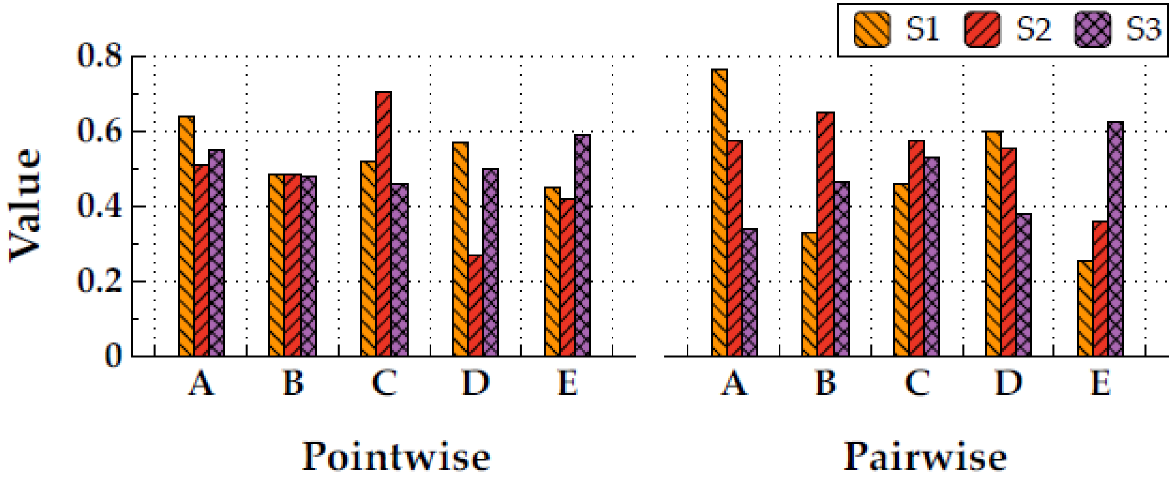
Including **more sampled pairs** can help the IRR-CDMs **promote the performance**, but the promotion has a **upper bound** due to the converge of information increment.

Diagnostic Results Analysis

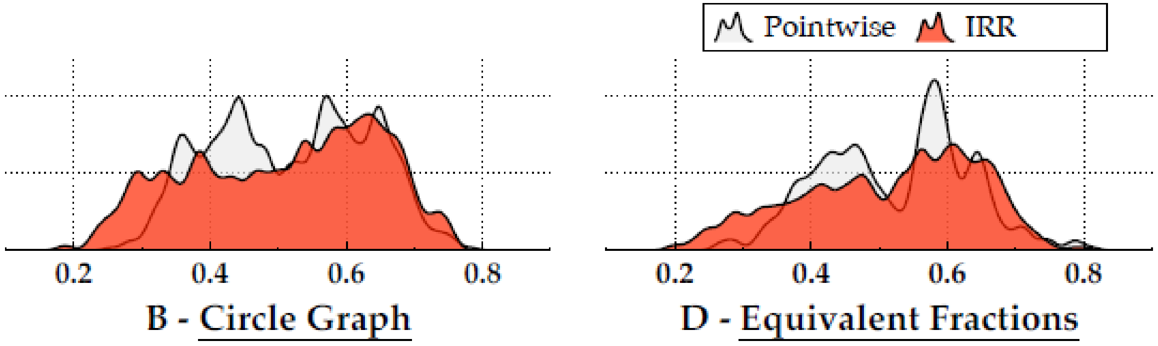
ID	Concepts	Student 1	Student 2	Student 3
Item 1	B, C	✗	✔	✗
Item 2	B, C, D, E	✗	✔	✔
Item 3	A, B, D	✔	✗	✗

A - Proportion B - Circle Graph C - Percent Of
D - Equivalent Fractions E - Finding Percents

(a) Response Logs diagram.



(b) Proficiency on knowledge concepts digram.



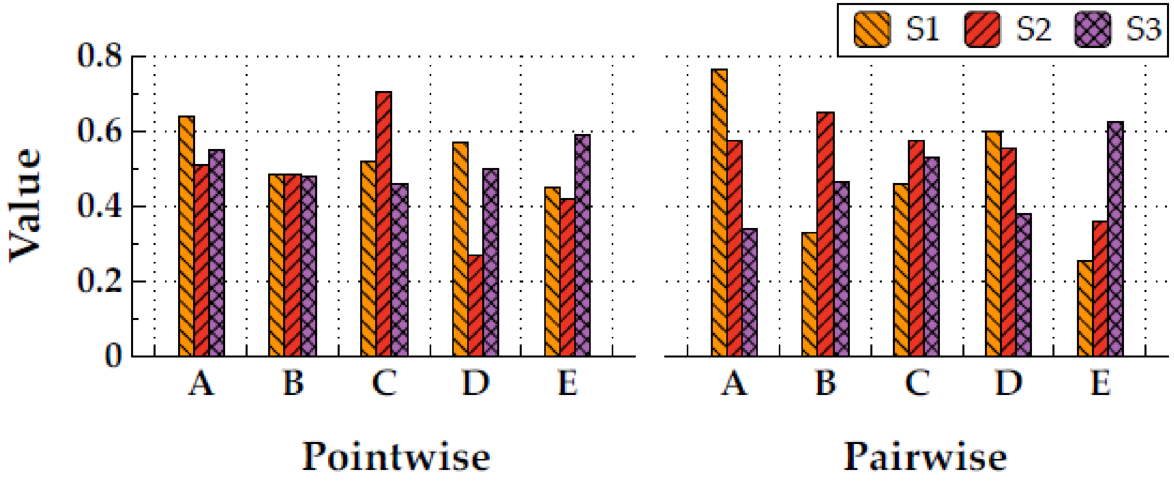
(c) Proficiency distribution digram.

Diagnostic Results Analysis

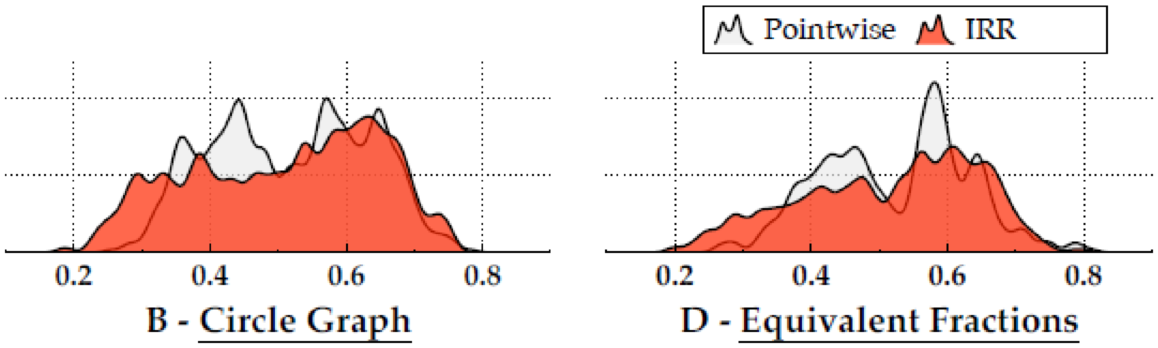
ID	Concepts	Student 1	Student 2	Student 3
Item 1	B, C	✗	✓	✗
Item 2	B, C, D, E	✗	✓	✓
Item 3	A, B, D	✓	✗	✗

A - Proportion B - Circle Graph C - Percent Of
D - Equivalent Fractions E - Finding Percents

(a) Response Logs diagram.



(b) Proficiency on knowledge concepts digram.



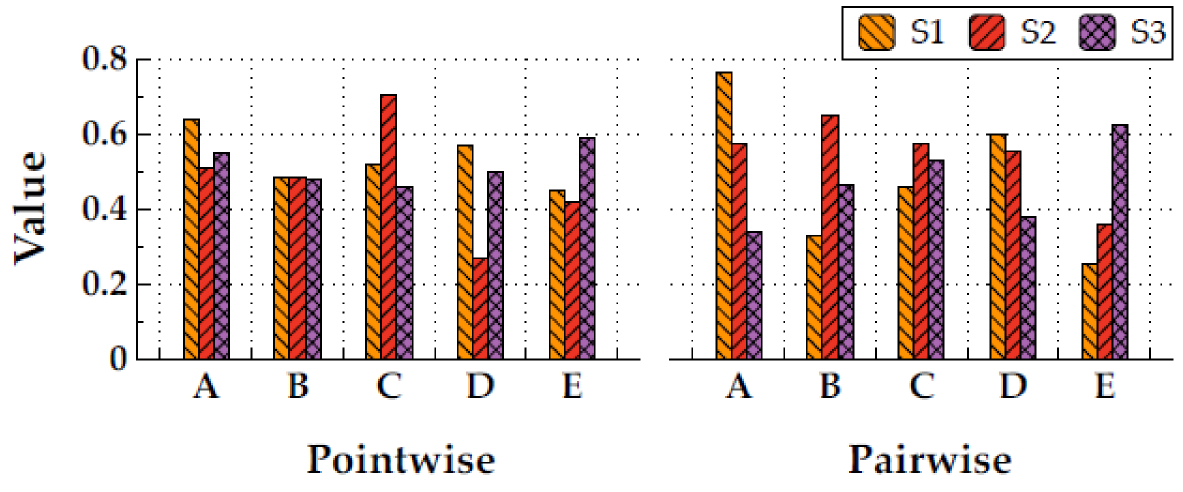
(c) Proficiency distribution digram.

Diagnostic Results Analysis

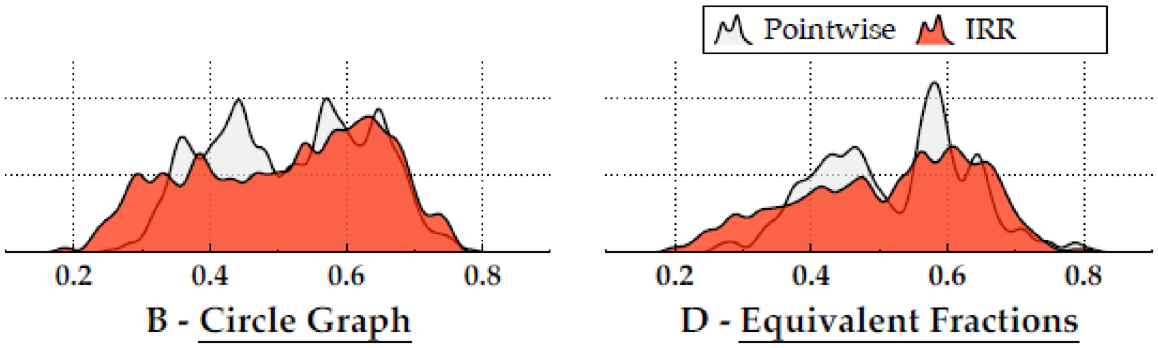
ID	Concepts	Student 1	Student 2	Student 3
Item 1	B, C	✘	✔	✘
Item 2	B, C, D, E	✘	✔	✔
Item 3	A, B, D	✔	✘	✘

A - Proportion B - Circle Graph C - Percent Of
D - Equivalent Fractions E - Finding Percents

(a) Response Logs diagram.



(b) Proficiency on knowledge concepts digram.



(c) Proficiency distribution digram.

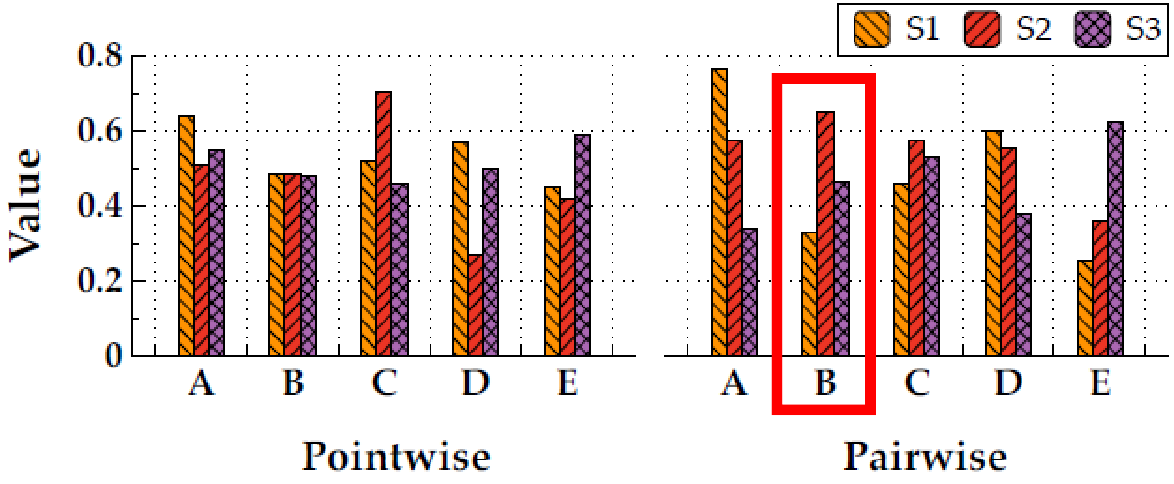
Diagnostic Results Analysis

IJCAI 2021 MONTREAL

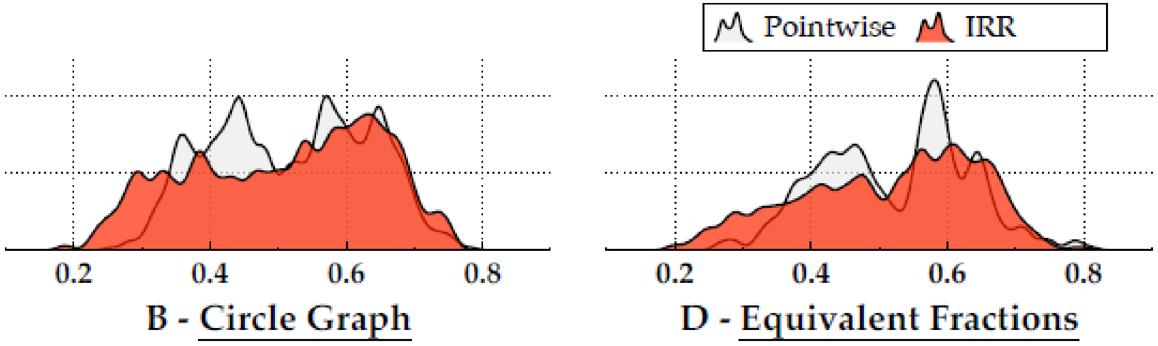
ID	Concepts	Student 1	Student 2	Student 3
Item 1	B, C	✗	✔	✗
Item 2	B, C, D, E	✗	✔	✔
Item 3	A, B, D	✔	✗	✗

A - Proportion B - Circle Graph C - Percent Of
D - Equivalent Fractions E - Finding Percents

(a) Response Logs diagram.



(b) Proficiency on knowledge concepts digram.



(c) Proficiency distribution digram.

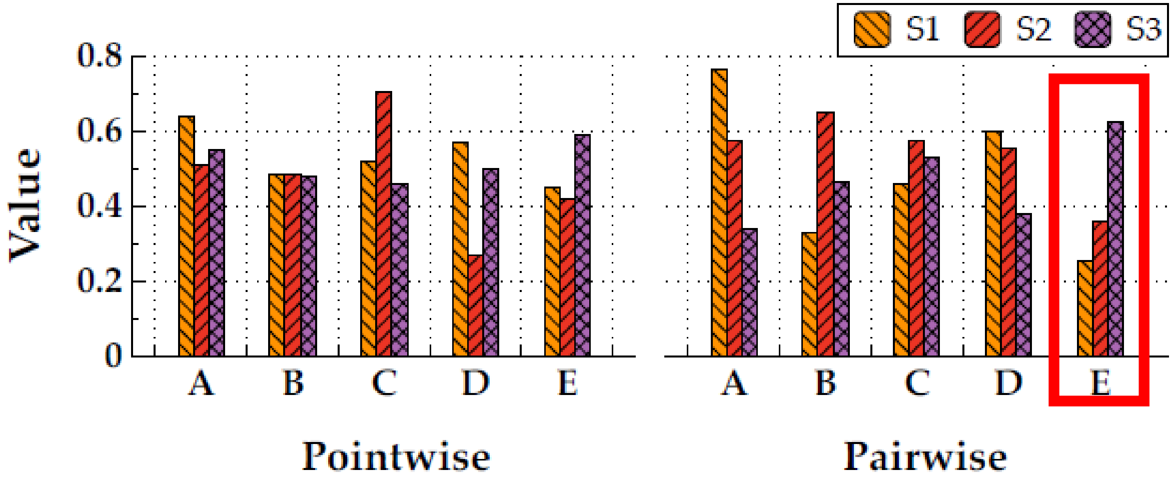
Diagnostic Results Analysis

ID	Concepts	Student 1	Student 2	Student 3
Item 1	B, C	✗	✓	✗
Item 2	B, C, D, E	✗	✓	✓
Item 3	A, B, D	✓	✗	✗

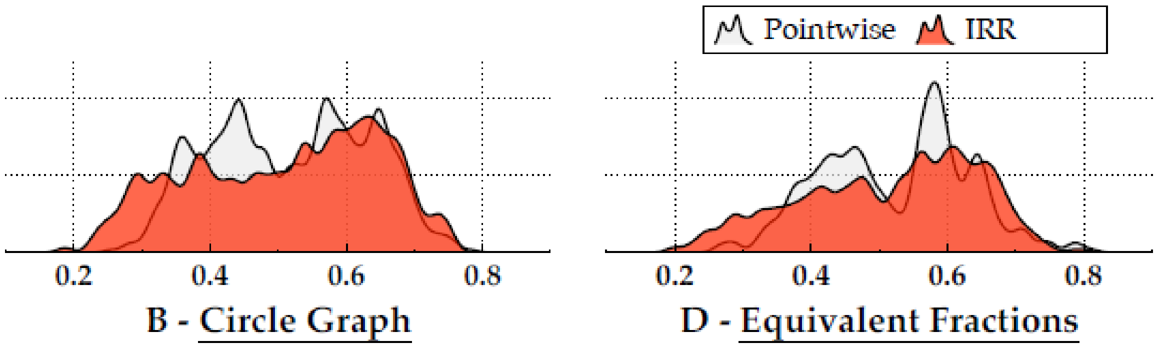
A - Proportion   B - Circle Graph   C - Percent Of

D - Equivalent Fractions   E - Finding Percents

(a) Response Logs diagram.



(b) Proficiency on knowledge concepts digram.



(c) Proficiency distribution digram.

Diagnostic Results Analysis

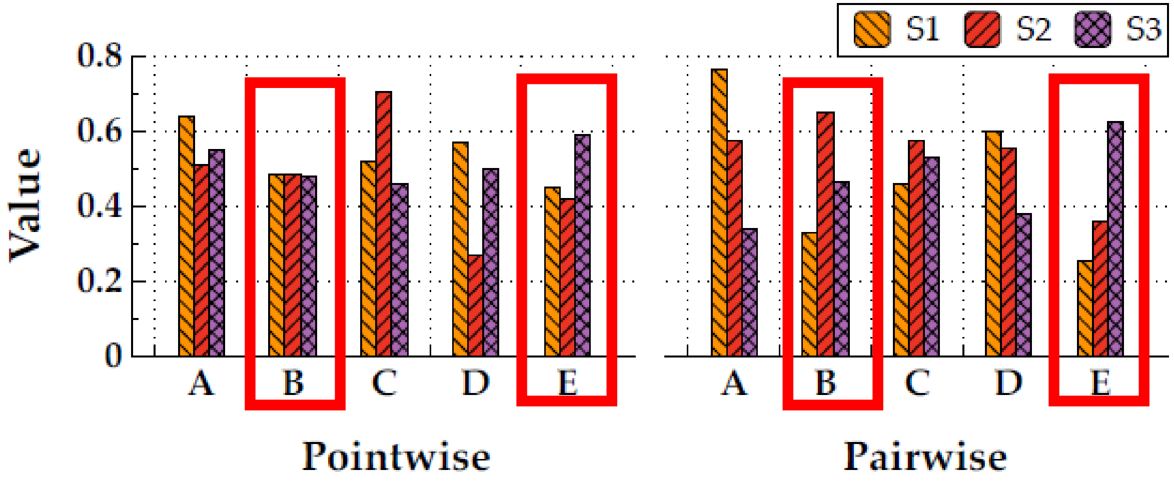
ID	Concepts	Student 1	Student 2	Student 3
Item 1	B, C	✘	✔	✘
Item 2	B, C, D, E	✘	✔	✔
Item 3	A, B, D	✔	✘	✘

A - Proportion B - Circle Graph C - Percent Of

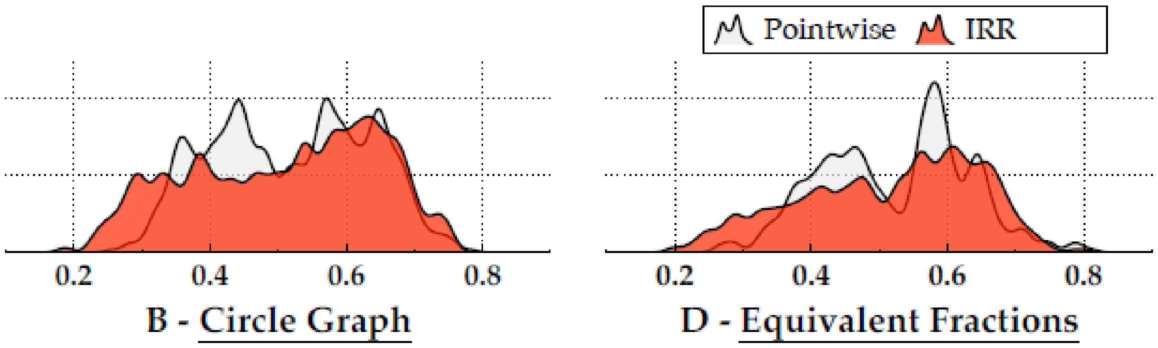
D - Equivalent Fractions E - Finding Percents

(a) Response Logs diagram.

- IRR-CDMs get a more precise and more discriminated ranking result



(b) Proficiency on knowledge concepts digram.



(c) Proficiency distribution digram.

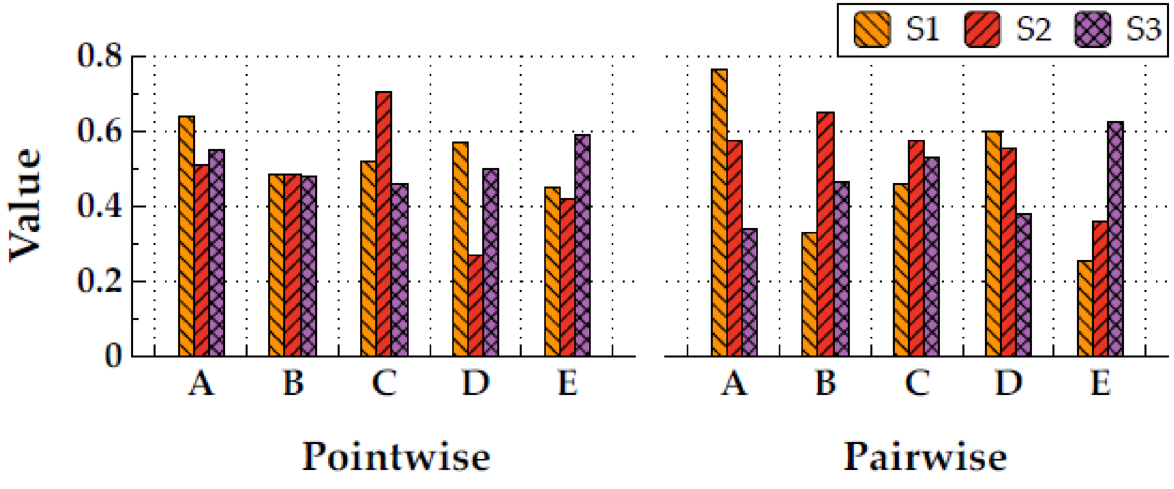
Diagnostic Results Analysis

ID	Concepts	Student 1	Student 2	Student 3
Item 1	B, C	✘	✔	✘
Item 2	B, C, D, E	✘	✔	✔
Item 3	A, B, D	✔	✘	✘

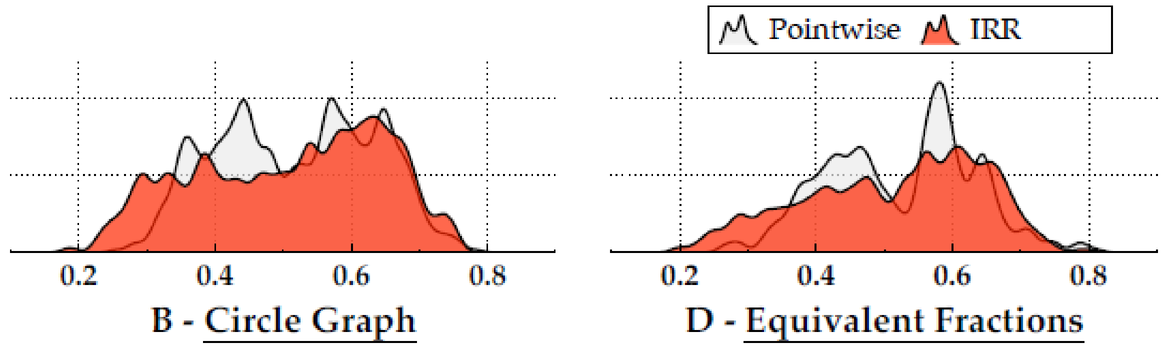
A - Proportion　B - Circle Graph　C - Percent Of
D - Equivalent Fractions　E - Finding Percents

(a) Response Logs diagram.

- IRR-CDMs get a more precise and more discriminated ranking result
- The distributions of the diagnosis proficiency values got by IRR-CDMs are more smooth



(b) Proficiency on knowledge concepts digram.



(c) Proficiency distribution digram.

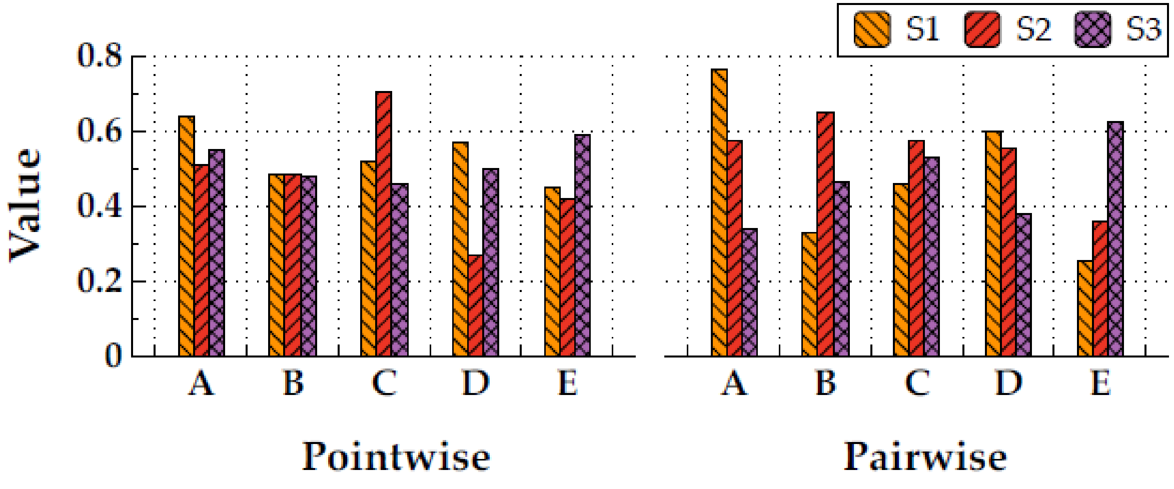
Diagnostic Results Analysis

ID	Concepts	Student 1	Student 2	Student 3
Item 1	B, C	✘	✔	✘
Item 2	B, C, D, E	✘	✔	✔
Item 3	A, B, D	✔	✘	✘

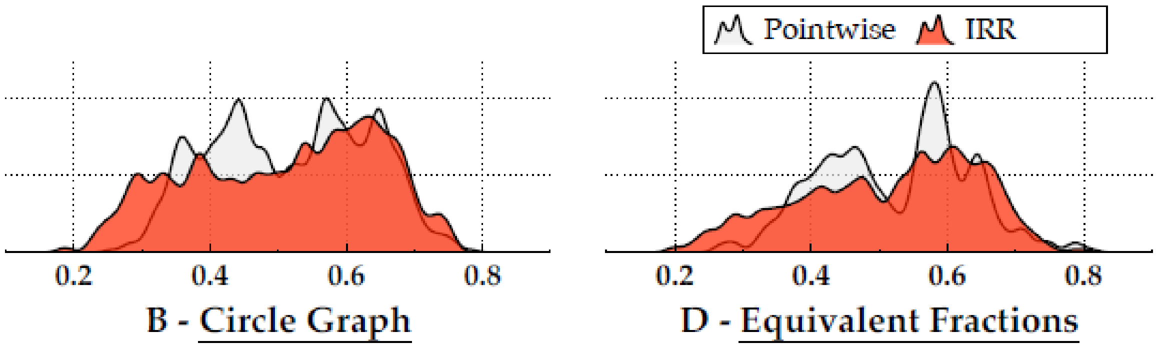
A - Proportion B - Circle Graph C - Percent Of
D - Equivalent Fractions E - Finding Percents

(a) Response Logs diagram.

- IRR-CDMs get a more precise and more discriminated ranking result
- The distributions of the diagnosis proficiency values got by IRR-CDMs are more smooth



(b) Proficiency on knowledge concepts digram.



(c) Proficiency distribution digram.

Conclusion

Our method can help CDMs get a **more precise and discriminated** diagnosis proficiency values.

Outline

30th International Joint Conference on Artificial Intelligence

Conclusion

- Cognitive diagnosis is a vital task in education area.
- Previous methods do not include monotonicity during optimization.
- We proposed the Item Response Ranking framework (IRR) to incorporate the monotonicity into the optimization objective.
  - promote the diagnosis **precision**;
  - better maintain the **monotonicity**;
  - get a more **discriminated** diagnosis proficiency values.
- Our codes are available in https://github.com/bigdata-ustc/EduCDM.

Thank you

Q&A

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

bigdata.ustc.edu.cn