Towards a New Generation of Cognitive Diagnosis

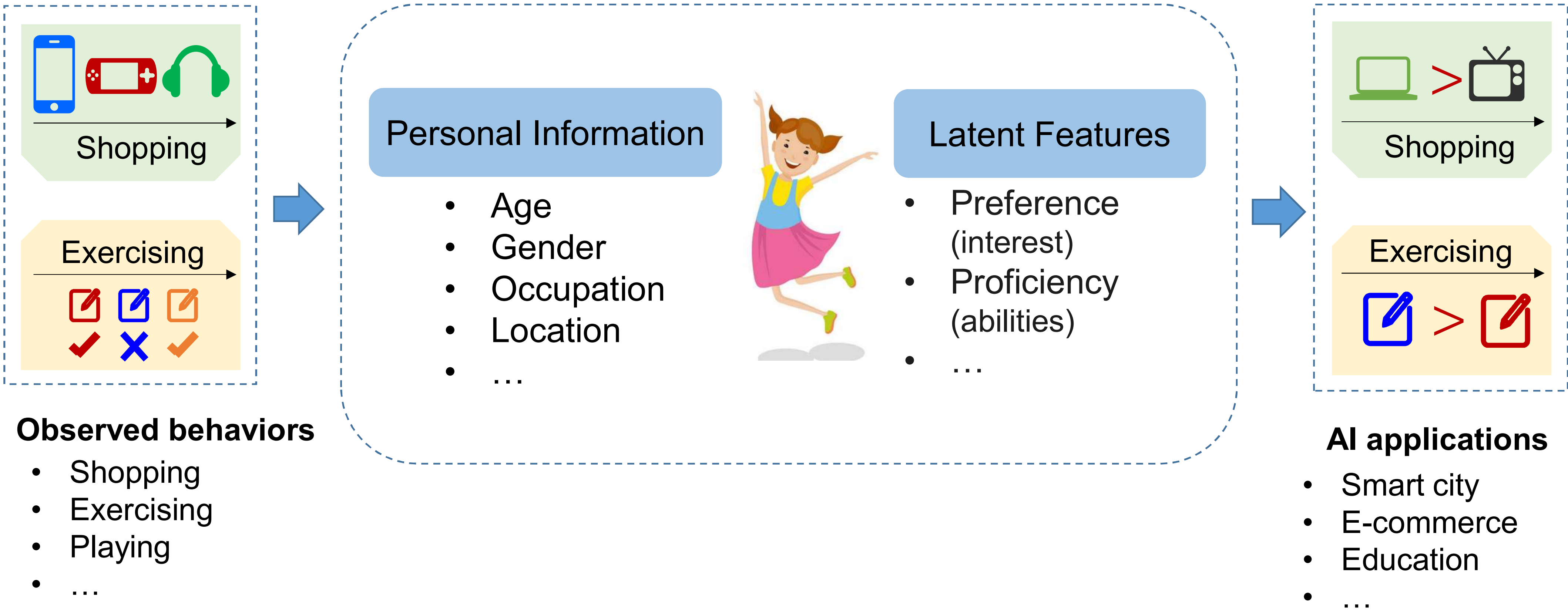
Qi Liu
qiliuql@ustc.edu.cn
University of Science and Technology of China

http://staff.ustc.edu.cn/~qiliuql/

Outline

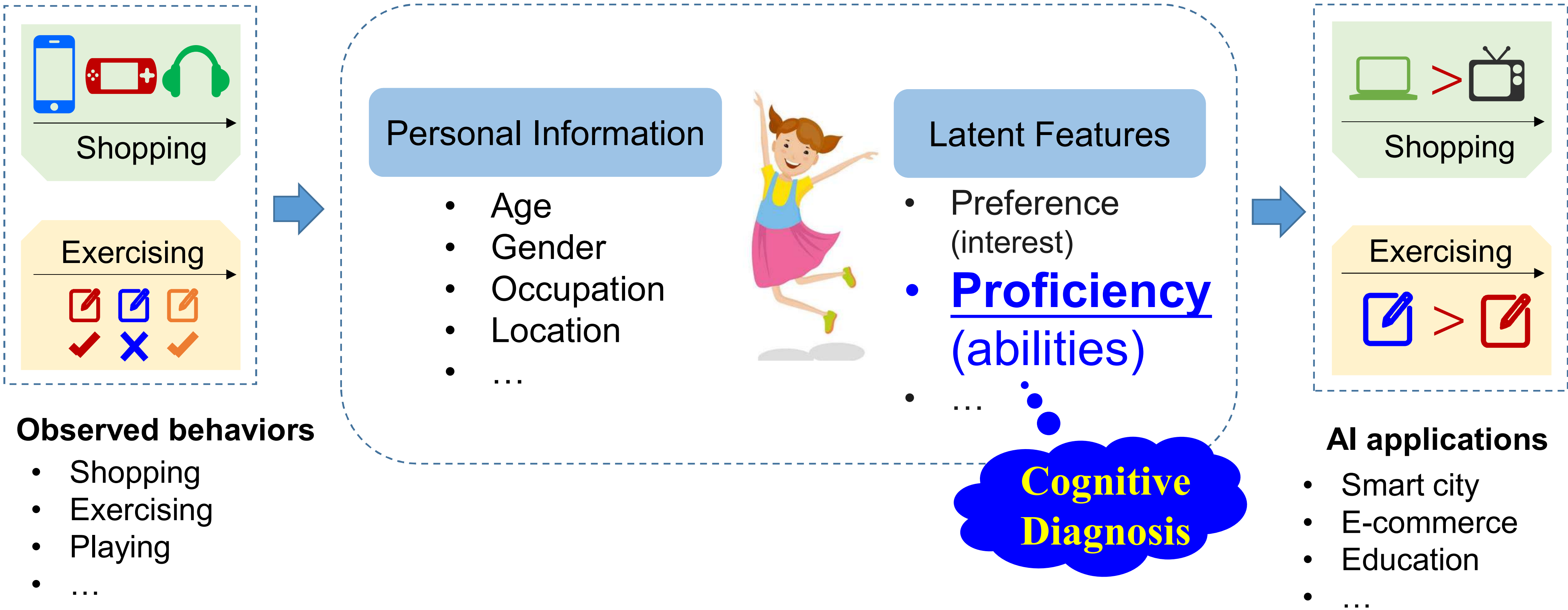
Introduction to Cognitive Diagnosis

- In the applications of artificial intelligence, we often need to characterize the difference of individuals in both personal information and latent features



Introduction to Cognitive Diagnosis

- In the applications of artificial intelligence, we often need to characterize the difference of individuals in both personal information and latent features



Introduction to Cognitive Diagnosis

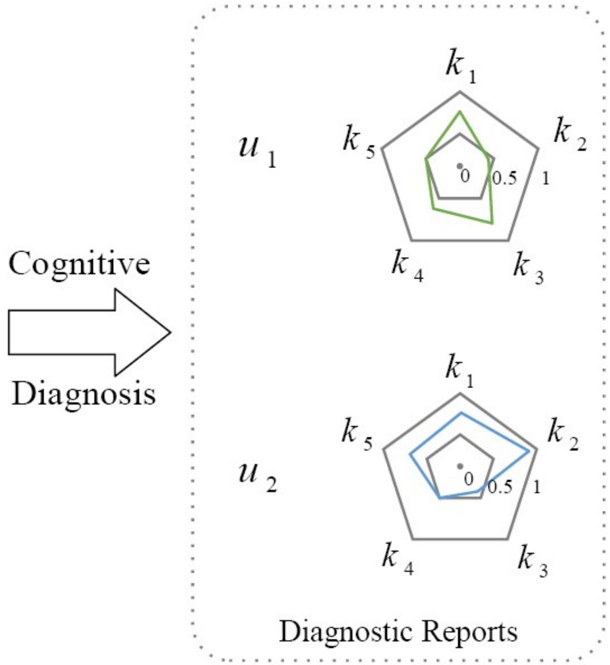
- An assessment for automatically measuring individuals' proficiency profiles
  - Quantifying the mastery level of examinees on specific knowledge concepts/skills

Tasks	Skills	Responses	
		u_1	u_2
e_1	k_1	✓	✓
e_2	k_2	✗	✓
e_3	k_3	✓	✗
e_4	k_2, k_5	✗	✓
e_5	k_3, k_4	✓	✗
Overall Score		**60**	**60**

Introduction to Cognitive Diagnosis

- An assessment for automatically measuring individuals' proficiency profiles
  - Quantifying the mastery level of examinees on specific knowledge concepts/skills

Tasks	Skills	Responses u_1	Responses u_2
e_1	k_1	✓	✓
e_2	k_2	✗	✓
e_3	k_3	✓	✗
e_4	k_2, k_5	✗	✓
e_5	k_3, k_4	✓	✗
Overall Score		**60**	**60**

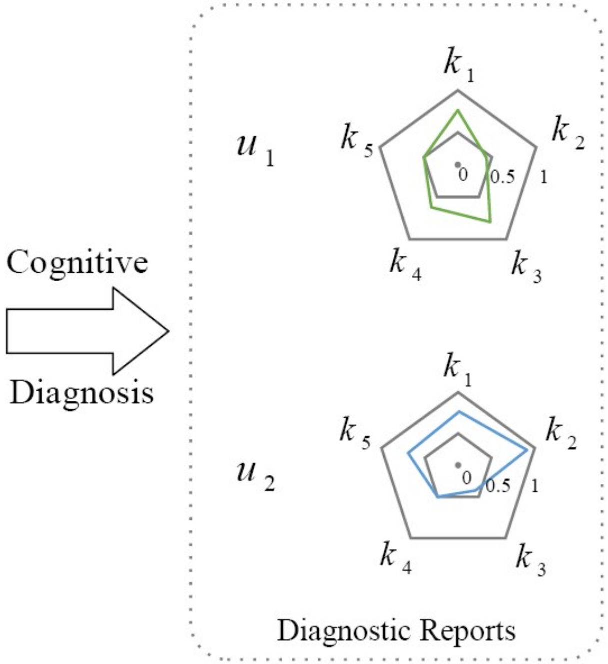


Users' proficiency on specific skills (cognitive states) are quite different

Introduction to Cognitive Diagnosis

- An assessment for automatically measuring individuals' proficiency profiles
  - Quantifying the mastery level of examinees on specific knowledge concepts/skills

Tasks	Skills	Responses	
		u_1	u_2
e_1	k_1	✓	✓
e_2	k_2	✗	✓
e_3	k_3	✓	✗
e_4	k_2, k_5	✗	✓
e_5	k_3, k_4	✓	✗
Overall Score		**60**	**60**



Assumptions:

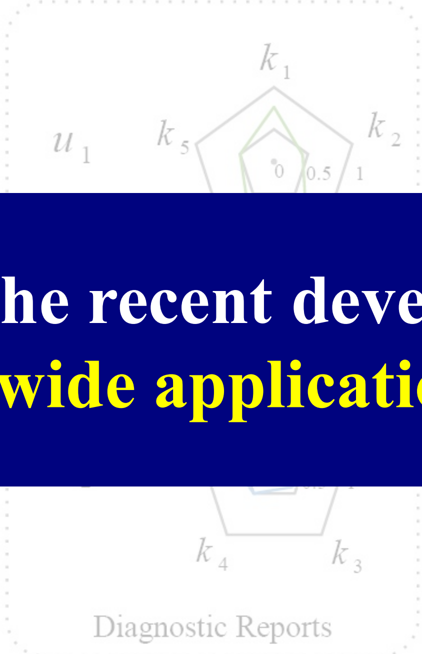
1. The observed interactive behaviors of users are determined by **their latent cognitive states**
2. The **cognitive state** of each user is stable in a short period of time thus can be diagnosed

Users' proficiency on specific skills (cognitive states) are quite different

Introduction to Cognitive Diagnosis

- An assessment for automatically measuring individuals' proficiency profiles
  - Quantifying the mastery level of examinees on specific knowledge concepts/skills.

Tasks	Skills	Responses u_1	Responses u_2
e_5	k_3, k_4	✓	✗
Overall Score		60	60



Assumptions:

user is stable in a short period of time thus can be diagnosed

This talk will briefly review the recent developments of Cognitive Diagnosis Models and show the wide applications of cognitive diagnosis

Users' proficiency on specific skills (cognitive states) are quite different

Outline

From Psychometric to Machine Learning

- Most of existing Cognitive Diagnosis Models (CDMs) are well designed based on Psychometric theories

- **Responses to the Items in a Scale**

Item Examples from Standford-Binet Intelligent Scale

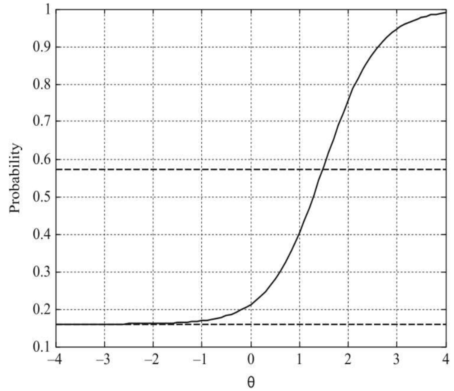
	Child	Adult
Criteria	Complete analogies	Describe differences between concepts
Item Example	*Brother is a boy, sister is a ___.*	*Describe the differences between laziness and idleness, poverty and misery, character and reputation.*

- **Item Response Theory (IRT)**

$$P(R_{ij} = 1 | \theta_i, a_j, b_j, c_j) = c_j + \frac{1 - c_j}{1 + \exp(-1.7 a_j(\theta_i - b_j))}$$

Discrimination | Guess factor | **Skill proficiency** | Difficulty

- The larger of $(\boldsymbol{\theta_i - b_j})$, the higher of the **probability** that correctly answer the item.



[1] Frederic Lord. A theory of test scores. Psychometric monographs, 1952.
[2] Roid G H, Pomplun M. The stanford-binet intelligence scales[M]. The Guilford Press, 2012.

From Psychometric to Machine Learning

- Most of existing Cognitive Diagnosis Models are well designed based on Psychometric theories

IRT

Unidimensional

$$P(R_{ij}=1|\theta_i,a_j,b_j,c_j)=c_j+\frac{1-c_j}{1+\exp(-1.7a_j(\theta_i-b_j))}$$

MIRT

Multidimensional

$$P(R_{ij}=1|\boldsymbol{\theta}_i,\boldsymbol{a}_j,d_j,c_j)=c_j+\frac{1-c_j}{1+\exp(-1.7(\boldsymbol{a}_j\boldsymbol{\theta}_i+d_j))}$$

Overall ability ✓

Mastery of specific skills ✗

Q-matrix

	Skill 1	Skill 2	Skill 3	Skill 4
Task 1	1	0	0	0
Task 2	0	1	1	0
Task 3	0	0	0	1

θ_i can possibly be inferred with the help of a task−skill relevancy matrix (Q−matrix)

From Psychometric to Machine Learning

- Traditional Cognitive Diagnosis Models

IRT

$$P\left(R_{ij}=1\middle|\theta_i,a_j,b_j,c_j\right)=c_j+\frac{1-c_j}{1+\exp(-1.7a_j(\theta_i-b_j))}$$

Characteristics		Pros	Cons
Input of the Model	➢ User responses to a small number of scale items	➢ High quality of test items	➢ Can only exploit users' numerical response records
Diagnose Function	➢ Manually designed ➢ Mostly linear	➢ Inspired from psychometric theories with good interpretability	➢ Labor intensive ➢ Limited approximation ability

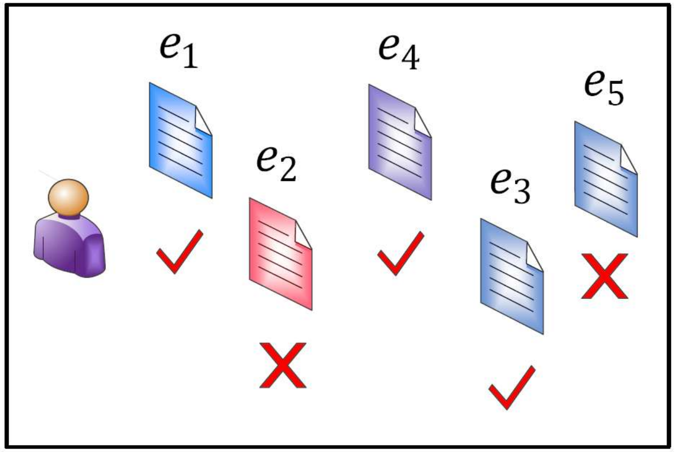
From Psychometric to Machine Learning

- **We are able to design cognitive diagnosis models from a machine learning perspective**

➢ Multiple types of data, such as **the images** and **the text description** about the task, are now available

Exercise/Task	Exercise / Task Content
e_1	If function $f(x) = x^2 - 2x + 2$ and $x \in [0,3]$, What is the range of $f(x)$?
e_2	If four numbers are randomly selected without replacement from set $\{1, 2, 3, 4\}$, what is the probability that the four numbers are selected in ascending order?
e_3	What is the y-intercept of the graph of equation $y = 2 \times \|4 \times x - 4\| - 10$?
e_4	What is the value of x If the inequality $\frac{2x-1}{x+2} \leq 3$?
e_5	If function $f(x) = 2x - 2$ and $x \in [-1,1]$, what is the range of $f(x)$?

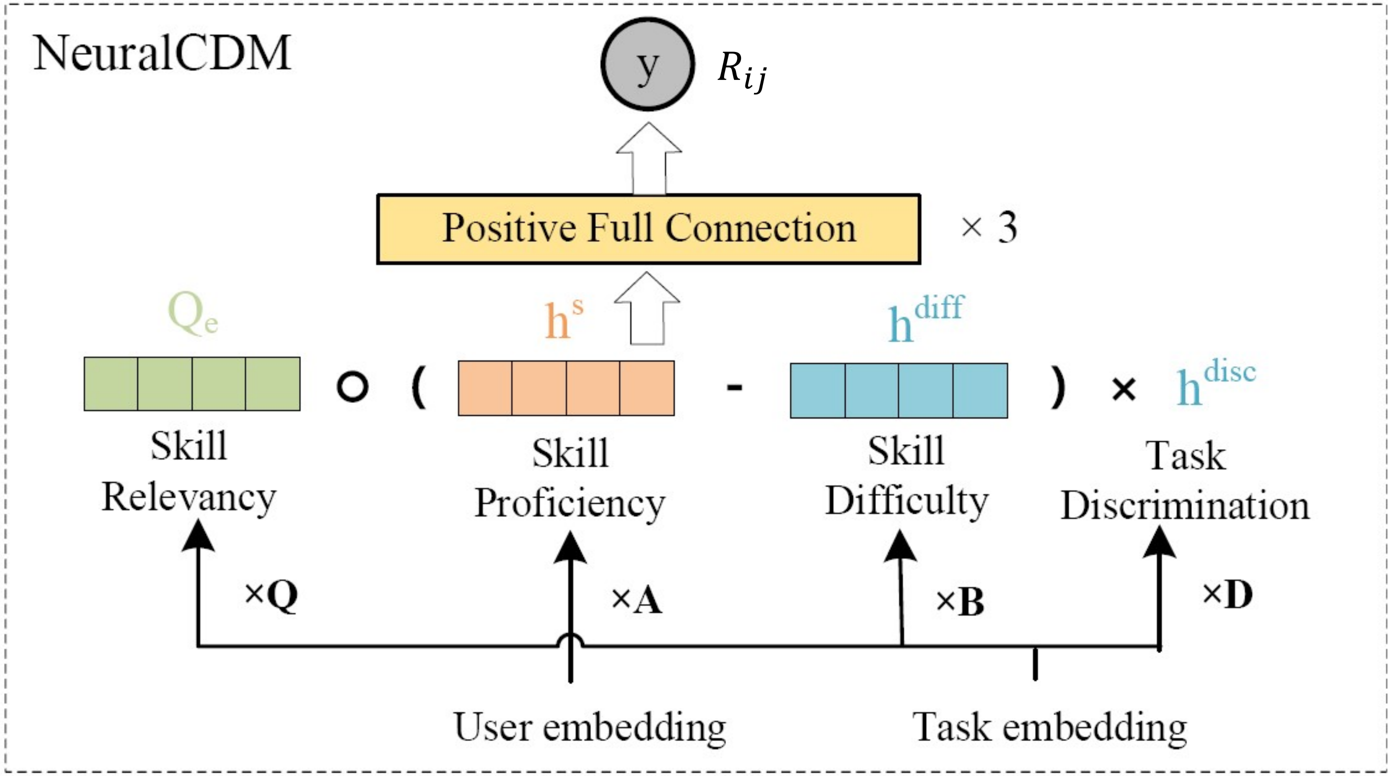
➢ Besides a well defined Scale, we can also diagnose the individuals from their **daily tasks/exercises**



Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, et al., EKT: Exercise-aware Knowledge Tracing for Student Performance Prediction, IEEE TKDE, 33(1): 100-115, 2021.

From Psychometric to Machine Learning

- Learn the diagnose function automatically from big data



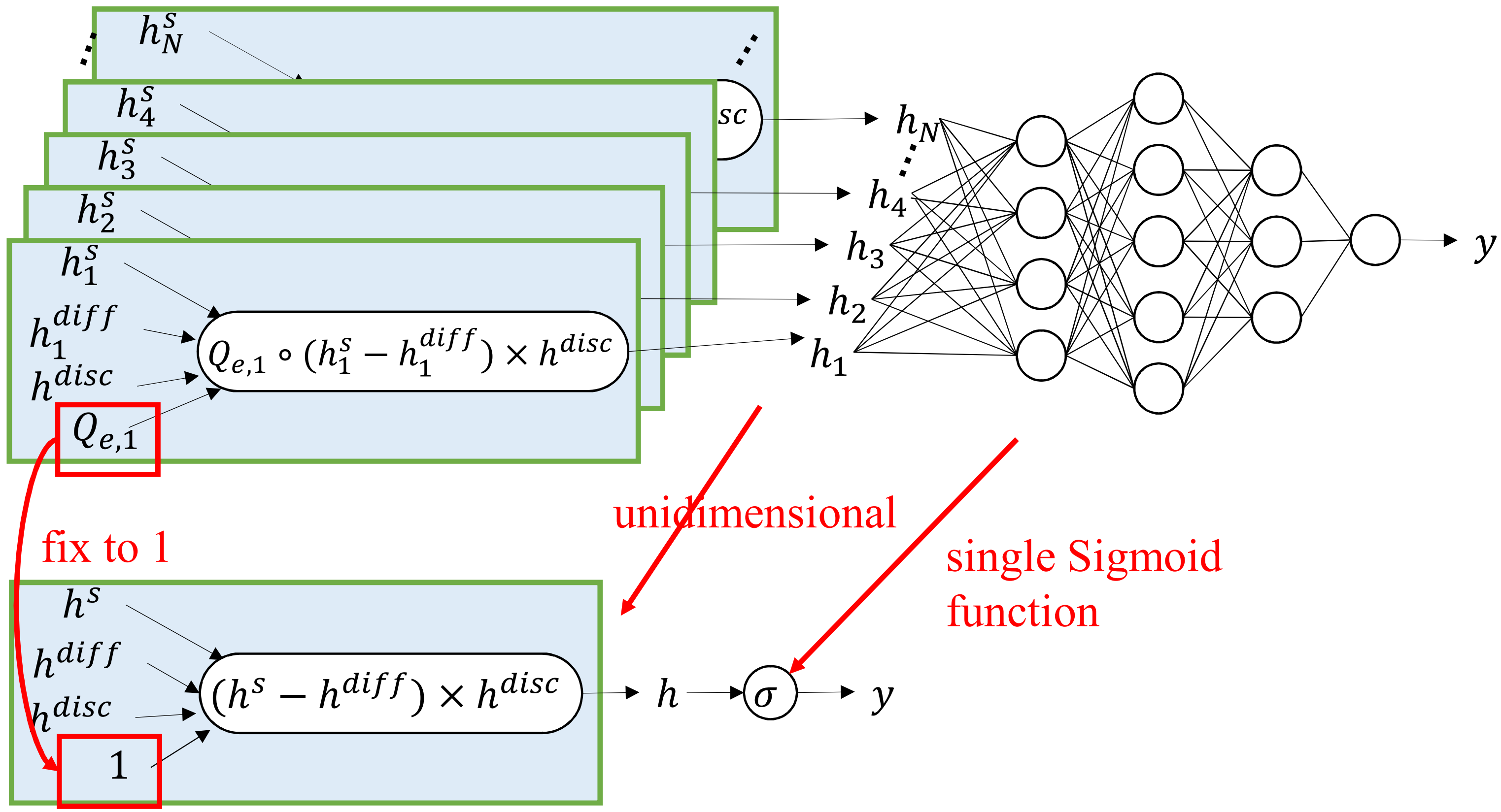
1. **Multidimensional parameters**
2. **Neural network captures the complex interactions**
3. **Skill relevancy vector ensures the interpretability**
4. **Maintain the monotonicity assumption**

Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, et al., Neural Cognitive Diagnosis for Intelligent Education Systems. AAAI 2020, 6153-6161.

From Psychometric to Machine Learning

- Our machine learning solution is a general framework

NeuralCDM



IRT

$$P(R_{uv} = 1 | \theta_u, a_v, b_v, c_v) = c_v + \frac{1 - c_v}{1 + \exp(-1.7 a_v (\theta_u - b_v))}$$

From Psychometric to Machine Learning

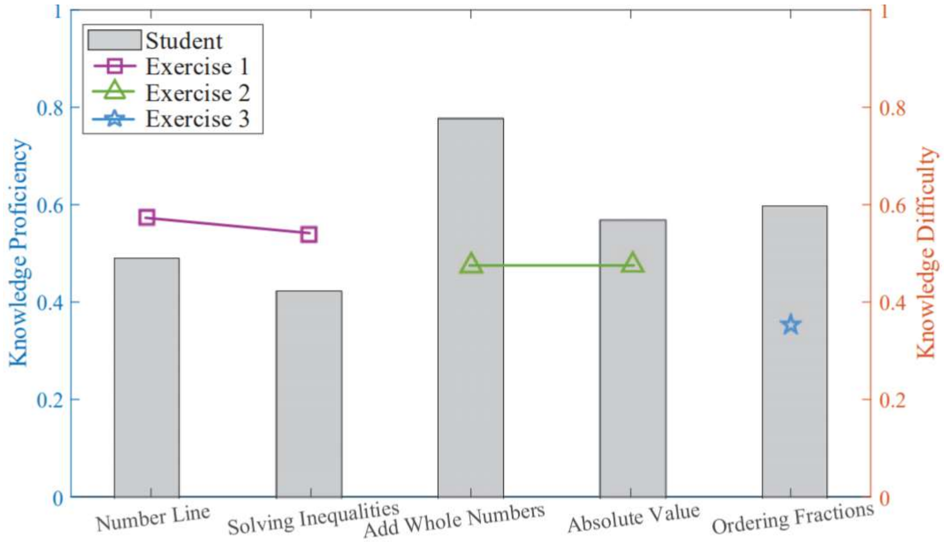
- Experimental Evaluation
  - Predicting users' (students') performance
    - **Math**: mathematical exercises (with texts) and logs
    - **ASSIST**: mathematical exercises (without texts) and logs

Model	Math Accuracy	Math RMSE	Math AUC	ASSIST Accuracy	ASSIST RMSE	ASSIST AUC
DINA	0.593±.001	0.487±.001	0.686±.001	0.650±.001	0.467±.001	0.676±.002
IRT	0.782±.002	0.387±.001	0.795±.001	0.674±.002	0.464±.002	0.685±.001
MIRT	0.793±.001	0.378±.002	0.813±.002	0.701±.002	0.461±.001	0.719±.001
PMF	0.763±.001	0.407±.001	0.792±.002	0.661±.002	0.476±.001	0.732±.001
NeuralCDM	0.792±.002	0.378±.001	0.820±.001	**0.719±.008**	**0.439±.002**	**0.749±.001**
NeuralCDM+	**0.804±.001**	**0.371±.002**	**0.835±.002**	-	-	-

Our machine learning solutions outperform the baselines

Q-matrix Responses

	Number Line	Solving Inequalities	Add Whole Numbers	Absolute Value	Ordering Fractions	Student Response
Exercise 1	1	1	0	0	0	✗
Exercise 2	0	0	1	1	0	✓
Exercise 3	0	0	0	0	1	✓



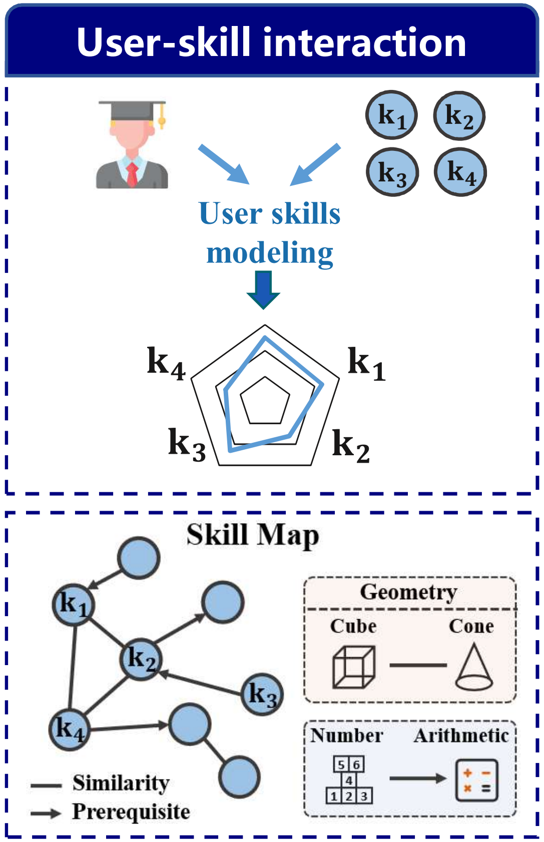
Knowledge Difficulty of Exercise (points)

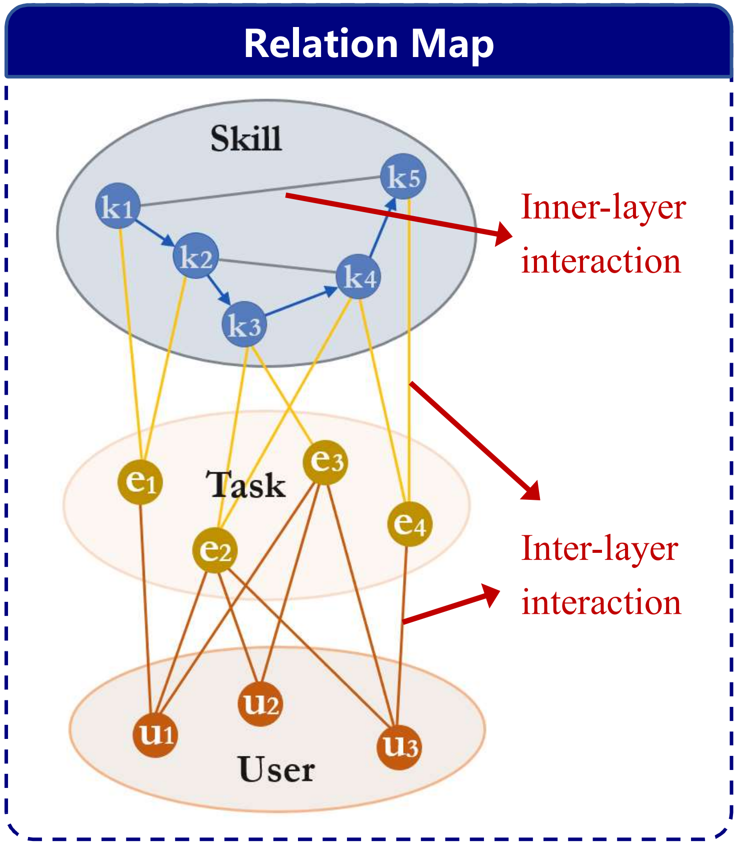
Knowledge Proficiency of Student (bars)

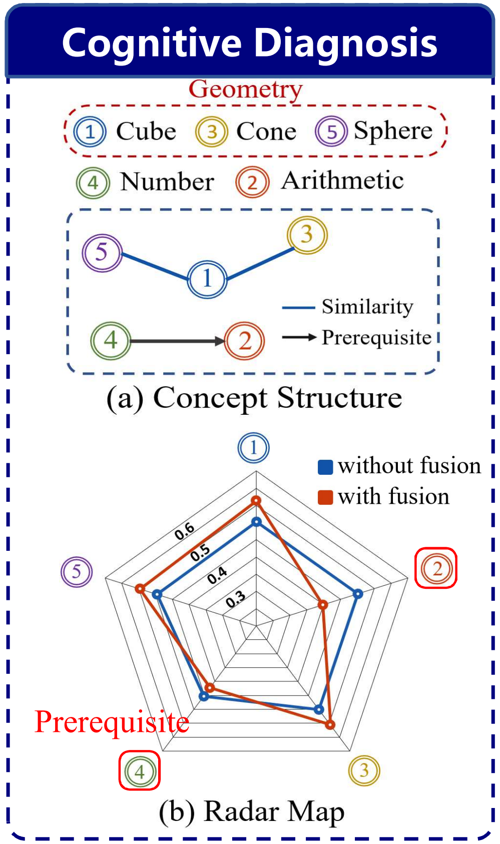
Outline

Our Extensions for CDMs

- Relation map-driven CDMs: Incorporating the interdependencies among skills







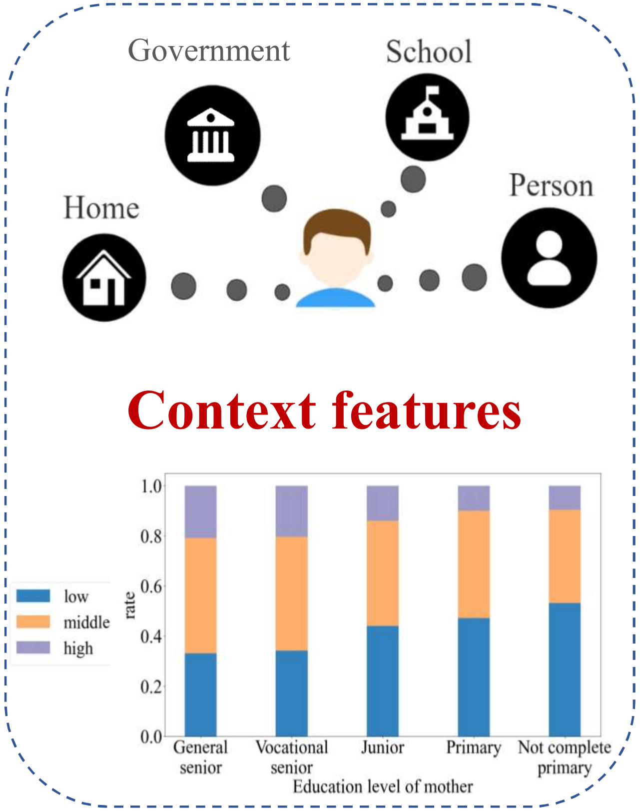
(a) Concept Structure

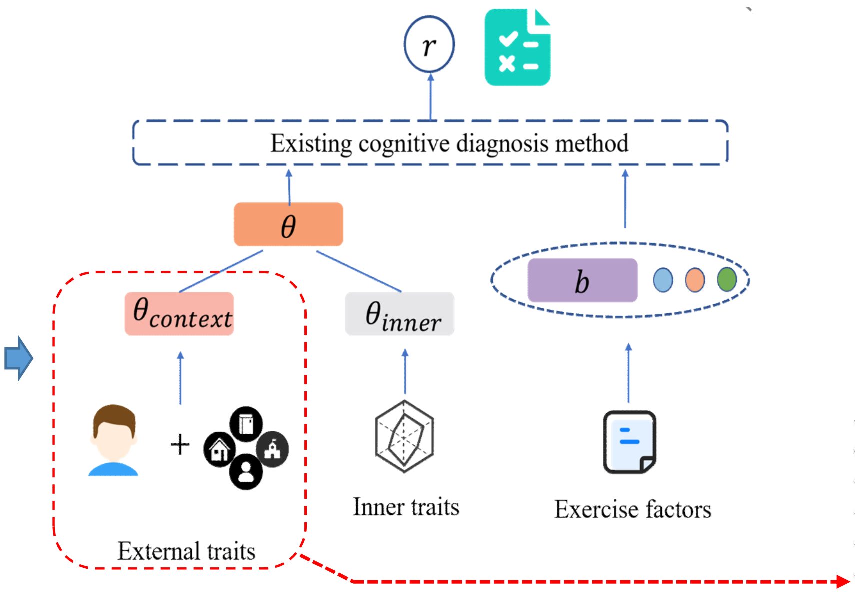
(b) Radar Map

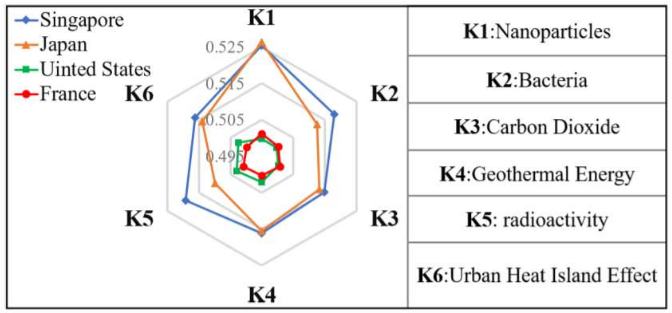
Weibo Gao, Qi Liu, Zhenya Huang, et al. Rcd: Relation map driven cognitive diagnosis for intelligent education systems. In ACM SIGIR, 2021.

Our Extensions for CDMs

- Context-aware CDMs: Considering the context-aware features of users







Region difference

Regions	Important Context					
Uinted States		"HE",	"SL", "TA",	–, –,		"SE"
Uinted Kindom		"HE",	"SL", "TA",	"SI", "IU",		"SE"
France		"HE",	"SL", "TA",	"SI", "IU"		
Germany		"HE",	"SL", "TA",	"SI", "IU"		
Italy		"HE",	"SL", "TA",	"SI", "IU"		
Singapore		"HE",		"SI", "IU",	"IonS",	"SE"
Japan		"HE",		"SI", "IU",		"SE"
Korea	"PE",	"HE",		"SI", "IU"		
China	"PE",	"HE",		"SI", "IU"		

"PE", "Parent Education", "HE", "Home ESCS", "SL", "School Learning", "TA", "Teach Attitude",
"SI", "School ICT", "IU" "ICT Usage" "IonS" "Interest on Science" "SE" "Self-efficacy"

Yuqiang Zhou, Qi Liu,et al., Modeling Context-aware Features for Cognitive Diagnosis in Student Learning, In ACM SIGKDD 2021.

Our Extensions for CDMs

- Item Response Ranking for CDMs: Modeling the partial orders between responses

General Form

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

Monotonicity

User's proficiency is monotonic with the probability of giving the right response to a task/item.

monotonic diagnose function

Objective Function

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

Our Solution

Pairwise Monotonicity

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

Monotonicity as the optimization target

Objective Function

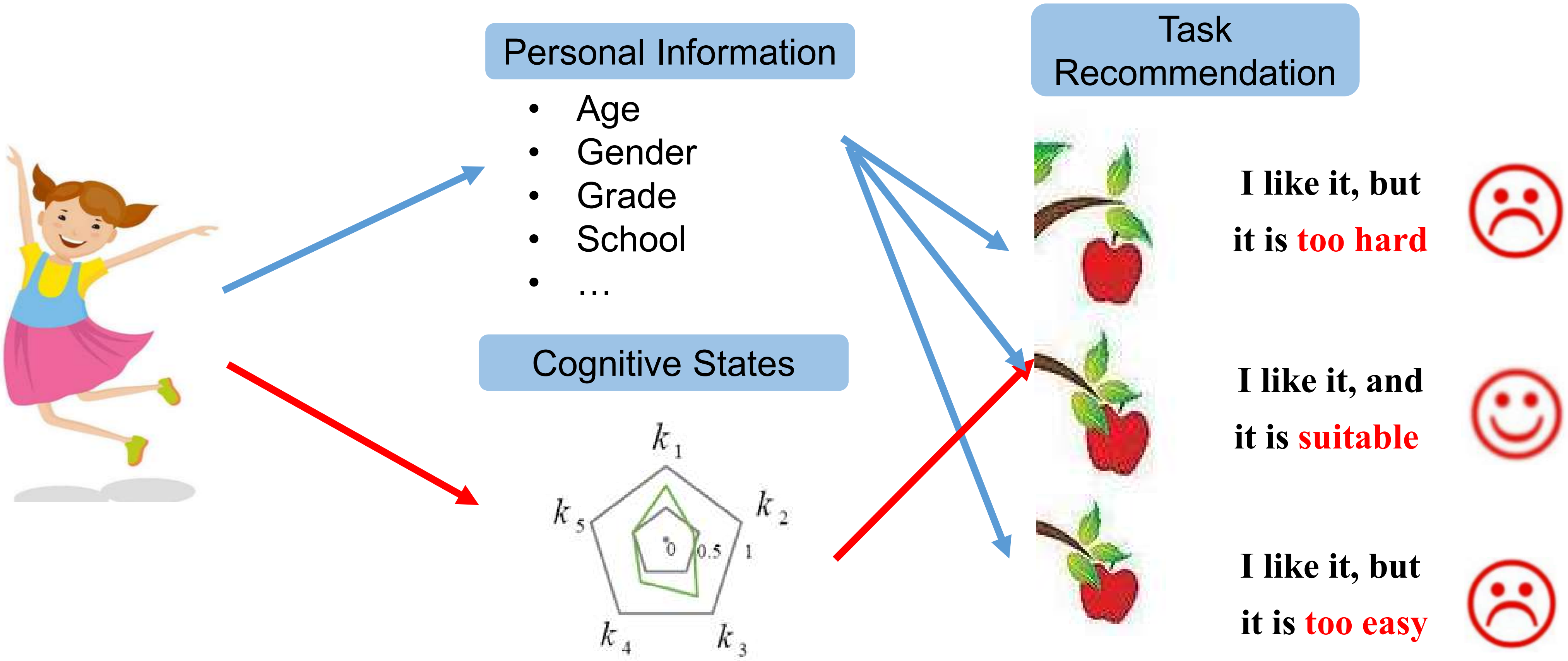
$$(r_{ie} - r_{je}) \leftarrow P(R_{ie} - \boldsymbol{R}_{je})$$

Shiwei Tong, Qi Liu, Runlong Yu, et al., Item response ranking for cognitive diagnosis, In IJCAI 2021.

Outline

Applications of Cognitive Diagnosis

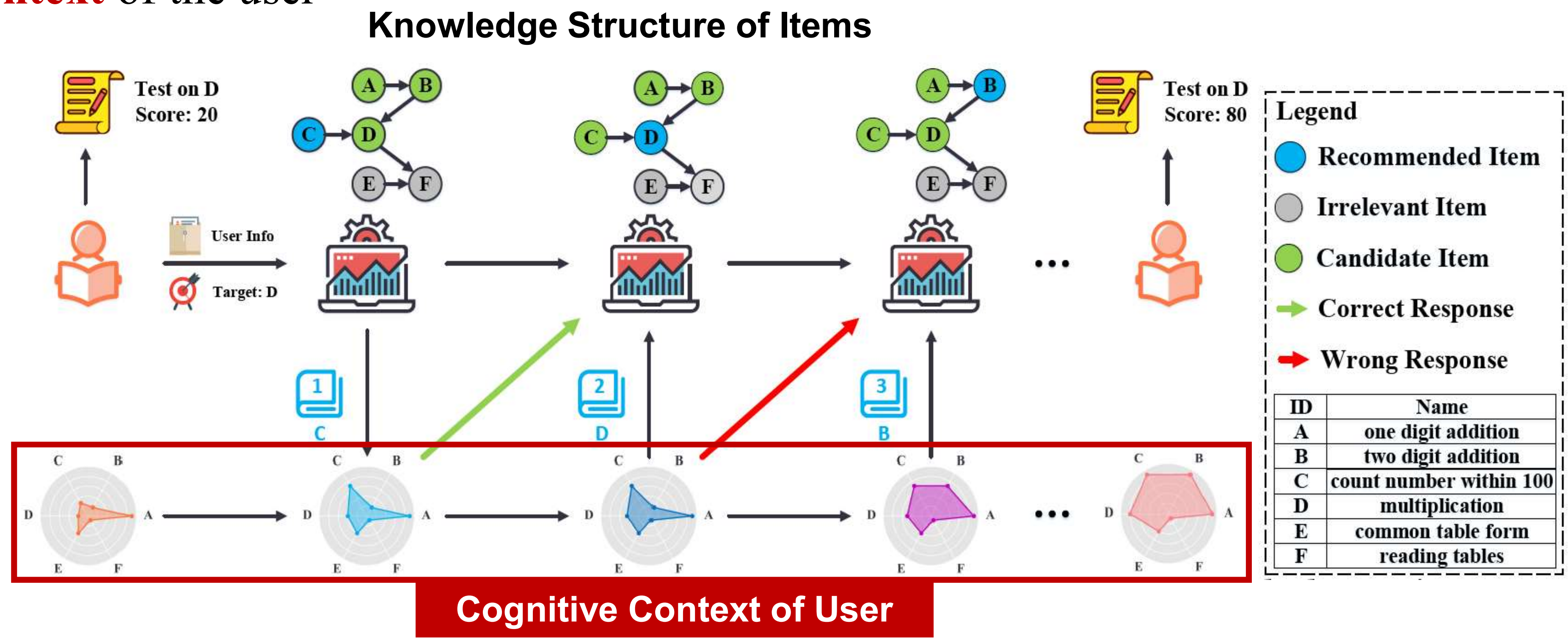
- Cognitive Diagnosis for Recommender Systems



Zhenya Huang, Qi Liu, Chengxiang Zhai, Yu Yin, and Guoping Hu. Exploring multi-objective exercise recommendations in online education systems. In ACM CIKM, 2019.

Applications of Cognitive Diagnosis

- Cognitive Diagnosis for Adaptive Learning
  - Recommending a learning/training path that is adaptive to the implicit evolving **cognitive context** of the user

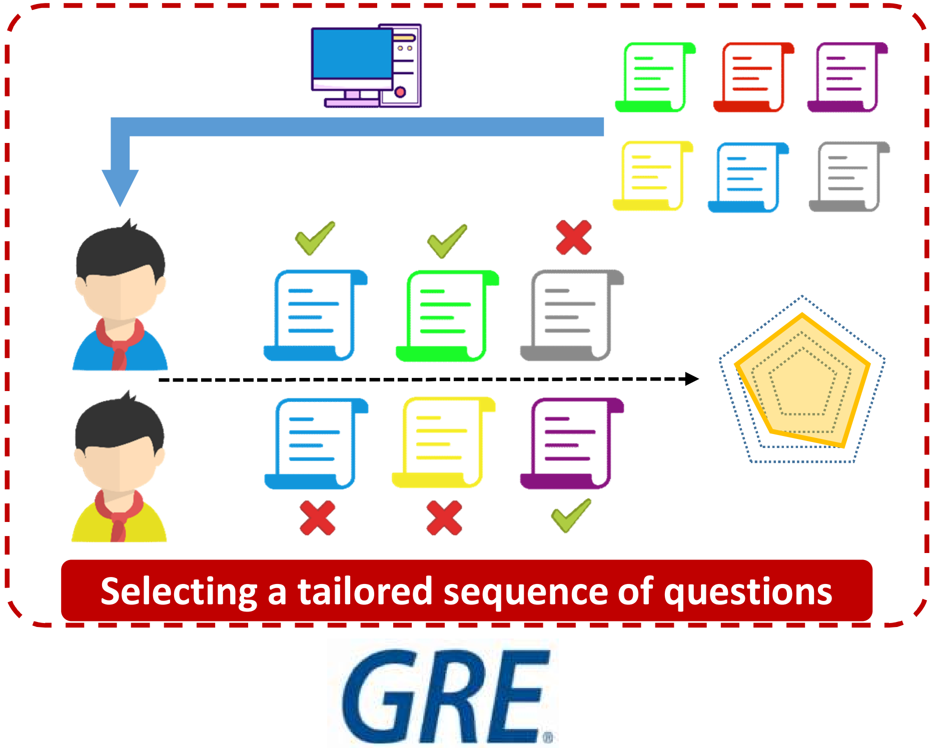


Qi Liu, Shiwei Tong, Chuanren Liu, et al.. Exploiting Cognitive Structure for Adaptive Learning. In ACM SIGKDD, 2019.

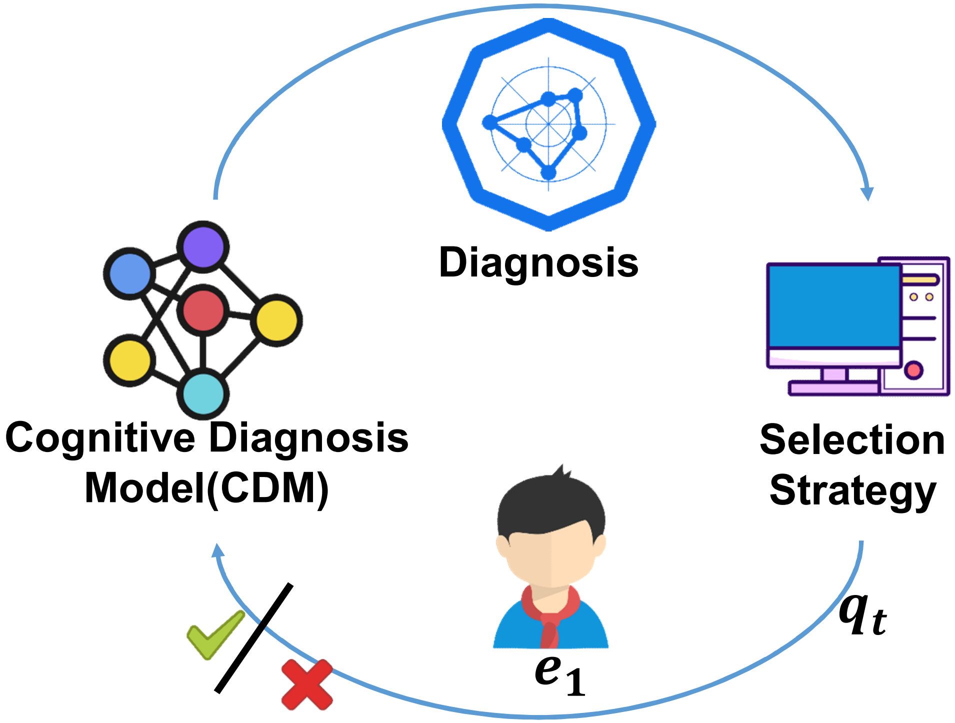
Applications of Cognitive Diagnosis

- Cognitive Diagnosis for Computerized adaptive testing (CAT)

Adaptive Testing



Diagnosis for selection



Selection for diagnosis

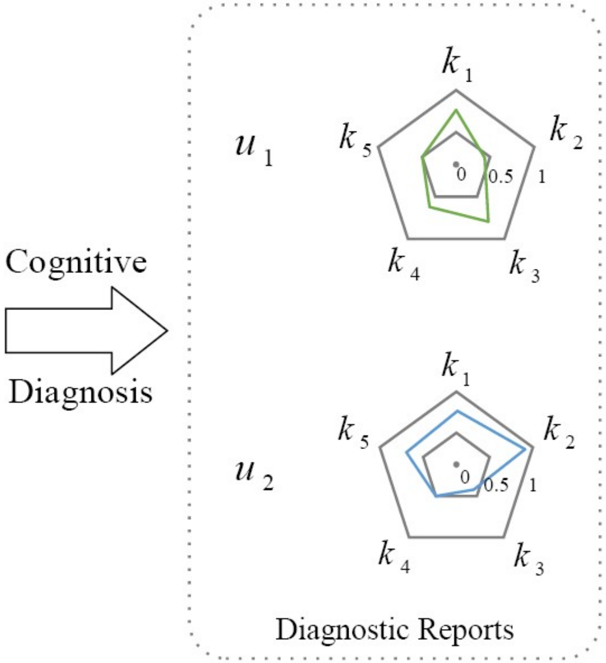
Haoyang Bi, Haiping Ma, Zhenya Huang, Yu Yin, Qi Liu, et al., Quality meets diversity: A model-agnostic framework for computerized adaptive testing. In IEEE ICDM, 2020.

Outline

Conclusion

- Comprehensive study on cognitive diagnosis mostly from a machine learning perspective
  - A new cognitive diagnosis model which learns user-task interactions from big data
  - Several extensions for basic cognitive diagnosis models
  - Applying cognitive diagnosis to provide users with better services

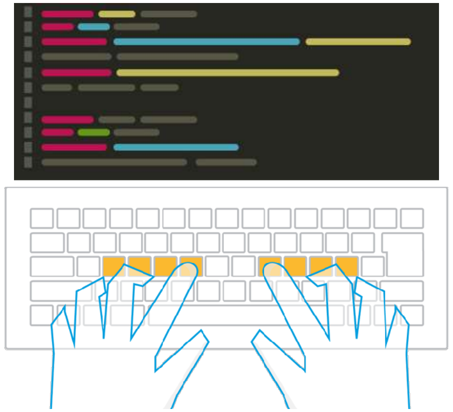
Tasks	**Skills**	**Responses** u_1	**Responses** u_2
e_1	k_1	✓	✓
e_2	k_2	✗	✓
e_3	k_3	✓	✗
e_4	k_2, k_5	✗	✓
e_5	k_3, k_4	✓	✗
Overall Score		**60**	**60**

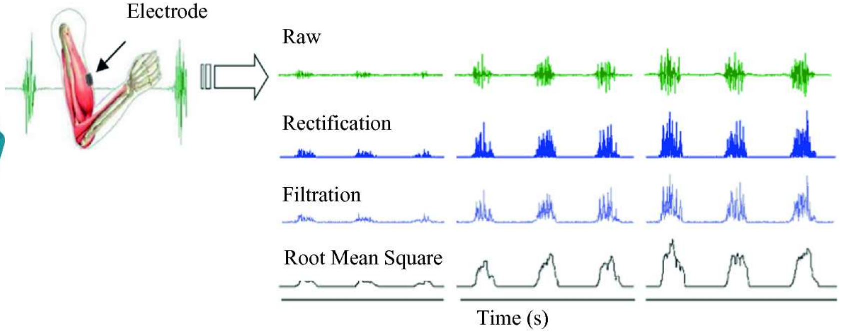


Future Research Directions

How to collect and exploit the more detailed user behaviors during their responses?

- Detailed logs
  - response time
  - multiple attempts
  - keystroke analysis (writing, coding, etc.)
- Physiological reaction
  - expression
  - EEG/EMG signals (sports, etc.)
- Behaviors in Teamwork
  - discussion
  - corporation
- etc.

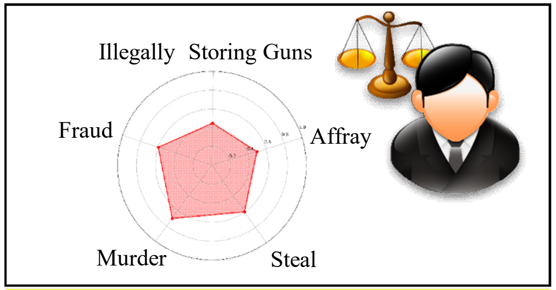




Future Research Directions

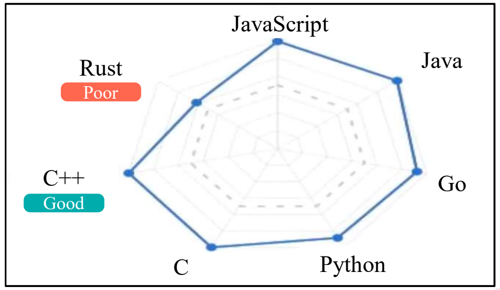
How to apply the idea of cognitive diagnosis to more scenarios besides the utilization in education?

- Legal Domain： Lawyer's Proficiency Assessment
  - The judgment results of the litigation is related to the lawyers' proficiency and the case difficulty
- Other domains： Sports、Job market、Game……

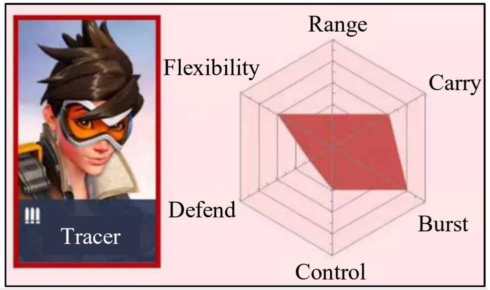


Lawyer's Proficiency Assessment

Athlete's Ability Index



Talent's Programming Skill



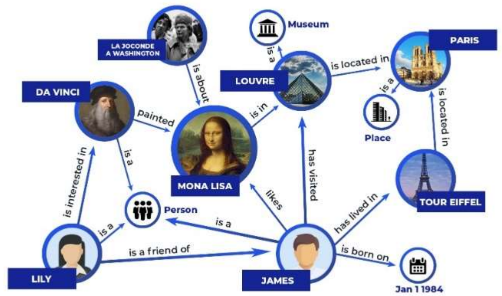
Game Player's Skill Level

Yanqing An, Qi Liu, Han Wu, et al., Lawyerpan: A proficiency assessment network for trial lawyers. In ACM SIGKDD, 2021.

Future Research Directions

How to combine the advantages of psychometric theories and machine learning for designing more reasonable cognitive diagnosis framework?

- **Theories** that might be beneficial
  - memory theories
  - cognitivism learning theory
- **Machine learning** methods that are suitable for the combining
  - memory networks
  - graph / social network analysis
  - knowledge graph



[1] Murre J M J, Dros J. Replication and analysis of Ebbinghaus' forgetting curve[J]. PloS one, 2015, 10(7): e0120644.
[2] Ertmer P A, Newby T J. Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective[J]. Performance improvement quarterly, 1993, 6(4): 50-72.
[3] Siemens G. Elearnspace. Connectivism: A learning theory for the digital age[J]. Elearnspace. org, 2004.

Source Code and Public Data

- EduCDM: https://github.com/bigdata-ustc/EduCDM

EduCDM

PyPI build passing codecov 100% downloads 218/month license Apache-2.0 DOI 10.5281/zenodo.5055201

The Model Zoo of Cognitive Diagnosis Models, including classic Item Response Ranking (**IRT**), Multidimensional Item Response Ranking (**MIRT**), Deterministic Input, Noisy "And" model(**DINA**), and advanced Fuzzy Cognitive Diagnosis Framework (**FuzzyCDF**), Neural Cognitive Diagnosis Model (**NCDM**) and Item Response Ranking framework (**IRR**).

Brief introduction to CDM

Cognitive diagnosis model (CDM) for intelligent educational systems is a type of model that infers students' knowledge states from their learning behaviors (especially exercise response logs).

Typically, the input of a CDM could be the students' response logs of items (i.e., exercises/questions), the Q-matrix that denotes the correlation between items and knowledge concepts (skills). The output is the diagnosed student knowledge states, such as students' abilities and students' proficiencies on each knowledge concepts.

Traditional CDMs include:

- IRT: item response theory, a continuous unidimensional CDM with logistic-like item response function.
- MIRT: Multidimensional item response theory, a continuous multidimensional CDM with logistic-like item response function. Mostly extended from unidimensional IRT.

- EduData: https://github.com/bigdata-ustc/EduData

EduData

PyPI Build Status codecov 79% downloads 223/month license Apache-2.0 DOI 10.5281/zenodo.4851317

Convenient interface for downloading and preprocessing datasets in education.

The datasets include:

- KDD Cup 2010 [Analysis] (TBA)
- ASSISTments [Analysis]
- OLI Engineering Statics 2011 [Analysis]
- JunyiAcademy Math Practicing Log [Analysis]
- slepemapy.cz
- synthetic
- math2015 [Analysis]
- EdNet [Analysis]
- pisa2015math [Analysis] (TBA)
- workbankr
- critlangacq
- math23k [Analysis]
- MOOCCube [Analysis]
- OpenLUNA

Selected References

- Shiwei Tong, Qi Liu, Runlong Yu, Wei Huang, Zhenya Huang, Zachary Pardos, and Weijie Jiang. Item response ranking for cognitive diagnosis. In IJCAI, 2021.
- Runze Wu, Qi Liu, Yuping Liu, Enhong Chen, Su Yu, Zhigang Chen, and Guoping Hu. Cognitive modelling for predicting examinee performance. In IJCAI, pages 1017–1024, 2015.
- Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, et al., Neural Cognitive Diagnosis for Intelligent Education Systems. AAAI 2020, 6153-6161.
- Yanqing An, Qi Liu, Han Wu, et al., Lawyerpan: A proficiency assessment network for trial lawyers. In ACM SIGKDD, 2021.
- Haoyang Bi, Haiping Ma, Zhenya Huang, Yu Yin, Qi Liu, et al., Quality meets diversity: A model-agnostic framework for computerized adaptive testing. In IEEE ICDM, pages 42–51. 2020.
- Jimmy De La Torre. Dina model and parameter estimation: A didactic. Journal of educational and behavioral statistics, 34(1):115–130, 2009.
- L. V DiBello, L. A Roussos, and W. Stout. 31a review of cognitively diagnostic assessment and a summary of psychometric models. Handbook of statistics, 26:979–1030, 2006.
- Weibo Gao, Qi Liu, Zhenya Huang, et al., Rcd: Relation map driven cognitive diagnosis for intelligent education systems. In ACM SIGIR, 2021.
- Zhenya Huang, Qi Liu, Chengxiang Zhai, Yu Yin, and Guoping Hu. Exploring multi-objective exercise recommendations in online education systems. In ACM CIKM, 2019.
- Qi Liu, Shiwei Tong, Chuanren Liu, et al.. Exploiting Cognitive Structure for Adaptive Learning. In ACM SIGKDD, 2019.
- Yuqiang Zhou, Qi Liu, Jinze Wu, Fei Wang, Zhenya Huang, et al., Modeling Context-aware Features for Cognitive Diagnosis in Student Learning, In ACM SIGKDD, 2021.
- Frederic Lord. A theory of test scores. Psychometric monographs, 1952.
- Mark D Reckase. Multidimensional item response theory models. In Multidimensional item response theory, pages 79–112. Springer, 2009.
- Qi Liu, R Wu, E Chen, et al. Fuzzy cognitive diagnosis for modelling examinee performance[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2018, 9(4): 1-26.
- Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, et al., EKT: Exercise-aware Knowledge Tracing for Student Performance Prediction, IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE), 33(1): 100-115, January, 2021.
- Qi Liu, Shuanghong Shen, Zhenya Huang, Enhong Chen, Yonghe Zheng. A Survey of Knowledge Tracing https://arxiv.org/abs/2105.15106

Thanks!

http://staff.ustc.edu.cn/~qiliuql/

qiliuql@ustc.edu.cn

Qi Liu would like to thank all the collaborators and the team members in BASE@USTC: http://base.ustc.edu.cn/