

# Towards a New Generation of Cognitive Diagnosis

Qi Liu

School of Computer Science and Technology & School of Data Science,  
 University of Science and Technology of China  
 qiliuql@ustc.edu.cn

## Abstract

Cognitive diagnosis is a type of assessment for automatically measuring individuals’ proficiency profiles from their observed behaviors, e.g. quantifying the mastery level of examinees on specific knowledge concepts/skills. As one of the fundamental research tasks in domains like intelligent education, a number of Cognitive Diagnosis Models (CDMs) have been developed in the past decades. Though these solutions are usually well designed based on psychometric theories, they still suffer from the limited ability of the handcrafted diagnosis functions, especially when dealing with heterogeneous data. In this paper, I will share my personal understanding of cognitive diagnosis and review our recent developments of CDMs mostly from a machine learning perspective. Meanwhile, I will show the wide applications of cognitive diagnosis.

## 1 Introduction

As is well known, individuals are different in both personal information (e.g., age, gender and location) and latent features (e.g., preference and proficiency). Therefore, in the applications of artificial intelligence, such as recommender systems, we need to discover these differences automatically for providing users with better services. Along this line, cognitive diagnosis is a type of assessment for measuring users’ proficiency profiles from their observed behaviors [DiBello *et al.*, 2006; Wu *et al.*, 2015].

Actually, cognitive diagnosis originates from the basic assumptions that (1) when finishing a task the observed interactive behaviors of users (e.g. the right or wrong responses) are determined by their latent cognitive states (e.g. the proficiency on the required skills), and (2) the cognitive state of each user is stable in a short period of time thus can be diagnosed [Gao *et al.*, 2021]. Let’s take Figure 1 as a toy example, two users  $u_1$  and  $u_2$  have just participated in a same examination with five tasks/exercises where different skills are needed for finishing each task. Though the overall scores got by these two users are the same (i.e. 60), with the help of cognitive diagnosis we can find from the radar maps that their proficiency on specific skills are quite different. Since

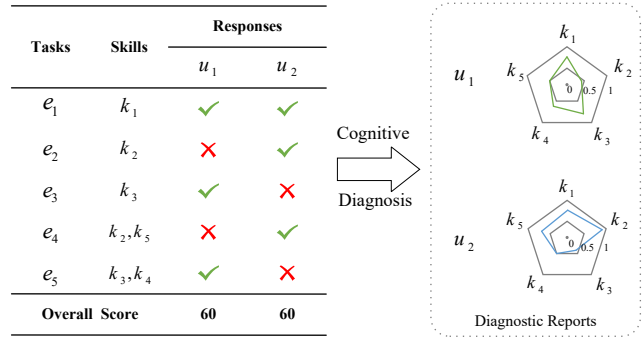


Figure 1: A toy example of the cognitive diagnosis process.

cognitive diagnosis is of significant importance for the domains like Games, Sports, Recruitment, E-commerce and Education, massive efforts have been devoted for this research problem, such as Deterministic Inputs, NoisyAnd gate model (DINA) [De La Torre, 2009], Item Response Theory (IRT) [Lord, 1952] and Multidimensional IRT (MIRT) [Reckase, 2009]. In fact, most of existing Cognitive Diagnosis Models (CDMs) are well designed based on psychometric theories, and the diagnostic results have a good interpretation. In spite of the importance of previous studies, the diagnose functions are usually handcrafted and can only exploit users’ numerical response records, leading to the insufficient ability in capturing the complex relationship among users, tasks and skills.

To address these problems, in the past few years we have conducted a comprehensive study on cognitive diagnosis and developed several cognitive diagnosis models mostly from a machine learning perspective. Our contributions include:

- We propose a new Neural Cognitive Diagnosis (NeuralCD) framework, which incorporates neural networks to learn the complex user-task interactions from heterogeneous data, for getting both accurate and interpretable diagnose results. We prove that the specific model NeuralCDM under NeuralCD is general enough to cover many traditional models such as IRT and MIRT.
- To diagnose the cognitive state of each user more precisely, we further propose several solutions considering the interdependencies among skills, the context-aware features of users, the partial orders between user responses and the data privacy issues, respectively.

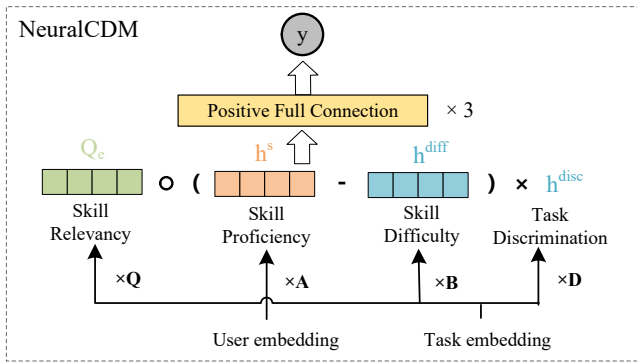


Figure 2: Neural cognitive diagnosis model.

- We show the way to combine cognitive diagnosis to improve the performance of personalized services in different applications, such as recommender systems, adaptive learning and computerized adaptive testing (CAT).

## 2 From Psychometric to Machine Learning

Input the interactive behaviors of users (e.g. right or wrong answers), cognitive diagnosis aims to discover their cognitive states (e.g. the proficiency on the required skills) based on these behaviors (a formal definition from the domain of education could be found in [Wang *et al.*, 2020]). For instance, the left part of Figure 1 demonstrates the input user behaviors and the right part demonstrates the diagnostic output, where user  $u_1$  is diagnosed to have the proficiency of 0.8 on skill  $k_3$ . Due to the fact that cognitive states are latent characteristics, cognitive diagnosis model is usually conducted and evaluated during the user performance/response prediction process.

Cognitive diagnosis was mostly developed from psychometric in the early years, among which item response theory (IRT) was one of the most typical and widely used solution [Lord, 1952]. In IRT, the probability that user  $u_i$  correctly finishes a task  $e_j$ , i.e.  $X_{ij} = 1$ , is assumed to increase with user ability while decrease with task difficulty, where user ability and task difficulty are described with two unidimensional and continuous parameters ( $\theta_i$  and  $\beta_j$  respectively). A typical item response function of IRT is  $P(X_{ij} = 1) = \text{Sigmoid}(a_j(\theta_i - \beta_j))$ , where  $a_j$  is the task discrimination parameter, and the interaction between the parameters was modeled in a logistic way. IRT builds a path from user ability to the response results, and measures user ability by fitting the corresponding response data. Extensions of IRT, such as Multidimensional IRT [Reckase, 2009], further enhance the fitting ability of IRT by updating the parameters to multidimensional ones. For instance,  $u_i$ 's proficiency on specific skill  $k_m$ ,  $\theta_{im}$ , can possibly be inferred with the help of a task-skill relevancy matrix. However, these methods rely on handcrafted diagnose functions that can only exploit users' numerical response records with insufficient ability.

Considering the limitations in psychometric solutions and the fact that multiple types of data, such as the images and the text description about the task, are now available in different applications, we have conducted a comprehensive study on leveraging machine learning approaches to design a

new type of cognitive diagnosis models [Wu *et al.*, 2015]. One important achievement is that we propose a new general and extendable Neural Cognitive Diagnosis (NeuralCD) framework which combines neural networks and psychometric assumptions to obtain both accurate and explainable diagnostic results [Wang *et al.*, 2020]. Specifically, NeuralCD uses multidimensional parameters to describe the cognitive states of users and the characteristics of tasks (e.g. difficulty), and neural networks are incorporated to learn the complex relationship between users and tasks from heterogeneous data. To ensure the interpretability of parameters, which is essential for cognitive diagnosis, we take two steps. The first step is using skill relevancy vector to align each dimension of user's cognitive state vector to a specific skill. The second step is applying the monotonicity assumption to the multi-layer neurons so as to make the cognitive state values variate in the same direction with the predicted probability of user successfully finishing the task. Inspired by IRT, an implemented model under this framework called NeuralCDM is shown in Figure 2, where  $Q_e$ ,  $h^s$ ,  $h^{diff}$  and  $h^{disc}$  denote skill relevancy of the task, skill proficiency of the user, skill difficulty and task discrimination respectively, and the monotonicity assumption is realized by positive full connection layers.

It is worth mentioning that NeuralCD is a general framework that can both learn from heterogeneous data and cover many traditional CDMs. For instance, as shown in Figure 3, NeuralCDM could be seen as the generalization of IRT by: (1) Representing the parameters from unidimensional to multidimensional; (2) Adding the skill relevancy vectors for the skill alignment; (3) Replacing the single Sigmoid interaction function to multilayer neural networks. These developments enhance the fitting abilities while preserving good interpretability of parameters [Wang *et al.*, 2020].

## 3 Our Extensions for CDMs

Given the basic cognitive diagnosis model like IRT and NeuralCDM, we further present the typical progress we have made in extensions for these models.

**Relation map-driven CDMs.** Most previous works consider cognitive diagnosis as an inter-layer interaction (e.g., user-task interactions or user-skill interactions) modeling problem, while the inner-layer structural relations, such as educational interdependencies among different knowledge concepts (e.g., the concept *Number* should be learned before concept *Arithmetic*), are still underexplored. Therefore, we propose a Relation map driven Cognitive Diagnosis (RCD) [Gao *et al.*, 2021] to uniformly capture both the inner-layer structures and the inter-layer interactions via a multi-layer relation map. We first represent users, tasks and skills as individual nodes in a hierarchical layout. We then construct three local maps to fully model the complex *user-task-skill* (e.g. *student-exercise-concept* in education) relations, which are *user-task* interaction map, *task-skill* correlation map and *skill dependency* map, respectively. The nodes can recursively aggregate information from their neighbors. Subsequently, we design a fusion layer to apply both node-level and map-level aggregation for each node and leverage an attention network to balance the multi-level information. The relation-enriched rep-

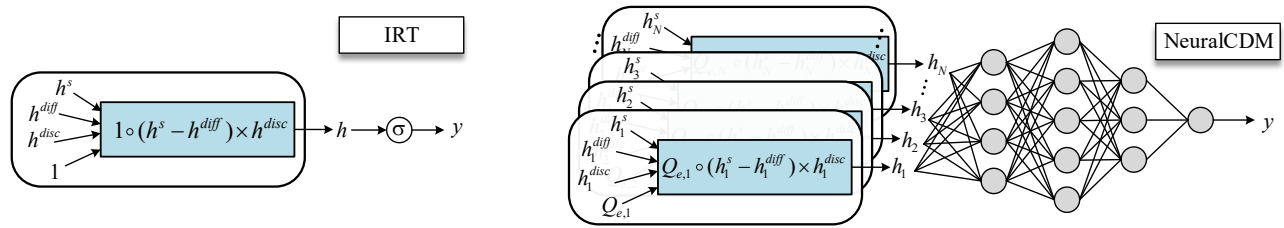


Figure 3: An illustration of the relationship between the traditional solution IRT and our NeuralCDM.

representations in RCD can be directly fused into existing CDMs and further improve their diagnose performance.

**Context-aware CDMs.** The context of users, such as the school, home and region of students, have important impacts on users’ implicit cognitive states. Therefore, users’ explicit response records, together with their implicit contexts, should be more able to reflect their proficiency profiles. We design a novel Educational context-aware Cognitive Diagnosis (ECD) [Zhou *et al.*, 2021] to incorporate plentiful educational context features into existing CDMs. Specifically, since educational contexts usually involve different types of contents, we first classify the diverse contexts into several fields. We then use an educational context modeling stage, which utilizes a hierarchical attentive network to represent the personalized influence from the contexts in each field and generate students’ external traits reflected by the contexts. Then, considering that educational contexts should have comprehensive influences on students, not limited to specific knowledge concepts, we design a diagnosis enhancement stage, in which the students’ external traits are adaptively integrated with their inner traits (i.e. the cognitive states reflected from the historical learning records). This general ECD framework is well defined to promote the performance of most existing CDMs. Experimental results also demonstrate interesting discoveries regarding the difference among different educational contexts in different countries and regions.

**Item response ranking for CDMs.** Monotonicity is a fundamental and critical condition in cognitive diagnosis theory, which assumes that user’s skill proficiency is monotonic with the probability of correct answers [Wang *et al.*, 2020]. However, the optimization process of current CDMs does not fully consider this property. Therefore, we introduce the pairwise learning strategy into cognitive diagnosis and propose a general Item Response Ranking framework (IRR) [Tong *et al.*, 2021] to directly model the monotonicity between item responses for CDMs. Specifically, as different items (tasks) usually have relations to non-overlapped skills/concepts, we design an item specific pair sampling method which samples responses from different users to the same item to keep the same related skills. Then, to incorporate the unobserved responses along with the observed responses, we conduct a two-branch sampling method, i.e., positive sampling and negative sampling. Based on these sampled pairs, we can introduce pairwise learning to model the partial order among response pairs, where a pairwise objective function is used to directly optimize the monotonicity. Extensive experiments indicate that CDMs with IRR can get significant improvements, and IRR can also provide some interpretable insights.

**Privacy-preserving CDMs.** Most of current CDMs have a common feature of aggregated data and centralized training, which is natural and convenient but suffers from the risk of privacy leakage. One possible solution is utilizing federated learning (FL) to build and aggregate CDMs while leaving private data isolated. Standard FL is an iterative framework with two core parts: client and server, where all clients train local models with consistent structures on user data individually and server aggregates the homogeneous local models to a global one. However, in cognitive diagnosis, various clients store inconsistent data since users tend to have different habits, thus FL is not suitable to be directly used. We present a Hierarchical Personalized Federated Learning (HPFL) framework [Wu *et al.*, 2021] that adapts to existing CDMs for isolated scenarios with three types of inconsistent clients: (1) Personal response data are non-independently identically distributed (Non-IID), which leads to *Statistical heterogeneity*; (2) Different information have different requirements of privacy, which results in *Privacy heterogeneity*; (3) Due to the different properties of the private data, the model structures among different clients should also be different, which causes *Model heterogeneity*. To address these problems, in the client stage, we define both public information and private information, and accordingly design a local model contains both public component and private component. Then, we propose a fine-grained personalized update strategy to fuse the corresponding components. In server stage, we utilize a differentiated component aggregation strategy to aggregate components received from clients. In this way, HPFL successfully applies FL to preserve the data privacy in cognitive diagnosis.

## 4 Applications of Cognitive Diagnosis

We show our explorations in applying cognitive diagnosis to provide users with better services, especially those applications in intelligent education.

**Recommender systems** aim to match users and proper items by capturing the preference and proficiency of users. Since cognitive diagnosis can measure the weaknesses and strengths of users, we have applied it to enhance user understanding in recommending new tasks (e.g. recommending exercises to students) [Huang *et al.*, 2019]. Specifically, we first design multiple long-term beneficial objectives, such as the smoothness of difficulty level of the recommended tasks. Then, we formalize these objectives as reward functions to optimize the recommendation strategy. As the most important part in the whole process, cognitive diagnosis is used to calculate the state of users at each time step.

**Adaptive learning** refers to providing different learning paths tailored to individual users, and the adaptiveness means the recommended items in the path will be dynamically adjusted based on users' abilities and preferences. Considering dynamic cognitive diagnosis<sup>1</sup> plays a key role in adaptive learning, we propose the CSEAL framework, short for Cognitive Structure Enhanced framework for Adaptive Learning [Liu *et al.*, 2019]. CSEAL views adaptive learning as a Markov Decision Process and can exploit both the knowledge level of users and knowledge structure (e.g., the prerequisite relations) of learning items for recommending both efficient and logical learning paths.

**Computerized adaptive testing (CAT)** is an emerging testing form in many standard tests, aiming to quickly and precisely diagnose the knowledge mastery levels of examinees via personalized testing procedure/exercises for each examinee. A typical CAT is made up of two iterative components: CDM and exercise selection strategy. At each step, after receiving examinee's answers, CAT uses a CDM to estimate her current knowledge level and then selects appropriate exercises for this examinee to answer in the next step. Existing CAT systems are usually inflexible as their exercise selection strategies combine with specific CDMs. Inspired by active learning in machine learning, we propose a novel Model-Agnostic Adaptive Testing framework (MAAT) [Bi *et al.*, 2020], which can adapt to different CDMs when selecting both high-quality and diverse exercises.

## 5 Conclusion and Future Research Directions

For getting both accurate and interpretable measurements of users' proficiency profiles from their observed behaviors, we developed a series of cognitive diagnosis models in the past several years. Actually, with the advent of big data, more and more efforts from the perspective of machine learning have been conducted. In order to facilitate the research on this area, we have published the source code of several typical CDMs<sup>2</sup> and collected the link of some public datasets<sup>3</sup>.

In the future, there still remain many open problems. For instance, how to collect and exploit the more detailed user behaviors during their responses; how to apply the idea of cognitive diagnosis to more scenarios besides the utilization in education [An *et al.*, 2021]; how to combine the advantages of psychometric theories and machine learning for designing more reasonable cognitive diagnosis framework.

**Acknowledgement.** Qi Liu would like to thank all the collaborators. This research was partially supported by grants from the National Key Research and Development Program of China (No. 2018YFC0832101), the National Natural Science Foundation of China (No. 61922073).

## References

[An *et al.*, 2021] Yanqing An, Qi Liu, Han Wu, Kai Zhang, Linan Yue, Mingyue Cheng, Hongke Zhao, and Enhong Chen. Lawyer-

pan: A proficiency assessment network for trial lawyers. In *ACM SIGKDD*, 2021.

[Bi *et al.*, 2020] Haoyang Bi, Haiping Ma, Zhenya Huang, Yu Yin, Qi Liu, Enhong Chen, Yu Su, and Shijin Wang. Quality meets diversity: A model-agnostic framework for computerized adaptive testing. In *IEEE ICDM*, pages 42–51. IEEE, 2020.

[De La Torre, 2009] Jimmy De La Torre. Dina model and parameter estimation: A didactic. *Journal of educational and behavioral statistics*, 34(1):115–130, 2009.

[DiBello *et al.*, 2006] Louis V DiBello, Louis A Roussos, and William Stout. 31a review of cognitively diagnostic assessment and a summary of psychometric models. *Handbook of statistics*, 26:979–1030, 2006.

[Gao *et al.*, 2021] Weibo Gao, Qi Liu, Zhenya Huang, Yu Yin, Haoyang Bi, Mu Chun Wang, Jianhui Ma, Shijin Wang, and Yu Su. Rcd: Relation map driven cognitive diagnosis for intelligent education systems. In *ACM SIGIR*, 2021.

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

[Liu *et al.*, 2019] Qi Liu, Shiwei Tong, Chuanren Liu, Hongke Zhao, Enhong Chen, Haiping Ma, and Shijin Wang. Exploiting cognitive structure for adaptive learning. In *ACM SIGKDD*, pages 627–635, 2019.

[Liu *et al.*, 2021] Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, Hui Xiong, Yu Su, and Guoping Hu. EKT: Exercise-aware knowledge tracing for student performance prediction. *IEEE TKDE*, 33(1):100–115, 2021.

[Lord, 1952] Frederic Lord. A theory of test scores. *Psychometric monographs*, 1952.

[Reckase, 2009] Mark D Reckase. Multidimensional item response theory models. In *Multidimensional item response theory*, pages 79–112. Springer, 2009.

[Shen *et al.*, 2020] Shuanghong Shen, Qi Liu, Enhong Chen, Han Wu, Zhenya Huang, Weihao Zhao, Yu Su, Haiping Ma, and Shijin Wang. Convolutional knowledge tracing: Modeling individualization in student learning process. *ACM SIGIR*, pages 1857–1860, 2020.

[Tong *et al.*, 2021] Shiwei Tong, Qi Liu, Runlong Yu, Wei Huang, Zhenya Huang, Zachary Pardos, and Weijie Jiang. Item response ranking for cognitive diagnosis. In *IJCAI*, 2021.

[Wang *et al.*, 2020] Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yuying Chen, Yu Yin, Zai Huang, and Shijin Wang. Neural cognitive diagnosis for intelligent education systems. In *AAAI*, volume 34, pages 6153–6161, 2020.

[Wu *et al.*, 2015] 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.

[Wu *et al.*, 2021] Jinze Wu, Qi Liu, Zhenya Huang, Yuting Ning, Hao Wang, Enhong Chen, Jinfeng Yi, and Bowen Zhou. Hierarchical personalized federated learning for user modeling. In *The Web Conference*, 2021.

[Zhou *et al.*, 2021] Yuqiang Zhou, Qi Liu, Jinze Wu, Fei Wang, Zhenya Huang, Wei Tong, Hui Xiong, Enhong Chen, and Jianhui Ma. Modeling context-aware features for cognitive diagnosis in student learning. In *ACM SIGKDD*, 2021.

<sup>1</sup>The knowledge tracing results [Liu *et al.*, 2021; Shen *et al.*, 2020] may be also used to represent the dynamic cognitive states.

<sup>2</sup><https://github.com/bigdata-ustc/EduCDM>

<sup>3</sup><https://github.com/bigdata-ustc/EduData>