MT-MCD: A Multi-task Cognitive Diagnosis Framework for Student Assessment

Tianyu Zhu[1], Qi Liu[1], Zhenya Huang[1], Enhong Chen[1](✉), Defu Lian[2], Yu Su[3], and Guoping Hu[4]

[1] Anhui Province Key Laboratory of Big Data Analysis and Application,
University of Science and Technology of China, Hefei 230026, China
{zhtianyu,huangzhy}@mail.ustc.edu.cn, {qiliuql,cheneh}@ustc.edu.cn

[2] University of Electronic Science and Technology of China, Chengdu 610054, China
dove.ustc@gmail.com

[3] Anhui University, Hefei 230039, China
firesysysy@163.com

[4] Anhui USTC IFLYTEK Co., Ltd., Hefei, China
gphu@iflytek.com

Abstract. Student assessment aims to diagnose student latent attributes (e.g., skill proficiency), which is a crucial issue for many educational applications. Existing studies, such as cognitive diagnosis, mainly focus on exploiting students' scores on questions to mine their attributes from an independent exam. However, in many real-world scenarios, different students usually participate in different exams, where the results obtained from different exams by traditional methods are not comparable to each other. Therefore, the problem of conducting assessments from different exams to obtain precise and comparable results is still underexplored. To this end, in this paper, we propose a Multi Task - Multidimensional Cognitive Diagnosis framework (MT-MCD) for student assessment from different exams simultaneously. In the framework, we first apply a multidimensional cognitive diagnosis model for each independent assessment task. Then, we extract features from the question texts to bridge the connections with each task. After that, we employ a multi-task optimization method for the framework learning. MT-MCD is a general framework where we develop two effective implementations based on two representative cognitive diagnosis models. We conduct extensive experiments on real-world datasets where the experimental results demonstrate that MT-MCD can obtain more precise and comparable assessment results.

Keywords: Student assessment · Cognitive diagonosis
Item Response Theory · Multi-task learning

1 Introduction

Educational Data Mining (EDM) is an emerging research field which seeks to develop methods for exploring data from educational settings (e.g., schools

J. Pei et al. (Eds.): DASFAA 2018, LNCS 10828, pp. 318–335, 2018.
https://doi.org/10.1007/978-3-319-91458-9_19

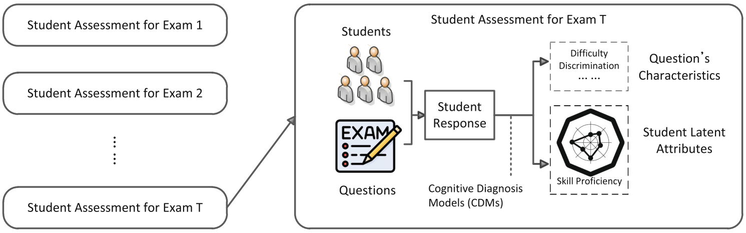

Student Assessment for Exam 1
Student Assessment for Exam 2
Student Assessment for Exam T
Student Assessment for Exam T
Students
EXAM
Questions
Student Response
Cognitive Diagnosis Models (CDMs)
Difficulty
Discrimination
… …
Question's Characteristics
Skill Proficiency
Student Latent Attributes

Fig. 1. Student assessment for exams

or learning systems). It contributes to learning theories, especially extracting instructive patterns from student learning, which helps understand students better and improve their learning [19,21].

One of the most important research issues in EDM is student assessment [4,17], where the goal is to discover student latent attributes (e.g., skill proficiency) based on their learning activities, such as exam scores [20] and feedback records in systems [26]. For better illustration, Fig. 1 shows a toy example of the general process of student assessment. From the figure, after collecting the responses of the students for each exam, the general goal of student assessment is to develop effective models to evaluate and diagnose student skills with the corresponding question characteristics (e.g., difficulty, discrimination). As the assessment results could be a fundamental task for various educational applications [22], such as targeted knowledge training and question recommendation, this issue has caused a great attention from both researchers and general publics [1].

In the literature, researchers have proposed many cognitive diagnosis models (CDMs) for the assessment along this line [12]. Existing CDMs have achieved a great success for student assessment in an independent exam, in which we argue that student A is more capable than student B if A gets a higher score than B. However, in most real-world scenarios, such as Graduate Record Examinations (GRE), students are allowed to take part in different exams [14]. If A gets a higher score than B when they participate in different exams, can we believe that A has a higher ability than B? In fact, educational psychologists claim that scores for students who participated in different exams could not compare directly [14]. Thus, in Fig. 1, it is not satisfied if we directly apply traditional CDM to conduct student assessment for all T exams. To this end, there is a urgent problem of conducting assessments from different exams simultaneously and it is necessary to propose an unified solution in such situation.

However, there are many challenges along this line. First, it is challenging to design a general unified framework to connect different exams for student assessment. Second, how to bridge connection with independent exams is a nontrivial problem. At last, in order to obtain comparable results for students, it is also difficult to find an appropriate way to estimate student latent attributes from independent exams simultaneously.

In this paper, inspired by the idea multi-task methods that can associate similar tasks together, we propose a Multi Task - Multidimensional Cognitive Diagnosis (MT-MCD) framework to conduct several independent student assessment tasks simultaneously. In this framework, given a set of exams containing response records of students and corresponding text information of questions, we first view the assessment in each exam as a single task and apply an existing CDM for each independent task. Then, we extract features from the question texts and develop an mapping matrix to bridge the connection with different tasks, which helps make tasks comparable. After that, we present a multi-task optimization method for the framework learning. Specifically, MT-MCD is a general framework and we propose two implementations based on two cognitive diagnosis models. i.e., M2PL model and M2PNO model. Finally, we conduct extensive experiments on real-world datasets, in which the experimental results demonstrate that MT-MCD can obtain more precise and comparable assessment results. The main contributions of this paper are summarized as follows:

- By conducting several independent student assessment tasks simultaneously, MT-MCD framework can estimate comparable student latent attributes. To the best of our knowledge, this is the first attempt to conduct several independent student assessment tasks at the same time.
- MT-MCD framework utilizes question's text information as supplemental material to bridge the connections among all assessment task, which ensures the comparability of student cognitive results.
- MT-MCD is a general framework which can apply many cognitive diagnosis models. Meanwhile, several student assessment tasks could be conducted simultaneously.

2 Related Work

In this section, we will introduce two aspects of related work: student assessment and multi-task learning.

2.1 Student Assessment

Student assessment is designed to measure specific knowledge structures and skills of students, which aims to find student latent attributes and provide information about their cognitive strength and weakness [5,12,16]. Educational psychologists have proposed a number of CDMs for student assessment [11].

Different CDMs are applied in specific occasions which can generate different types of student latent attributes (e.g. skill proficiency, guessing and slip factors) [25]. According to the assessment result, CDMs could be classified into two main categories: unidimensional CDM and multidimensional CDM. Unidimensional CDMs represent student latent attribute by a single dimensional variable [8]. For example, Item Response Theory (IRT) applies a mathematical expression that shows the relation between characteristics of a student (e.g., a latent trait) and the characteristics of the questions [13]. IRT provides a collection of models such as Two-Parameter Normal Ogive (2PNO) model and Two-Parameter Logistic (2PL) model [18]. One of the violates assumptions is the uni-dimensionality in the latent trait structure [15]. When the single dimensional variable is insufficient to indicate the complex and diverse student latent attributes, multidimensional CDMs would be necessary. Multidimensional Item Response Theory (MIRT) is a nature extension of IRT [19], and also contains a collection of models such as multidimensional extension of the 2PNO model (M2PNO [24]) and multidimensional extension of 2PL model (M2PL [18]). These MIRT models represent student latent attributes by a vector [18]. Multidimensional CDMs can assess a more complex student latent attributes.

However, most traditional CDMs aimed to do student assessment for an individual exam. In many real-word situations, student in different schools usually participate in different exams. So, it is eager to considered a framework which can conduct several independent student assessment simultaneously.

2.2 Multi-task Learning

Multi-task Learning (MTL) is a subfield of machine learning, in which several learning tasks are solved simultaneously by exploiting commonalities and differences across tasks [29].

MTL aims to improve the performance of each task by learning them jointly, which is different from single task learning. When adopting multi-task learning methods, independent tasks are learned simultaneously by utilizing shared information through tasks [28]. Multi-task learning has been applied in many different research fields, which utilizes the similarity information to conduct several tasks simultaneously to get higher performance [3,28], especially for those research problems where the amount of data per task is small. For example, Bansal et al. used multi-task method in text recommendations which a combination of content recommendation is trained by the text encoder network [2]. Yu et al. conducted image privacy protection by a deep multi-task learning algorithm to jointly learn more representative deep convolutional neural networks and more discriminative tree classifier [27].

In the research field of student assessment, it suffered from the problem that records available for each exam are limited. Therefore, applying MTL in student assessment may expand the sample size and generate more accuracy estimation. Therefore, it is necessary to consider a multi-task framework to optimize several independent student assessment tasks together based on the shared information.

3 Multi Task - Multidimensional Cognitive Diagnosis

In this paper, we propose a Multi Task - Multidimensional Cognitive Diagnosis (MT-MCD) framework which can implement several independent student assessment tasks simultaneously to generate more comparable and accurate student latent attributes than traditional CDMs. First, we formulate our problems in Sect. 3.1. Then we describe our MT-MCD framework in Sect. 3.2. At last, we illustrate wo implementations on the basis of MT-MCD in Sect. 3.3.

3.1 Problem Formulation

Given a set of exams $E = \{E_1, E_2, \cdots, E_T\}$, and student set $U_t = \{U_{t1}, U_{t2}, \cdots, U_{tU}\}$, question set $V_t = \{V_{t1}, V_{t2}, \cdots, V_{tV}\}$ for each exam $E_t (t = 1, 2, \cdots, T)$, we consider each student assessment on exam E_t as an independent task T_t $(t = 1, 2, \cdots, T)$. Note that, none of these students or questions sets overlaps among different tasks. In this paper, independent tasks are implemented simultaneously to generate comparable results.

Students' responses to questions are represented by matrix $\boldsymbol{Y_t}$ for task t, where Y_{tuv} is the student U_{tu}'s response on question V_{tv}. Usually, in traditional CDMs, Y_{tuv} equals 1 when U_{tu} answered V_{tv} correctly, and equals 0 otherwise. Therefore, each student response matrix $\boldsymbol{Y_t}$ is a binary matrix composed of 0 and 1. In addition, we also collect corresponding question's text information as a supplement to connect independent assessment tasks. For each task t, we have questions' text feature $\boldsymbol{F_t}$ which is generated from text information. Specifically, $\boldsymbol{F_t} = (\boldsymbol{F_{t1}}, \boldsymbol{F_{t2}}, \cdots, \boldsymbol{F_{tV}})$ is composed of row vector $\boldsymbol{F_{tv}}$ which represent the text feature for question V_{tv}. Therefore, our problem could be defined as:

Problem Definition: *Given a set of exams $E = \{E_1, E_2, \cdots, E_T\}$, student set U_T and question set V_t for each exam E_t, student response matrix $\boldsymbol{Y_t}$ and question information matrix $\boldsymbol{F_t}$ for each exam E_t, the main propose of our MT-MCD framework is: (1) Implement T independent student assessment tasks for each exam simultaneously to obtain comparable and accurate student latent attributes and question's characteristics (e.g., discrimination, difficult); (2) Predict student's performance on questions based on the student latent attributes and question's characteristics assessed by MT-MCD.*

For better illustration, Table 1 shows some important math notations.

Table 1. Some important notations

Notation	Description
T	Task number
$\boldsymbol{U}_t$, $\boldsymbol{V}_t$	Students and questions in task t
$\boldsymbol{Y}_t$	Students' response matrix for task t
$\boldsymbol{F}_t$	Questions' text feature matrix for task t
$\boldsymbol{\Xi}_t$	Questions' parameter matrix for task t
$\boldsymbol{\xi}_{tv}$	Parameters for vth question in task t
$\boldsymbol{\Theta}_t$	Student latent attributes for task t
$\boldsymbol{\theta}_{tu}$	Latent attributes for uth student in task t
$\boldsymbol{W}_t$	Mapping matrix for questions in task t
M	Dimension of student latent attributes
D	Dimension of question's text feature

3.2 Framework

We propose the MT-MCD framework to conduct several independent student assessment tasks simultaneously. Figure 2 illustrate MT-MCD framework.

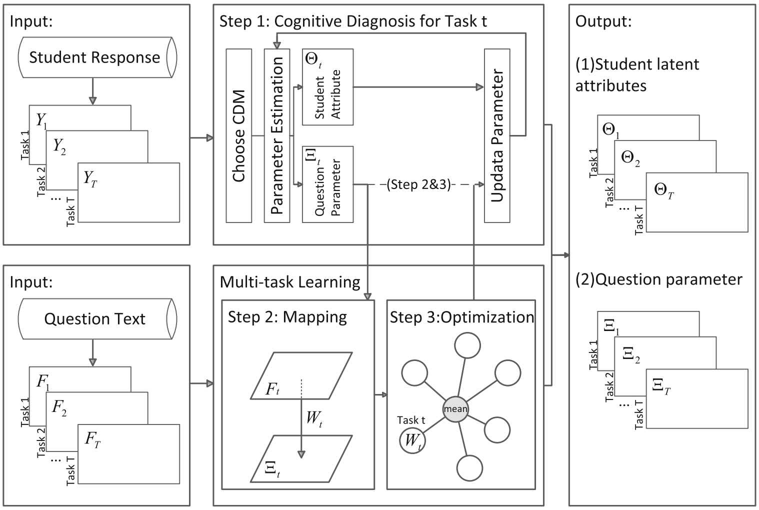


Fig. 2. MT-MCD framework

3.2.1 Step 1: CDM for Single Assessment Task

Step 1 of our proposed MT-MCD framework is to apply an existing multidimensional CDM to each individual exam. Therefore, we need to select a basic multidimensional CDM. The CDM which could be applied in MT-MCD framework can be constructed in the following way:

$$P(Y_{uv} = 1|\boldsymbol{\theta_u}, \boldsymbol{\xi_v}) \equiv f(\boldsymbol{\theta_u}, \boldsymbol{\xi_v}), \tag{1}$$

where $\boldsymbol{\theta_u}$ is a M-dimensional column vector which represents the latent attributes of student u (we will discuss the effectiveness of hyperparameter M experimentally in Sect. 4.4), and $\boldsymbol{\xi_v}$ is a row vector which represents the parameters of question v. When there are several independent tasks for different exams to be assessed simultaneously, function of CDM that t assessment tasks givens as (2):

$$f(\boldsymbol{\theta_u}, \boldsymbol{\xi_v}) = \prod_{t=1}^{T} f(\boldsymbol{\theta_{tu}}, \boldsymbol{\xi_{tv}}). \tag{2}$$

There are many existing CDMs which could be applied for each task separately to assess student latent attributes, however, the student latent attributes estimated from different individual assessment task are not comparable to each other. To solve this problem, step 2 and step 3 of MT-MCD framework connect independent tasks, and conduct these assessment tasks simultaneously to estimate accurate and comparable student latent attributes.

3.2.2 Step 2: Connecting Questions in Different Tasks

Questions play a significant role in student assessment, and it can be a great help to bridge the difference among diverse tasks. In our proposed MT-MCD framework, question's text feature are utilized as a supplement to those separate assessment tasks because it is easy to obtain and remain unchanged.

In order to connect question's text feature to its' parameters, we suppose there is a mapping matrix $\boldsymbol{W_t} \in \mathbb{R}^{D \times M}$ for each task t. The question's parameter $\boldsymbol{\xi_{tv}}$ could be represented by it's feature $\boldsymbol{F_{tv}}$ and mapping matrix $\boldsymbol{W_t}$:

$$\boldsymbol{\xi_{tv}} = m(\boldsymbol{F_{tv}}, \boldsymbol{W_t}, \boldsymbol{\xi_{tv}}), \tag{3}$$

where $\boldsymbol{F_{tv}}$ is a $1 \times D$ row vector represent the feature of question V_{tv}. The questions's parameters $\boldsymbol{\xi_{tv}}$, appear on both sides of the function m because in a specific implementation, part of the question's parameters may not represented by it's feature and mapping matrix. Therefore, the probability of student U_{tu}'s response to question V_{tv} is defined as follow:

$$P(Y_{tuv} = 1 | \boldsymbol{\theta_{tu}}, \boldsymbol{\xi_{tv}}) \equiv f(\boldsymbol{\theta_{tu}}, m(\boldsymbol{F_{tv}}, \boldsymbol{W_t}, \boldsymbol{\xi_{tv}})). \tag{4}$$

In step 2, we introduce the question's text feature as a supplement and connect it to the question's parameter. Therefor, we can obtain the interaction between students and question's text feature based on the selected CDM.

3.2.3 Step 3: Multi-task Learning Optimization

After we applied question's text feature matrix $\boldsymbol{F_t}$ into each assessment task t in step 2, we need to connect these individual tasks. There are two basic assumptions in our framework:

Assumption 1. Similar questions are similar in text feature.

Assumption 2. Similar questions should have similar parameters.

Based on these two assumptions, questions which have similar text feature should have similar parameters even in different tasks. Therefore, we assume that mapping matrix $\boldsymbol{W_t}$ for all tasks are close to each other.

Evgeniou and Pontil presented a multi-task learning method based on the minimization of regularization functionals, which is a natural extension to existing methods for single task learning [9]. Inspired by this, we define the optimization function of the mapping matrix $\boldsymbol{W_t}$ for each task who's regularization function penalizes the deviation from the mean:

$$\begin{aligned}
&\min_W \frac{1}{2} \sum_{t=1}^{T} \parallel \hat{\boldsymbol{Y}}_t - \boldsymbol{Y}_t \parallel_F^2 + \lambda \sum_{t=1}^{T} \parallel \boldsymbol{W_t} - \frac{1}{T} \sum_{s=1}^{T} \boldsymbol{W}_s \parallel_F^2 \\
&= \min_W \frac{1}{2} \sum_{t=1}^{T} (\sum_{u,v} (f(\boldsymbol{\theta}_{tu}, m(\boldsymbol{F}_{tv}, \boldsymbol{W}_t, \boldsymbol{\xi}_{tv})) - Y_{tuv})^2) + \lambda \sum_{t=1}^{T} \parallel \boldsymbol{W_t} - \frac{1}{T} \sum_{s=1}^{T} \boldsymbol{W}_s \parallel_F^2 .
\end{aligned} \tag{5}$$

The first part of Eq. (5) is the loss function which ensures the accuracy of estimation. The second part is the regularization which tries to make $\boldsymbol{W_t}$ closer to each other. If question's text feature mapping matrix $\boldsymbol{W_t}$ from different tasks are made close, then question's parameters for similar questions in different tasks will be closer.

We apply gradient descent (GD) method to optimize the Eq. (5). The gradient $\boldsymbol{W_t}$ is as follow, where f' is the first derivative of f:

$$\nabla \boldsymbol{W_t} = \sum_{u,v} (f(\boldsymbol{\theta_{tu}}, m(\boldsymbol{F_{tv}}, \boldsymbol{W_t}, \boldsymbol{\xi_{tv}})) - Y_{tuv}) f'(\boldsymbol{\theta_{tu}}, m(\boldsymbol{F_{tv}}, \boldsymbol{W_t}, \boldsymbol{\xi_{tv}}))(\boldsymbol{\theta_{tu}} \boldsymbol{F_{tv}})^T + \lambda((-\frac{1}{T}) \sum_{s=1}^{T} \boldsymbol{W_s} + \frac{2T-1}{T} \boldsymbol{W_t}). \tag{6}$$

By optimizing the mapping matrix $\boldsymbol{W_t}$, we can update the question parameter $\boldsymbol{\xi_t}$ for each question V_{tv} according to Eq. (3), and new question parameters are used in next estimation epoch in Step 1.

3.2.4 Output and Predicting

After processing by the 3-steps framework, MT-MCD, we could generate latent attributes for each student and parameters for every single question.

The outputs of MT-MCD framework are student latent attribute matrix $\boldsymbol{\Theta_t}$ and question parameters $\boldsymbol{\Xi_t}$ for each task t. For the students, we get latent attribute matrix $\boldsymbol{\Theta_t} = (\boldsymbol{\theta_{t1}}, \boldsymbol{\theta_{t1}}, \cdots, \boldsymbol{\theta_{tU}})$ for task t, which is composed of column vector $\boldsymbol{\theta_{tu}}$ represent the student U_{tu}'s latent attribute. For questions, we estimate its' parameter matrix $\boldsymbol{\Xi_t} = (\boldsymbol{\xi_{t1}}, \boldsymbol{\xi_{t2}}, \cdots, \boldsymbol{\xi_{tV}})^T$ for task t where $\boldsymbol{\xi_{tv}}$ is a row vector represented the question V_{tv}'s parameter.

Since similar questions gain close parameters, student latent attributes assessment corresponding to the questions they have answered would be more comparable. Thus, MT-MCD framework not only guarantee the accuracy of student latent attributes and question parameters, but also make questions and students from independent task more comparable.

The second purpose of MT-MCD is to predict student's performance on questions. MT-MCD helps us to assess comparable student latent attributes and corresponding question's parameters. Since the basic CDM we choose in step 1 describes the interaction between students and questions. We could easily adopt the output of MT-MCD to predict student performance by utilizing the probabilistic function for the selected basic CDM.

3.3 MT-MCD Implementation

As we mentioned before, many existing CDMs could be applied in MT-MCD framework to generate comparable student assessment result.

There are many existing CDMs proposed for student assessment, among which, Multidimensional Item Response Theory (MIRT) provides a collection of models that describe how questions and students interact to produce probabilistic response of correct or incorrect [8,23]. MIRT model is assumed to be a continuous probability function relating the student latent attribute $\boldsymbol{\theta}$ to the probability of correct response to a question with specified structural parameters. In this section, we illustrate MT-MCD framework with two MIRT models.

3.3.1 MT-MCD with M2PL Model

Multidimensional extension of the two-parameter logistic (M2PL) model [7] is a widely used MIRT model. First, we use M2PL model to illustrate how MT-MCD work.

When we select M2PL model as basic CDM, the cognitive diagnosis function Eq. (1) would be replaced by Eq. (7), which defines the probability that student U_{tu} answered question V_{tv} correctly by the changing shape of the standard logistic function [18] as:

$$f(\boldsymbol{\theta_{tu}}, \boldsymbol{\xi_{tv}}) = f(\boldsymbol{\theta_{tu}}, (\boldsymbol{\alpha_{tv}}, \beta_{tv})) = \frac{e^{(\boldsymbol{\alpha_{tv}} \boldsymbol{\theta_{tu}} + \beta_{tv})}}{1 + e^{(\boldsymbol{\alpha_{tv}} \boldsymbol{\theta_{tu}} + \beta_{tv})}}, \tag{7}$$

where the question's parameter $\boldsymbol{\xi_{tv}} = (\boldsymbol{\alpha_{tv}}, \beta_{tv})$ is composed of discrimination parameters $\boldsymbol{\alpha_{tv}} = (\alpha_{tv1}, \alpha_{tv2}, \cdots, \alpha_{tvM})$ and difficulty parameter β_{tv} [7]. We suppose that the mapping matrix $\boldsymbol{W_t}$ is connecting question's text feature $\boldsymbol{F_{tv}}$ and discrimination parameters $\boldsymbol{\xi_{tv}}$ as:

$$m(\boldsymbol{F_{tv}}, \boldsymbol{W_t}, \boldsymbol{\xi_{tv}}) = (\boldsymbol{F_{tv}} \boldsymbol{W_t}, \beta_{tv}), \tag{8}$$

Thus, the probability of student U_{tu}'s response to question V_{tv} correctly (Eq. (4)) could be replaced by Eq. (9) when selecting M2PL model as basic CDM:

$$f(\boldsymbol{\theta_{tu}}, (\boldsymbol{F_{tv}} \boldsymbol{W_t}, \beta_{tv})) = \frac{e^{(\boldsymbol{F_{tv}} \boldsymbol{W_t} \boldsymbol{\theta_{tu}} + \beta_{tv})}}{1 + e^{(\boldsymbol{F_{tv}} \boldsymbol{W_t} \boldsymbol{\theta_{tu}} + \beta_{tv})}}. \tag{9}$$

Then, for the multi-task learning optimization in step 3, the first derivative f' in gradient descent (Eq. (6)) could be replace by Eq. (10):

$$f'(\boldsymbol{\theta_{tu}}, (\boldsymbol{F_{tv}} \boldsymbol{W_t}, \beta_{tv})) = \frac{e^{(\boldsymbol{F_{tv}} \boldsymbol{W_t} \boldsymbol{\theta_{tu}} - \beta tv)}}{(1 + e^{(\boldsymbol{F_{tv}} \boldsymbol{W_t} \boldsymbol{\theta_{tu}} - \beta tv)})^2}. \tag{10}$$

3.3.2 MT-MCD with M2PNO Model

Besides M2PL model, there are many other forms of MIRT model. Multidimensional extension of the two-parameter normal ogive (M2PNO) model [24] derives from the assumption of normally distributed measurement error an is theoretically appealing on that bia s [6]. M2PNO is another widely used MIRT model, and When M2PNO is selected as basic CDM, the cognitive diagnosis function Eq. (1) would be replace by Eq. (11):

$$f(\boldsymbol{\theta_{tu}}, \boldsymbol{\xi_{tv}}) = f(\boldsymbol{\theta_{tu}}, (\boldsymbol{\alpha_{tv}}, \beta_{tv})) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\boldsymbol{\alpha_{tv}} \boldsymbol{\theta_{tu}} - \beta tv} e^{-\frac{t^2}{2}} dt = \Phi(\boldsymbol{\alpha_{tv}} \boldsymbol{\theta_{tu}} - \beta tv), \tag{11}$$

where the question's parameter $\boldsymbol{\xi_{tv}} = (\boldsymbol{\alpha_{tv}}, \beta_{tv})$ is also composed of discrimination parameter $\boldsymbol{\alpha_{tv}}$ and difficulty parameter β_{tv} [24], and $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z} e^{-\frac{x^2}{2}} dx$ is the normal cumulative density function (normal CDF). After the question's parameter $\boldsymbol{\xi_{tv}}$ is replaced by Eq. (8), the response probability for student U_{tu} on question V_{tv} (Eq. (4)) could be replaced by Eq. (12):

$$f(\boldsymbol{\theta_{tu}}, (\boldsymbol{F_{tv}} \boldsymbol{W_t}, \beta_{tv})) = \Phi(\boldsymbol{F_{tv}} \boldsymbol{W_t} \boldsymbol{\theta_{tu}} - \beta tv). \tag{12}$$

Correspondingly, the first derivative f' in gradient descent (Eq. (6)) is replaced by the following equation:

$$f'(\boldsymbol{\theta_{tu}}, (\boldsymbol{F_{tv}} \boldsymbol{W_t}, \beta_{tv})) = \varphi(\boldsymbol{F_{tv}} \boldsymbol{W_t} \boldsymbol{\theta_{tu}} - \beta tv) = \frac{1}{\sqrt{2\pi}} e^{\frac{(\boldsymbol{F_{tv}} \boldsymbol{W_t} \boldsymbol{\theta_{tu}} - \beta tv)^2}{2}}. \tag{13}$$

3.3.3 Conclusion

As we can see, many different existing CDMs could be applied in MT-MCD framework. Apart from this, there are many existing CDM could by applied in MT-MCD framework such as multidimensional partial credit model and multidimensional extension of Rasch model [18]. Therefore, MT-MCD framework could implement several independent student assessment tasks simultaneously and improve the accuracy and comparability of traditional CDMs. The effectiveness of MT-MCD would be proved in Sect. 4.

4 Experiment

In this section, we conduct extensive experiments to demonstrate the effectiveness of MT-MCD framework. Specifically, we use two implementations, which denoted as MT-MCD(M2PL) and MT-MCD(M2PNO), introduced in Sect. 3.3.

In the following section, we first introduce our experimental datasets and setups in Sect. 4.1. Then, we report experimental results of MT-MCD framework from the following four aspects:

- **Student Score Prediction**: Evaluate the accuracy of student assessment for each task in Sect. 4.2.
- **Student Attribute Evaluation**: Comparability evaluation of student attributes in Sect. 4.3.
- **Dimension Sensitivity of Student Attributes**: Evaluate the accuracy of MT-MCD with different dimensions of student attribute M in Sect. 4.4.
- **Question Parameter Evaluation**: Question analysis via the learned parameters in Sect. 4.5.

4.1 Dataset and Setups

4.1.1 Experimental Dataset

In the experiments, we use two real-world datasets supplied by iFLYTEK Co., Ltd., i.e., $MATH1$ and $MATH2$, to evaluate the effectiveness of MT-MCD

Table 2. Task statistics

(a) MATH1

T	E	School	U_t	V_t	Records
T_1	E_1	S_1	910	40	36,400
T_2	E_2	S_2	892	40	35,680
T_3	E_3	S_3	885	20	17,700
T_4	E_4	S_4	628	20	12,560
T_5	E_5		3,315[1]	40	132,600
Total:			3,315	160	243,940

(b) MATHA2

T	E	School	U_t	V_t	Records
T_6	E_6	S_5	698	22	15,356
T_7	E_7	S_6	437	51	22,287
T_8	E_8	S_7	711	42	29,862
T_9	E_9	S_8	842	19	15,998
T_{10}	E_{10}	S_9	849	50	42,450
T_{11}	E_{11}	S_{10}	1,575	46	72,450
T_{12}	E_{12}	S_{11}	726	21	15,246
T_{13}	E_{13}	S_{12}	523	21	10,983
T_{14}	E_{14}		2,839[2]	20	56,780
Total:			6,361	292	281,412

[1] All students in School S_1 to S_4

[2] Part of students in School S_5 to S_{12}

framework. Both datasets are about mathematics exam records for high school students collected from different schools in China.

In both datasets, students of the same school take the same exam, and each exam is taken by at least one school's students. Specifically, in $MATH1$, there are 4 senior high school students (S_1, S_2, S_3, S_4) participating in 5 different exams $(E_1, E_2, \cdots, E_5)$. In $MATH2$, 8 senior high schools $(S_5, S_6, \cdots, S_{12})$ participates in 9 different exams $(E_6, E_7, \cdots, E_{14})$. For task partition, we take each exam as a student assessment task in our MT-MCD framework. Therefore, there are 5 (9) tasks in $MATH1$ and $MATH2$, respectively. Table 2 shows the statistics of both datasets. In the following experiments, we take the first 4 (8) tasks for training, and the remaining one for testing.

We collect student records and the original texts of questions in all exams. For preprocessing, we first utilize the open source software *Jieba*[1] tool to segment each question's original text into a word sequence. Then, we extract question features by averaging the word embedding vector in the dimensions of $D = 60$.

4.1.2 Setups

We select the M2PL model and M2PNO model to illustrate MT-MCD framework, which have been introduced in Sect. 3.3.

When selecting M2PL as basic model, we apply a Maximum Likelihood Estimation (MLE) method in step 1 of MT-MCD framework [15]. In the following experiments, we set the numbers of MLE iterations to 1,500 for each task. When applying M2PNO model in MT-MCD, we apply a 5-step Gibbs Sampler [24] in step 1. In the following experiments, we set the number of iterations of gibbs sampler to 1,500 and estimate the parameter based on the last 1,000 samples to guarantee the convergency of the Markov Chain. Besides, we set regularization parameter λ in Eq. (5) to 0.001 in all of the following experiments.

[1] https://github.com/fxsjy/jieba.

4.1.3 Baseline Approaches

To demonstrate the effectiveness of MT-MCD framework, we compare two implementations i.e., MT-MCD(M2PL) and MT-MCD(M2PNO), with many models from various perspectives. First, we consider the traditional CDMs without MT-MCD framework on multiple tasks to evaluate whether MT-MCD improve the performance, we introduce M2PL_m and M2PNO_m method. Then, to evaluate the effectiveness of MT-MCD framework by applying a multi-task learning method in multiple tasks, we introduce M2PL_s and M2PNO_s method. At last, introduce a traditional multi-task learning (MTL) method from data mining area as the baseline. The details of them are as follows:

(1) M2PL_m [15]: Use M2PL model (Eq. (7)) on each task independently to generate parameters of students and questions.
(2) M2PNO_m [18,24]: Conduct the M2PNO model (Eq. (11)) on each task independently to generate parameters of students and questions.
(3) M2PL_s: Consider all tasks as a whole and applied M2PL model to do student assessment.
(4) M2PNO_s: Consider all tasks as a whole and apply M2PNO model to do student assessment.
(5) MTL [9,28]: A multi-task learning method to optimize several related classification task simultaneously. In this baseline approach, we use $(\bar{Y_t}u, \bar{Y_t}v, \boldsymbol{F_t v})$ as a feature vector or each response record for student u on question v in task t.

4.2 Student Score Prediction

One of the problems to be solved by MT-MCD is to obtain accurate student latent attributes and corresponding question's parameters. In this section, we evaluate the accuracy of the results assessed by MT-MCD. We compare the performance on predicting student's score against the baseline approaches. In other words, we evaluate the precision of predicting the students response to prove the accuracy of parameter estimation [25].

In this experiment, we evaluate the performance of MT-MCD from both regression and classification perspectives. For regression, we adopt *root mean square error* (RMSE) and *mean absolute error* (MAE) to quantify the distance between predicted scores and the actual ones. The smaller these values are, the better the results have. For classification, we consider the predicted scores which bigger than 0.5 as 1 and those less than 0.5 as 0, to compute *precision*, *recall* and $F1$, and the larger, the better.

Figure 3 shows the predicting results of our MT-MCD framework and baseline approaches on dataset $MATH1$ and $MATH2$. First, we construct different size of training sets with 90%, 80%, 70% and 60% of records for each student to observe how MT-MCD behave at different sparsity levels. Then, we set the dimensions of student latent attributes $M = 3$ to observe the effectiveness of MT-MCD framework. From this figure, we observe that, MT-MCD framework could improve the accuracy of the basic CDM which demonstrates that improve the accuracy of estimation for students and questions of basic CDM. This is because MT-MCD framework introduces the question's text feature as a supplement to do a multi-task optimization on several independent student assessment tasks. Second, the performance of MT-MCD frame work beats MTL method, this is because MT-MCD framework applied student assessment method to observe student latent attributes and question's parameters.

In many real-world occasions, students usually participate in different test, thus, MT-MCD helps to improve the student assessment accuracy.

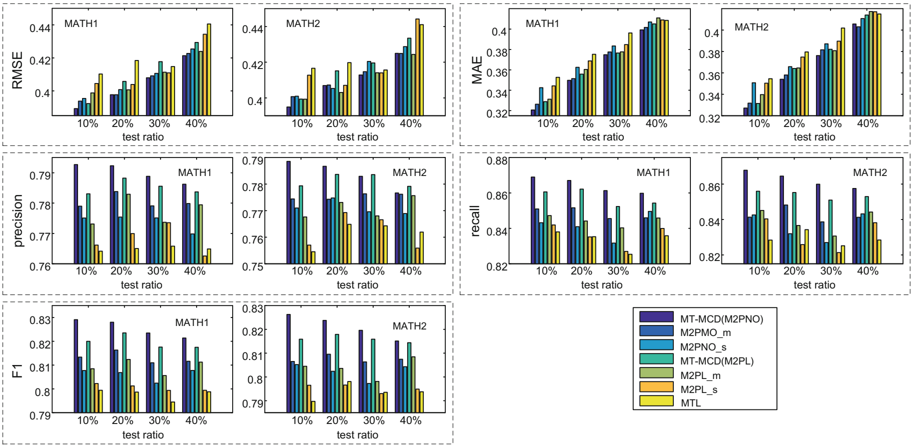


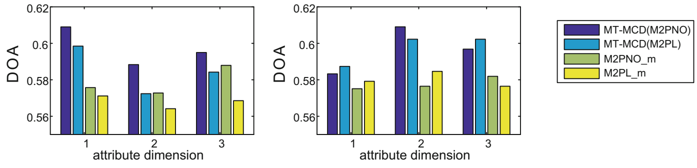
Fig. 3. Predicting student performance

4.3 Student Attribute Evaluation

In this subsection, we evaluate the comparability of student latent attributes assessed by MT-MCD framework. Intuitively, if student a masters better than student b on a specific dimension of latent attributes, a will have a higher probability to get larger score than student b when they participated in the same exam. We adopt $Degreeof Agreement(DOA)$ [10] metric for a specific dimension m, which is defined as:

$$DOA(m) = \sum_{a=1}^{U} \sum_{b=1}^{U} \frac{\delta(\theta_{am}, \theta_{bm}) \cap \delta(Sum_a, Sum_b)}{\delta(\theta_{am}, \theta_{bm})}, \tag{14}$$

where m refers to the ability dimensions, θ_{im} represent student ith ability on dimension m which assessed from task T_1 to T_4 in dataset $MATH1$ or T_6 to T_9 in dataset $MATH2$. Besides, Sum_i is the total score for student i in task T_5 or T_{14}. The higher the DOA value, the stronger comparability of student latent attributes.



(a)MATH1

(b)MATH2

Fig. 4. DOA

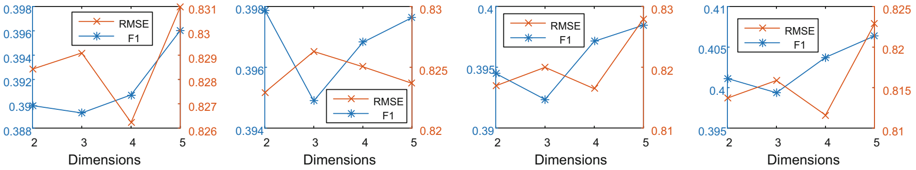
Figure 4 shows the result of *DOA* for MT-MCD(M2PNO), MT-MCD (M2PL), and comparative approach M2PNO_m, M2PL_m when the dimension of students is set to 3. As can be seen from the figure, the comparability of student latent attributes assessed by MT-MCD framework is higher than the basic CDM.

4.4 Dimension Sensitivity of Student Attributes

In this subsection, we apply MT-MCD(M2PNO) and MT-MCD(M2PL) to evaluate when the dimension of student latent attributes M is set to different values.

We set the dimension of student latent attribute M equals 2 to 5. Then, construct the size of training sets with 90% of records in dataset $MATH1$ and $MATH2$ in this experiment.

Figure 5 shows the results of MT-MCD framework whit different dimensions M. As we can see from this figure, as dimensions of student latent attributes increases, the performance of MT-MCD framework firstly increases but decreases when dimensions surpasses 3 with both MT-MCD(M2PNO) and MT-MCD (M2PL) in both datasets $MATH1$ and $MATH2$. Therefore, we can summarize that performance of $M = 3$ is better and more stable, and set $M = 3$ in Sects. 4.2 and 4.3 to obtaining the best results.



(a)MT-MCD(M2PNO) / MATH1

(b)MT-MCD(M2PNO) / MATH2

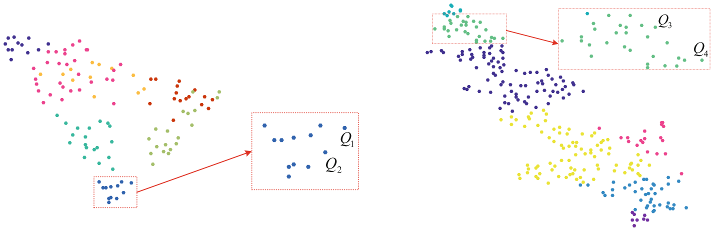
(c)MT-MCD(M2PL) / MATH1

(d)MT-MCD(M2PL) / MATH2

Fig. 5. Results of different dimensions

4.5 Question Parameter Evaluation

We emphasize that MT-MCD framework can make parameters of similar questions in different tasks closer, therefore, we evaluate the question parameters (discrimination, difficulty) estimated by MT-MCD framework with M2PNO basic CDM to prove the effectiveness in this section.



(a) MATH1

(b) MATH2

Fig. 6. Clustering result

For this experiment, we cluster the question's parameters estimated by MT-MCD to illustrate that similar question's parameters are closer. Specifically, we set the dimension parameter $M = 3$. Then we use the K-means clustering method to cluster the result of question's parameters into 7 categories. Finally, we adopt *t-SNE*[2] program to visualize these questions. Figure 6(a) and (b) shows the clustering result of question's parameter for dataset $MATH1$ and $MATH2$. Each dot in Fig. 6(a) represent a question in task T_1 to task T_4, and each dot in Fig. 6(b) represent a question in task T_6 to task T_{13}. The dots of same color belong to the same class clustered by K-means.

We check all the categories clustered by K-means, and questions in same category are similar to each other. For example, we find that all questions corresponding to these blue dots in Fig. 6(a) are about 'function' knowledge point, and all these green dots in Fig. 6(b) are about 'triangle' knowledge point. Further, a case study of several questions in these two categories are listed in Table 3. This experiment proves that MT-MCD framework makes parameters of similar questions in different student assessment tasks closer.

Table 3. Case study

	Task	Question description	Parameters $(\boldsymbol{\alpha},\beta)$
Q_1	T_1	The number of zero points for function $f(x) = 3x^2 + 2x - 4$ is?	((0.25, 0.66, 0.89), −1.77)
Q_2	T_2	The range of function $f(x) = x + \sqrt{1-2x}$ is?	((0.23, 0.50, 0.94), −1.21)
Q_3	T_8	A, B, C is the inner corner of the triangle, therefore, $sin(A + B) = sinC$?	((1.56, 1.69, 2.54), 0.99)
Q_4	T_{11}	$A = (5.1)$, $B = (1, 1)$, $C = (2.3)$, The shape of triangle $\triangle ABC$ is?	((1.45, 1.38, 2.07), 1.05)

[2] https://lvdmaaten.github.io/tsne/.

5 Conclusion and Future Work

In this paper, we proposed a MT-MCD framework to conduct several independent student assessment task simultaneously to generate accurate and comparable student latent attributes for students who participated in different exams. Specifically, we first applied an existing multidimensional cognitive diagnosis model to each independent student assessment task to estimate student latent attributes and corresponding question's parameters (e.g., discrimination, difficulty). Second, we introduced question's text information as a bridge to connect each independent assessment tasks. Then, we and employed a multi-task optimization method to make parameters of similar questions closer. New question's parameters updated by multi-task learning method will be adopted in cognitive diagnosis model for each student assessment task to obtain comparable student latent attributes. Extensive experiments on the real-world datasets clearly demonstrated the effectiveness of our propose framework MT-MCD which can assess accurate and comparable student latent attributes and question's parameters from independent student assessment tasks.

In the future, there are some directions for further studies. First, we will consider to find more relatedness between independent student assessment tasks. For example, student is an important aspect in assessment, the characteristics of students may connect individual student assessment together. Second, many natural language processing (NLP) method could be used for the pre-processing of question's text information.

Acknowledgments. This research was partially supported by grants from the National Natural Science Foundation of China (Grants No. 61672483, U1605251 and 91546103), and the Youth Innovation Promotion Association of CAS (No. 2014299).

References

1. Baker, R.S.J.D., Yacef, K.: The state of educational data mining in 2009: a review and future visions. JEDM-J. Educ. Data Min. **1**(1), 3–17 (2009)
2. Bansal, T., Belanger, D., McCallum, A.: Ask the GRU: multi-task learning for deep text recommendations. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 107–114. ACM (2016)
3. Bickel, S., Bogojeska, J., Lengauer, T., Scheffer, T.: Multi-task learning for HIV therapy screening. In: Proceedings of the 25th International Conference on Machine Learning, pp. 56–63. ACM (2008)
4. Cox, K., Imrie, B.W., Miller, A.: Student Assessment in Higher Education: A Handbook for Assessing Performance. Routledge, London (2014)
5. Cui, Y., Li, J.: Evaluating person fit for cognitive diagnostic assessment. Appl. Psychol. Meas. **39**(3), 223–238 (2015)
6. De La Torre, J., Minchen, N.: Cognitively diagnostic assessments and the cognitive diagnosis model framework. Psicología Educativa **20**(2), 89–97 (2014)
7. DiBello, L.V., Roussos, L.A., Stout, W.: 31A review of cognitively diagnostic assessment and a summary of psychometric models. Handb. Stat. **26**, 979–1030 (2006)

8. DiBello, L.V., Stout, W.: Guest editors' introduction and overview: IRT-based cognitive diagnostic models and related methods. J. Educ. Meas. **44**(4), 285–291 (2007)
9. Evgeniou, T., Pontil, M.: Regularized multi-task learning. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 109–117. ACM (2004)
10. Fouss, F., Pirotte, A., Renders, J.-M., Saerens, M.: Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. IEEE Trans. Knowl. Data Eng. **19**(3), 355–369 (2007)
11. Huebner, A.: An overview of recent developments in cognitive diagnostic computer adaptive assessments. Pract. Assess. Res. Eval. **15**(3), n3 (2010)
12. Huff, K., Goodman, D.P.: The demand for cognitive diagnostic assessment (2007)
13. Klaus, D., Kubinger, K.D.: On the revival of the rasch model-based LLTM: from constructing tests using item generating rules to measuring item administration effects. Psychol. Sci. **50**(3), 311 (2008)
14. Kuncel, N.R., Hezlett, S.A., Ones, D.S.: A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. Psychol. Bull. **127**(1), 162 (2001)
15. Lee, J.: Multidimensional Item Response Theory: An Investigation of Interaction Effects Between Factors on Item Parameter Recovery Using Markov Chain Monte Carlo. Michigan State University, Measurement and Quantitative Methods (2012)
16. Leighton, J., Gierl, M.: Cognitive Diagnostic Assessment for Education: Theory and Applications. Cambridge University Press, Cambridge (2007)
17. Liu, Q., Runze, W., Chen, E., Guandong, X., Yu, S., Chen, Z., Guoping, H.: Fuzzy cognitive diagnosis for modelling examinee performance. ACM Trans. Intell. Syst. Technol. (TIST) **9**(4), 48 (2018)
18. Reckase, M.: Multidimensional Item Response Theory, vol. 150. Springer, New York (2009). https://doi.org/10.1007/978-0-387-89976-3
19. Romero, C., Ventura, S., Pechenizkiy, M., d Baker, R.S.J.: Handbook of Educational Data Mining. CRC Press, Boca Raton (2010)
20. Saxon, P.D., Morante, E.A.: Effective student assessment and placement: challenges and recommendations. J. Dev. Educ. **37**(3), 24 (2014)
21. Scheuer, O., McLaren, B.M.: Educational data mining. In: Seel, N.M. (ed.) Encyclopedia of the Sciences of Learning, pp. 1075–1079. Springer, Boston (2012). https://doi.org/10.1007/978-1-4419-1428-6
22. Serrano-Laguna, Á., Torrente, J., Moreno-Ger, P., Fernández-Manjón, B.: Tracing a little for big improvements: application of learning analytics and videogames for student assessment. Procedia Comput. Sci. **15**, 203–209 (2012)
23. Sheng, Y.: Markov chain Monte Carlo estimation of normal ogive IRT models in MATLAB. J. Stat. Softw. **25**(8), 1–15 (2008)
24. Sheng, Y., Headrick, T.C.: A gibbs sampler for the multidimensional item response model. ISRN Appl. Math. **2012**, 14 (2012)
25. Wu, R., Liu, Q., Liu, Y., Chen, E., Su, Y., Chen, Z., Hu, G.: Cognitive modelling for predicting examinee performance. In: IJCAI, pp. 1017–1024 (2015)
26. Wu, R., Xu, G., Chen, E., Liu, Q., Ng, W.: Knowledge or gaming?: cognitive modelling based on multiple-attempt response. In: Proceedings of the 26th International Conference on World Wide Web Companion, pp. 321–329. International World Wide Web Conferences Steering Committee (2017)

27. Jun, Y., Zhang, B., Kuang, Z., Lin, D., Fan, J.: iPrivacy: image privacy protection by identifying sensitive objects via deep multi-task learning. IEEE Trans. Inf. Forensics Secur. **12**(5), 1005–1016 (2017)
28. Zhou, J., Chen, J., Ye, J.: Malsar: multi-task learning via structural regularization. Arizona State University, vol. 21 (2011)
29. Zhou, J., Chen, J., Ye, J.: Multi-task learning: theory, algorithms, and applications. In: https://www.siam.org/meetings/sdm12/zhou_chen_ye.pdf. Citeseer (2012)