

# Classification of Attribute Mastery Patterns Using Deep Learning

Dezhi Chen\*, Congcong Yan

Zhejiang Normal University, Hangzhou, China

Email: \*cdezhi@zjnu.cn

**How to cite this paper:** Chen, D.Z. and Yan, C.C. (2021) Classification of Attribute Mastery Patterns Using Deep Learning. *Open Journal of Modelling and Simulation*, 9, 198-210.

<https://doi.org/10.4236/ojmsi.2021.92013>

**Received:** March 18, 2021

**Accepted:** April 27, 2021

**Published:** April 30, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

It is very important to identify the attribute mastery patterns of the examinee in cognitive diagnosis assessment. There are many methods to classify the attribute mastery patterns and many studies have been done to diagnose what the individuals have mastered and or Montel Carl Computer Simulation is used to study the classification of the attribute mastery patterns by Deep Learning. Four results were found. Firstly, Deep Learning can be used to classify the attribute mastery patterns efficiently. Secondly, the complication of the structures will decrease the accuracy of the classification. The order of the influence is linear, convergent, unstructured and divergent. It means that the divergent is the most complicated, and the accuracy of this structure is the lowest among the four structures. Thirdly, with the increasing rates of the slipping and guessing, the accuracy of the classification decreased in verse, which is the same as the existing research results. At last, the results are influenced by the sample size of the training, and the proper sample size is in need of deeper discussion.

## Keywords

Cognitive Diagnosis Assessment, Deep Learning, Attribute Mastery Pattern, Classification

---

## 1. Introduction

Cognitive Diagnosis Assessments (CDAs) are used to evaluate the strengths and weaknesses of subjects in terms of cognitive skills learned [1]. Different from the other methods, CDAs not only provide the general results of evaluation, but also show detailed information of individual cognitive skills. Then, according to the evaluation results of CDAs, the subjects were remedied or instructed. The cognitive skills are also called attributes in the CDAs. The core of the CDAs is to iden-

tify or classify the attributes or cognitive skills which the subjects had mastered; it is the attribute mastery pattern of the subjects. The accurate and effective identification of the attribute mastery pattern will directly affect the results of the evaluation in the CDAs.

Psychometric models are often used to identify or classify attribute mastery patterns. In the past several years, the psychometric models of CDAs have been developed rapidly and variously. For example, a series of models have been explored based on the DINA model [2]: HODINA (Higher-order DINA) [3], GDINA (Generalized DINA) [4], RDINA (Reparameterized DINA) [4], HORDINA (higher-order reparameterized DINA) [5], P-DINA [6], time series G-DINA (sequential G-DINA) [7] and multi-level GDINA model [8]. However, these psychometric methods usually make strict assumptions about the specific probability function form of the subjects' item responses. There will be poor classifications if the observed data do not fit the model well [9], and the current psychometric methods usually work well for large-scale assessments and unfit for the small-scale at the classroom level [10] [11] [12]. And then nonparametric methods have been proposed. Those methods are HaiMing distance discrimination method [13], clustering method [10] [11] [14] and the general nonparametric classification method [12].

With the development of Artificial Intelligence (AI), great progress has been made in the core algorithms of AI. The Artificial Neural Network (ANN) algorithm in AI was once used to classify the attribute mastery pattern. Current researches indicate that neural networks can be used for CDAs to classify the attribute mastery patterns, and it has some advantages with assumptions that do not depend on the distribution of subjects and can minimize the error of classifications [15]. The results of some research showed that different parameters and attribute numbers of artificial neural network had an impact on classification accuracy. Compared with parametric methods, the performance of the ANN approach was obviously better, especially when model-data misfits were present [9]. The later study found that ANN was more accurate than the DINA model in recovering skill prerequisite relations [16].

Although the ANN has been used in CDAs and has more advantages than parametric methods under some conditions, there are still some unknown problems to be studied. How the structures of the attributes might affect the ANN method in CDAs, and whether the parameters of guessing and slipping would impact the accuracy of the classification? This paper does some exploration in the above aspects and contains the following parts. At first, the Deep Learning algorithm of the ANN is introduced. In the second part, the DINA model is shown, as it is the frame work of this study, which is used to generate the real response matrix. Thirdly, about the simulation study, the structures of the attributes, the parameters of guessing and slipping, and the sample of training in ANN, are discussed, focusing on how they may affect the accuracy of the DL in the CDAs. At last, the results of simulation are summarized and discussed.

## 2. Background Technology

### 2.1. Background and Motivation

DL is applied in learning evaluation, especially in online learning which needs to know the individual learning effect in real time, and CDAs is the most appropriate method at present. The application of DL algorithm in CDAs can not only evaluate individuals diagnostically, but also analyze and process a large number of data, which provides a technical method for cognitive diagnosis and evaluation of online learning in the future. Especially in recent years, with the continuous development and innovation of AI technology, discussion on related topics is not only the direction of education, psychological statistics and assessment technology, but also the only way to realize real AI through online learning.

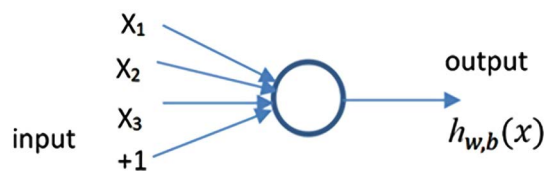
Classification accuracy is one of the important contents of CDAs. Previous studies have shown that the structure of attributes and the parameters of guessing and slipping are one of the important factors affecting the pattern of mastering attributes [17] [18] [19]. Although ANN has been used in cognitive diagnosis, the regularity of influencing factors of classification accuracy has not been reported. For example, with the increase of the number of attributes, the complexity of attribute relations, how the accuracy of classification is affected, and how the size of guessing and slipping parameters will bring about changes to the accuracy of classification. This paper discusses the regularity of the factors that affect the accuracy of classification diagnosis of DL in CDAs. On the one hand, it is to explore the application of DL in CDAs and expand the method of CDAs; on the other hand, it is to apply AI algorithm to education and psychological evaluation, so as to provide methods and basis for real-time evaluation of online learning effect in the future.

### 2.2. The introduction of the Deep Learning

Deep Learning (DL) [20] [21] [22] [23] [24] is one of the algorithms in ANN and it has been used widely to face recognition, natural language processing and image processing. In order to get to know the DL, the basic concept will be derived from the one simple neuron icon, shown in **Figure 1**. The left is the input quantities including  $x_1$ ,  $x_2$ ,  $x_3$  and the constant (+1). The middle is the node, the right is the output function  $h_{w,b}(x)$ .

The expression of the relationship between the input quantities and the output quantities is as follows (1).

$$h_{w,b}(x) = f(W^T X) = f(\sum W_i X_i + b) \quad (1)$$



**Figure 1.** One neuron icon.

The parameter  $b$  in the above is the intercept, shown as +1 in the **Figure 1**.  $f$  is the activation function. The sigmoid function is often chosen as the activation function, which is shown in the following (2).

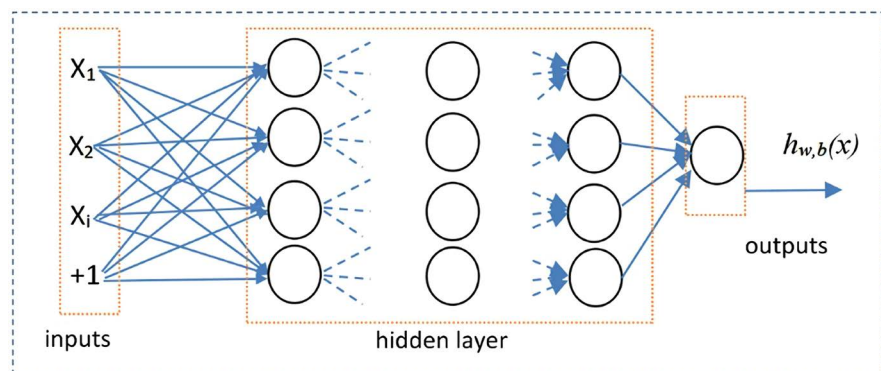
$$f(z) = \frac{1}{1 + \exp(-z)} \quad (2)$$

There is only one neuron and only one layer from the input to the output in **Figure 1**. It is the DL of one layer. The multi-layer is shown in **Figure 2**.

$n_l$  represents the number of layers, from left to right as shown in the figure above, which includes the input layer, the middle hidden layer, and the right most output layer. The parameter  $w, b = (w^{(1)}, b^{(1)}, w^{(2)}, b^{(2)}, \dots, w^{(n_l-1)}, b^{(n_l-1)})$ . It is the connection parameter between the  $j$ th unit of the  $l$ st layer and the  $(i + 1)$ th unit of the next layer; and it is the offset term. The initial values of the parameters  $w$  and  $b$  parameters are randomly generated from  $N(0, 1)$ . The back propagation algorithm is widely used to estimate the parameters  $w$  and  $b$  [23] [24]. And it has been compiled and realized in the MATLAB program. In the paper, it will be used to classify the mastery pattern of the subjects in CDAs, as the focus of this study is the application of the DL in CDAs, not the back propagation algorithm to estimate the parameters  $w$  and  $b$ . More information about the back propagation algorithm can be got from Rumerlhart, Hinton & Williams [23] and Lecun [24].

As an algorithm in Artificial Intelligence, DL has many advantages. For example, there is no overfitting in DL. When a model learns the details and noises in the training data to the extent, overfitting will occur, which will have a negative impact on the performance of the model on the new data. Regulation, Dropout and Early-stopping are often done to prevent overfitting. Another advantage is weak assumption that means it doesn't care about the distribution of the parameters during the estimation. The initial value of  $w$  and  $b$  can be generated randomly, and then they are estimated by the back propagation algorithm based on the sample data.

All the parameters including  $w$  and  $b$  have no strong distribution assumed and some are outside of the models themselves. But the application of DL is based on large-scale data, whether DL can be used on the small-scale data and



**Figure 2.** Multi-layer neuron icon.

the performance of DL with small-scale is still unknown. The necessary of the large-scale may be a shortcoming for DL.

### 2.3. Deep Learning in CDAs

The process from the input to the output quantities is the training or learning in DL. As the output quantities is the object, which is existed or classified. Take the diagnosis imagination of a dog for an example, the output is a picture of a dog, and the input is the imagination to be identified. The parameters  $w$  and  $b$  will be estimated by the process of training or learning using the back propagation algorithm. Then there is a question: what's the input and output quantities in the CDAs with DL. In the CDAs, the attribute mastery patterns of the subjects are classified by their response matrix on the examination, and then the cognitive diagnosis model is the connection of the skill mastered and the real response. In the DL used in the CDAs, the real response is the input, the real skills mastered are the output, and the connection between them is the hidden layers. But the real skills mastered of the subjects are unknown at all.

To overcome this limitation, we can use the ideal mastery patterns based on the attribute hierarchy, which are also the ideal item response data, consisting of response patterns that can be fully accounted for by the presence or absence of attributes without random errors or slips [9].

### 2.4. DINA Model

DINA model [2] is the basic cognition diagnosis model, and its item response function is as the following.

$$P(Y_{ij} = 1 | \alpha_i) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}} \quad (3)$$

subject to:  $0 < g_j < 1 - s_j < 1$

The item parameters  $s_j = P(Y_{ij} = 0 | \eta_{ij} = 1)$ ,  $g_j = P(Y_{ij} = 1 | \eta_{ij} = 0)$  are the probabilities of slipping and guessing respectively. In the paper, the individuals' real response matrix was generated under the frame of the DINA model. If the subject had gotten all the skills of the item  $i$ , then the  $\eta_{ij} = 1$  or  $\eta_{ij} = 0$ . For example, if the  $s_{ij} = g_{ij} = 0.05$ , the  $p(\alpha_i)$  can be computed. Then the random number  $u$  from the normal  $U(0, 1)$ , was compared with the  $p(\alpha_i)$ . If the  $p(\alpha_i)$  was larger than  $u$ , then the response of the subject was 1 and 0 otherwise. The simulation program of the real response matrix was compiled in the program of the MATLAB.

### 2.5. Q-Matrix

In the CDAs, the Q-Matrix was a  $J \times K$  binary matrix, which was used to relate attributes to categories [25] [26].  $J$  was the test length and  $K$  was the numbers of attributes. The element  $q_{jk}$  in row  $j$  and column  $k$  was equal to 1, if attribute  $k$  was needed by item  $j$  and 0 otherwise. Thus, given  $K$  attributes, there were at most  $2^K - 1$  distinct item attribute profiles. For example, there were 3 attributes as

shown in the following **Figure 3**, and then there were seven kinds of distinct item attribute profiles: 111, 110, 100, 101, 011, 010, and 001. Because if individuals got to the A2 attribute, they must master the A1 first, and then there were only three kinds of item attribute profiles (111, 110, 100), which can describe the relationship of the attribute structures. The 111, 110, 100 are the reduced  $Q_r$  matrix, which was used to relate the different attribute structure to the categories. The  $Q_r$  matrix can be computed based on the method developed by Tatsuoka and Ding [25] [26] [27], and more information could be found from their papers.

### 3. Simulation Study

#### 3.1. Attribute Structure in the Simulation

Four kinds of different structures of six attributes were compared in the simulation. The structures were shown in the following figures. The four kinds of structures were Linear, Convergent, Divergent and Unstructured respectively as shown in the following **Figure 4**.

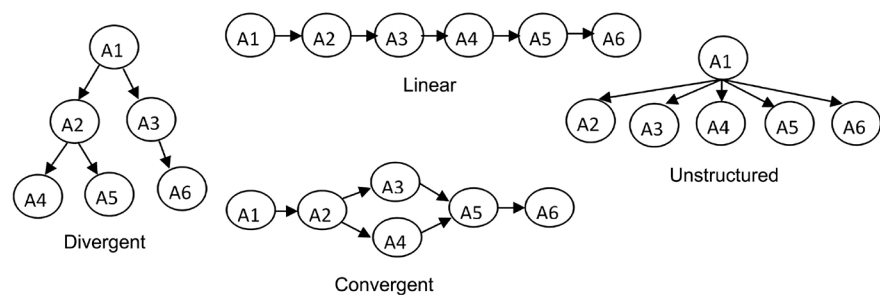
As **Figure 4** showed, the Linear of the hierarchy attributes was that if the individuals wanted to master the attribute A2, they must master the attribute A1 first. After the attribute A1 and A2 had both been mastered, the A3 could be mastered, which meant the later attributes were mastered based on the skills ahead. In the Divergent, after the A1 had been mastered, then the A2 and A3 could be mastered, but there was no correlation between the A2 and A3. The relationship among the attributes in the Convergent, was that the A6 mastered was based on the hierarchy of the A1, A2, A3, and A5 or that of the A1, A2, A4, and A5. In the Unstructured, the A1 was the prerequisites of the A2, A3, A4, A5 and A6, but there was no correlation among the A2, A3, A4, A5 and A6. In this paper, the performance of the DL in CDAs, was discussed under the four different kinds of hierarchy.

#### 3.2. Q-Matrix of the Attribute Structures

The ideal attributes mastery patterns from **Figure 4** are shown in **Table 1**.



**Figure 3.** The hierarchy of three attributes.



**Figure 4.** The four kinds of hierarchy with six attributes.

**Table 1.** The ideal attribute mastery patterns of **Figure 4**.

Divergent						Convergent						Linear						Unstructured						
1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	
1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	
1	1	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	
1	0	1	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	0	1	0	0	0	
1	1	0	1	0	0	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0	1	0	0	
1	0	1	0	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	1	0	
1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	
1	1	1	0	0	0	1	0	0	0	0	0	1	1	1	1	0	0	1	1	1	0	0	0	
1	1	1	0	1	0	1	1	0	0	0	0							1	1	0	1	0	0	
1	1	1	0	0	1	1	1	1	0	0	0							1	1	0	0	1	0	
1	1	1	1	0	0	1	1	1	1	0	0							1	1	0	0	0	1	
1	1	1	1	1	0	1	1	1	1	1	0							1	0	1	1	0	0	
1	1	1	1	0	1	1	1	1	1	1	1							1	0	1	0	1	0	
1	0	1	0	1	1													1	0	1	0	0	1	
1	1	1	0	1	1													1	1	1	1	0	0	
1	1	1	1	1	1													1	1	1	0	1	0	
																			1	1	1	0	0	1
																			1	0	0	1	1	0
																			1	0	0	1	0	1
																			1	1	0	1	1	0
																			1	0	0	1	1	1
																			1	1	0	0	1	1
																			1	0	1	0	1	1
																			1	1	1	0	1	1
																			1	0	0	1	1	1
																			1	1	0	1	1	1
																			1	0	1	1	1	1
																			1	1	1	1	1	1

followed. They were computed based on the method of  $Q_r$  matrix [26] [27]. There are 15, 12, 7 and 32 different kinds of ideal attributes master patterns respectively in the Divergent, Convergent, Linear and Unstructured structures.

### 3.3. The Simulation of Classification with DL in CDAs

The hierarchy of the cognitive attributes, the random error and guessing parameters of the examinee, and the frequency of the training and testing in DL are all considered in the paper. Four kinds of hierarchy with six attributes were studied, which were Divergent, Convergent, Linear, and Unstructured. The structures of the attributes were shown in **Figure 4**. The value of slipping ( $s$ ) and guessing ( $g$ ) was as the following:  $s = g = 0.05, 0.1, 0.15$  or  $0.2$ . The sample size was 1000, which was divided into two cases: one training number 500, testing number 500, and another training number 800, and testing number 200. The whole study had a total of  $4 \times 4 \times 2 = 32$  factors. The testing has the same length of 35 items. The credibility and feasibility of classifying the attribute mastery patterns based on DL were compared in different conditions.

In this simulation study, the response matrix of the examinee was the input layer of DL, and the attribute mastery pattern was the output layer. The identification was processed by training the existing data, including the simulated response matrix and ideal attribute mastery pattern. And then the subjects left were to be classified. For example, if the whole sample of simulation is 1000, then the response matrix and attribute mastery pattern of the 500 samples will be trained, the remaining 500 samples will be tested and classified based on the results of the previous sample trained.

In order to explore the value setting of model hyperparameters and optimizer hyperparameters, some trails were carried out. When set to the following values, the result is ideal in comparison: the number of the layers is 5, the input layer is the response matrix and its value is the number of the item response, the output is the attribute mastery patterns, and the value is the number of the mastery patterns; the middle is the hidden layer, and the number of the hidden neurons are 80, 80, and 60 respectively. As for the optimizer hyperparameter, the learning rate is 1, the number of epochs is 150, and the batch size is 100. Of course, there may be other settings that can make the results more accurate, and this requires constant experimentation.

The process of simulation consists of the following sections. First, when the number and structure of the attributes had been identified as the above, then the  $Q_r$  matrix was computed as shown in **Table 1**, and the ideal mastery pattern was also presented in **Table 1**. Second, the subjects were generated based on the ideal mastery pattern, the sample of the subject was 1000. The real response matrix was simulated under the DINA model. In addition, the sample of subject was divided into two groups, one group was for training and the other group was for testing or classification with the DL in CDAs. At last, the factors were discussed and summarized, which influenced the accuracy of the identification of the DL in CDAs.

### 3.4. Index of the Results

Pattern Match Ration ( $PMR$ ) and Marginal Match Ratio ( $MMR$ ) are used to



evaluate the accuracy and bias of the results of the different conditions [28] [29]. The formulations of the indicators are shown as the following.

$$PMR = \frac{\sum_{f=1}^F (M_k / N)}{F} \quad (4)$$

$$MMR = \frac{\sum_{f=1}^F (n_k / N)}{F} \quad (5)$$

$F$  is the frequencies of the simulation and  $M_k$  is the number that the attribute mastery patterns estimated are the same with the true patterns.  $k$  is the attribute  $k$ ,  $n_k$  is the number  $n_k$  examinees have been estimated correctly, if the sample of examinee is  $N$ . The mean  $PMR$  and the mean  $MMR$  across the  $F = 100$  replications were then obtained and reported as percentages respectively, under each data set.

#### 4. Results

The results of the  $PMR$  are shown in **Table 2**. It can be seen from the table that the  $PMR$  decreases with the increasing of the value of  $s$  and  $g$  in different attribute structures. What's more, the attribute structures affect the results of classification by DL. The accuracy of the classification by DL decreases with the complication of the structures of the attributes. For example, the Divergent structure is relatively the most complicated, and the accuracy of this structure is the lowest. On the contrary, the Linear is relatively the simplest and the accuracy of this structure is the highest. The sample size of the training in DL is also another factor to influence the results of  $PMR$ . It was found that the value of  $PMR$  will increase with the increasing of the sample size of the training.

The results of  $MMR$  are shown in **Table 3**. It is shown that the value of  $MMR$  will change with the difference of the attribute structure. The hierarchy was more complex, the value of  $MMR$  was much lower. For example, the divergent structure was the most complicated, the  $MMR$  of this structure was the lowest. The Linear was the simplest, and its'  $MMR$  was the highest. Secondly, the value of the  $MMR$  will decrease with the increasing of  $s$  and  $g$ . In addition, the sample

**Table 2.** The results of  $PMR$ .

Sample size of training	$s = g$	Divergent	Convergent	Line	Unstructured
500	0.05	0.9490	0.9923	0.9962	0.9794
	0.10	0.8847	0.9675	0.9924	0.9295
	0.15	0.7524	0.9256	0.9712	0.9021
	0.20	0.6306	0.8393	0.9629	0.8787
800	0.05	0.9518	0.9952	0.9978	0.9882
	0.10	0.8955	0.9737	0.9953	0.9868
	0.15	0.7652	0.9347	0.9753	0.9728
	0.20	0.6495	0.8513	0.9735	0.9672

**Table 3.** The results of *MMR*.

Attribute structure	$s = g = 0.05$		$s = g = 0.1$		$s = g = 0.15$		$s = g = 0.2$	
	Sample size of training							
	500	800	500	800	500	800	500	800
Divergent	0.9906	0.9965	0.9768	0.9798	0.8707	0.9524	0.9105	0.9156
Convergent	0.9968	0.9991	0.9944	0.9953	0.9867	0.9884	0.9681	0.9705
Linear	0.9995	0.9997	0.9989	0.9994	0.9954	0.9961	0.9931	0.9955
Unstructured	0.9945	0.9964	0.9724	0.9870	0.9647	0.9918	0.9751	0.9950

size of the training in DL will also have an impact on the *MMR*. The size of sample was much larger, and the value of *MMR* was much greater.

### 5. Discussion

Classifying the attribute mastery patterns is of great importance in CDAs. The accuracy of classification will directly affect the credibility of CDAs. A number of studies have made a fine comparison of the existing methods, some were used to develop the existing methods, and some were to explore the new methods. As an algorithm of ANN, DL algorithm has been widely used in industrial realization of AI. DL, as a nonparametric classification method, has also been used in CDAs, and it has been found that the performance of the DL was better than that of the parametric methods when the observed data do not fit the model well [9]. Based on the above background, this study is of great value and significance for the application of DL in CDAs, especially for the evaluation of online learning in the future.

The results appeared the more complicated the attribute structure, the lower the classification accuracy; the larger the slipping and guessing value, the less the classification accuracy; and the larger the training sample size, the higher the classification accuracy. But different from the previous methods [18] [30], the results in this paper was that DL had the lowest value on the Divergent structure, which was the most complicated. The order of accuracy from good to bad is linear, divergent, convergent and unstructured. These research results and rules suggest that when the guessing and slipping parameters are large and the attribute structure relationship is complex, we can consider increasing the sample size of learning and training to improve the accuracy of classification; secondly, when the attribute structure relationship is linear, we can choose DL as the first choice for classification.

The focus of this paper was to discuss the factors that influence the identification accuracy of the DL in CDAs. Only six attributes, four structures, four guessing and slipping rates are discussed. There are still many problems to be discussed in the near future. First of all, the comparison of the DL with the other exiting methods should be studied deeply to find out the advantages and disadvantages of the DL in CDAs. The DL is usually used on a large scale; whether it

can be used on a small scale and how it will perform? These will be explored later. Secondly, there are still some deficiencies in the practical application of this paper, and in the future, the application of the DL in CDAs should also be explored. Some factors are discussed in this paper, and there are still some issues that need to be further studied, such as the current number of the attributes was only 6, but with the increase of the number of attributes, how the complexity of attribute structure may affect the diagnostic classification of the DL; when the  $Q$  matrix is misplaced, what will happen in the DL of the CDAs with different hierarchy; how will the parameters of DL affect the classification, such as the number of the layers and neuron, and etc. These issues will be studied further. The above conclusion is only limited to the simulation answer, using DINA model, when attribute relationship is 6 attributes, with 4 different structures, and 4 different guessing and slipping rate conditions.

### Funding

This work was supported by the fund of China Scholarship Council (201608330425) and the Project of Educational Science Planning in Zhejiang Province (2019SCG313).

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Leighton, J.P., Gierl, M.J. and Hunka, S.M. (2004) The Attribute Hierarchy Method for Cognitive Assessment: A Variation on Tatsuoka's Rule-Space Approach. *Journal of Educational Measurement*, **41**, 205-236. <https://doi.org/10.1111/j.1745-3984.2004.tb01163.x>
- [2] Haertel, E.H. (1989) Using Restricted Latent Class Models to Map the Skill Structure of Achievement Items. *Journal of Educational Measurement*, **26**, 301-321. <https://doi.org/10.1111/j.1745-3984.1989.tb00336.x>
- [3] De la Torre, J. and Douglas, J.A. (2004) Higher-Order Latent Trait Models for Cognitive Diagnosis. *Psychometrika*, **69**, 333-353. <https://doi.org/10.1007/BF02295640>
- [4] De la Torre, J. (2011) The Generalized DINA Model Framework. *Psychometrika*, **76**, 179-199. <https://doi.org/10.1007/s11336-011-9207-7>
- [5] Decarlo, L.T. (2011) On the Analysis of Fraction Subtraction Data: The DINA Model, Classification, Latent Class Sizes, and the Q-Matrix. *Applied Psychological Measurement*, **35**, 8-26. <https://doi.org/10.1177/0146621610377081>
- [6] Tu, D.B., Cai, Y., Dai, H.Q. and Ding, S.L. (2010) A Polytomous Cognitive Diagnosis Model: P-DINA Model. *Acta Psychologica Sinica*, **42**, 1011-1020. <https://doi.org/10.3724/SP.J.1041.2010.01011>
- [7] Ma, W.C. and De la Torre, J. (2016) A Sequential Cognitive Diagnosis Model for Polytomous Responses. *British Journal of Mathematical and Statistical Psychology*, **69**, 253-275. <https://doi.org/10.1111/bmsp.12070>
- [8] Huang, H.-Y. (2017) Multilevel Cognitive Diagnosis Models for Assessing Changes

- in Latent Attributes. *Journal of Educational Measurement*, **54**, 440-480.  
<https://doi.org/10.1111/jedm.12156>
- [9] Cui, Y., Gierl, M. and Guo, Q. (2016) Statistical Classification for Cognitive Diagnostic Assessment: An Artificial Neural Network Approach. *Educational Psychology*, **36**, 1065-1082. <https://doi.org/10.1080/01443410.2015.1062078>
- [10] Kang, C.H., Ren, P. and Zhen, P.F. (2015) Nonparametric Cognitive Diagnosis: A Cluster Diagnostic Method Based on Grade Response Items. *Acta Psychologica Sinica*, **47**, 1077-1088. <https://doi.org/10.3724/SP.J.1041.2015.01077>
- [11] Kang, C.H., Ren, P. and Zhen, P.F. (2016) The Influence Factors of Grade Response Cluster Diagnostic Method. *Acta Psychologica Sinica*, **48**, 891-902.  
<https://doi.org/10.3724/SP.J.1041.2016.00891>
- [12] Chiu, C.Y., Sun, Y. and Bian, Y.H. (2018) Cognitive Diagnosis for Small Educational Programs: The General Nonparametric Classification Method. *Psychometrika*, **83**, 355-375. <https://doi.org/10.1007/s11336-017-9595-4>
- [13] Luo, Z.S., Yu, X.F., Gao, C.L. and Peng, Y.F. (2015) A Simple Cognitive Diagnosis Method Based on Q-Matrix Theory. *Acta Psychologica Sinica*, **47**, 264-272.  
<https://doi.org/10.3724/SP.J.1041.2015.00264>
- [14] Guo, L., Yang, J. and Song, N.Q. (2018) Application of Spectral Clustering Algorithm under Various Attribute Hierarchical Structures for Cognitive Diagnostic Assessment. *Journal of Psychological Science*, **41**, 735-742.
- [15] Gierl, M.J., Cui, Y. and Hunka, S. (2008) Using Connectionist Models to Evaluate Examinees' Response Patterns on Tests. *Journal of Modern Applied Statistical Methods*, **7**, 234-245. <https://doi.org/10.22237/jmasm/1209615480>
- [16] Guo, Q., Cutumisu, M. and Cui, Y. (2017) A Neural Network Approach to Estimate Student Skill Mastery in Cognitive Diagnostic Assessments. *Proceedings of the 10th International Conference on Educational Data Mining*, Wuhan, 25-28 June 2017, 370-371.
- [17] Tu, D.B., Cai, Y. and Dai, H.Q. (2013) Comparison and Selection of Five Noncompensatory Cognitive Diagnosis Models Based on Attribute Hierarchy Structure. *Acta Psychologica Sinica*, **45**, 243-252. <https://doi.org/10.3724/SP.J.1041.2013.00243>
- [18] Cai, Y., Tu, D.B. and Ding, S.L. (2013) A Simulation Study to Compare Five Cognitive Diagnostic Models. *Acta Psychologica Sinica*, **45**, 1295-1304.  
<https://doi.org/10.3724/SP.J.1041.2013.01295>
- [19] Gao, X.L., Wang, D.X., Yan, C. and Tu, D.B. (2018) Comparison of CDM and Its Selection: A Saturated Model, a Simple Model or a Mixed Method. *Journal of Psychological Science*, **41**, 727-734.
- [20] Bengio, Y., Lamblin, P., Popovici, D. and Larochelle, H. (2007) Greedy Layer-Wise Training of Deep Networks. In: *Advances in Neural Information Processing Systems* 19, MIT Press, Cambridge, 153-160.
- [21] Schmidhuber, J. (2015) Deep Learning in Neural Networks: An Overview. *Neural Networks*, **61**, 85-117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- [22] Hinton, G.E., Osindero, S. and Teh, Y.-W. (2006) A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, **18**, 1527-1554.  
<https://doi.org/10.1162/neco.2006.18.7.1527>
- [23] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) Learning Representation by Back-Propagating Errors. *Nature*, **323**, 533-536.  
<https://doi.org/10.1038/323533a0>
- [24] Lecun, Y. (1987) Modeles connectionistes de l'apprentissage. Ph.D. Thesis, University

de Paris VI, Paris.

- [25] Tatsuoka, K.K. (1995) Architecture of Knowledge Structures and Cognitive Diagnosis: A Statistical Pattern Recognition and Classification Approach. In: Nichols, P.D., Chipman, S.F. and Brennan, R.L., Eds., *Cognitively Diagnostic Assessment*, Erlbaum, Hillsdale, 327-359.
- [26] Tatsuoka, K.K. (2009) Cognitive Assessment—An Introduction to the Rule Space Method. Routledge, New York. <https://doi.org/10.4324/9780203883372>
- [27] Ding, S.L., Wang, W.Y. and Luo, F. (2012) The Q Matrix and the Q Matrix Theory in the Cognitive Diagnosis Assessment. *Journal of Jiangxi Normal University (Natural Science Edition)*, **36**, 441-445.
- [28] Hu, J.X., Miller, M.D., Corinne, A. and Chen, Y.-H. (2016) Evaluation of Model Fit in Cognitive Diagnosis Models. *International Journal of Testing*, **16**, 119-141. <https://doi.org/10.1080/15305058.2015.1133627>
- [29] Cohen, A.S., Kane, M.T. and Kim, M.S.H. (2001) The Precision of Simulation Study Results. *Applied Psychological Measurement*, **25**, 136-145. <https://doi.org/10.1177/01466210122031966>
- [30] Wang, W.Y., Song, Y.H. and Ding, S.L. (2016) Application of Neural Networks and Support Vector Machines to Cognitive Diagnosis. *Journal of Psychological Science*, **39**, 777-782.