A Method of Cognitive Diagnosis for Objective Problems Based on a Bayesian Network

Huanran Lin, Dongfeng Liu*

Guangdong University of Technology, Guangzhou, China

1608823150@qq.com, *liudf@gdut.edu.cn

Abstract. With the popularity of online education in recent years, intelligence tutoring systems (ITSs) have developed rapidly. One of the key technologies to embody ITS intelligence is its evaluation mechanism. However, the existing evaluation mechanism either stays at the ability level and cannot go deep into the analysis of students' skills or can only conduct cognitive analysis and cannot give ability evaluation. To this end, an interpretable cognitive diagnosis model (CDM) based on a Bayesian network has been proposed for a particular kind of problem, an objective problem. Specifically, we first approximate the ability of examinees empirically, and then item response theory (IRT) is introduced to model the examinees' proficiency in some skills. Finally, educational hypotheses and slip and guess factors were combined to infer the examinees' scores on a problem. At the end of the paper, a specific example was presented to show the good interpretability of the model, and a cognitive diagnostic result can be easily derived. Experiments on real-world datasets prove that the CDM we proposed can reasonably estimate students' ability level distribution, and the prediction task proves the effectiveness of CDM in-depth diagnosis.

Keywords: BAYESIAN NETWORK, COGNITIVE DIAGNOSIS MODEL, EDUCATIONAL ASSESSMENT, ITEM RESPONSE THEORY.

1 Introduction

Cognitive diagnosis is an intelligent education task that aims to learn students' cognitive states on knowledge concepts based on historical answering logs over questions[1].Psychometrically, educational achievement tests have been deeply studied. Tatsuoka et al. creatively proposed and developed rule space model theory and Q-matrix theory. Dibello proposed a unified model considering the completeness of the Q-matrix and properties of the problem tested. This model is fine but complex and cannot be used in actual operation. Matrix factorization (MF) is a classical modelling technique that is widely used to model examinees by latent factors[2], but the latent factors in MF are not helpful to explain students' cognitive state. Although the previous methods have solved many problems, most of them can only consider one aspect of the process of cognitive diagnosis and cannot have both ability level evaluation and cognitive diagnosis at the same time.

Y. Feng et al. (eds.), *Proceedings of the 4th International Conference on Internet, Education and Information Technology (IEIT 2024)*, Atlantis Highlights in Social Sciences, Education and Humanities 26,
https://doi.org/10.2991/978-94-6463-574-4_13

There are two assessment tasks we need to handle. First, students are given a general ability score to show how far they have mastered a knowledge field. Second, model the knowledge structure and introduce slip and guess factors that precisely infer the proficiency of specific skills in a certain field of knowledge.

To address these tasks, this text proposes a three-layer cognitive diagnosis model for modelling the cognitive process of examinees. CDM is a generative model to capture the relationship between examinees' inner ability level and external performance on problems. First, an a priori probability distribution describing the ability levels of examinees is empirically determined based on the knowledge test results. Then, with the help of IRT theory, the parameter relationship between ability and skill proficiency is determined. Finally, we assumed that the skill interactions on objective problems satisfy a conjunctive relation, and considering two exceptions, slip and guess, we simulated the examinees' score generation.

2 Related Work

Cognitive diagnosis is the judgment of the student's cognitive ability, is a wide-spread concern in educational science[3].Many cognitive diagnosis models (CDMs) have been developed to profile students. Tatsuoka creatively put forward the matrix theory, and then the theory was further promoted, including representative models such as the DINA (deterministic inputs, noisy and gate) model, NIDA (noisy inputs, deterministic "and" gate) model, and G-DINA (generalized DINA) model. The diagnostic model represented by DINA was subsequently improved and further applied in many scenarios. Hou et al. [4] studied the impact of unified or nonunified item functions on learning achievement evaluation and found that the unified item function is better. Liu et al. [5] explored the method of modelling students' proficiency in various skills with a cognitive diagnostic model to divide them into different learning teams according to their learning characteristics. Wang et al. [6] propose two implementations of NeuralCD by specializing the required concepts of each exercise, i.e., the NeuralCDM with traditional Q-matrix and the improved NeuralCDM+ exploring the rich text content. Therefore, the CDM proposed in this paper ensures the accuracy of evaluation and simplicity of parameter estimation, which is beneficial to real-time interaction with students.

3 Preliminaries

We introduce the fundamental ideas of cognitive diagnostics in this section. demonstrate their application in educational achievement testing and introduce the characteristics of objective problems in preparation for later model construction.

Cognitive diagnosis is the core of the new generation of test theory and belongs to the category of the "cognitive level paradigm". The process of cognitive diagnosis is shown in Fig. 1. Examinees are required to complete a test paper after they have completed a phase of study to evaluate their performance. Based on their feedback results (in the form of a score matrix, each line corresponds to an examinee, and each column corresponds to a problem), psychologists have proposed a probabilistic model to infer their level of knowledge mastery (the proportion of mastery of each knowledge point involved in the problems in the test paper); this model is called the cognitive diagnosis model. Finally, a cognitive outcome report is generated for each student to help them understand their own strengths and weaknesses to promote their long-term progress. In the process of cognitive diagnosis, in addition to the score matrix, the Q-matrix is also needed to identify the skills necessary for each problem, and a Q-matrix is prepared by examination designers or education experts in advance. 1 indicates that the skill is needed, and 0 does not. The meaning of skills is very broad, which can be problem-solving skills or knowledge.

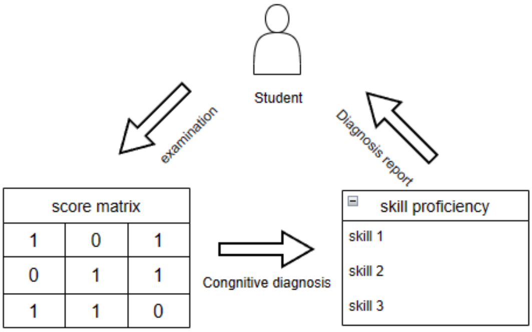

Fig. 1. cognitive diagnosis

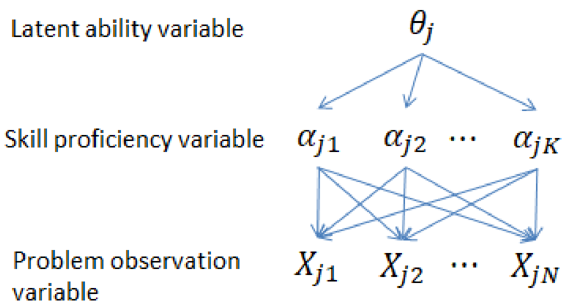

Fig. 2. Bayesian network assessment model

A good cognitive diagnosis includes two types of questions: subjective and objective problems. However, due to the limitations of the discrete nature of usual Bayesian networks, we only consider objective problems. Indeed, whether a question is subjective or objective is not naturally determined; it is determined by the type of answer and the method of scoring. The answer to an objective question should be deterministic and unique, and its scoring results should only be right or wrong, but not partially correct as the subjective question. For example, in choice or fill-in-the-blank questions, subjective questions may be treated as objective questions in some cases. As a result, a series of classic psychometrical models based on objective problems is proposed.

4 Cognitive Diagnosis Model Framework and Construction

In this section, the three-layer network model was proposed, which takes the traditional evidence-centered design (ECD), and evidence reasoning using a Bayesian network in the ECD context has been proven to be very effective. ECD provides a guiding ideology for analysing and developing diagnostic mechanisms. The variables to be measured are one or more variables related to the knowledge, skills, and abilities we want to measure. They are unknown variables and are a subset of variables in Bayesian networks. The observed variables are the results of the students' responses, which are known variables. Observation variables accumulate evidence (observation variable values) in the task to update their value distribution.

The Bayesian network diagnosis model we designed is shown in Fig.1.One of the most important tasks for our diagnosis is to infer the proficiency of student j in mastering skill k. We first formalize the concept of S_{jk}. Use α_{jk} to indicate whether student j has mastered skill k, α_{jk}=1 indicate mastered, α_{jk}=0 indicate not mastered, and $\alpha_{jk} \in \{0,1\}$. From the meaning of proficiency, we define the following formula 1:

$$S_{jk} = E(\alpha_{jk}) \tag{1}$$

because α_{jk} is a binary variable, so S_{jk}=P (α_{jk}=1), that is, the proficiency S_{jk} of student j mastering skill k is equivalent to the probability of student j correctly mastering skill k. The three-layer Bayesian network evaluation model is shown in Figure 2. For better illustration, Table 1 shows some important math concepts and symbols.

Table 1. Some important mathematical codes

Mathematical Codes	Description
θ_j	the ability level parameter of examinee j
S_{jk}	the proficiency of examinee j master skill k
X_{ji}	score of examinee j on problem i
α_{jk}	whether examinee j has mastered skill k
η_{ji}	whether examinee j has mastered all the skills measured by problem i

Next, consider layer 2 and layer 3 of the network. They were intended to collect proof of more in-depth understanding and knowledge. Each task observable (X_{ji}) was linked to appropriate proficiency nodes (α_{jk}), so the observable evidence could be conversely propagated appropriately through the network to update proficiencies and latent ability. The structural relationship between the second and third layers in the network is determined by the Q-matrix. In addition to determining the relevant skills involved in the item, we also need to consider how the interaction between skills affects the score of students on an item. There are a series of interaction types between parent nodes and observable nodes for developers to choose. The skill's interaction with problems can be mainly categorized into conjunctive and compensatory. Conjunctive distribution means that all skills are necessary to find the proper way to solve the problem. Based on compensatory distribution, having more talent will "compensate" for having less talent, and the probability of success is determined by the weighted sum of the skills. Examinees must possess all necessary skills without leaving any out to appropriately respond to an objective problem, which has a single standard response. Therefore, it is typically considered that the skill's interaction with objective problems is conjunctive. Therefore, the second assumption we proposed is as follows: The interaction of skills on objective problems is conjunctive.

According to the assumption 2, η_{ji} indicates whether student j has mastered all the skills of item i, $\eta_{ji} \in \{0,1\}$, $\eta_{ji} = 1$ means have mastered, otherwise have not. $q_i = (q_{i1}, q_{i2}, \cdots, q_{iK})$, $q_{ik} \in \{0,1\}$ is the measurement mode of item i defined by Q matrix.

$$\eta_{ji} = \prod_{k=1}^{K} \alpha_{jk}^{q_{ik}} \tag{2}$$

In addition to accurately simulating future exams, experts are also interested in the actual factors that influence how well test takers perform. In fact, the score of an examinee on a problem is also affected by exceptions, so in the last layer, slip and guess are introduced. Slip refers to a situation where an examinee has the ability to answer correctly but carelessly makes mistakes. Guess implies that an examinee who does not master the required skills is likely to answer correctly. These well-marked values are equal to 0.1 in the literature. Thus, the simulated score of student j on item i is shown in formula 3.

$$\mathrm{P}\left(X_{ji} = 1 | \alpha_j\right) = 0.9^{\eta_{ji}} 0.1^{1-\eta_{ji}} \tag{3}$$

After restricting the type of questions to objective questions, score X_{ji} of student j on item i is also a binary variable, $X_{ji} \in \{0,1\}$, where 0 means wrong answer and 1 means correct answer.

Another question to be considered is the prior probability distribution of ability level θ. Typically, in IRT, latent trait θ is a continuous variable that cannot be applied directly to a discrete Bayesian network. The method of approximating the continuous ability variable,θ, with a discrete variable, restricting $\theta \in \{-2, -1, 0, 1, 2\}$, has been widely adopted, making all variables discrete and creating a Bayesian network. This approximation has been proven may not even be that bad. The typical value of difficulty parameter β is between [- 3, 3] predefined by the expert.

The a priori probability distribution of θ depends on the examinees and their evaluation method. There are two kinds of evaluation methods: the norm-referenced test and the criterion-referenced test. The meaning of a standard reference test is simple. Those with a score of more than 0.9 have excellent ability, and those with a good ability level have a score of [0.8, 0.9).

According to the proportion of people in each scoring interval, the probability distribution of θ can be estimated.

$$\mathrm{P}\left(\theta_j = i\right) = \frac{N_i}{N}, \ \mathrm{N} = \sum_{i=-2}^{2} N_i \tag{4}$$

After collecting the evidence (e.g., test results), a weighted average of θ can be obtained as the final ability score, as shown in formula 5.

$$\mathrm{E}(\theta_j) = \sum_{i=-2}^{2} \theta_j P(\theta_j = i) \tag{5}$$

5 Experiments

We validate our model using real-world datasets. A total of 536 participants were tested on 20 objective questions involving 8 skills. The raw data are visualized as shown in Fig.3.The upper picture shows the total score distribution, and the lower picture shows the density distribution of the total score. Obviously, it does not obey the normal distribution.

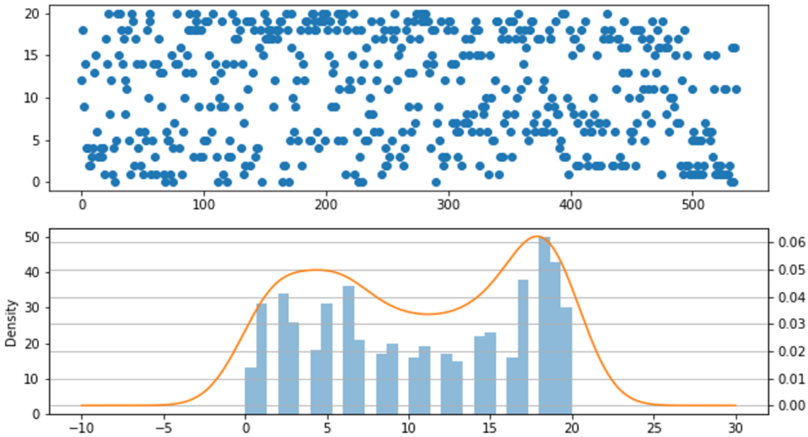

Fig. 3. The distribution of the original score

5.1 Estimation of Students' Ability Level

The correlation between the student's actual test result and the ability level that was determined is explored in this section. Specifically, the distribution of students' ability level is estimated by formula 4. Then, the third layer node of the diagnostic model collects students' answer data to update the whole Bayesian network to infer a new ability distribution. Finally, the expectation of students' ability level is obtained by formula5.The results of the evaluation of the data using CDM are shown in Fig.4. Overall, the evaluation ability increases with increasing scoring rate. For examinees with the same total score, due to different weights assigned to skills, the evaluated skill is also different. To a certain extent, the model can accurately reflect the students' ability level.

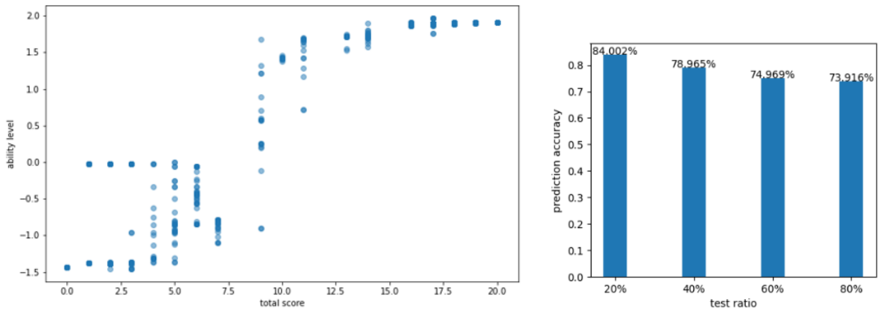

Fig. 4. Evaluated ability level using CDM and prediction task accuracy

After a period of study, we can easily obtain students' proficiency in related skills through the estimated distribution of students' ability level. Then, we can predict students’ score X^_ji on the test questions by combining the Q matrix with formula 2 and formula 3.

$$\hat{X}_{ji} = \begin{cases} 1 & if\ P(\hat{X}_{ji} = 1) > Th_R \\ 0 & otherwise \end{cases} \tag{6}$$

Here, we set the threshold value Th_R to 0.5. If the probability of answering correctly is greater than 0.5, it is predicted that students can do the problem correctly, and if it is less than 0.5, they will make a mistake. The prediction results are shown in Fig.4. When 80% of the evidence is collected, the prediction accuracy reaches 84%, which is within the acceptable range.

6 Conclusion

In this article, we designed a cognitive diagnosis model that integrates the two functions of overall ability assessment and skill diagnosis and has the good characteristics of a simple parameter estimation method and strong interpretability. First, the ability distribution is estimated according to the actual examinees population. Then, the parameter relationship between ability and skill proficiency is determined with IRT theory. Finally, by assuming that the skill interactions on objective problems satisfy the conjunctive relation and introduce two exceptions, slip and guess, we modelled the generation of problem scores. Furthermore, we used an example to illustrate the effectiveness of the model in detailed skill diagnosis, and an experiment on a real dataset illustrates the rationality of the ability assessment.

References

1. Chen, X., Feng, S., Yang, M., Zhao, K., Xu, R., Cui, C., & Chen, M. (2024). Modeling question difficulty for unbiased cognitive diagnosis: A causal perspective. Knowledge-Based Systems, 294, 111750.
2. Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., & Schmidt-Thieme, L. (2010). Recommender system for predicting student performance. *Procedia Computer Science*, *1*(2), 2811-2819.
3. Liu, Y., Zhang, T., Wang, X., Yu, G., & Li, T. (2023). New development of cognitive diagnosis models. *Frontiers of Computer Science*, *17*(1), 171604.
4. Hou, L., la Torre, J. D., & Nandakumar, R. (2014). Differential item functioning assessment in cognitive diagnostic modeling: Application of the Wald test to investigate DIF in the DINA model. *Journal of Educational Measurement*, *51*(1), 98-125.
5. Liu, Y. et al.: Collaborative Learning Team Formation: A Cognitive Modeling Perspective. In: Navathe, S., Wu, W., Shekhar, S., Du, X., Wang, S., Xiong, H. (eds) Database Systems for Advanced Applications 2016, DASFAA, vol. 9643, pp. 383-400. Springer, Cham(2016). https://doi.org/10.1007/978-3-319-32049-6_24
6. Wang, F., Liu, Q., Chen, E., Huang, Z., Chen, Y., Yin, Y., ... & Wang, S. (2020, April). Neural cognitive diagnosis for intelligent education systems. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 04, pp. 6153-6161).