Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Cognitive Modelling for Predicting Examinee Performance

Runze Wu[1], Qi Liu[1,*], Yuping Liu[1], Enhong Chen[1], Yu Su[2], Zhigang Chen[2], Guoping Hu[2]
[1]School of Computer Science and Technology, University of Science and Technology of China
{wrz179,liuyup}@mail.ustc.edu.cn, {qiliuql,cheneh}@ustc.edu.cn
[2]Anhui USTC iFLYTEK Co., Ltd., China
{yusu,zgchen,gphu}@iflytek.com

Abstract

Cognitive modelling can discover the latent characteristics of examinees for predicting their performance (i.e. scores) on each problem. As cognitive modelling is important for numerous applications, e.g. personalized remedy recommendation, some solutions have been designed in the literature. However, the problem of extracting information from both objective and subjective problems to get more precise and interpretable cognitive analysis is still underexplored. To this end, we propose a *fuzzy cognitive diagnosis framework* (FuzzyCDF) for examinees' cognitive modelling with both objective and subjective problems. Specifically, to handle the partially correct responses on subjective problems, we first fuzzify the skill proficiency of examinees. Then, we combine fuzzy set theory and educational hypotheses to model the examinees' mastery on the problems. Further, we simulate the generation of examination scores by considering both slip and guess factors. Extensive experiments on three real-world datasets prove that FuzzyCDF can predict examinee performance more effectively, and the output of FuzzyCDF is also interpretative.

1 Introduction

Recently, there are a lot of researches for better education, e.g., on massive open on-line course (MOOC) [Anderson *et al.*, 2014], intelligent tutoring systems [Burns *et al.*, 2014] and cognitively diagnostic assessment [Nichols *et al.*, 2012].

A crucial step of these educational researches is the cognitive modelling of examinees, which aims at discovering the latent factors/characteristics of examinees. Usually, the effectiveness of cognitive analysis is validated by predicting the possible scores from these examinees to each problem by collaboratively modelling a group of examinees, i.e. *predicting examinee performance* (PEP). Since PEP could be further applied to numerous applications, such as the personalized remedy recommendation and the teaching plan improvement, massive efforts of both data mining and educational psychology have been made to the solutions of cognitive modelling for PEP: In terms of data mining, Matrix factorization (MF) [Koren *et al.*, 2009] is a classical prediction technique, which is widely used to model examinees [Toscher and Jahrer, 2010; Thai-Nghe *et al.*, 2010; Desmarais, 2012]. In educational psychology, most of the existing studies focus on cognitive diagnosis. With the cognitive diagnosis models (CDMs) [DiBello *et al.*, 2006], examinees are characterized by the proficiency on specific skills (e.g. problem-solving skills like calculation), where a **Q**-matrix [Tatsuoka, 1984] is previously given as the prior knowledge from education experts for denoting which skills are needed for each problem.

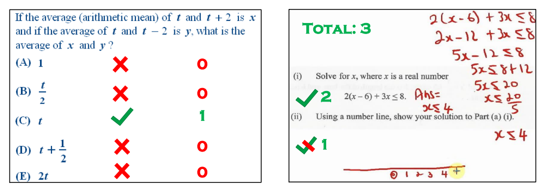


Figure 1: An example of objective and subjective problems.

In spite of the importance of previous studies, there are still some limitations in existing methods. For instance, the latent factors in MF are unexplainable for describing the knowledge state of examinees' cognition. Comparatively, the results of CDMs could lead to a better interpretation. However, CDMs can only analyse examinees based on the simple objective problems, and the information of the subjective problems is largely underexplored. As shown in Figure 1, the objective (e.g. chosen) problems have dichotomous scores with either right or wrong response and the subjective (e.g. free-response) problems have polytomous scores with correct, incorrect or partially correct response. Obviously, for subjective problems, it is harder for examinees to get a correct response by guessing an answer or get a wrong response by carelessness (e.g. a slip of a pen). Thus, these subjective problems measure the examinees much better, and it is of significant importance to extract information from both objective and subjective problems for cognitive modelling rather than simply ignore the subjective problems or treat them as objective ones [Samejima, 1972]. Towards this goal, there are several challenges: How to handle the dichotomous scores of objective problems and the polytomous scores of subjective problems simultaneously? How to get both precise and in-

*Corresponding author.

terpretable cognitive analysis? How to predict the examinee performance based on this cognitive modelling?

To conquer these challenges, in this paper, we propose a *fuzzy cognitive diagnosis framework* (FuzzyCDF) for examinees' cognitive modelling. Since the response to a subjective problem may be partially correct, we adopt a solution inspired by fuzzy systems. Specifically, we first fuzzify the skill proficiency of examinees using fuzzy set theory. Then we suppose the skill interactions on objective and subjective problems satisfy two different hypotheses: conjunctive and compensatory [Pardos *et al.*, 2008], and fuzzify the problem mastery of examinees based on these two hypotheses. Next, we model the generation of problem scores (predict examinee performance) by considering two exceptions: slip and guess. In this way, FuzzyCDF is represented as a four-tier hierarchical model and we propose a Monte Carol Markov chain (MCMC) sampling algorithm to estimate parameters, and thus we could predict examinee performance (i.e. scores). The main contributions of this paper are outlined as follows:

- To the best of our knowledge, this is the first comprehensive attempt for extracting information from both objective and subjective problems to get more precise and interpretable cognitive analysis for PEP task.
- We propose a four-tier hierarchical cognitive model FuzzyCDF, which combines educational assumptions and fuzzy theories to redefine skill proficiency and problem mastery, and it predicts the examinee performance by considering the slip and guess factors.
- We design a simple but effective MCMC sampling algorithm for parameter estimation and conduct extensive experiments on real-world datasets to demonstrate the effectiveness of FuzzyCDF.

2 Related Work

We introduce existing modelling methods for PEP from two aspects: data mining methods and cognitive diagnosis.

Data Mining Methods. Recently, more and more studies demonstrate the effectiveness of MF for examinee modelling and predicting examinee performance by factorizing the score matrix. For instance, [Toscher and Jahrer, 2010] utilized singular value decomposition (SVD) and other factor models to model examinees. In [Thai-Nghe *et al.*, 2010], MF technique was compared with regression methods for PEP. Besides, [Desmarais, 2012; Sun *et al.*, 2014] applied nonnegative matrix factorization to infer the **Q**-matrix. However, the latent factors inferred by the MF model are usually unexplainable, i.e. each dimension of the factor vector cannot correspond to a specific skill. In this work, we will adopt cognitive diagnosis on exam data to obtain interpretative results.

Cognitive Diagnosis. In educational psychology, many cognitive diagnosis models (CDMs) [DiBello *et al.*, 2006] have been developed to mine examinees' skill proficiency, which can be used for PEP. Figure 2 shows a toy process of cognitive diagnosis. CDMs can be roughly divided into two categories: continuous ones and discrete ones. The fundamental continuous CDMs are *item response theory* (IRT) models [Rasch, 1961; Birnbaum, 1968; Embretson and Reise, 2013], which characterize examinee by a continuous variable, i.e. latent trait, and use a logistic function to model the probability that an examinee correctly solves a problem. For discrete CDMs, the basic method is *deterministic inputs, noisy "and" gate model* (DINA) [Haertel, 1984; Junker and Sijtsma, 2001; De La Torre, 2011]. DINA describes an examinee by a latent binary vector which denotes whether she masters the skills required by the problem, and a given **Q**-matrix is used to guarantee the interpretation of the diagnosis results. Though discrete CDMs can model examinees interpretatively, their diagnosis results are usually not accurate enough. Furthermore, existing methods cannot handle the subjective problems.

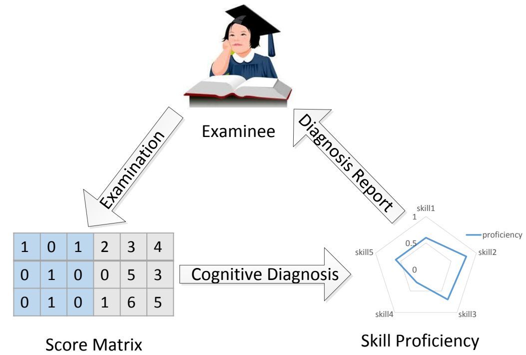


Figure 2: A toy process of cognitive diagnosis.

Table 1: Some improtant notations.

Notation	Description
R_{ji}	the score of examinee j on problem i
θ_j	the high-order latent trait of examinee j
α_{jk}	the proficiency of examinee j on skill k
a_{jk}, b_{jk}	the discrimination, difficulty of examinee j on skill k
μ_k	the membership function of fuzzy set related to skill k
η_{ji}	the mastery of examinee j on problem i
s_i, g_i	the slip and guess factors of problem i

3 Fuzzy Cognitive Diagnosis

In this section, we will introduce our *fuzzy cognitive diagnosis framework* (FuzzyCDF). As shown in Figure 3 (from up to bottom), our proposed method is a generation process which starts with examinees' latent trait, then determines exminees' skill proficiency, next computes examinees' problem mastery and generates examinees' observable scores on problems by considering slip and guess factors. And we propose a MCMC sampling algorithm to infer the unobservable parameters of FuzzyCDF. For better illustration, Table 1 shows some math notations and each step of FuzzyCDF will be specified in the following subsections.

3.1 Fuzzifying Skill Proficiency

In this subsection, we show the way to get the proficiency of examinees on specific skills (e.g. problem-solving skills like calculation), which is the first step in cognitive diagnosis

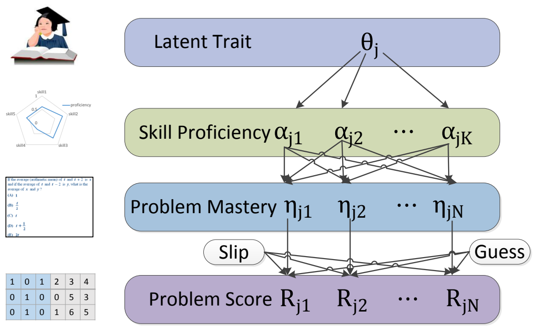


Figure 3: 4-tier fuzzy cognitive diagnosis framework.

models. Here, the challenge is that we have to handle both objective problems and subjective problems, simultaneously.

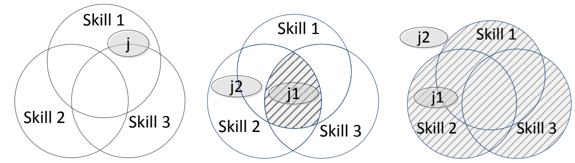
In DINA-based CDMs, the examinees' skill proficiency is assumed to be absolutely mastered (i.e. 1) or nonmastered (i.e. 0) so that this modelling can fit objective problems with absolutely right or wrong responses, and meanwhile, this skill proficiency is influenced by the high-order latent trait of the examinee and the properties of the skill [De La Torre and Douglas, 2004]. However, for a subjective problem which has a partially correct response, the above "absolutely" modelling on skill proficiency cannot fit well as shown in the experiments. Thus, to address this issue, we adopt fuzzy systems [Jantzen, 2013; Chrysafiadi and Virvou, 2014] in cognitive modelling and redefine the proficiency of an examinee on a skill by the following assumption:

Assumption 1 *The proficiency of an examinee on a skill is* ***the grade of membership of the examinee in a fuzzy set*** *which is related to the skill.*

Here, we assume that a skill k is related to a fuzzy set (J, μ_k), where J is the set of examinees and $\mu_k : J \to [0, 1]$ is the membership function. Then, for each $j \in J$, we define the proficiency of examinee j on skill k, α_{jk}, as the grade of membership of j in (J, μ_k), $\mu_k(j)$. Thus, if examinee j masters skill k to some extent then the element j is a fuzzy member of the fuzzy set, i.e. $0 \leq \alpha_{jk} = \mu_k(j) \leq 1$. In this way, we can fuzzify the proficiency of an examinee on a skill into a continuous variable with a value in [0,1]. For instance, as shown in Figure 4(a), examinee j is fully included, fully excluded and partially included by the fuzzy sets which is related to Skill 1, Skill 2 and Skill 3, respectively. It also means, examinee j has completely mastered, completely nonmastered and partially mastered Skill 1, Skill 2 and Skill 3, respectively. Formally, following an IRT-like high-order logistic model, α_{jk} and $\mu_k(j)$ is defined as:

$$\alpha_{jk} = \mu_k(j) = \frac{1}{1 + exp[-1.7a_{jk}(\theta_j - b_{jk})]}. \tag{1}$$

The implication of this definition is that the proficiency of an examinee on a specific skill (α_{jk}) depends on the difference between the examinee's high-order trait (θ_j) and the properties of the skill: difficulty (b_{jk}) and discrimination (a_{jk}) of skill k for examinee j. Here, the coefficient 1.7 is an empirical scaling constant in logistic cognitive models [Hulin *et al.*, 1983; Camilli, 1994]. Generally, the major



(a) Fuzzy Set (b) Intersection (c) Union

Figure 4: An example of the fuzzy set.

difference between this definition and the traditional models [De La Torre and Douglas, 2004] is that we redefine the skill proficiency from binary variables to the probabilistic ones.

3.2 Fuzzifying Problem Mastery

Based on the fuzzified skill proficiency in Section 3.1, we can further fuzzify the problem mastery of examinees (i.e. the probability of being able to solve the problem). CDMs assume that problem mastery is a result of interaction of examinees' proficiency on required skills of this problem [Pardos *et al.*, 2008]. Thus, we will first discuss the skills' interaction on objective and subjective problems, then introduce a method to model the two kinds of interactions according to *fuzzy logic* and fuzzify the mastery of an examinee on a problem.

The skill's interaction on problems can be mainly categorized into *conjunctive* and *compensatory* [Pardos *et al.*, 2008]. Conjunctive means that an examinee must master all the required skills to solve a problem, while compensatory means that an examinee is possible to solve a problem as long as she masters any skill required by the problem. As for examinations, an objective problem has a unique standard answer and cannot be answered correctly unless the examinee masters all the required skill without any omission. Thus the skill's interaction on objective problems is usually assumed to be conjunctive [Pardos *et al.*, 2008]. In contrast, a subjective problem is a free-response one and the examinees can write not only the final answers but also the solving process including writing equations, deducing, calculating and so on (a toy example is shown in the right part of Figure 1). That is, given a subjective problem and the required skills, the more skills an examinee masters, the higher score she will get on this problem. Therefore in this research we assume that the skill's interaction on subjective problems is compensatory, the effectiveness of which can be demonstrated experimentally. To sum up, we propose an assumption about skills' interaction on the problems as the following:

Assumption 2 *The skills' interaction on objective (or subjective) problems is* ***conjunctive*** *(or* ***compensatory****).*

Now, we model these two kinds of interactions in a fuzzy way to fuzzify the problem mastery. Specifically, given the set of examinees J, suppose we have a problem i requiring Skill 1, Skill 2 and Skill 3 with their fuzzy sets (J, μ_1), (J, μ_2) and (J, μ_3). As shown in Figure 4(b), examinee j_1 is fully included by the fuzzy sets related to Skill 1, Skill 2 and Skill 3 and examinee j_2 is not included by the fuzzy set related to Skill 3, which means that if we adopt the conjunctive assumption then examinee j_1 has mastered problem i because she has mastered all the skills needed by problem i and examinee j_2 has not; as shown in Figure 4(c), examinee j_1 is fully included by the fuzzy sets related to Skill 1 and Skill 2 and examinee j_2 is not included by any of the fuzzy set. Here, if we adopt the compensatory assumption then examinee j_1 is possible to master problem i because she has mastered at least one of the skills needed by problem i. In summary, the set of the examinees who master all and any of the skills required for problem i is the intersection and union of the fuzzy sets related to the skills, respectively. Thus, we propose an assumption to fuzzify the problem mastery as follows:

Assumption 3 *If the skills' interactions between each other on a problem is conjunctive (or compensatory), the mastery of an examinee on this problem is the grade of membership of this examinee in the* ***intersection*** *(or* ***union****) set of the fuzzy sets related to the skills required by the problem.*

Formally, given a **Q**-matrix with K skills, the mastery of an examinee j on a problem i, η_{ji}, is defined as the following equation under conjunctive assumption:

$$\eta_{ji} = \mu_{\bigcap_{1 \le k \le K, q_{ik}=1} k}(j). \tag{2}$$

And η_{ji} for subjective problems is defined as the following equation under compensatory assumption:

$$\eta_{ji} = \mu_{\bigcup_{1 \le k \le K, q_{ik}=1} k}(j). \tag{3}$$

Here, q_{ik} from the **Q**-matrix indicates whether problem i requires skill k (1 means required and 0 means non-required). Without loss of generality, we adopt the simplest fuzzy intersection and union operation[1] (*standard fuzzy intersection and union* [GEORGE J and Bo, 2008]) as follows:

$$\mu_{A \cap B}(x) = min(\mu_A(x), \mu_B(x)). \qquad \mu_{A \cup B}(x) = max(\mu_A(x), \mu_B(x)). \tag{4}$$

In this way, we could get the mastery of each examinee on every problem (η_{ji}), either objective or subjective problem.

3.3 Modelling Generation of Examinees' Scores

With the problem mastery defined in Section 3.2, we can now determine examinees' scores on problems. Specifically, we take two exceptions, slip and guess [d Baker *et al.*, 2008], into account and adopt two distributions to simulate the generation of scores of objective and subjective problems, respectively.

In a real-world examination, the score of an examinee on a problem not only depends on the examinee's problem mastery. For instance, an examinee who is unable to solve the problem can get a correct response by guessing an answer (e.g. choosing "C" as the final answer somehow); Meanwhile, she who is able to do it right may get a wrong response as a consequence of carelessness (e.g. a slip of a pen) [Embretson, 1985]. Here, we assume that each problem has its own slip and guess factors and consider these two exceptions to model the generation of examinees' scores.

Meanwhile, we handle the different score patterns of objective and subjective problems. With either right or wrong response, the score of an examinee on a objective problem can be coded to a binary variable with a value in $\{0,1\}$. Thus, we adopt a *Bernoulli* distribution to model the scores of examinees on objective problems. Considering different score scales of subjective problems, we normalize the scores on a subjective problem by dividing the full score of the problem into a continuous variable with a value in [0,1]. Then we assume that the score of examinees on subjective problems follow a *Gaussian* distribution, which is widely used in the literature [Mnih and Salakhutdinov, 2007].

Formally, combining the problem mastery of the examinees and the exceptions of slip and guess, we simulate the generation of the scores as follows:

$$P(R_{ji} = 1 | \eta_{ji}, s_i, g_i) = (1 - s_i)\eta_{ji} + g_i(1 - \eta_{ji}). \tag{5}$$

$$P(R_{ji} | \eta_{ji}, s_i, g_i) = \mathcal{N}(R_{ji} | [(1 - s_i)\eta_{ji} + g_i(1 - \eta_{ji})], \sigma^2). \tag{6}$$

Equations (5) and (6) stand for objective problems and subjective problems, respectively. Here, R_{ji} denotes the score (normalized score for subjective problem) of the examinee j on problem i, η_{ji} is the mastery of an examinee j on the problem i, s_i and g_i denote the slip and guess factors of problem i, and σ^2 is the variance of the normalized score of an examinee on a subjective problem. Thus, $(1 - s_i)\eta_{ji}$ means this examinee masters the problem and answers it successfully (i.e. without carelessness), while $g_i(1 - \eta_{ji})$ represents the examinee guesses a right response without mastering. That is, these are the two ways for an examinee to give a correct response.

Summary. To better understand our proposed FuzzyCDF method, we represent it using a graphic model as shown in Figure 5. Here, what we can observe are the score matrix **R** with M examinees, N_o objective and N_s subjective problems and the **Q**-matrix with K skills (if problem i requires skill k, then $q_{ik} = 1$). An examinee j is related to skill proficiency $\alpha_{jk}, k = 1, 2, \cdots, K$, which depends on high-order latent trait θ_j and skill parameters $a_{jk}, b_{jk}, k = 1, 2, \cdots, K$. A problem mastery η_{ji} is determined by required skill proficiency $\alpha_{jk} | q_{ik} = 1$ and a problem score R_{ji} is influenced by η_{ji} and problem parameters s_i, g_i. It is noted that σ is only used for modelling the generation of the score of an examinee on a subjective problem.

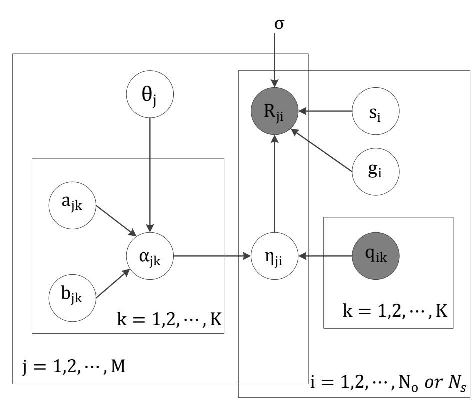


Figure 5: The graphic model of FuzzyCDF.

[1]The effects of different kinds of fuzzy set operation are out the scope of this paper.

3.4 MCMC Training Algorithm

In this section, we will introduce an effective training algorithm using MCMC method for the proposed FuzzyCDF model, i.e. to infer the unshaded variables in Figure 5. Specifically, we assume that the prior distributions of the parameters in FuzzyCDF as the following:

$$
\begin{aligned}
&\theta_j \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2), a_{jk} \sim ln\mathcal{N}(\mu_a, \sigma_a^2), b_{jk} \sim \mathcal{N}(\mu_b, \sigma_b^2);\\
&s_i \sim Beta(v_s, w_s, min_s, max_s),\\
&g_i \sim Beta(v_g, w_g, min_g, max_g);\\
&1/\sigma^2 \sim \Gamma(x_\sigma, y_\sigma).
\end{aligned} \tag{7}
$$

where $Beta(v, w, min, max)$ is a four-parameter *Beta* distribution which has two shape parameters v and w and is supported on the range $[min, max]$. The functional forms of the prior distributions are chosen out of convenience, and the associated hyperparameters are selected to be reasonably vague within the range of realistic parameters. Then, the joint posterior distribution of $\theta, \mathbf{a}, \mathbf{b}, \mathbf{s}, \mathbf{g}$ and σ^2 given the score matrix R is as follows:

$$
\begin{aligned}
&P(\theta, \mathbf{a}, \mathbf{b}, \mathbf{s}, \mathbf{g}, \sigma^2|\mathbf{R}) \qquad (8)\\
&\propto L(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b})P(\theta)P(\mathbf{a})P(\mathbf{b})P(\mathbf{s})P(\mathbf{g})P(\sigma^2),
\end{aligned}
$$

where L is the joint likelihood function of FuzzyCDF:

$$
L(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b}) = L_o(\mathbf{s}, \mathbf{g}, \theta, \mathbf{a}, \mathbf{b})L_s(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b}), \tag{9}
$$

where L_o and L_s denote the joint likelihood functions of objective and subjective problems respectively, and are defined as the following equations:

$$
L_o(\mathbf{s}, \mathbf{g}, \theta, \mathbf{a}, \mathbf{b}) = \prod_j^M \prod_i^{N_o} (X_{ji})^{R_{ji}}(1 - X_{ji})^{1-R_{ji}}, \tag{10}
$$

$$
L_s(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b}) = \prod_j^M \prod_i^{N_s} \mathcal{N}(R_{ji}|X_{ji}, \sigma^2), \tag{11}
$$

where, $X_{ji} = (1 - s_i)\eta_{ji} + g_i(1 - \eta_{ji})$. Then, the full conditional distributions of the parameters given the data and the rest of parameters as the following:

$$
P(\mathbf{a}, \mathbf{b}|\mathbf{R}, \theta, \mathbf{s}, \mathbf{g}, \sigma^2) \propto L(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b})P(\mathbf{a})P(\mathbf{b}), \tag{12}
$$

$$
P(\theta|\mathbf{R}, \mathbf{a}, \mathbf{b}, \mathbf{s}, \mathbf{g}, \sigma^2) \propto L(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b})P(\theta), \tag{13}
$$

$$
P(\mathbf{s}, \mathbf{g}|\mathbf{R}, \theta, \mathbf{a}, \mathbf{b}, \sigma^2) \propto L(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b})P(\mathbf{s})P(\mathbf{g}), \tag{14}
$$

$$
P(\sigma^2|\mathbf{R}, \theta, \mathbf{a}, \mathbf{b}, \mathbf{s}, \mathbf{g}) \propto L_s(\mathbf{s}, \mathbf{g}, \sigma^2, \theta, \mathbf{a}, \mathbf{b})P(\sigma^2). \tag{15}
$$

Finally, we propose a Metropolis-Hastings (M-H) based MCMC algorithm [Hastings, 1970] for parameter estimation by Algorithm 1. To be specific, we first randomize all the parameters as the initial values. Then, using observed score matrix $\mathbf{R}$ and the $\mathbf{Q}$-matrix, we compute the full conditional probability of skill discrimination $\mathbf{a}$, skill difficulty $\mathbf{b}$, examinee latent trait θ, problem slip factor $\mathbf{s}$ and guess factor $\mathbf{g}$ and the variance of normalized score of subjective problem σ^2. Next, the acceptance probability of samples can also be calculated based on M-H algorithm. In this way, we could estimate the parameters with the MCMC formed through sampling.

Predicting Examinee Performance. After the training stage, we can easily obtain fuzzified proficiency of the examinees on each skill based on estimated latent trait, skill discrimination and difficulty (Equation (1)). Then, we can further compute fuzzified mastery of the examinees on each problem according to $\mathbf{Q}$-matrix and problem type (objective or subjective) by Equations (2) and (3). Combining the estimated slip and guess factors of each problem, we can now complete the PEP task, i.e. predict examinees' performance (i.e. score) on each problem, by Equations (5) and (6). Note that, the output of Equations (5) is continuous, and we can discretize them for predicting on objective problems.

Algorithm 1 Sampling algorithm for FuzzyCDF.

Input: score matrix $\mathbf{R}$, experts' knowledge $\mathbf{Q}$
Output: samples of θ,$\mathbf{a}$,$\mathbf{b}$,$\mathbf{s}$,$\mathbf{g}$,σ^2

```
Initialize θ_0, a_0, b_0, s_0, g_0, σ_0^2 with random values
for t = 1, 2, ···, T do
    Draw a_t ~ U(a_{t-1} − δ_a, a_{t-1} − δ_a), b_t ~ U(b_{t-1} − δ_b, b_{t-1} − δ_b), and accept a_t, b_t with the probability:
        min{1, L(s_{t-1},g_{t-1},σ_{t-1}^2,θ_{t-1},a_t,b_t)P(a_t)P(b_t) / L(s_{t-1},g_{t-1},σ_{t-1}^2,θ_{t-1},a_{t-1},b_{t-1})P(a_{t-1})P(b_{t-1})}.
    Draw θ_t ~ U(θ_{t-1} − δ_θ, θ_{t-1} + δ_θ), and accept θ_t with the probability:
        min{1, L(s_{t-1},g_{t-1},σ_{t-1}^2,θ_t,a_t,b_t)P(θ_t) / L(s_{t-1},g_{t-1},σ_{t-1}^2,θ_{t-1},a_t,b_t)P(θ_{t-1})}.
    Draw s_t ~ U(s_{t-1} − δ_s, s_{t-1} + δ_s), g_t ~ U(g_{t-1} − δ_g, g_{t-1} + δ_g), and accept s_t, g_t with the probability:
        min{1, L(s_t,g_t,σ_{t-1}^2,θ_t,a_t,b_t)P(s_t)P(g_t) / L(s_{t-1},g_{t-1},σ_{t-1}^2,θ_t,a_t,b_t)P(s_{t-1})P(g_{t-1})}.
    Draw σ_t^2 ~ U(σ_{t-1}^2 − δ_σ, σ_{t-1}^2 + δ_σ), and accept σ_t^2 with the probability:
        min{1, L(s_t,g_t,σ_t^2,θ_t,a_t,b_t)P(σ_t^2) / L(s_t,g_t,σ_{t-1}^2,θ_t,a_t,b_t)P(σ_{t-1}^2)}.
    if convergence criterion meets then
        return
    end if
end for
return
```

4 Experiments

We first compare the performance of FuzzyCDF against the baseline approaches on the PEP task; then, we conduct a *case study* to assess the interpretation of the cognitive diagnosis.

Experimental Setup. The public dataset in our experiment comprises of scores of middle school students on fraction subtraction objective problems [Tatsuoka, 1984; DeCarlo, 2010]. The two private datasets[2] are collected from two final mathematical exams for high school students including both objective and subjective problems. We denote the three datasets as *FrcSub, Math1* and *Math2*. Each of the datasets is represented by a score matrix and a given $\mathbf{Q}$-matrix by education experts. The brief summary of these datasets is shown in Table 2. Further, Figure 6 shows the preview of these three datasets, where each column for each subfigure stands for a problem and each row above and below represents an examinee and a skill, respectively. Specifically, the three subfigures in the above show FrcSub's score matrix with only dichotomous scores, Math1's and Math2's normalized score matrix with both dichotomous and polytomous scores; The three subfigures in the below show the three $\mathbf{Q}$-matrices.

[2]The two private datasets we use have been publicly available on http://staff.ustc.edu.cn/%7Eqiliuql/data/math2015.rar.

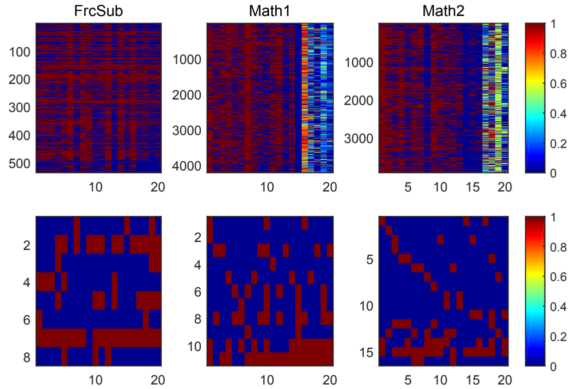


Figure 6: The preview of the datasets.

Table 2: Datasets Summary.

Dataset	# Examinee	# Skill	# Problem Obj.	Sub.
FrcSub	536	8	20	0
Math1	4,209	11	15	5
Math2	3,911	16	16	4

For the prior distributions of parameters in FuzzyCDF, we set the hyperparameters as follows:

$$\mu_\theta = 0, \sigma_\theta = 1; \mu_a = 0, \sigma_a = 1; \mu_b = 0, \sigma_b = 1;$$
$$v_s = 1, w_s = 2, min_s = 0, max_s = 0.6;$$
$$v_g = 1, w_g = 2, min_g = 0, max_g = 0.6;$$
$$x_\sigma = 4, y_\sigma = 6.$$

In this experiments, we set the number of iterations of Algorithm 1 to 5,000 and estimate the parameters based on the last 2,500 samples to guarantee the convergency of the Markov chain. Both our FuzzyCDF and other baseline approaches are implemented by using Matlab on a Core i5 3.1Ghz machine with Windows 7 and 4 GB memory.

PEP Task. To demonstrate the effectiveness of FuzzyCDF, we conduct experiments on PEP task, i.e. predicting the scores from the examinees on each problem. To observe how the methods behave at different sparsity levels, we construct different sizes of training sets, with 20%, 40%, 60% and 80% of score data of each examinee, and the rest for testing, respectively. We use *root mean square error* (RMSE) and *mean absolute error* (MAE) as the evaluation metrics. Then, we consider baseline approaches as follows:

- *IRT*:[Rasch, 1961; Birnbaum, 1968] a cognitive diagnosis method modelling examinees' latent trait and the parameters of problems like difficulty and discrimination.
- *DINA*:[Junker and Sijtsma, 2001] a cognitive diagnosis method modelling examinees' skill proficiency and the slip and guess factors of problems with a **Q**-matrix.
- *PMF*:[Mnih and Salakhutdinov, 2007] probabilistic matrix factorization is a latent factor model projecting examinees and problems into a low-dimensional space.
- *NMF*:[Lee and Seung, 2001] non-negative matrix factorization is a latent non-negative factor model and can be viewed as a topic model.

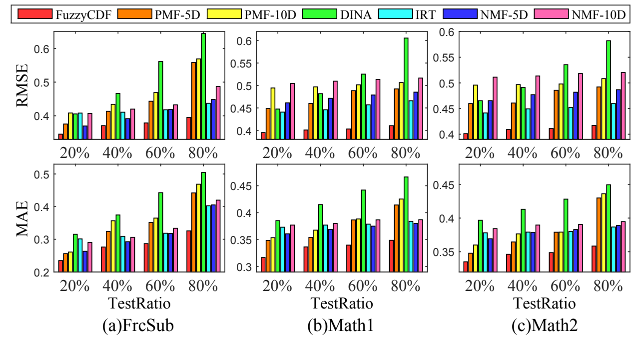


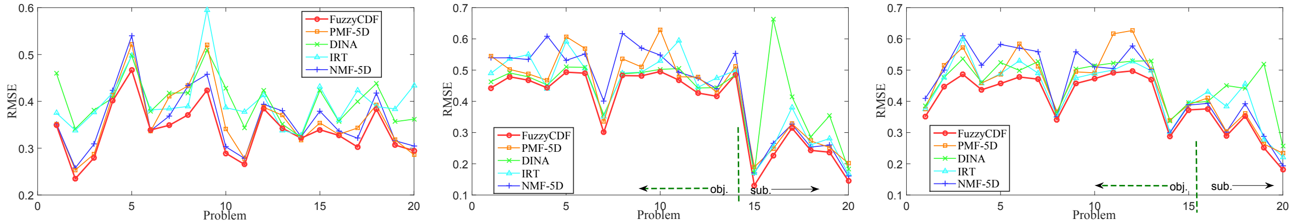
Figure 7: Prediction task performance.

For the purpose of comparison, we record the best performance of each algorithm by tuning their parameters, and Figure 7 shows the PEP results of our FuzzyCDF and baseline approaches on three datasets. Here, we consider two implementations of the matrix factorization methods, PMF and NMF. That is, PMF-5D and PMF-10D (NMF-5D and NMF-10D) represent the PMF (NMF) with 5 and 10 latent factors, respectively. Thus, there are totally seven results in each split.

From this figure, we observe that, over all the datasets, FuzzyCDF performs the best. Specifically, by combining educational hypotheses it beats PMF and NMF, and by quantitatively analysing examinees from a fuzzy viewpoint, it beats IRT and DINA. More importantly, with the increasing of the sparsity of the training data (training data ratio declines from 80% to 20%), the superiority of our FuzzyCDF method becomes more and more significant. For instance, when the training data is 20% and under the metric of MAE, the improvement of FuzzyCDF compared to the best baseline method IRT could reach 19%, 10%, and 8% on each data, respectively. In summary, FuzzyCDF captures the characteristics of examinees more precisely and it is also more suitable for the real-world scenarios, where the data is sparse and the examinees/problems are cold-start.

Fixing the training data ratio equal to 80%, Figure 8(a), 8(b) and 8(c) show the prediction performance for each specific problem in these three datasets. For simplicity, we only give the results of FuzzyCDF and four baselines that have better performance: PMF-5D, DINA, IRT and NMF-5D. From each subfigure, we can observe that FuzzyCDF outperforms almost all the baselines on all the problems. Specifically, in Math1 and Math2 datasets (the last five and four problems are subjective problems, respectively), FuzzyCDF method can obtain the best performance for both objective and subjective problems, which in turn proves the reasonability of our Assumption 2 about skill interaction on objective and subjective problems. However, matrix factorization methods (PMF and NMF) cannot fit the scores of objective problems very well and the normal psychometric methods (IRT and DINA) are unsuitable for subjective problems. Moreover, Table 3 shows the runtime of each method under this setting.

Case Study. Here, we present an example of the diagnosis results of an examinee on each skill in Dataset FrcSub using DINA and FuzzyCDF, and the results are shown in Figure 9. We can observe that both FuzzyCDF and DINA can



(a) FrcSub (b) Math1 (c) Math2

Figure 8: Prediction performance for each problem.

Table 3: Runtime Results (seconds).

Datasets	FuzzyCDF	PMF	DINA	IRT	NMF
FrcSub	111.26	0.14	11.19	1.77	0.07
Math1	885.63	1.27	239.56	78.87	0.344
Math2	1096.00	1.13	6184.21	68.14	0.256

obtain interpretatively meaningful diagnosis results with the well-designed **Q**-matrix. However, DINA can only distinguish whether an examinee masters a skill (1 or 0) while our FuzzyCDF can tell the extent to which the examinee masters a skill. Thus, based on our diagnosis results, an examinee can find out the true strength and shortcomings of hers at present, and furthermore, educators or tutoring systems can give her targeted remedy plans for improvement. We should note that traditional IRT and matrix factorization methods describe an examinee with latent variables, which cannot provide intuitive and interpretative results for each examinee.

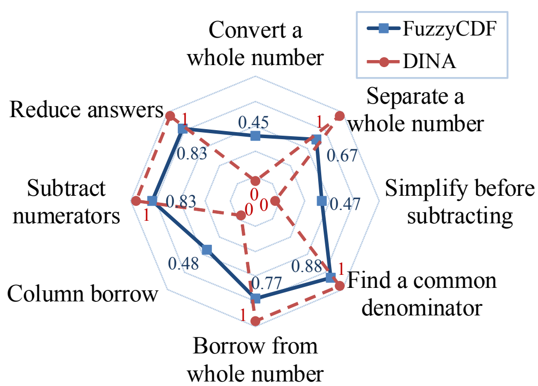


Figure 9: Diagnosis results of an examinee on each skill.

Discussion. From the experimental results of PEP task, we can observe that FuzzyCDF outperforms the baselines on both objective and subjective problems. The case study demonstrated that FuzzyCDF could obtain interpretative cognitive analysis results for examinees, which can be used for composing a detailed and human-readable diagnosis report.

On the other hand, there is still some room for improvement. First, FuzzyCDF currently confronts the problem of high computational complexity, and we will try to design an efficient sampling algorithm in the future. Second, we can test more fuzzy set operation functions. Third, there may be some other problem types beyond objective and subjective problems that should be considered for cognitive modelling.

5 Conclusion

In this paper, we designed a fuzzy cognitive diagnosis framework, FuzzyCDF, to explore the scores of both objective and subjective problems for cognitive modelling. Specifically, we first fuzzified the skill proficiency of examinees based on a fuzzy set assumption, then fuzzified the problem mastery by mapping conjunctive and compensatory interactions into the fuzzy set operations, and next modelled the generation of the two kinds of problems with different distributions by considering slip and guess factors. Finally, extensive experimental results demonstrated that FuzzyCDF could quantitatively and interpretatively analyse the characteristics of each examinee and thus obtained a better performance for the PEP task. We hope this work could lead to more future studies.

Acknowledgements

This research was partially supported by grants from the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61325010), the Natural Science Foundation of China (Grant No. 61403358) and the Fundamental Research Funds for the Central Universities of China (Grant No. WK2350000001). Qi Liu gratefully acknowledges the support of the Youth Innovation Promotion Association, CAS.

References

[Anderson *et al.*, 2014] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide web*, pages 687–698. International World Wide Web Conferences Steering Committee, 2014.

[Birnbaum, 1968] Allan Birnbaum. Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*, 1968.

[Burns *et al.*, 2014] Hugh Burns, Carol A Luckhardt, James W Parlett, and Carol L Redfield. *Intelligent tutoring systems: Evolutions in design*. Psychology Press, 2014.

[Camilli, 1994] Gregory Camilli. Teachers corner: Origin of the scaling constant d= 1.7 in item response theory. *Jour-*

nal of Educational and Behavioral Statistics, 19(3):293–295, 1994.

[Chrysafiadi and Virvou, 2014] K Chrysafiadi and M Virvou. Fuzzy logic for adaptive instruction in an e-learning environment for computer programming. 2014.

[d Baker *et al.*, 2008] Ryan SJ d Baker, Albert T Corbett, and Vincent Aleven. More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *Intelligent Tutoring Systems*, pages 406–415. Springer, 2008.

[De La Torre and Douglas, 2004] Jimmy De La Torre and Jeffrey A Douglas. Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3):333–353, 2004.

[De La Torre, 2011] Jimmy De La Torre. The generalized dina model framework. *Psychometrika*, 76(2):179–199, 2011.

[DeCarlo, 2010] Lawrence T DeCarlo. On the analysis of fraction subtraction data: The dina model, classification, latent class sizes, and the q-matrix. *Applied Psychological Measurement*, 2010.

[Desmarais, 2012] Michel C Desmarais. Mapping question items to skills with non-negative matrix factorization. *ACM SIGKDD Explorations Newsletter*, 13(2):30–36, 2012.

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

[Embretson and Reise, 2013] Susan E Embretson and Steven P Reise. *Item response theory for psychologists*. Psychology Press, 2013.

[Embretson, 1985] Susan E Embretson. Multicomponent latent trait models for test design. *Test design: Developments in psychology and psychometrics*, pages 195–218, 1985.

[GEORGE J and Bo, 2008] KLIR GEORGE J and Yuan Bo. Fuzzy sets and fuzzy logic, theory and applications. 2008.

[Haertel, 1984] Edward Haertel. An application of latent class models to assessment data. *Applied Psychological Measurement*, 8(3):333–346, 1984.

[Hastings, 1970] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[Hulin *et al.*, 1983] Charles Lee Hulin, Fritz Drasgow, and Charles K Parsons. *Item response theory: Application to psychological measurement*. Dorsey Press, 1983.

[Jantzen, 2013] Jan Jantzen. *Foundations of fuzzy control: a practical approach*. John Wiley & Sons, 2013.

[Junker and Sijtsma, 2001] Brian W Junker and Klaas Sijtsma. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3):258–272, 2001.

[Koren *et al.*, 2009] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, (8):30–37, 2009.

[Lee and Seung, 2001] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in NIPS*, pages 556–562, 2001.

[Mnih and Salakhutdinov, 2007] Andriy Mnih and Ruslan Salakhutdinov. Probabilistic matrix factorization. In *Advances in NIPS*, pages 1257–1264, 2007.

[Nichols *et al.*, 2012] Paul D Nichols, Susan F Chipman, and Robert L Brennan. *Cognitively diagnostic assessment*. Routledge, 2012.

[Pardos *et al.*, 2008] Zachary A Pardos, Joseph E Beck, Carolina Ruiz, and Neil T Heffernan. The composition effect: Conjunctive or compensatory? an analysis of multi-skill math questions in its. In *The 1st International Conference on Educational Data Mining*. Citeseer, 2008.

[Rasch, 1961] Georg Rasch. On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 4, pages 321–333. University of California Press Berkeley, CA, 1961.

[Samejima, 1972] Fumiko Samejima. A general model for free-response data. *Psychometrika Monograph Supplement*, 1972.

[Sun *et al.*, 2014] Yuan Sun, Shiwei Ye, Shunya Inoue, and Yi Sun. Alternating recursive method for q-matrix learning. *Educational Data Mining (EDM 2014)*, 2014.

[Tatsuoka, 1984] KK Tatsuoka. Analysis of errors in fraction addition and subtraction problems (final report for grant no. nie-g-81-0002). urbana: University of illinois. *Computer-based Education Research Laboratory*, 1984.

[Thai-Nghe *et al.*, 2010] Nguyen Thai-Nghe, Lucas Drumond, Artus Krohn-Grimberghe, and Lars Schmidt-Thieme. Recommender system for predicting student performance. *Procedia Computer Science*, 1(2):2811–2819, 2010.

[Toscher and Jahrer, 2010] A Toscher and Michael Jahrer. Collaborative filtering applied to educational data mining. *KDD Cup*, 2010.