



A generalized multi-skill aggregation method for cognitive diagnosis

Suojuan Zhang¹ · Song Huang¹ · Xiaohan Yu¹ · Enhong Chen² · Fei Wang² · Zhenya Huang²

Received: 27 July 2021 / Revised: 21 October 2021 / Accepted: 17 December 2021 /

Published online: 14 May 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Online education brings more possibilities for personalized learning, in which identifying the cognitive state of learners is conducive to better providing learning services. Cognitive diagnosis is an effective measurement to assess the cognitive state of students through response data of answering the problems (e.g., right or wrong). Generally, the cognitive diagnosis framework includes the mastery of skills required by a specified problem and the aggregation of skills. The current multi-skill aggregation methods are mainly divided into conjunctive and compensatory methods and generally considered that each skill has the same effect on the correct response. However, in practical learning situations, there may be more complex interactions between skills, in which each skill has different weight impacting the final result. To this end, this paper proposes a generalized multi-skill aggregation method based on the Sugeno integral (SI-GAM) and introduces fuzzy measures to characterize the complex interactions between skills. We also provide a new idea for modeling multi-strategy problems. The cognitive diagnosis process is implemented by a more general and interpretable aggregation method. Finally, the feasibility and effectiveness of the model are verified on synthetic and real-world datasets.

Keywords Cognitive diagnosis · Fuzzy measure · Sugeno integral · Multi-skill aggregation · Multi-skill interactions · Multiple strategies

1 Introduction

Under accelerated globalization and the rapid development of information technology, online education has gradually developed. To restrain the spread of COVID-19, most schools around the world have been closed. Online learning on a large scale started, which further promoted the development of online education [35]. Online education brings more possibilities for the personalized learning. However, facing many learners, how to better provide learning services for learners is a challenging issue. The premise is to identify

✉ Song Huang
huangsong@aeu.edu.cn

Extended author information available on the last page of the article

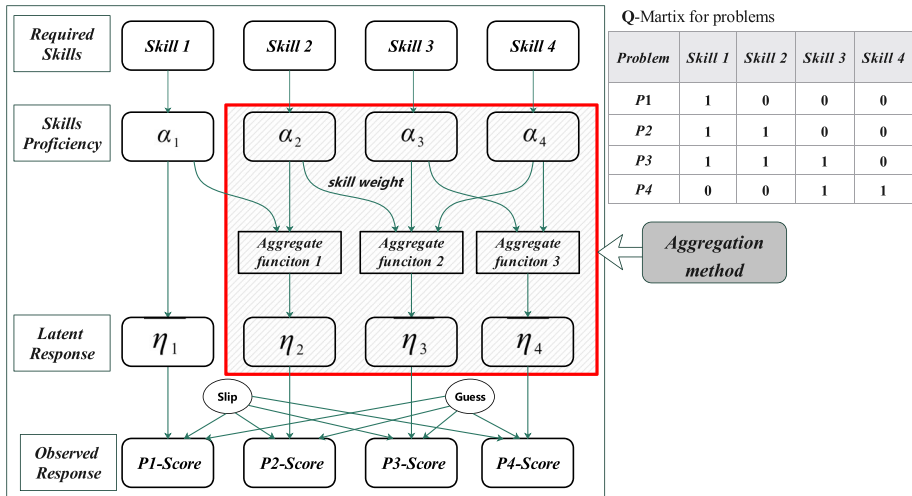


Fig. 1 A general cognitive diagnosis model framework

learners' current cognitive state. However, learners' cognitive states can not be directly observed, which needs to be reflected through learning behavior. Cognitive diagnosis models (CDMs) can be used to identify students' potential cognitive states and alleviate this problem to a certain extent. It has been viewed as a psychometric tool in educational assessments to estimate students' proficiency [33]. CDMs provide specific information in response patterns to aid in acquiring the knowledge state that cannot be observed directly. In terms of students' strengths and weaknesses, better instruction and possible interventions to address individual and group needs are implemented.

Several CDMs of various formulations have been proposed in the literature. Most CDMs consist of three major components: (1) **Q-matrix** [42], which is the prior knowledge assigned by education experts to denote which skills are needed; (2) the diagnosis vector of students' proficiency of required skills; and (3) the condensation rules [34], which can be represented by the aggregate function. A general CDM Framework is shown in Figure 1. The **Q-matrix** can build a link between problems and the corresponding skills. As shown in the table on the right, Problem 1 only needs a single skill, and Problem 2 requires two skills to be answered correctly. More skills are involved in Problem 3. The first layer of the CDM framework consists of several related skill nodes. Students' proficiency in various skills is described as vector α , which can be implied by the response to the problem (e.g., correct or wrong). The third layer reflects the aggregation when multiple skills are involved in a problem. Then, the latent response (η) is acquired by combining the mastery of skills (α) and specific aggregate functions. Finally, the success probability is computed by the latent response with the slip and guess factors.

From Figure 1, we can see the aggregation method of cognitive diagnosis should consider the weight of skills and the determination of aggregate function in the multi-skill learning scenario. Although skill weight have been mentioned in a few studies, skill vectors are often assumed to have the same probability of answering a problem correctly. Tatsuoaka [41] claimed that certain skills would be more critical than others in solving a problem or completing a task. The G-DINA model [13] decomposed the sum of the effects due to the presence of specific skills and their interactions. Lei et al. [21]

Table 1 Example of multi-skill interactions

x_1	x_2	x_3	x_4	P(Correct)
1	0	0	0	0.3
0	1	0	0	0.4
1	0	0	1	0.95
0	1	1	0	0.5
1	1	1	1	1

developed a polytomous cognitive diagnostic model that considers that different skills indeed have different weight. When a student with certain required skills or the subset of the skills' coalition has a higher probability of answering the problem correctly, it can be considered that the skills are more important for solving the problem. For instance, a problem requires four skills $X = \{x_1, x_2, x_3, x_4\}$ to be answered correctly. If the probability of correct response is similar to possessing all skills when a student only masters a partition of skills (e.g., $\{x_1, x_4\}$ as shown in Table 1). The reason for this may be that each skill vector varies in its degrees of effect. Mastering the coalition of skills x_1 and x_4 is more important than knowing the skills x_2 and x_3 , and their probabilities of success are not identical. Here, the importance of each skill affects the response. Hence, we need to describe the weight of a single skill and, more importantly, represent the weight of the subset of the skills' coalition.

On the other hand, the functions employed to aggregate multiple skills are divided into two categories [36]: conjunctive and compensatory approaches. Especially, the conjunctive approach is widely used in CDM models, such as the deterministic inputs, noisy, 'and' gate model (e.g., DINA model) [11, 22, 27] and the generalized DINA (G-DINA) model [13]. In these models, the interaction of multiple skills is conjunctive, which means that only a student who has all skills can answer the problem correctly. In the compensatory approach, some skills may make up for the lack of other skills, including the deterministic input, noisy-or-gate model (DINO) [28, 43], and the compensatory reparametrized unified model [23]. Based on these two approaches, Wu et al. [16, 49] further distinguished objective and subjective problems under this assumption: the skills' interaction on objective (or subjective) problems is conjunctive (or compensatory). Moreover, Zhang et al. [51] modeled the structure of knowledge components and capture the correlations between submissions to evaluate student knowledge states. The difference between the conjunctive and the compensatory approach is the selection of the multi-skill aggregate functions. The more skills involved in a problem, the more complicated it is to determine which aggregation function should be used, and this is not a trivial task. Neither the conjunctive nor the compensatory approach can effectively express the fusion between multiple skills. Generally, regardless of which cognitive diagnosis model is selected, the multi-skills aggregation method is considered. Thus, this paper focuses on a generalized multi-skill aggregation method to reflect the skill weight and aggregate functions.

The multi-strategy problem also increases the difficulty of aggregation method modeling. Most CDMs ignore the diversity of strategies and assume that a single strategy that corresponds to a \mathbf{Q} -matrix may be used to solve a problem. As shown in Table 2, a student may achieve success with Strategy A (using the combination of x_1 and x_2) or Strategy B (using the combination of x_3 and x_4). To address the issue of multiple strategies, De la Torre and Douglas [10] introduced a model that extends the DINA model to allow for multiple strategies of problem-solving. The multi-strategy DINA model may be coded

Table 2 Example of multiple strategies

Strategy	x_1	x_2	x_3	x_4	P(Correct)
A	1	1	0	0	1
B	0	0	1	1	1

by constructing M different matrices. Based on this idea, Tu et al. [46] took into account polytomous scoring items, and then Ma et al. [32] proposed a generalized multi-strategy CDM. These studies adopt multiple matrices to represent multiple strategies, while skill aggregation methods still only consider a single strategy [10]. In addition, in practice, the \mathbf{Q} -matrix is typically assigned by domain experts, which is certainly affected by the subjective tendency of experts and, to a large extent, may consist of some misspecifications [47]. Considering multiple strategies in the multi-skill aggregation method, the model will become more general.

To further explore the above challenges, We introduce a general multi-skill aggregation method to reflect skill weight, accommodate various aggregate functions (e.g., conjunction, compensation methods), and different strategy selections. The main contributions of this paper are outlined as follows:

1. Illustrate the interactions between multiple skills and model the skills' weight.
2. Develop a general multi-skill aggregation method that is independent of any specific cognitive diagnosis model.
3. Put forward a new approach to explore alternative strategies in cognitive diagnosis.

The rest of the paper is as follows: Section 2 introduces the relevant background knowledge, including the cognitive diagnosis model, fuzzy measure, and Sugeno integral. Section 3 describes the modeling process of the multi-skill aggregation method in cognitive diagnosis. Firstly, we discuss the representation of skill weight based on the interactions. Secondly, the Sugeno integral is introduced to realize the aggregation of multi-skill in cognitive diagnosis. Section 4 is compared with the representative models to illustrate the generality of our proposed method. In Section 5, the effectiveness of the aggregation method in solving multi-strategy problems is discussed and proved. Section 6 presents the experiment and results. Finally, the conclusions and discussion of future research are given in Section 7.

2 Background

2.1 Cognitive diagnosis models

Cognitive diagnostic models (CDMs) aim to identify cognitive processing and to evaluate whether a student has mastered or possesses specific cognitive skills or knowledge [16]. CDMs assume a relationship between the students' knowledge mastery and the skills required to solve a problem. A massive effort has been made about cognitive diagnosis models, such as the DINA model, Item Response Theory (IRT) [6], G-DINA model [13], Polytomous Hierarchical DINA model (PH-DINA model) [4]. Specially, IRT is one of the most basic and classic psychological and educational theories which roots in psychological measurement [37]. With

the student's latent trait θ , problem discrimination a , and difficulty b as parameters, IRT can predict the probability that the student answers a specific problem correctly with a logistic function. The function is defined as follows:

$$P(\theta) = \frac{1}{1 + e^{-Da(\theta-b)}}$$

where $P(\theta)$ is the correct probability, D is a constant which often as 1.7. However, the latent traits of students (such as mathematics comprehensive ability) in the IRT model do not consider the ability of students in each skill, it is unnecessary to discuss the multi-skill aggregation issue.

In this section, the details of cognitive diagnosis models and involved aggregation methods are discussed. DINA is considered to be one of the most widely used and representative cognitive diagnostic models. A series of extended models such as HO-DINA (Higher-Order DINA) [8], NIDA (the noisy inputs, deterministic “and” gate model)[27], and P-DINA (Polytomous deterministic inputs, noisy “and” gate[14] have been developed based on the DINA model, which greatly enriched the application of the DINA model. Firstly, we take the DINA model as a prototype to understand the process of cognitive diagnosis.

2.1.1 The DINA model

DINA is a dichotomous cognitive diagnosis model in which the skill's mastery is expressed by $\{0, 1\}$. At this time, the calculated answer probability is either 1 or 0. Under DINA model [9], each item has a corresponding **Q**-matrix, which represents the skills involved. The **Q**-matrix specifies the skills required for each item and is a $J \times K$ matrix of zeros and ones; the element in the j th row and k th column of the matrix is named q_{jk} .

$$q_{jk} = \begin{cases} 0, & \text{if skill } x_k \text{ is not required by item } j \\ 1, & \text{if skill } x_k \text{ is required by item } j \end{cases}$$

Let $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iK})$ denote the proficiency of K skills that are needed to solve the items for the i^{th} student. For a specific student, α_i can be simplified as α . The aggregate function between the student's knowledge mastery and the item specification defines the latent response variable:

$$\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}} \quad (1)$$

Here, without the slipping and guessing factors, the variable η_{ij} can be regarded as an ideal response or latent response, which is also known as the mastery of the i^{th} student on the j^{th} problem [40]. In this paper, η_{ij} can be simplified as η , and the two symbols are not distinguished in the following description and can be written alternately. From (1), only a student who has all the skills from the **Q**-matrix may answer a problem correctly. The skill weight are viewed as the same. Moreover, noise is introduced in the cognitive diagnostic process due to slip and guess parameters. To account for the probabilistic nature of the observed response, slip and guessing parameters are considered on the latent response [12], and are defined as $s_j = P(X_{ij} = 0 | \eta_{ij} = 1)$ and $g_j = P(X_{ij} = 1 | \eta_{ij} = 0)$ respectively. Here, $X_{ij} = 1$ means that the answer is correct, otherwise it is wrong. The factor s_j denotes the probability that a student fails to answer the problem correctly when he possesses all the

Fig. 2 The process of cognitive diagnosis

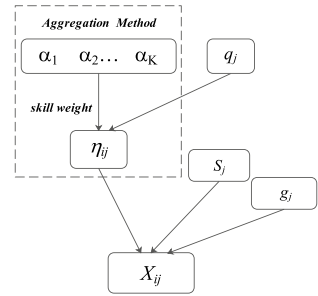
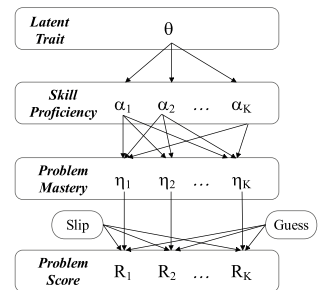


Fig. 3 The fuzzy cognitive diagnosis framework



required skills. The factor g_j is the probability of correctly answering without one or more required skills. Therefore, the probability of the i^{th} student with skills mastery vector α_i answering the j^{th} problem correctly is given by

$$P(X_{ij} = 1 | \alpha_i) = (1 - s_i)^{\eta_{ij}} g_j^{1 - \eta_{ij}} \quad (2)$$

No matter the type of cognitive diagnosis model, the specified skill proficiency, skill weight, aggregate functions, and slip and guessing factors need to be considered. In the cognitive diagnosis process, the latent response is determined by the skills proficiency α , skill weight, and aggregate function(in Figure 2). The aggregation method can realize the generation of the latent response. In the DINA model, the weight of each skill is the same, and the latent response is obtained by multiplying the proficiency of multiple skills, which can be regarded as a conjunctive aggregate function. As long as one skill's mastery equals 0, the latent answering response is also 0.

2.1.2 The fuzzy cognitive diagnosis framework

Based on the general cognitive diagnosis process, the fuzzy cognitive diagnosis framework (FuzzyCDF) considered both subjective and objective exercise types of balancing precision and interpretability of the diagnosis results [31]. Without considering the weight difference, FuzzyCDF models two multiple skills aggregate functions according to the characteristics of subjective and objective problems. As shown in Figure 3 [30], the framework is a generic process that starts with the student's latent traits (e.g., a general ability in math) and then get the student's skill proficiency as follows:

$$\alpha_k = \frac{1}{1 + \exp[-1.7a_k(\theta - b_k)]} \quad (3)$$

Here, we can see that the proficiency of a student on a specific skill (α_k) depends on the difference between the student's high-order latent trait (θ) and the properties of the skill: the difficulty (b_k) and discrimination (a_k) of skill k for student j [45]. The coefficient 1.7 is an empirical scaling constant in logistic cognitive models [26]. By (3), we can determine a student's proficiency in specific skills from his latent trait. Next, the student's problem mastery (i.e., the latent response) is computed by specific aggregation methods, and the observable scores are generated by considering slip and guess factors.

In FuzzyCDF, each skill is deemed as equally important. The skills' aggregation on objective problems is conjunctive, and the minimum operator has been used for the aggregate function. The latent response η is defined as the following formulation under a conjunctive assumption:

$$\eta = \min\{\alpha_k | q_k = 1, k = 1, 2, \dots, K\} \quad (4)$$

For subjective problems, the skills' aggregation is under the compensatory assumption. This assumption means that a student could solve the problem by mastering at least one required skill. The maximum operator forms the aggregate function. Similarly, the latent response η can be written as follows:

$$\eta = \max\{\alpha_k | q_k = 1, k = 1, 2, \dots, K\} \quad (5)$$

In DINA and FuzzyCDF, skills share the same weight, and different functions are used for skills fusion to form a multi-skill aggregation method. However, in the practical learning scenario, the weight between skills are often unequal. We also expect to establish a unified expression of conjunctive and complementary aggregate functions. Therefore, the determination of the aggregation method should consider the weight of multiple skills and the selection of the aggregate function.

2.2 Fuzzy measure

The classic measure theory stems from measuring the size of sets, such as the length of objects, the area of regions, or the weighted average method; these are additive measures. However, in many cases, additivity cannot be satisfied. For example, the efficiency of two workers' cooperation cannot be simply equal to the sum of their efficiencies. Sugeno [38] first proposed weak monotonicity instead of additivity for set functions and called fuzzy measures. Thus, fuzzy measures can be used to handle non-additive issues. The fuzzy measures are an extension of the classic measure theory, enabling one to represent the non-additive property efficiently [2]. The definition is as follows:

Definition 1 A set function $\mu : 2^X \rightarrow [0, 1]$ is a fuzzy measure if it satisfies the following axioms:

1. $\mu(\emptyset) = 0, \mu(X) = 1$ (normalized)
2. $\mu(A) \leq \mu(B)$ whenever $A \subseteq B$ for $A, B \in 2^X$ (monotonicity)

Here, X is denoted as a finite set of elements, such as players in a cooperative game, criteria in a decision problem, attributes, experts, or voters in an opinion pooling problem, etc. Specifically, if we have $A \cup B = C \Rightarrow \mu(A) + \mu(B) = \mu(C)$, $\forall A, B, C \subseteq 2^X$, the fuzzy measure is called a probability measure [17], which is an additive measure. The non-additive feature makes fuzzy measures reflect the interdependence among the elements in detail by measuring importance. Fuzzy measures μ usually denote importance, reliability, satisfaction, or similar concepts [44].

2.3 Sugeno integral

The Sugeno integral is one of the most important fuzzy integrals, and it has been proved useful in several applications [20], notably in some game theory problems and multi-criteria decision-making [18]. A decision-making problem is to find a desirable solution from a finite number of feasible alternatives assessed on multiple skills [48]. Here, the answer response depends on the proficiency of multiple skills, which is regarded as a decision-making problem. The properties of the Sugeno integral are well known, as well as its relationship with classical aggregation operators.

Definition 2 Given a finite set $X = \{x_1, x_2, \dots, x_K\}$, a fuzzy measure $\mu : 2^X \rightarrow [0, 1]$ and a function $f : X \rightarrow [0, 1]$, also written as $f(x_k) = \alpha_k$, the Sugeno integral with respect to fuzzy measures μ is defined by

$$(s) \int f d\mu = \bigvee_{A \subseteq X} [\min_{x_k \in A} f(x_k) \wedge \mu(A)] = \bigvee_{k=1}^K [\alpha_k \wedge \mu\{x_k | f(x_k) \geq \alpha_k\}] \quad (6)$$

Here, $A \subseteq X$ for all it holds $\mu(A) \in [0, 1]$. Without losing generality, we assume that $f(x_k)$ is monotonically increasing with respect to k . Here, $f(x_k)$ is treated as a function that is ordered by $0 \leq f(x_1) \leq f(x_2) \leq \dots \leq f(x_K) \leq 1$. The Sugeno integral is solely defined based on aggregate operators \vee and \wedge as opposed to other integrals, it can be applied when set elements are in a complex relationship. Note that $\vee(a, b) = a$, if and only if $a \geq b$, and $\wedge(a, b) = b$, if and only if $a \geq b$.

3 The generalized multi-skill aggregation method based on the Sugeno integral

This section models the generalized multi-skill aggregation method based on the Sugeno integral (hereinafter referred to as SI-GAM) independent of any specific cognitive diagnosis model. SI-GAM combines the mastery supplied by required skills according to weight of multi-skill to obtain the latent response. Therefore, the modeling of SI-GAM consists of two parts: the representation of skill weight and the selection of aggregate function.

3.1 The representation of skill weight

Firstly, we discuss the interactions between multiple skills. The interactions exist not only between single skills but also between skill sets. Then, we introduce fuzzy measures for modeling the interactions quantitatively.

3.1.1 The interactions between multiple skills

Most entities in the real world are interconnected, which form complex networks [50]. In the field of learning, there are complex interactions between skills. More often than not, mastering certain skills increases the probability of success than getting other skills. As mentioned in Jimmy's research [13], mastering three of the four required skills for the problem, namely, performing a basic fraction subtraction operation (x_1), separating the whole number from the fraction (x_3), and borrowing one from the whole number (x_4), provided a markedly higher probability of success compared to any other subsets of skills. This means that different skills may have different weight for solving a problem due to the influence of the skill or interactions with multiple skills. Thus, the importance of correctly identifying the skills cannot be overstated. The aggregation method should reflect the skill weight for correct answers.

The first step in making inferences with the Attribute Hierarchy Method depends on accurately identifying the psychological ordering of cognitive competencies required to solve test problems. Identifying skill hierarchies serves a critical function. In this paper, we discuss the possible interactions between multiple skills under the hierarchical structures (including linear, convergent, divergent, and independent hierarchies [29, 39]).

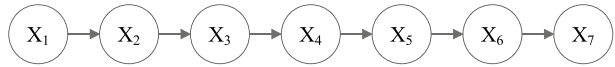
1. Redundancy If the contribution of a pair of skills x_i, x_j is not greater than the sum of their individual supports, there is a negative synergistic relationship between x_i and x_j . That is to say, there is no enhancement by combining the skills x_i and x_j . The general formula is defined as:

$$\omega_{\{x_i, x_j\}} \leq \omega_{x_i} + \omega_{x_j}$$

Here, ω denotes a certain weight, that is, the degree of support or contribution of the skill to the correct answer. Redundancy can be regarded as a particular case of negative synergies, such as the prerequisite relationship between skills. In terms of the hierarchical structure between skills, the prerequisite can be viewed as a linear structure. For example, the skill x_1 is a prerequisite to skill x_2 (shown in Figure 4); this implies that a student is not expected to possess skill x_2 unless they already possess skill x_1 . Thus, skill x_2 is more important than x_1 for answering the problem correctly, and thus skill x_2 has a larger weight: $\omega_{x_2} > \omega_{x_1}$. Then, we define the interaction of skills x_1 and x_2 , $\omega_{\{x_1, x_2\}} = \omega_{x_2}$, that means that the support strength of the combination of x_1 and x_2 is equal to the skill weight with more support. The skill x_1 may not make any contribution to improving the correct answer rate. The redundancy relationship between x_i and x_j is defined by:

$$\omega_{\{x_i, x_j\}} = \max(\omega_{x_i}, \omega_{x_j}) \quad (7)$$

Hence, the supports are not additive because of the existing hierarchy relationship. Other skills follow a similar interpretation. Mastering skill x_2 is a prerequisite for

Fig. 4 Linear structure between skills

obtaining x_3 . Therefore, it can be considered that there is redundancy between x_3 and x_2 , that is, $\omega_{x_2} + \omega_{x_3} \geq \omega_{\{x_2, x_3\}} = \omega_{x_3}$. Other skill pairs (x_3 and x_4 , x_4 and x_5 , etc.) also follow this relationship.

2. **Positive synergy** The performance of the pair of skills x_i, x_j is better than the sum of their individual performances. That is to say, the combination of x_i and x_j directly impacts the probability of a correct response as opposed to using each skill separately. The skills interact through compensation, and the expression is defined by:

$$\omega_{\{x_i, x_j\}} > \omega_{x_i} + \omega_{x_j} \quad (8)$$

Combined with the hierarchical structure between skills, the divergent structure[29] (Figure 5) can be considered as a manifestation of positive synergy. If a student has mastered both the skills x_4 and x_6 , the probability of answering the problem correctly is higher than the sum of the success probabilities of mastering skill x_4 and skill x_6 alone. The coalition of x_4 and x_6 may produce more chemical effects, such as creating a new latent skill, which is important for solving the problem. This can be denoted as $\omega_{\{x_4, x_6\}} > \omega_{x_4} + \omega_{x_6}$. The contribution of the combination of x_4 and x_6 is greater than the sum of the two skills separately. Mastering both x_4 and x_6 can achieve a dramatic increase in the probability of success on a problem. The interaction of positive synergy is also non-additive.

3. **Independency** Intermediate case, where each skill contributes to the probability of a correct response:

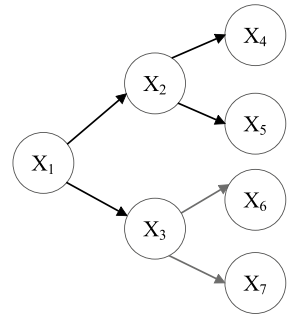
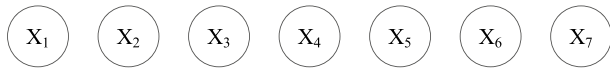
$$\omega_{\{x_i, x_j\}} = \omega_{x_i} + \omega_{x_j} \quad (9)$$

As shown in Figure 6 [29], the skills under the structure are independent. There are additive relationships since the multiple skills do not interfere with each other. The expression can be written as: $\omega_{\{x_1, x_2\}} = \omega_{x_1} + \omega_{x_2}$.

From the assumption that skills are hierarchically organized, it is essential to note that the structures in Figures 4, 5 and 6 can be combined to form increasingly complex networks of hierarchies where the complexity varies with the cognitive task. That is, there may be non-additive interactions between skills in the problem. Commonly used weight calculation methods (such as the weighted average) cannot reflect this interactive relationship between skills. Thus, our paper tries to characterize non-additive weight, which involves the effect of singletons and the interactions of multiple skills.

3.1.2 Using fuzzy measures for the non-additivity weight of skills

In this paper, skills' weight can be interpreted as certain importance, i.e., the support strength of skills for the correct response. Fuzzy measures are introduced to construct the non-additive weight for skills, which are powerful in characterizing the interactions of multiple skills [35]. For instance, we have a set of skills $X = \{x_1, x_2, x_3, x_4\}$ that solve a problem, and we assume the fuzzy measures $\mu(\{x_1\}) = 0.2$ and $\mu(\{x_2\}) = 0.3$. Here, the skill x_2 has more

Fig. 5 Divergent structure between skills**Fig. 6** Independent structure with skills

support than x_1 for the correct response. That is to say, the skill x_2 is more critical than x_1 in solving the problem. As noted earlier, we cannot apply the weighted calculation directly to add the support measures. If we suppose $\mu(\{x_1, x_2\}) = 0.32$, joining x_1 with x_2 does not yield a markedly higher probability of success than having a grasp of x_2 alone. Obviously, $\mu(\{x_1, x_2\}) \neq \mu(\{x_1\}) + \mu(\{x_2\})$. The weak monotonicity is in line with the characteristics of skills. We can use a non-negative monotone set function defined on its set to describe the importance of each information source (i.e., the support strength of the skill) and their varied combinations. Thus, the non-additivity of fuzzy measures can satisfy the non-additive weight for multi-skill aggregation. Fuzzy measures enable us to express the interactions better.

It is not comprehensive to determine the global importance of each skill only by the measure of a single skill; we must also consider the measured value of the subset involved in the skill. Therefore, the importance of the skills coalition should be discussed. Let us denote X as a set of skills containing all possible skills needed to solve the problem. Then $\forall A \subseteq X$, $\mu(A)$ represents the importance or strength of the coalition A for the particular skills group under consideration. To have a flexible representation of complex interactions between skills (e.g., positive synergy or redundancy between skills), it is necessary to consider fuzzy measures in their full generality and not restrict oneself. In multi-skill aggregation problems, X denotes the set of all involved skills.

Let $X = \{x_1, x_2, \dots, x_K\}$ be a set of skills. The number of fuzzy measures is 2^X : $\mu(\{x_1\}), \dots, \mu(\{x_K\}), \mu(\{x_1, x_2\}), \dots, \mu(\{x_{k-1}, x_k\}), \dots, \mu(\{x_1, x_2, \dots, x_K\})$. We suppose that there are four skills required for a specific problem, the fuzzy measures are shown in Table 3.

According to Definition 1 in Section 2, $\mu(A) \leq \mu(B)$ whenever $A \subseteq B$ for $A, B \in 2^X$ (monotonicity). It is easy to see that these corresponding values satisfy the properties of fuzzy measures. Table 3 shows that the interaction of the skill x_1 and x_2 is redundant, that is $\mu(\{x_1, x_2\}) < \mu(\{x_1\}) + \mu(\{x_2\})$, and the combination of x_1 and x_3 enhances the support of success, that is $\mu(\{x_1, x_3\}) > \mu(\{x_1\}) + \mu(\{x_3\})$. Moreover, we infer that the skill x_4 plays an important role in this problem. Using fuzzy measures instead of the equally weighted status can more effectively represent the interactions between skills.

Table 3 A sample of fuzzy measures

Fuzzy measure set	Corresponding value	Fuzzy measure sets	Corresponding value
\emptyset	0	$\mu(\{x_2, x_3\})$	0.55
$\mu(\{x_1\})$	0.2	$\mu(\{x_2, x_4\})$	0.75
$\mu(\{x_2\})$	0.3	$\mu(\{x_3, x_4\})$	0.78
$\mu(\{x_3\})$	0.1	$\mu(\{x_2, x_3, x_4\})$	0.85
$\mu(\{x_4\})$	0.6	$\mu(\{x_1, x_2, x_3\})$	0.87
$\mu(\{x_1, x_2\})$	0.32	$\mu(\{x_1, x_2, x_4\})$	0.9
$\mu(\{x_1, x_3\})$	0.52	$\mu(\{x_1, x_3, x_4\})$	0.92
$\mu(\{x_1, x_4\})$	0.72	$\mu(\{x_1, x_2, x_3, x_4\})$	1

3.2 The aggregation of multiple skills in cognitive diagnosis

Based on the importance of aggregate function selection, our idea is to propose a multi-skill aggregation method. This method combines the representation of skill weight and expresses the function of conjunctive and compensatory in a uniform way.

3.2.1 The importance of aggregate function selection

In practical learning scenarios, the aggregation of multiple skills is a complex issue. For an arithmetic problem, such as $4\frac{4}{12} - 2\frac{7}{12}$ (Figure 7), four skills $\{x_1, x_2, x_3, x_4\}$ are required based on the **Q**-matrix.

Assuming the skill proficiency of a student is given in Table 4, the value is a binary variable in $\{0, 1\}$ (Dichotomous value), such as in the DINA model and DINO model. If the aggregation method is conjunctive, the latent response η equals zero; the latent response η equals one under the compensatory approach.

The selection of an aggregate function also has a dramatic impact on the polytomous situation. In FuzzyCDF [49], the skill proficiency is a fuzzy variable in $[0, 1]$ (polytomous value). From Table 5, if the method is regarded as the conjunctive, using the min operator can compute the latent response $\eta = 0.3$ based on (4). Then the approach is the compensatory type, and we use the max operator to calculate $\eta = 0.9$ based on (5).

Different aggregate functions lead to different results. In other words, the selection of aggregate functions may influence the latent response. Specifically, if the lack of any skill means the student will not answer correctly, we should choose the conjunctive approach. Conversely, if any one of these skills can make up for the lack of other skills so that the student can provide the correct answer, the compensatory approach is a more suitable choice. However, using only the conjunctive or compensation function may not reflect the fusion of practical problems.

3.2.2 Using Sugeno integral for multi-skill aggregation

In cognitive diagnosis, multiple skills may be independent or interact with each other. In this paper, the multi-skill aggregation should consider non-additive interaction. Furthermore, we should define an appropriate function for skills fusion. Sugeno integrals are

<i>Problem</i>	X_1	X_2	X_3	X_4
$4\frac{4}{12}-2\frac{7}{12}$	1	1	1	1

- x_1 : performing basic fraction subtraction operation
- x_2 : simplifying/reducing
- x_3 : separating whole number from fraction
- x_4 : borrowing one from whole number to fraction

Fig. 7 Q-matrix for the fraction subtraction data

Table 4 Different aggregation for dichotomous

Aggregation Method	Skill Proficiency (dichotomous)			
	$\alpha_1 = 1$	$\alpha_2 = 0$	$\alpha_3 = 0$	$\alpha_4 = 0$
Conjunctive	$\eta = 0$			
Compensatory	$\eta = 1$			

Table 5 Different aggregation for polytomous

Aggregation Method	Skill Proficiency (polytomous)			
	$\alpha_1 = 0.9$	$\alpha_2 = 0.3$	$\alpha_3 = 0.4$	$\alpha_4 = 0.6$
Conjunctive	$\eta = \min(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = 0.3$			
Compensatory	$\eta = \max(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = 0.9$			

known to be one of the most influential and flexible functions. They permit the fusion of the information under different assumptions on the independence of the information sources. In particular, they can be used to model situations in which sources are independent and in cases where such independence cannot be assured. Thus, we introduce the Sugeno integral into the multi-skill aggregation method (SI-GAM) to represent the interactions and fusion of skills. The Sugeno integral models the aggregation method among globally important criteria. Here, the criteria denote skill proficiency, which affects problem-solving performance. The global importance of criteria depends on the non-additive weight, i.e., the fuzzy measures. From (2), we explore an alternative way to calculate η using Sugeno integral:

$$\eta = \bigvee_{k=1}^K [\alpha_k \wedge \mu(\{x_i | \alpha_i \geq \alpha_k\})] \tag{10}$$

In SI-GAM, the latent response is defined over an aggregate function on the reference set of skills X . Then, given x_k in X , the variable α_k corresponds to the mastery of the skills

on X . Next, sort α_k , and the sorted result $\alpha_{(k)}$ refers to the ordered inputs generated by $\alpha_{(1)} \leq \alpha_{(2)} \leq \dots \leq \alpha_{(K)}$, that is

$$\alpha_{(1)} = \min(\alpha_1, \alpha_2, \dots, \alpha_K) \quad (11)$$

$$\alpha_{(K)} = \max(\alpha_1, \alpha_2, \dots, \alpha_K) \quad (12)$$

Here, skill weight is expressed by the fuzzy measures μ , which are considered the support strengths for the correct response, and thus have some importance to the corresponding problem. The fuzzy measure $\mu(A)$:

$$\mu(A) \subseteq \{\mu(\{x_1\}), \dots, \mu(\{x_K\}), \mu(\{x_1, x_2\}), \dots, \mu(\{x_1, x_2, \dots, x_K\})\}$$

Corresponding with the chain nested subsets induced by the ordering permutation of x_k . There are two operators involved in (10), which generalizes max-min operators. The SI-GAM can also be defined as:

$$\eta = \bigvee_{k=1}^K [\alpha_{(k)} \wedge \mu(A_k)] \quad (13)$$

where $A_k = \{x_{(k)}, x_{(k+1)}, \dots, x_{(K)}\}$, and $x_{(k)}$ is the skill corresponding to the ranked skill's mastery $\alpha_{(k)}$.

Example 1 There are four skills required for solving the problem in Figure 7. The latent response η can be computed by the mastery of skills and corresponding fuzzy measures. A student's proficiency in the four skills $\{x_1, x_2, x_3, x_4\}$ is $\alpha_1 = 0.9$, $\alpha_2 = 0.3$, $\alpha_3 = 0.4$, and $\alpha_4 = 0.6$. α_k is the vector of skill proficiency expressed on some ordinal scale. After reordering in ascending order, this is expressed as $\alpha_{(1)} = \alpha_2 = 0.3$, $\alpha_{(2)} = \alpha_3 = 0.4$, $\alpha_{(3)} = \alpha_4 = 0.6$, $\alpha_{(4)} = \alpha_1 = 0.9$.

Assuming the fuzzy measures are known (seen in Table 3), it is sufficient to define the mapping of the proficiency of skills α_k to the latent response η :

$$\begin{aligned} \eta &= \bigvee_{k=1}^4 [\alpha_{(k)} \wedge \mu(A_k)] \\ &= (\alpha_{(1)} \wedge \mu(\{x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}\})) \bigvee (\alpha_{(2)} \wedge \mu(\{x_{(2)}, x_{(3)}, x_{(4)}\})) \\ &\quad \bigvee (\alpha_{(3)} \wedge \mu(\{x_{(3)}, x_{(4)}\})) \bigvee (\alpha_{(4)} \wedge \mu(\{x_{(4)}\})) \\ &= \max(\min(0.3, 1), \min(0.4, 0.92), \min(0.6, 0.72), \min(0.9, 0.2)) \\ &= 0.6 \end{aligned}$$

The solution of multi-skill aggregation was inspired by the Sugeno integral, one of the typical representatives for fuzzy integrals that are very useful for fusing information from various sources. The SI-GAM can take partial support for the result from the standpoint of each information source and fuse it with the worth of each subset of X in a non-additive fashion. This worth is encoded into a fuzzy measure [1]. In our proposed SI-GAM, the interactions between skills are expressed by fuzzy measures, and the min-max operator expresses the aggregate function to realize the modeling of the multi-skill aggregation method.

4 Comparison with other cognitive diagnosis models

In this section, SI-GAM will be compared to other aggregation methods in existing models. Through derivation, it can be proved that SI-GAM is also applicable to current models. Firstly, we discuss the non-compensatory /conjunctive model (DINA) and compensatory model (DINO) for dichotomous response. Secondly, we compare with FuzzyCDF, which can express the skill fusion of conjunctive and compensatory in polytomous response simultaneously. Since most cognitive diagnosis models consider the equal weight of each skill, without loss of generality, we simplify the value of fuzzy measures μ to $\{0, 1\}$, that is, if the skill sets (A_K) has support for correct answering, the $\mu(A_K) = 1$, otherwise $\mu(A_K) = 0$.

4.1 Comparison with DINA model

In the DINA model, the aggregate function is conjunctive. The fuzzy measures can be defined by:

$$\mu(A) = \begin{cases} 1, & \text{if } A = X \\ 0, & \text{others} \end{cases} \quad (14)$$

Using SI-GAM, we can well describe the skills' fusion by conjunctive approach as follows (14):

$$\begin{aligned} \eta &= \bigvee_{k=1}^K [\alpha_{(k)} \wedge \mu(A_k)] \\ &= (\alpha_{(1)} \wedge 1) \bigvee (\alpha_{(2)} \wedge 0) \bigvee (\alpha_{(3)} \wedge 0) \cdots \bigvee (\alpha_{(K)} \wedge 0) \\ &= \alpha_{(1)} \end{aligned}$$

With SI-GAM, we can infer that the skill with the lowest mastery determines the value of the latent response η . Thus, if $\alpha_k = 0$, the skill x_k is not mastered by the student. The latent responses η obtained from both approaches are equal to 0. Our SI-GAM can be viewed as a variant of the conjunctive approach.

Example 2 Assuming the skill proficiency $\alpha_1 = 1, \alpha_2 = 1, \alpha_3 = 0, \alpha_4 = 1$ from Table 4, the response can be computed. The condensation rule in the DINA model typically assumes that a student that has mastered all required skills may solve the problem. Hence, the latent response $\eta = 0$ (1). x_3 is the skill of the worst mastery among the four skills. Since the DINA model is dichotomous, we can define $\alpha_{(1)} = \alpha_3 = 0, \alpha_{(2)} = \alpha_1 = 1, \alpha_{(3)} = \alpha_2 = 1, \alpha_{(4)} = \alpha_4 = 1$. Combined with the value of fuzzy measures from (14), the latent response is $\eta = 0$ under SI-GAM. Table 6 shows that the two approaches obtain the same results. In other words, the DINA model's conjunctive approach can be considered a special case of SI-GAM.

Table 6 Different aggregation methods for dichotomous

Aggregation method	Skill proficiency (dichotomous)			
	$\alpha_{(1)} = 0$	$\alpha_{(2)} = 1$	$\alpha_{(3)} = 1$	$\alpha_{(4)} = 1$
Conjunctive approach	$\eta = 0 \times 1 \times 1 \times 1 = 0$			
SI-GAM	$\eta = \max(\min(0, 1), \min(1, 0), \min(1, 0), \min(1, 0)) = 0$			

4.2 Comparison with DINO model

Moreover, the DINO model is the compensation version of the DINA model. It is assumed that a student can answer the problem correctly as long as they master any skill required by the problem, that is, the skills can replace or compensate each other. Thus, in the DINO model, the latent response can be defined as follows:

$$\eta = 1 - \prod_{k=1}^K (1 - \alpha_k)^{q_{jk}}$$

where q_{jk} specifies whether skill x_k is measured by item j . q_{jk} takes the binary values of 0 or 1. In case that skill k is measured by item j , q_{jk} would take a value of 1, otherwise, q_{jk} equals 0. If the students master the skill x_k , then $\alpha_k = 1$, and thereby $1 - \alpha_k$ would be 0. However, if the student does not master the skill α_k , $1 - \alpha_k = 1$. Thus, the latent response of the DINO model is rewritten as:

$$\eta = \begin{cases} 1, & \text{if } \max(\alpha_{(k)}) = 1 \\ 0, & \text{if } \max(\alpha_{(k)}) = 0 \end{cases} \quad (15)$$

$$= \alpha_{(K)}$$

Here, $\alpha_{(k)}$ is the representation of α_k sorting the skills' mastery in ascending order, that means $\alpha_{(1)} \leq \alpha_{(2)} \leq \dots \leq \alpha_{(K-1)} \leq \alpha_{(K)}$. Hence, we can also illustrate the latent response variable η using SI-GAM, which is written as:

$$\begin{aligned} \eta &= \bigvee_{k=1}^K [\alpha_{(k)} \wedge \mu(A_k)] \\ &= (\alpha_{(1)} \wedge 1) \bigvee (\alpha_{(2)} \wedge 1) \dots \bigvee (\alpha_{(l)} \wedge 1) \dots \bigvee (\alpha_{(K)} \wedge 1) \\ &= \alpha_{(K)} \end{aligned}$$

Under SI-GAM, we can deduce that the skill with the highest mastery determines the value of the latent response η . The results are consistent with (15). Thus, it can be seen that the compensation approach in the DINO model can also be regarded as a special case of SI-GAM.

4.3 Comparison with FuzzyCDF

Similarly, SI-GAM can realize the unified modeling of aggregation under conjunctive and compensatory functions. Two types of fuzzy interactions express the problem mastery in

(4) and (5). Suppose the skills' interaction is conjunctive (or compensatory). In that case, a student's mastery of this problem is the degree of membership of this student in the intersection (or union) set of the fuzzy sets related to the skills required by the problem. This framework fuzzifies the skill proficiency, which redefines the original binary variable (i.e., mastered or non-mastered) to a fuzzy one valued in $[0, 1]$ [30].

1. **Conjunctive aggregation** For objective problems, the conjunctive aggregate function is defined as:

$$\eta = \min(\alpha_{x_1}, \alpha_{x_2}, \dots, \alpha_{x_K}) = \alpha_{(1)}$$

In our SI-GAM, we first determine the fuzzy measures by the skill's interactions. If the interaction for the skills has a conjunctive feature, we can argue that there is sufficient support to answer correctly if the fuzzy measure set includes all skills. We can infer the fuzzy measures using (14), and then, the fusion of α_k with respect to μ for each skill will give the success probability for a specified problem. For objective problems, the lowest mastery of skills has a marked impact on the outcome.

Example 3 If the problem requires four skills, we make the following assumption: $\alpha_1 = 0.9$, $\alpha_2 = 0.3$, $\alpha_3 = 0.4$, $\alpha_4 = 0.6$ from Table 5. In FuzzyCDF, the latent response under the fuzzy interaction equals $\min(\alpha_k) = 0.3$. In SI-GAM, the value of fuzzy measures is given from (14), the computation of the latent response (from 13) is then expressed as:

$$\eta = \max(\min(0.3, 1), \min(0.4, 0), \min(0.6, 0), \min(0.9, 0)) = 0.3$$

Now, two methods obtain consistent results. The aggregation method for objective problems in FuzzyCDF is a special case of SI-GAM as well.

2. **Compensatory aggregation** In this case, a student could obtain a higher probability of success for the problem when she has mastered at least one of the required skills $x_k \in \{x_1, x_2, \dots, x_K\}$. In FuzzyCDF, the latent response $\eta = \max(\alpha_{(1)}, \alpha_{(2)}, \dots, \alpha_{(K)}) = \alpha_{(K)}$ from (5) and (12). According to Definition 1, for any $\mu(\{x_k\}) = 1$, the fuzzy measures as $\mu(B) = 1$, if $x_k \in B, B \subseteq X$. Therefore, the latent response is

$$\begin{aligned} \eta &= \bigvee_{k=1}^K [\alpha_{(k)} \wedge \mu(A_k)] \\ &= (\alpha_{(1)} \wedge 1) \bigvee (\alpha_{(2)} \wedge 1) \dots \bigvee (\alpha_{(K)} \wedge 1) \\ &= \alpha_{(K)} \end{aligned} \quad (16)$$

It is easy to conclude that the highest proficiency of skills $\alpha_{(K)}$ is the most critical influence.

Example 4 Given the four skills' proficiency as in Table 5, the latent response is $\eta = \max(\alpha_{(k)}) = 0.9$ with the compensatory interaction in FuzzyCDF. We can also compute the result $\eta = \alpha_{(K)} = 0.9$ in SI-GAM. Here, SI-GAM achieves the same values with the aggregation method for subjective problems. Therefore, the conjunctive and compensatory aggregation for objective and subjective problems in FuzzyCDF can be regarded as special cases of SI-GAM.

In the above description, we have shown more of a generalized aggregation method to obtain a more precise representation of responses for each student. The fuzzy measures in SI-GAM can model the complex interactions between skills, reflect the support strength of skills and skill sets for the correct response, and combine the min-max operator to generate a more general aggregation method.

5 Using SI-GAM for multiple strategies problems

In practical applications, two or more strategies may provide similar support strength to solve the same problem. At present, the main idea behind solving multi-strategy problems suppose that each item has as many as distinct strategies that would suffice. In the multi-strategy DINA model [10], each problem has M distinct strategies. Multiple strategies may be coded by constructing M different matrices, $\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_M$. Let $\eta_{ijm} = \prod_{k=1}^K \alpha_{ik}^{q_{jkm}}$, for $m = 1, 2, \dots, M$, where q_{jkm} denotes the element in the j^{th} row and k^{th} column of \mathbf{Q}_M . The variable η_{ijm} indicates whether i^{th} student has the skills to apply the m^{th} strategy to the j^{th} item. Here, the latent response η_{ij} is denoted as:

$$\eta_{ij} = \max\{\eta_{ij1}, \eta_{ij2}, \dots, \eta_{ijM}\} \quad (17)$$

Suppose there are two strategies for a problem. For Strategy A, there are n skills required to solve the problem, the set of involved skills is $X_1 = \{x_{l(1)'}, x_{l(2)'}, \dots, x_{l(n)'}\}$, where $l(k)' \in \{1, 2, \dots, K\}$. The \mathbf{Q} -matrix can be expressed as $\mathbf{Q}_1 = (q_{11}, q_{12}, \dots, q_{1K})$. Similar, the skills set $X_2 = \{x_{l(1)''}, x_{l(2)''}, \dots, x_{l(m)''}\}$, that is, for Strategy B, m skills are needed for the problem, where $l(k)'' \in \{1, 2, \dots, K\}$. The corresponding matrix is $\mathbf{Q}_2 = (q_{21}, q_{22}, \dots, q_{2K})$. Here, $X_1 \subseteq X$, $X_2 \subseteq X$ and $X_1 \neq X_2$, and we have

$$q_{ik} = \begin{cases} 1, & \text{if } x_k \in X_i, i = 1, 2 \\ 0, & \text{else} \end{cases} \quad (18)$$

Based on this assumption, following Eqs. 17 and 18, the latent response can be written as:

$$\begin{aligned} \eta &= \max\{\eta_1, \eta_2\} \\ &= \max\left\{\prod_{k=l(1)'}^{l(n)'} \alpha_k, \prod_{k=l(1)''}^{l(m)''} \alpha_k\right\} \\ &= \max\{\alpha_{(1)'}, \alpha_{(1)''}\} \end{aligned} \quad (19)$$

Since the skill proficiency α_k of DINA model only takes 0 or 1, the result of multiplication operation depends on the lowest value of α_k .

Next, we deal with multi-strategy problem using our SI-GAM. Obviously, $\mu(X_1) = 1$ and $\mu(X_2) = 1$. For the aggregate function under the DINA model is conjunctive, we can determine fuzzy measures as $\mu(A) = 1$, for $\forall A \subseteq X$ satisfying $X_1 \subseteq A$ or $X_2 \subseteq A$, and $\mu(B) = 0$ for other subsets $B \subseteq X$.

Let $L_k = \{x_k, x_{k+1}, \dots, x_K\}$, we have

$$\begin{aligned} \mu(L_k) &= 1, \text{ if } X_1 \subseteq L_k \text{ or } X_2 \subseteq L_k \\ \mu(L_{k+1}) &= 0, \text{ if } X_1 \not\subseteq L_{k+1} \text{ and } X_2 \not\subseteq L_{k+1} \end{aligned} \quad (20)$$

From (13), the latent response is given by

$$\eta = (\alpha_{(1)} \wedge \mu(L_1)) \bigvee (\alpha_{(2)} \wedge \mu(L_2)) \cdots \bigvee (\alpha_{(K)} \wedge \mu(L_K)) \\ = \alpha_{(k)} \quad (21)$$

with $\mu(L_k) = 1$, $\mu(L_{k+1}) = 0$, and $\alpha_{(1)} \leq \alpha_{(2)} \cdots \leq \alpha_{(K)}$. To compute the latent response, we need to determine the value of fuzzy measures and discuss the different cases that satisfy these constraints.

- Case 1. If $\alpha_{(1)'} < \alpha_{(1)'}$, then $\eta = \alpha_{(k)} = \alpha_{(1)'}$.
- Case 2. If $\alpha_{(1)'} = \alpha_{(1)'}$, then $\eta = \alpha_{(k)} = \alpha_{(1)'} = \alpha_{(1)'}$.
- Case 3. If $\alpha_{(1)'} > \alpha_{(1)'}$, then $\eta = \alpha_{(k)} = \alpha_{(1)'}$.

For readability, the derivation of the latent response under three assumptions is presented in the Appendix 1. Here, it is not hard to notice that two methods can get the same result. The SI-GAM can also solve the multi-strategy problem in cognitive diagnosis. In addition, the SI-GAM can also deal with the polytomous scoring for the multiple strategies by changing the value range of the proficiency of skills α_k to the interval of $[0, 1]$ expands the existing multiple-strategy DINA Model.

Example 5 Two alternative strategies both support solving the problem; e.g., $4\frac{4}{12} - 2\frac{7}{12}$. Let X be a finite set $X = \{x_1, x_2, x_3, x_4, x_5, x_6, \dots, x_K\}$. One strategy, Strategy A, requires students to perform fraction subtraction with mixed numbers and involves skills x_1, x_2, x_3, x_4 . The other strategy, Strategy B, requires subtraction of fractions where mixed numbers are first changed to improper fractions and involves skills x_1, x_2, x_6 . When using Strategy A, which requires skills x_1, x_2, x_3, x_4 , for the problem, arriving at the correct answer would involve the steps given in Figure 8. Alternatively, Strategy B, which requires skills x_1, x_2 , and x_6 for this problem, involves the steps outlined in Figure 9. Substituting the expression for the multi-strategy DINA model (19), we get:

$$\eta = \max\{\eta_1, \eta_2\} \\ = \max\{\alpha_{(1)'}, \alpha_{(1)'}\}$$

Suppose that the proficiency of each skill involved is $\alpha_1 = 1, \alpha_2 = 1, \alpha_3 = 0, \alpha_4 = 1, \alpha_6 = 1$, then for Strategy A, $\alpha_{(1)'} = 0$ and for Strategy B, $\alpha_{(1)'} = 1$. The latent response is calculated as $\eta = \alpha_{(1)'} = 1$,

Next, we use the SI-GAM to solve the multi-strategy problem. According to the given Q-matrix from Figures 8 and 9, the fuzzy measures can be determined. The fuzzy measures meet the following constraints: $\mu(A) = 1$, for $\forall A \subseteq X$ satisfying $\{x_1, x_2, x_3, x_4\} \subseteq A$ or $\{x_1, x_2, x_6\} \subseteq A$. In addition, for $\forall B \subset X$, $\mu(B) = 0$ if $\{x_1, x_2, x_3, x_4\} \not\subseteq B$ and $\{x_1, x_2, x_6\} \not\subseteq B$. From (20), SI-GAM can be used to calculate the latent response as follows:

$$\eta = (\alpha_{(1)} \wedge \mu(L_1)) \bigvee (\alpha_{(2)} \wedge \mu(L_2)) \cdots \bigvee (\alpha_{(K)} \wedge \mu(L_K)) \\ = \alpha_{(k)}$$

where $\mu(L_k) = 1$ and $\mu(L_{k+1}) = 0$. Since $\alpha_{(1)'} = 0$, and $\alpha_{(1)'} = 1$, it is clear that $\alpha_{(1)'} < \alpha_{(1)'}$. We can directly use the conclusion from Case 1, that is, $\eta = \alpha_{(1)'} = 1$. Clearly, the result demonstrates the effectiveness of SI-GAM in solving multi-strategy problems. By the

Skill required	
$4\frac{4}{12} - 2\frac{7}{12}$ $= 2\frac{4}{12} - \frac{7}{12}$ $= 1\frac{16}{12} - \frac{7}{12}$ $= 1\frac{9}{12}$ $= 1\frac{3}{4}$	X ₃ : separating whole number from fraction
	X ₄ : borrowing one from whole number to fraction
	X ₁ : performing basic fraction subtraction operation
	X ₂ : simplifying/reducing

Fig. 8 Solving a fraction subtraction problem using Strategy A

Skill required	
$4\frac{4}{12} - 2\frac{7}{12}$ $= \frac{52}{12} - \frac{31}{12}$ $= \frac{21}{12}$ $= 1\frac{9}{12} = 1\frac{3}{4}$	X ₆ : converting mixed number to fraction
	X ₁ : performing basic fraction subtraction operation
	X ₂ : simplifying/reducing

Fig. 9 Solving a fraction subtraction problem using Strategy B

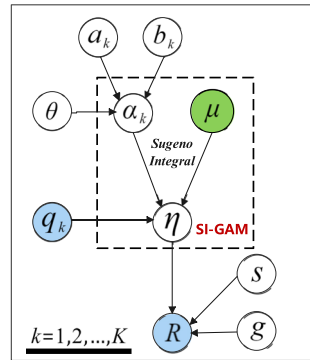
specific example, it can be seen that SI-GAM can solve the multi-strategy problem instead of the multi-strategy DINA model.

6 Experiment

Since multi-skill aggregation is a part of cognitive diagnosis, it needs to be implemented in a specific cognitive diagnosis model. In the framework of FuzzyCDF, the generation from skill proficiency to problem mastery is the process of multi-skill aggregation. Only the skill aggregation process in the model is modified to more clearly observe the performance brought by the aggregation method. To better understand our proposed aggregation method, we put our SI-GAM into FuzzyCDF as an aggregation method to form an improved model (FuzzyCDF-SI-GAM) and represent FuzzyCDF-SI-GAM using a graphic model, as shown in Figure 10.

Here, what we can observe are the score matrix R , and the Q -matrix with K skills (if the problem requires skill k , then $q_k = 1$). A student is related to skill proficiency α_k ,

Fig. 10 The graphic model of FuzzyCDF-SI-GAM



$k = 1, 2, \dots, K$, which depends on latent trait θ and skill parameters $a_k, b_k, k = 1, 2, \dots, K$ (3). The latent response η is determined by required skill proficiency α_k and the skill weight μ (13), and a problem score R is influenced by η and problem parameters s, g . The generation of the scores is simulated as follows:

$$P(R = 1|\eta, s, g) = (1 - s)\eta + g(1 - \eta)$$

In FuzzyCDF[30], the fuzzy intersection and fuzzy union are introduced to calculate the subjective and objective problems, respectively. However, SI-GAM is a more generalized aggregation method that does not distinguish between objective problems and subjective problems. Regardless of objective or subjective problems, SI-GAM adopts a unified formulation to calculate the latent response.

6.1 DataSet description

To verify the effectiveness of the proposed SI-GAM in dealing with skills aggregation and multi-strategy problems, we set up experiments to predict students' learning performance by synthetic and real-world datasets¹. The specific work involves mapping relations from skill mastery to the latent response with different aggregation methods. First, the synthetic datasets ($D1$ and $D2$) are presented in this paper. In dataset $D1$, we design 10 problems that include three interactions between skills, i.e., redundant, enhanced, and independent. The interactive relationship between skills is reflected by the value of fuzzy measures. For example, Problem 3 involves skills x_1, x_3, x_4 , where $\mu(\{x_1\}) = 0.2$, which means that the support strength of skill x_1 for the problem is 0.2. From the Table 7, it is shown that $\mu(\{x_1, x_4\}) = \mu(\{x_1\}) + \mu(\{x_4\})$. As can be seen from the (9), there is an independent relationship between skills x_1 and x_4 . In addition, $\mu(\{x_1, x_3\}) > \mu(\{x_1\}) + \mu(\{x_3\})$. According to the (8), the interaction between skills x_1 and x_3 is positive synergy. In $D1$, which contains complex interactions, we got the scores of 5980 students on 10 problems. While considering the possibility of multiple strategies for specified problem, $D2$ generates by 10 problems as well. In $D2$ dataset, there are only two strategies for each problem. Similarly, we got the scores of 6209 students on 10 problems. Taking Problem 6 as an example, a student who has mastered the coalition of skills x_1, x_2, x_4 or only mastered skill x_3 and x_4 can solve the problem. Hence, the first strategy can be represented by fuzzy measure $\mu(\{x_1, x_2, x_4\}) = 1$, another strategy is expressed as $\mu(\{x_3, x_4\}) = 1$.

¹ <https://github.com/kathy-sj/SI-GAM>

Table 7 A sample of fuzzy measures

Fuzzy measure sets	Corre- sponding value	Fuzzy measure sets	Corre- sponding value
\emptyset	0	$\mu(\{x_2, x_3\})$	0.2
$\mu(\{x_1\})$	0.2	$\mu(\{x_2, x_4\})$	0.3
$\mu(\{x_2\})$	0	$\mu(\{x_3, x_4\})$	0.7
$\mu(\{x_3\})$	0.2	$\mu(\{x_2, x_3, x_4\})$	0.7
$\mu(\{x_4\})$	0.3	$\mu(\{x_1, x_2, x_3\})$	0.5
$\mu(\{x_1, x_2\})$	0.2	$\mu(\{x_1, x_2, x_4\})$	0.5
$\mu(\{x_1, x_3\})$	0.5	$\mu(\{x_1, x_3, x_4\})$	1
$\mu(\{x_1, x_4\})$	0.5	$\mu(\{x_1, x_2, x_3, x_4\})$	1

Table 8 describes the details of the datasets. Data sets *D1* and *D2* respectively contain the response data on 10 problems, of which Problem 1 – 6 are objective problems and the rest are subjective problems. Then, the fuzzy measures with respect to specified skills are constructed, which follow the properties of the fuzzy measures described in Section 3. The real-world dataset (Math1) is collected from two final mathematical exams from high school students. There are twenty problems in Math1, including 16 objective and 4 subjective items. Each problem in the Math1 dataset involves more than two skills by the **Q**-matrix. The FuzzyCDF-SI-GAM model obtains students' mastery on each skill as the input, and the output is the response score of students. From the datasets, input-output label data can be obtained to form a training set, which is used to train the parameters in the FuzzyCDF-SI-GAM model.

6.2 Experimental setup

This experiment introduces an effective training algorithm using MCMC method to estimate model parameters [15]. The parameters involved are randomly generated by a certain distribution, so as to calculate and generate students' probability of correct answer. Specifically, following the settings adopted in the HO-DINA model [45], we assume the prior distributions of the parameters based on the as:

$$\begin{aligned}
 \theta &\sim \mathcal{N}(\mu_\theta, \sigma_\theta^2), \quad a \sim \ln\mathcal{N}(\mu_a, \sigma_a^2), \quad b \sim \mathcal{N}(\mu_b, \sigma_b^2), \\
 s &\sim \text{Beta}(v_s, \omega_s, \min_s, \max_s), \\
 g &\sim \text{Beta}(v_g, \omega_g, \min_g, \max_g), \\
 1/\sigma^2 &\sim \Gamma(x_\sigma, y_\sigma).
 \end{aligned} \tag{22}$$

where $\text{Beta}(v, w, \min, \max)$ is a four-parameter *Beta* distribution that has two shape parameters v and w and is supported on the range $[\min, \max]$. In the training algorithm for FuzzyCDF-SI-GAM, the joint posterior distribution of θ, a, b, s, g , and σ^2 given the score matrix R is:

$$\begin{aligned}
 &P(\theta, a, b, s, g, \sigma^2 | R) \\
 &\propto L(s, g, \sigma^2, \theta, a, b) P(\theta) P(a) P(b) P(s) P(g) P(\sigma^2)
 \end{aligned} \tag{23}$$

where L is the joint likelihood function of FuzzyCDF-SI-GAM:

Table 8 Dataset details

Dataset	Students	Skills	Problem	
			Obj.	Sub.
D1 (Complex Interaction)	5980	4	6	4
D2 (Multiple Strategies)	6209	4	6	4
Math1 (Real-world Dataset)	4209	11	16	4

$$L(s, g, \sigma^2, \theta, a, b) = L_o(s, g, \theta, a, b)L_s(s, g, \sigma^2, \theta, a, b) \quad (24)$$

where L_o and L_s denote the joint likelihood functions of objective and subjective problems, respectively, and they can be defined as follows:

$$L_o(s, g, \theta, a, b) = \prod (X)^R (1 - X)^{1-R} \quad (25)$$

$$L_s(s, g, \sigma^2, \theta, a, b) = \prod \mathcal{N}(R|X, \sigma^2) \quad (26)$$

where $X = (1 - s)\eta + g(1 - \eta)$. Note that η , i.e., the latent response can be calculated given the \mathbf{Q} -matrix by using (13). Then, the full conditional distributions of the parameters (a, b) given the observed score matrix R and the rest of the parameters are as follows:

$$P(a, b|R, \theta, s, g, \sigma^2) \propto L(s, g, \sigma^2, \theta, a, b)P(a)P(b) \quad (27)$$

Similarly, the the full conditional distributions of the parameters, e.g., θ, s, g, σ^2 are given by this way.

In our FuzzyCDF-SI-GAM, we first randomize the parameters ($a, b, \theta, s, g, \sigma^2$) as the initial values. In this framework of FuzzyCDF-SI-GAM, fuzzy measures are known as prior knowledge or learned from data by other model. For each iteration, each parameter is uniformly and randomly sampled from a predefined interval. Then, given the observable R and the \mathbf{Q} -matrix, we compute the full conditional probability of skill discrimination a , skill difficulty b , student' trait θ , problem slip factor s , and guess factor g and the variance of normalized scores of subjective problems σ by using (23), (24), (26), (25) and (27). Next, the acceptance probability of samples can also be calculated using a Metropolis-Hastings-(M-H) based MCMC algorithm. In this way, we could estimate the parameters after T iterations of sampling.

6.3 Experimental results and analysis

In this section, we conduct experiments on the task of predicting the scores of students over each subjective or objective problem. To observe how the methods behave at different sparsity levels, we construct different sizes of training sets, with 20%, 50%, 80% of the response data of each student and the rest for testing, respectively. We use the root mean square error (RMSE) and mean absolute error (MAE) as the evaluation metrics. Then, we consider baseline approaches as follows:

- IRT: a cognitive diagnosis method modeling students' latent traits and the parameters of problems like difficulty and discrimination [5, 37].

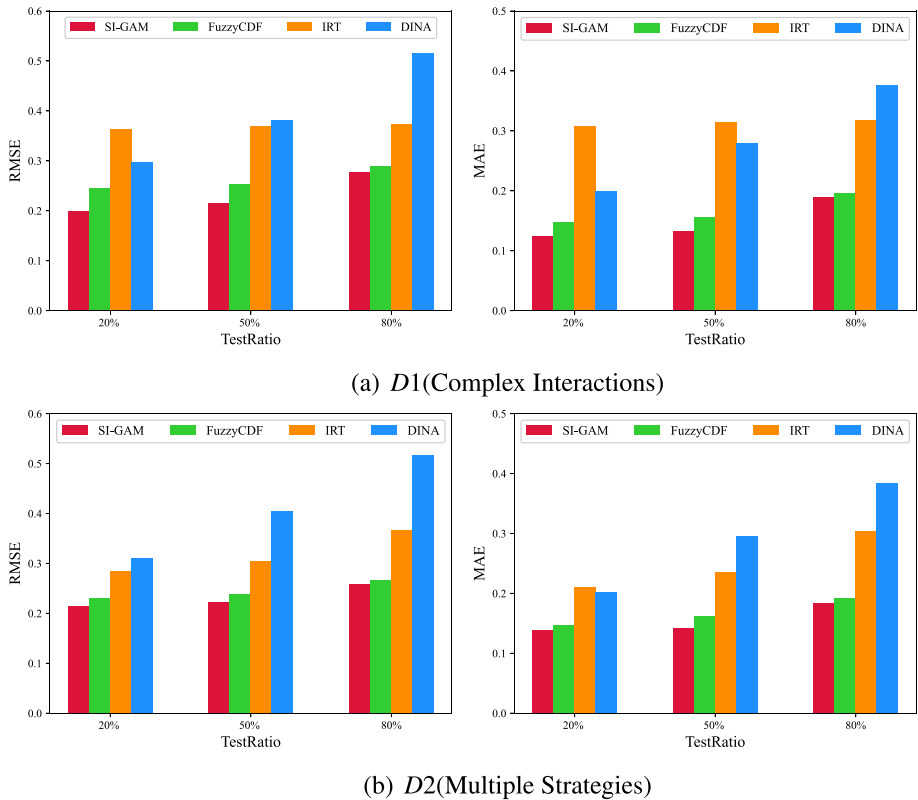


Fig. 11 The performance of SI-GAM model and baselines

- DINA: a cognitive diagnosis method modeling students' skill proficiency and the slip and guess factors of problems with a **Q**-matrix. [27].
- FuzzyCDF: a fuzzy cognitive diagnosis framework for students' cognitive modeling with both objective and subjective problems [30].

In order to ensure fairness of comparison, we record the best performance of each algorithm by tuning their parameters. Figure 11 shows the performance of our SI-GAM model and other baseline approaches on two datasets when dealing with complex interactions and multi-strategy problems. The results prove the effectiveness of our SI-GAM. In particular, compared with the FuzzyCDF model, the performance is improved only by using SI-GAM under the same parameter training model. More importantly, as the sparsity of training data increases (the training data ratio decreases from 80% to 20%), our SI-GAM still has advantages over other baselines. In *D1* and *D2*, the performance of SI-GAM surpasses each baseline method on evaluation metrics (RMSE and MAE). In addition, it should be noted that SI-GAM is more explainable, it can more accurately understand the cognitive state of students, and better identify the support strength of multiple skills in problem-solving.

To enhance the evaluation of our SI-GAM, we also conduct the same experiment on the Math1(Real-World Dataset) dataset [3]. Because the dataset is not labeled with fuzzy measures, we learn the fuzzy measures by a neural network. In this experiment, the fuzzy

Table 9 Comparisons with baselines in Math1 (Real-World Dataset)

Dataset	Test Ratio	Model	MAE	AUC
Math1	20%	SI-GAM	0.285	0.687
		FuzzyCDF	0.322	0.678
		IRT	0.330	0.648
		DINA	0.375	0.633
	80%	SI-GAM	0.311	0.658
		FuzzyCDF	0.337	0.649
		IRT	0.361	0.623
		DINA	0.416	0.501

measures are known by the Fuzzy Integral Neural Network. According to the definition 1, the fuzzy measures can be expressed as:

$$\mu(A) = \bigvee_{B \subset A} \mu(B) + \Delta\mu(A)$$

Here, the monotonicity of the fuzzy measures can be ensured. According to the monotonicity constraint of fuzzy measures, the Fuzzy Integral Neural Network designs an optimization method based on random gradient descent to realize the training of fuzzy measures [25]. In the experiment, AUC (Area Under the Curve) and MAE were used to compare the prediction of cognitive diagnostic models. Firstly, 80% of the data were randomly selected for training and 20% for testing. In order to further observe the effect under the different sparsity, 20% of data were randomly selected for training, and the rest was test data. Table 6 shows that compared with the FuzzyCDF model, the MAE error of the SI-GAM decreases after introducing the fuzzy measures and has obvious advantages compared with the IRT and DINA model. On the other hand, we calculate the AUC metric for objective questions in the Math1 data set. From Table 9, the accuracy of SI-GAM is higher than all other baselines. In addition, with the increase of training data sparsity, the advantages of SI-GAM still exist. As shown from the results, our multi-skill aggregation method (SI-GAM) can achieve a better performance in predicting students' performance .

Furthermore, we set the sensitivity experiment of the parameters. We observe that fuzzy measure is an important parameter in SI-GAM, reflecting the weight of skills. However, it is unknown how the effectiveness of our SI-GAM varies with the precision of fuzzy measures. Therefore, we tune up the known fuzzy measures within a certain range and observe the corresponding latent response. Take Problem 1 in dataset *D1* as an example, which involves three skills, x_1 , x_2 and x_3 , where $\mu(\{x_1\}) = 0.4$, $\mu(\{x_2\}) = 0.3$, $\mu(\{x_3\}) = 0.1$. We choose skill x_1 , x_2 with large weight for this experiment. Firstly, the value of the fuzzy measure $\mu(\{x_1\})$ is set to fluctuate within the range of [0.38,0.42], and other fuzzy measures of supersets containing x_1 are randomly generated under the monotonicity assumption. Then, we calculate the average absolute error between the latent result of adjusted by $\mu(\{x_1\})$ and the real result when $\mu(\{x_1\}) = 0.4$. It can be seen from Figure 12 that the average absolute error is less than 2%, when $\mu(\{x_1\}) = 0.4$ changes within the range of [0.38, 0.42]. Similarly, we set the value of $\mu(\{x_2\})$ within the range of [0.28, 0.32], and the average absolute error is less than 2.1%. In dataset *D2*, We randomly selected a problem involving four skills, where $\mu(\{x_2\}) = 0.7$ and $\mu(\{x_3\}) = 0.6$. We set up fluctuation values of up and down 5% for the fuzzy measures of skills x_2 and x_3 , respectively. As can be seen

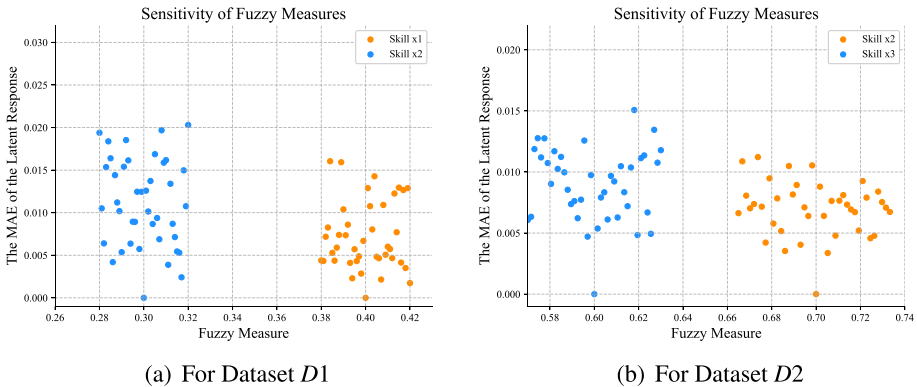


Fig. 12 Sensitivity of Fuzzy measure

	x_1	x_2	x_3	x_4	<i>Student Response</i>
<i>Q-matrix</i> (used by FuzzyCDF)	0	0	1	1	
<i>Proficiency</i>	0.48	0.69	0.41	0.65	
					✓

Fig. 13 The Q -matrix of Problem 6 and corresponding skills' mastery and response of a student in D2 for multiple strategies problems

from the Figure 12, the average absolute error is still less than 2%. With the decline of the accuracy of fuzzy measure(within 6%), the performance degradation begins a gradual drop, but the error remains at a low value, thus the SI-GAM has good stability in cognitive diagnosis.

6.4 Case study

Here we present an example of a student's diagnostic result of SI-GAM on D2 for multi-strategy problems. The Figure 13 shows the proficiency of the student on four skills. According to the Q -matrix, the strategy used by the FuzzyCDF includes skills 3 and 4. Problem 6 is an objective problem, involving skills x_3 and x_4 . From (4), we can calculate the latent response based on FuzzyCDF $\eta = \min(\alpha_{x_3}, \alpha_{x_4}) = 0.41$. Further, considering the slip and guess, the prediction probability is 0.42. Then, we can discretize the result for prediction on objective problems by the predefined threshold (usually set to be 0.5). Thus, the prediction equals 0. Using SI-GAM, the latent response $\eta = 0.65$. With the slipping and guess factors, the prediction probability is 0.66, the result equals 1. The result of SI-GAM model is consistent with the observed data. From the fuzzy measures, we can better explain the diagnostic result. Since there are two strategies for solving the Problem 6, that is $\mu(\{x_1, x_2, x_4\}) = 1$ and $\mu(\{x_3, x_4\}) = 1$. Therefore, we can explain why the student answer correctly. Although he has a low level of skill x_3 , he has a better mastery of skills

x_1 , x_2 and x_4 . Therefore, he has a higher probability of success when he uses another strategies (using skills x_1 , x_2 and x_4).

From the above experiments, we can observe that the aggregation method (SI-GAM) proposed in this paper provides more accurate prediction results for complex skill interactions and multi-strategy problems. From the perspective of the aggregation method itself, the fuzzy measures construct a more explanatory way of modeling the interactions between skills to represent the weight. In summary, SI-GAM captures the characteristics of students more precisely, and it is more suitable for the real-world scenarios where complex interactions and multiple strategies exist.

7 Conclusion

This paper proposes a generalized multi-skill aggregation method based on the Sugeno integral (SI-GAM). The proposed method incorporates non-additive weight, that is, fuzzy measure. Through the representation of fuzzy measures, the interactions among multiple skills are considered. In addition, the max-min operator of Sugeno integral can also describe different aggregate functions, including conjunctive and compensatory approaches. At the same time, it also proves that our SI-GAM is suitable for multi-strategy issues, which provides a new idea for solving problems of multiple strategies. Our future work is to consider the learning of fuzzy measures efficiently in the cognitive diagnosis model. We can also view the following two aspects for future research:

1. The application of the fuzzy measures can be further expanded. The global importance of skill is not only determined by its own fuzzy measure, but also needs to consider the fuzzy measure value that includes all subsets of the skill. The shapely value and interaction index[19] can be used to measure each skill's importance and mutual influence. The calculation of these two indices depends on the determination of the fuzzy measure value.
2. How to learn fuzzy measures is an important issue. In this paper, the prerequisite of SI-GAM is that the fuzzy measure is known as prior knowledge or learned from data by a neural network separately. Constructing an effective cognitive diagnosis model that can learn fuzzy measures simultaneously is a focus for future work. At the same time, we also need to consider the dimensional space of the fuzzy measure. By definition, if n skills are involved, 2^n fuzzy measures are required. If there are multiple skills, such as $n = 10$, the number of possible fuzzy measures is 1024. As the number of skills increases, the number of fuzzy measures will increase exponentially, bringing dimensional disaster. How to balance dimensional space and accuracy is also a question worth considering. Next, we will establish a unified expression framework for cognitive diagnosis based on neural networks and realize fuzzy measures learning. However, they are often regarded as a black box, i.e., their predictions cannot be explained [7], which makes our future work face more challenges.

Appendix

The derivation of the latent response under three assumptions:

- Case 1. If $\alpha_{(1)'} < \alpha_{(1)'}$, then
 - If $\alpha_{(k)} = \alpha_{(1)'}$, then $X_1 \subseteq L_k$ and $X_1 \not\subseteq L_{k+1}$, thus $\mu(L_k) = 1$. Since $\alpha_{(1)'} < \alpha_{(1)'}$, $X_2 \subseteq L_k$ and $X_2 \subseteq L_{k+1}$. It is easy to show that the result is contrary to the (20), it does not meet the constraints.
 - Else if $\alpha_{(k)} = \alpha_{(1)'}$, then $X_2 \subseteq L_k$, and $X_2 \not\subseteq L_{k+1}$. As $\alpha_{(1)'} < \alpha_{(1)'}$, we can infer that $X_1 \not\subseteq L_k$ and $X_1 \not\subseteq L_{k+1}$. Hence, we use $\alpha_{(k)} = \alpha_{(1)'}$ to substitute into the calculation formula, which satisfies the (20), that is $\mu(L_k) = 1$ and $\mu(L_{k+1}) = 0$.

Thus, the result of the latent response is given by $\eta = \alpha_{(k)} = \alpha_{(1)'}$, when $\alpha_{(1)'} < \alpha_{(1)'}$.
- Case 2. If $\alpha_{(1)'} = \alpha_{(1)'}$, in other words, $\alpha_{(k)} = \alpha_{(1)'} = \alpha_{(1)'}$, then $X_1 \subseteq L_k$, $X_2 \subseteq L_k$, $X_1 \not\subseteq L_{k+1}$ and $X_2 \not\subseteq L_{k+1}$. This inference meets the constraints of (20), thus $\mu(L_k) = 1$ and $\mu(L_{k+1}) = 0$. Moreover, the latent response equals to $\eta = \alpha_{(k)} = \alpha_{(1)'} = \alpha_{(1)'}$, when $\alpha_{(1)'} = \alpha_{(1)'}$.
- Case 3. If $\alpha_{(1)'} > \alpha_{(1)'}$, then
 - If $\alpha_{(k)} = \alpha_{(1)'}$, then $X_1 \subseteq L_k$, $X_1 \not\subseteq L_{k+1}$. For $\alpha_{(1)'} > \alpha_{(1)'}$, it is clear that $X_2 \not\subseteq L_k$ and $X_2 \not\subseteq L_{k+1}$. According to (20), we can conclude that $\mu(L_k) = 1$. Moreover, the latent response is $\eta = \alpha_{(k)} = \alpha_{(1)'}$.
 - Else if $\alpha_{(k)} = \alpha_{(1)'}$, then $X_2 \subseteq L_k$, obviously, $X_2 \not\subseteq L_{k+1}$. Since $\alpha_{(1)'} > \alpha_{(1)'}$, we can infer that $X_1 \not\subseteq L_k$ and $X_1 \subseteq L_{k+1}$. Hence, it is impossible that make $\mu(L_k) = 1$ and $\mu(L_{k+1}) = 0$. Hence, the latent response is computed by $\eta = \alpha_{(k)} = \alpha_{(1)'}$ when $\alpha_{(1)'} > \alpha_{(1)'}$.

Acknowledgements The research was partially supported by National Key R& D Program of China (No: 2018YFB1403400), the National Education Scientific Planning Projects (Grants No. BCA190081).

Funding The project is supported by National Key R& D Program of China (No: 2018YFB1403400) and National Education scientific Planning Projects (Grants No. BCA190081).

Declarations

Conflicts of interest The authors have no financial or proprietary interests in any material discussed in this article.

References

1. Anderson, D.T., Keller, J.M., Havens, T.C.: Learning fuzzy-valued fuzzy measures for the fuzzy-valued sugeno fuzzy integral. In: International Conference on Information Processing and Management of UncertaintyIPMU 2010 (2010)
2. Beliaikov, G., Divakov, D.: On representation of fuzzy measures for learning choquet and sugeno integrals. *Knowl Based Syst* **189**(10), 105134 (2019)
3. bigdata-ustc: EduData. <https://github.com/bigdata-ustc/EduData>. Accessed 2022-04-22
4. Bin, Q., Hongxia, Z., Yu, W., Jixing, L.: Questions recommendation based on collaborative filtering and cognitive diagnosis. *Computer Science* **46**(11), 235–240 (2019)
5. Birnbaum, A.: Some latent trait models and their use in inferring an examinee's ability (1968)
6. Carlson, J.E., von Davier, M.: Item response theory. *Ets Research Report* **2013**(2), i–69 (2013)
7. Chakraborty, M., Biswas, S.K., Purkayastha, B.: Rule extraction from neural network trained using deep belief network and back propagation. *Knowl Inf Syst* **62**(17) (2020)

8. Cheng, Y.: Computerized adaptive testing—new developments and applications. Dissertations & Theses - Gradworks (2008)
9. De La Torre, J.: Dina model and parameter estimation: A didactic. *J Educ Behav Stat* **34**(1), 115–130 (2009)
10. De La Torre, J., Douglas, J.A.: Model evaluation and multiple strategies in cognitive diagnosis: an analysis of fraction subtraction data. *Psychometrika* **73**(4), 595–624 (2008)
11. De La Torre, J., Lee, Y.S.: A note on the invariance of the dina model parameters. *J Educ Meas* **47**(1), 115–127 (2010)
12. De La Torre, J., Minchen, N.: Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicol Educ* **20**(2), 89–97 (2014)
13. Delatorre, J.: The generalized dina model framework. *Psychometrika* **76**(2), 179–199 (2011). <https://doi.org/10.1007/s11336-011-9207-7>
14. DONGbo, T.: A polytomous cognitive diagnosis model: P-dina model. *Acta Psychologica Sinica* **42**(10), 1011–1020 (2010)
15. Geyer, C.J.: Markov chain monte carlo maximum likelihood. American Cancer Society (2005)
16. Gierl, M.J.: Making diagnostic inferences about cognitive attributes using the rule-space model and attribute hierarchy method. *J Educ Meas* **44**(4), 325–340 (2007)
17. Grabisch, M., Murofushi, T., Sugeno, M.: Fuzzy measures and integrals—theory and applications. Medical Tribune (2000)
18. Grabisch, M.: Set functions, games and capacities in decision making. Université Paris1 Panthéon Sorbonne (2016)
19. Grabisch, M.: The representation of importance and interaction of features by fuzzy measures. *Pattern Recognit Lett* **17**(6), 567–575 (1996)
20. Grabisch, M., Labreuche, C.: A decade of application of the choquet and sugeno integrals in multi-criteria decision aid. *Ann Oper Res* **175**(1), 247–286 (2010)
21. Guo, L., Bao, Y., Wang, Z., Bian, Y.: Cognitive diagnostic assessment with different weight for attribute: based on the dina model. *Psychol Rep* **114**(3), 802–822 (2014)
22. Haertel, E.H.: Using restricted latent class models to map the skill structure of achievement items. *J Educ Meas* **26**(4), 301–321 (1989)
23. Hartz, S.M.: A bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality. Ph.D. thesis, ProQuest Information & Learning (2002)
24. Hastings, W.K.: Monte carlo sampling methods using markov chains and their applications. *Biometrika* **1**, 97–109 (1970)
25. Islam, M.A., Anderson, D.T., Pinar, A.J., Havens, T.C., Scott, G., Keller, J.M.: Enabling explainable fusion in deep learning with fuzzy integral neural networks. *IEEE Transactions on Fuzzy Systems* (2019)
26. Item response theory - application to psychological measurement. *British Journal of Surgery* (1983)
27. Junker, B.W., Sijtsma, K.: Cognitive assessment models with few assumptions, and connections with non-parametric item response theory. *Appl Psychol Meas* **25**(3), 258–272 (2001)
28. Kaya, Y., Leite, W.L.: Assessing change in latent skills across time with longitudinal cognitive diagnosis modeling: An evaluation of model performance. *Educ Psychol Meas* **77**(3) (2016)
29. Leighton, J.P., Gierl, M.J., Hunka, S.M.: The attribute hierarchy method for cognitive assessment: a variation on tatsuoaka's rule-space approach. *J Educ Meas* **41**(3), 205–237 (2004)
30. Liu, Q., Wu, R., Chen, E., Xu, G., Su, Y., Chen, Z., Hu, G.: Fuzzy cognitive diagnosis for modelling examinee performance. *ACM Trans Intell Syst Technol* **9**(4), 1–26 (2018)
31. Liu, Q., Huang, Z., Yin, Y., Chen, E., Xiong, H., Su, Y., Hu, G.: Ekt: Exercise-aware knowledge tracing for student performance prediction. *IEEE Transactions on Knowledge and Data Engineering* **33**(1), 100–115 (2021)
32. Ma, W., Guo, W.: Cognitive diagnosis models for multiple strategies. *Br J Math Stat Psychol* (2019)
33. Ma, W., Guo, W.: Cognitive diagnosis models for multiple strategies. *Br J Math Stat Psychol* **72**(2), 370–392 (2019)
34. Maris, E.: Estimating multiple classification latent class models. *Psychometrika* **64**(2), 187–212 (1999)
35. Murillo, J., Guillaume, S., Bulacio, P.: k-maxitive fuzzy measures: A scalable approach to model interactions. *Fuzzy Sets Syst* **324**, 33–48 (2017)
36. Pardos, Z.A., Heffernan, N.T., Ruiz, C., Beck, J.E.: The composition effect: conjunctive or compensatory? an analysis of multi-skill math questions in its. In: International Conference on Educational Data Mining (2008)
37. Rasch, G.: Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests (1960)
38. Sugeno, M.: Theory of fuzzy integrals and its applications. Ph.D. thesis, Institute of Technology (1974)
39. Sun, J., Xin, T., Zhang, S., de la Torre, J.: A polytomous extension of the generalized distance discriminating method. *Appl Psychol Meas* **37**(7), 503–521 (2013)

40. Tatsuoaka, K.K.: Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In: Presented at the Office of Naval Research Contractors Conference in Iowa City, IA, May 1993. Lawrence Erlbaum Associates, Inc (1995)
41. Tatsuoaka, K.K.: Boolean algebra applied to determination of universal set of knowledge states. ETS Research Report Series **1991**(2), i–36 (1991)
42. Tatsuoaka, K.K.: Rule space: An approach for dealing with misconceptions based on item response theory. *J Educ Meas* **20**(4), 345–354 (2005)
43. Templin, J., Henson, R.: Measurement of psychological disorders using cognitive diagnosis models. *Psychol Methods* **11**(3), 287–305 (2006)
44. Torra, V., Narukawa, Y.: The interpretation of fuzzy integrals and their application to fuzzy systems. *International Journal of Approximate Reasoning* **41**(1), 43–58 (2006)
45. Torre, J.D.L., Douglas, J.A.: Higher-order latent trait models for cognitive diagnosis. *Psychometrika* **69**(3), 333–353 (2004)
46. Tu, D.B., Cai, Y., Ding, S.L.: A new multiple-strategies cognitive diagnosis model: the mscd method. *Acta Psychologica Sinica* **44**(011), 1547–1553 (2012)
47. Wang, D., Cai, Y., Tu, D.: Q-matrix estimation methods for cognitive diagnosis models: based on partial known q-matrix. *Multivariate Behav Res* pp. 1–13 (2020)
48. Wei, G.W., Wang, H.J., Lin, R.: Application of correlation coefficient to interval-valued intuitionistic fuzzy multiple attribute decision-making with incomplete weight information. *Knowl Inf Syst* (2011). <https://doi.org/10.1007/s10115-009-0276-1>
49. Wu, R., Liu, Q., Liu, Y., Chen, E., Su, Y., Chen, Z., Hu, G.: Cognitive modelling for predicting examinee performance. In: Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI’15, pp. 1017–1024. AAAI Press (2015)
50. Yue, Kou: Derong, Shen, Hongbin, Xu, Menger, Lin, Ge, Yu: Two-level interactive identification and derivation of topic clusters in complex networks. *World Wide Web* **18**(4), 1093–1122 (2015)
51. Zhang, M., Zhu, J., Wang, Z., Chen, Y.: Providing personalized learning guidance in moocs by multi-source data analysis. *World Wide Web* **22**, 1189–1219 (2019)

Authors and Affiliations

Suojuan Zhang¹ · Song Huang¹ · Xiaohan Yu¹ · Enhong Chen² · Fei Wang² · Zhenya Huang²

Suojuan Zhang
suojuanzhang@aeu.edu.cn

Xiaohan Yu
yuxh@aeu.edu.cn

Enhong Chen
cheneh@ustc.edu.cn

Fei Wang
wf314159@mail.ustc.edu.cn

Zhenya Huang
huangzhy@ustc.edu.cn

¹ College of Command & Control Engineering, Army Engineering University of PLA, 210000 Nanjing, Jiangsu, China

² Anhui Province Key Laboratory of Big Data Analysis and Application, School of Computer Science and Technology, University of Science and Technology of China, 230000 Hefei, Anhui, China