mathematics

MDPI

Article

A Student Performance Prediction Model Based on Hierarchical Belief Rule Base with Interpretability

Minjie Liang, Guohui Zhou *, Wei He, Haobing Chen and Jidong Qian

School of Computer Science and Information Engineering, Harbin Normal University, Harbin 150025, China; liangminjie@stu.hrbnu.edu.cn (M.L.); hewei@hrbnu.edu.cn (W.H.); isla_000@163.com (H.C.); qianjidong@stu.hrbnu.edu.cn (J.Q.)

* Correspondence: zhouguohui@hrbnu.edu.cn

Abstract: Predicting student performance in the future is a crucial behavior prediction problem in education. By predicting student performance, educational experts can provide individualized instruction, optimize the allocation of resources, and develop educational strategies. If the prediction results are unreliable, it is difficult to earn the trust of educational experts. Therefore, prediction methods need to satisfy the requirement of interpretability. For this reason, the prediction model is constructed in this paper using belief rule base (BRB). BRB not only combines expert knowledge, but also has good interpretability. There are two problems in applying BRB to student performance prediction: first, in the modeling process, the system is too complex due to the large number of indicators involved. Secondly, the interpretability of the model can be compromised during the optimization process. To overcome these challenges, this paper introduces a hierarchical belief rule base with interpretability (HBRB-I) for student performance prediction. First, it analyzes how the HBRB-I model achieves interpretability. Then, an attribute grouping method is proposed to construct a hierarchical structure by reasonably organizing the indicators, so as to effectively reduce the complexity of the model. Finally, an objective function considering interpretability is designed and the projected covariance matrix adaptive evolution strategy (P-CMA-ES) optimization algorithm is improved. The aim is to ensure that the model remains interpretable after optimization. By conducting experiments on the student performance dataset, it is demonstrated that the proposed model performs well in terms of both accuracy and interpretability.

Keywords: education; student performance; behavior prediction; belief rule base; hierarchical structure; interpretability

MSC: 68T37

check for updates

Citation: Liang, M.; Zhou, G.; He, W.; Chen, H.; Qian, J. A Student Performance Prediction Model Based on Hierarchical Belief Rule Base with Interpretability. *Mathematics* **2024**, *12*, 2296. https://doi.org/10.3390/math12142296

Academic Editor: Andrea Scozzari

Received: 25 June 2024
Revised: 14 July 2024
Accepted: 19 July 2024
Published: 22 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Behavior prediction refers to the prediction of possible future behaviour or performance by analyzing previous and current data about an individual or group [1,2]. The study predicts how students will perform in their future study through their prior academic performance and their current state of learning stage. Thus, in education, student performance prediction is considered as a typical behavior prediction problem [3,4]. By predicting the achievement of index at the next stage of the learning process, education experts can detect the risk of students' academic failure in advance. Then, based on techniques such as sensitivity analysis, they can identify the main factors affecting students' performance and thus improve educational strategies to elevate education quality. Thus predicting student performance plays an early warning, monitoring and guiding role in the teaching process [5]. Therefore, it is important to know how to predict student performance accurately and reliably.

A number of student performance prediction methods have been proposed. These methods can be broadly categorized into data-driven methods, probabilistic statistical

methods, and time series analysis methods. Data-driven models involve adjusting the parameters in the model with a large amount of data so that they can learn and recognize patterns in the data. Song and Luo et al. proposed a synthetically forgetting behavior knowledge tracing (SFBKT) model considering the influence of forgetting in the learning process. The model first extracted the forgetting information in the learning record, and then this feature information was inputted into a continuous-time long short-term memory network (CTLSTM) and finally predicted whether the students could answer the questions correctly or not [6]. Xiao et al. analyzed the factors affecting performance in online education and then constructed a prediction model based on BP neural network for monitoring students' learning status and abnormal behavior [7]. Imhof et al. combined subjective and objective factors and extracted indicators from questionnaires and students' learning logs, respectively, and then constructed a predictive model based on random forest (RF). This study was used to predict assignment submission delays and assignment completion effects [8]. Cui et al. proposed a new Tri-Branch CNN model that records student behavior in an end-to-end manner in terms of time, location, and frequency. The model took the behavioral information recorded in the student's campus card as input, and the output was the prediction of academic performance [9]. The effectiveness of such methods depends on the amount of data. However, more importantly, the results cannot be trusted due to their black-box nature, which is difficult to accept in educational scenarios.

Probability statistical methods infer the state of the system by probability theory and statistical methods based on current observable data. Mubarak et al. proposed a method for early student dropout prediction by recognizing the problems of poor course continuity and high dropout rates in online learning. The method extracted the most salient features of student activities from logs and then used an input–output hidden Markov model (IOHMM) and sequential logistic regression, respectively, to obtain the risk of students dropping out in each learning cycle [10]. Chen et al. predicted academic performance based on students' reading behavior in the context of a university course using e-books as instructional support. They constructed classifiers based on logistic regression and Gaussian naïve Bayes to accurately identify students at academic risk [11]. Hao et al. constructed a Bayesian network through hill-climbing and maximum likelihood estimation, which was used to predict students' performance in MOOC. A genetic algorithm was then used to provide opinions for at-risk students based on their current learning and past experiences [12]. This type of method usually requires certain assumptions to be set by the researcher, and it is difficult to identify causal relationships. How to reasonably utilize information such as experts' experience to make predictions is a problem that it is difficult to solve.

Time series analysis is a method of predicting future trends based on data containing temporal information. Li et al. proposed an end-to-end deep neural network model. The model captured the temporal features of each type of behavioral sequence using LSTM, and the correlation features between behaviors using a two-dimensional convolutional neural network (2D-CNN). These two features were used as inputs to predict academic performance [13]. Shou et al. analyzed the shortcomings of traditional machine learning algorithms in extracting temporal features and proposed to construct multiple parallel fully convolutional networks (FCNs) to extract the multi-scale temporal features of student behaviors. The noisy information was then removed using a variational information bottleneck (VIB), and the likelihood of student dropout was predicted based on the remaining main features [14]. Mumtaz et al. used the quality of student interactions with a learning management system as a key indicator for prediction. They used exponential smoothing (ETS) and autoregressive integrated moving average (ARIMA) methods as predictive models to predict the quality of students' weekly interactions [15]. These types of methods focus primarily on the information present in the data itself, which leads to two problems. One is the heavy reliance on the quality of the data, and the other is the difficulty of incorporating external factors into the model prediction process [16].

Summarizing the existing studies, it has been observed that student performance prediction methods still suffer from the problems of difficult-to-understand black-box models, the difficulty of using expert decision making effectively, and high data quality requirements. Therefore, the study of student performance prediction in this paper focuses on the following problems. The first is the interpretability of the model. The prediction results of the model are used to regulate and predict students' performance, allowing educators to adapt their strategies promptly based on these results. A lack of interpretability means that educators are completely unable to judge whether the prediction results are reliable [17] and whether a new teaching policy is really reasonable, which may cause serious teaching accidents. Therefore, the transparency and interpretability of the model are the primary considerations. Secondly, education follows the principle of human-centeredness. The accumulated experience of education experts should be integrated into the model mechanism to provide decision making for the prediction process. Hence, the model must facilitate the integration of expert knowledge. Thirdly, there are various forms of data collection. The questionnaire data may be influenced by subjectivity, and data from logs and other sources can be distorted by outliers. These uncertainties will interfere with the prediction results [18]. Therefore, prediction models are required to have the ability to handle uncertain information.

To solve these problems, this research develops a prediction model based on a belief rule base (BRB). BRB consists of rules in the form of IF-THEN, a logical expression based on conditional statements, offering clarity and intuitive knowledge representation [19]. BRB represents qualitative or quantitative data via belief distribution, effectively managing diverse types of uncertain information. Moreover, BRB incorporates the evidential reasoning (ER) algorithm as its inference machine. ER, based on D-S evidence theory, offers a transparent inference process [20], providing probabilistic and holistic reasoning for the ultimate outcome. Consequently, BRB has good interpretability, and experts are able to understand how the prediction results are derived.

However, in student performance prediction, the method based on BRB still needs to deal with the following three challenges. Firstly, the initial BRB model is constructed based on expert knowledge, incorporating subjective perspectives potentially deviating from objective reality. Therefore, an optimization algorithm needs to be used to adjust the parameters of the BRB, but the interpretability of the model will be destroyed. To maintain interpretability, Cao et al. summarized the general interpretability criterion for BRB, which provides a theoretical basis for the construction and optimization of interpretable BRB [21]. Secondly, decision making in rule-based systems hinges on explicit rule sets. Rules constructed by relying on expert knowledge can be understood, but with average accuracy. Rules generated by learning from data have high accuracy but low interpretability [22]. Therefore, striking a balance between accuracy and interpretability is crucial. Finally, predicting student performance relies on numerous indicators, resulting in a combination explosion within the BRB. Complex systems can greatly affect the modeling process, thus the size of the rule base needs to be simplified to reduce the complexity. To solve the above problems, hierarchical BRB with interpretability (HBRB-I) is proposed. The model uses a hierarchical structure to aggregate data layer by layer from the bottom up to effectively reduce the number of rules. Furthermore, the attribute grouping method of sub-BRBs is designed for rational construction of HBRB. During model optimization, the improved projection covariance matrix adaptive evolution strategy (P-CMA-ES) is used as the optimization algorithm, and a new objective function is proposed, which integrates the accuracy and interpretability of the model.

The main contributions of this paper are as follows: (1) A new student performance prediction model based on HBRB-I is proposed. (2) An attribute grouping method is designed to construct a hierarchical model with a reasonable structure. (3) An objective function and optimization algorithm considering interpretable constraints are designed to ensure the interpretability of the optimized model.

This paper is organized as follows. In Section 2, the problems in student performance prediction are summarized, and then the HBRB-I-based prediction model is proposed. In Section 3, the interpretability criteria and optimization strategies for the HBRB-I model are proposed. In Section 4, the inference and optimization process of the HBRB-I prediction model is described. In Section 5, the validity of the HBRB-I model is verified through experiments. In Section 6, the research of this paper is summarized.

2. Problem Formulation

In Section 2.1, the problems in the student performance prediction model are analyzed. In Section 2.2, the prediction model based on HBRB-I is constructed.

2.1. Problems with Student Performance Prediction

Experts have accumulated experience over long-term teaching practice, enabling them to accurately predict student performance based on observable conditions. The experience is distilled by experts into logical expressions, transforming them into the form of belief rules. In this way, expert knowledge can be added to the BRB construction as a priori knowledge. Even when an expert's factual judgment may be slightly constrained, the overall trend remains accurate. Therefore, the initial rules are considered credible in the study of this paper and interpretability is based on this. The construction of a student performance prediction model based on HBRB-I focuses on the following three problems:

Problem 1. *Interpretability criteria for models. Interpretability is described as the ability of a model to represent the behavior of a system in an understandable way [23]. In student behavior prediction, interpretability facilitates the model to provide more valid information [24]. For users, interpretability means that the prediction results of the model are more credible and therefore the right decisions can be taken. Non-interpretable models are not trusted because their internal mechanisms are not known and it is not even possible to judge whether the results are reasonable or not. Therefore, it is important to first identify interpretability criteria to guide model construction.*

$$\text{Interpretability criteria} : \{C|(C_1, C_2, \ldots, C_m)\} \tag{1}$$

where C is the set of interpretable criteria. m denotes the number of interpretable criteria.

Problem 2. *Constructing an interpretable student performance prediction model. Initial rules set up in combination with expert knowledge are the basis for the interpretability of the model. The antecedents of the rules are indicators that include multiple aspects of the student's past and recent past. However, multiple indicators can exceed the fitting ability of expert knowledge and reduce the reliability of the rules [25], and the combination explosion of rules will increase the complexity of the system, which in turn affects the interpretability of the model. Therefore, the model constructed should have simple rules and reasonable structure.*

$$y = M(X, C, \Omega_{EK}) \tag{2}$$

where $M(\cdot)$ denotes the modeling process of HBRB-I. X is the attribute input to the prediction model. Ω_{EK} denotes the initial model parameters determined based on expert knowledge. y is the prediction result of student performance.

Problem 3. *Methods for ensuring the interpretability of optimization models. The initial parameters need to be optimized because of the uncertainty in expert knowledge. However, optimization algorithms are random in nature, and the optimized rules differ significantly from those based on expert knowledge, and the interpretability of the model is thus undermined. In addition, most BRB models are optimized with the minimum mean square error (MSE) as the objective function [26,27],*

neglecting the requirement of interpretability. Therefore, the optimization algorithm and objective function need to be improved to ensure the interpretability of the model.

$$\Omega_{obj} = O(X, Y, \Omega_{EK}, P) \tag{3}$$

where $O(\cdot)$ denotes the model optimization process. X and Y are the indicators and the actual student performance in the training data respectively. P denotes the parameters required for the optimization algorithm. Ω_{EK} denotes the model parameters to be optimized. Ω_{obj} denotes the optimized model parameters, which are the outcome of the optimization.

2.2. Subsection Construction of HBRB-I Prediction Model

In the HBRB-I model, there are one or several BRBs in each layer. The prediction model starts with the aggregation of attributes from the sub-BRBs in the bottom layer, and then the outputs of the sub-BRBs are used as the inputs of the BRBs of the upper layer until the inference is reached at the BRB in the top layer. Each of these BRBs consists of a set of belief rules, each of which has the form of Equation (4):

$$\begin{aligned} R_k \quad & : IF(x_1 \text{ is } A_1^k) \wedge (x_2 \text{ is } A_2^k) \wedge \ldots \wedge (x_{T_k} \text{ is } A_{T_k}^k) \\ & THEN\ y \text{ is } \{(D_1, \beta_{1,k}), (D_2, \beta_{2,k}), \ldots, (D_N, \beta_{N,k})\}, \left(\sum_{n=1}^{N} \beta_{n,k} \leq 1\right) \\ & WITH \text{ rule weight } \theta_k \\ & AND \text{ attribute weight } \delta_1, \delta_2, \ldots, \delta_{T_k} \\ & in\ C_1, C_2, \ldots, C_m \end{aligned} \tag{4}$$

where R_k (k = 1,2,..., L) denotes the kth rule in the BRB and L is the total number of rules. $x_1, x_2, \ldots, x_{T_k}$ are the input data for a set of attributes in X, and T_k is the number of attributes in this set. $A_1^k, A_2^k, \ldots, A_{T_k}^k$ are the reference values corresponding to each attribute. y is the inference result of the BRB. $D_1, D_2, \ldots, D_N$ are a set of predetermined outcome levels, and $\beta_1, \beta_2, \ldots, \beta_N$ are the corresponding belief degrees. θ_k is the weight of the kth rule. $\delta_1, \delta_2, \ldots, \delta_{T_k}$ are the attribute weights describing the importance of the attributes. $C_1, C_2, \ldots, C_m$ are the interpretability criteria of the model. The structure of the prediction model based on HBRB is shown in Figure 1.

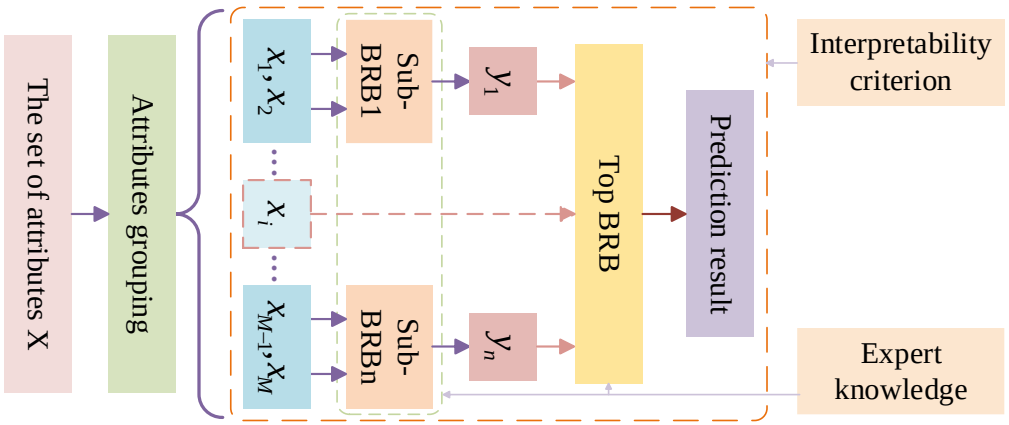


Figure 1. Structure of the HBRB prediction model.

Therefore, the construction process of HBRB-I and the solution of the problems can be represented by Figure 2.

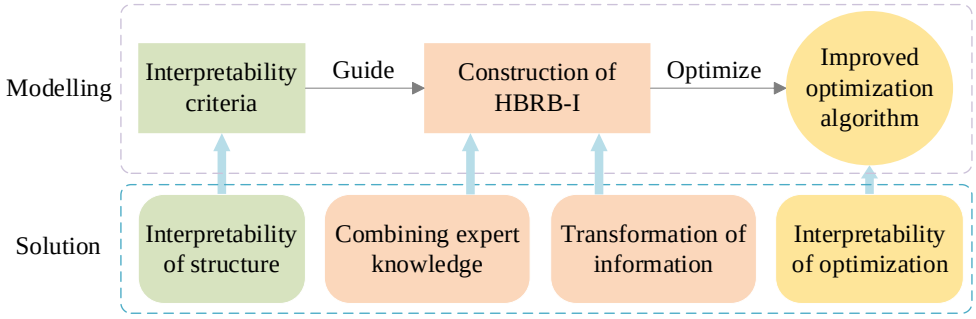


Figure 2. Construction of student performance prediction model based on HBRB-I.

3. Interpretability of Student Performance Prediction Model Based on HBRB-I

In this section, based on the general interpretability guidelines for BRB proposed in the existing study [19], the interpretability criteria for the HBRB-I prediction model in this paper are summarized. Then the HBRB construction method based on attribute grouping is proposed. In order that the interpretability of the BRB is not destroyed, a strategy of model optimization is proposed for improving the optimization algorithm. In Section 3.1, the interpretability criteria of the model are summarized. In Section 3.2, the construction method of the model is described. In Section 3.2, the interpretability strategies followed by the optimization algorithm are proposed.

3.1. Interpretability Criteria of HBRB-I Prediction Model

Criterion 1. *Reference values for attributes can be clearly distinguished. Reference values are used to describe different levels or states of an attribute, and distinguishable reference values make the semantics of the rule formulation clearer. In student performance prediction, the range of reference values should be set equal to or greater than the range of attribute values, and each reference value should correspond to a level or state of the attribute.*

Criterion 2. *The rule base is complete. Completeness means that the rules in the BRB should contain all the working states of the system. Specifically, the antecedents of rules need to traverse every reference value of all attributes [28]. This way, any input can correspond to at least one reference value, thus activating at least one rule.*

Criterion 3. *The matching degrees should be normalized. The matching degrees describe the degree of affiliation of the input data to the reference value, and they should sum to one. This is useful for distinguishing the distributional characteristics of the data, and is important in terms of which rules the data can match.*

Criterion 4. *The rule base should be simple enough. Rules in student performance prediction models are statements that abstract expert knowledge into the form of IF-THEN. If the rules have too many antecedent attributes, they will exceed the expert's ability to fit the actual situation. Therefore, the number of attributes and reference values in the BRB should be moderate, so that the number of rules is appropriate and the readability of the BRB is better. In addition, the consistency of the rules in the general interpretability criterion is implicitly included in the criterion. This is because when the rules are simple enough, expert knowledge ensures that the rules are consistent, meaning that there is no conflict between the rules.*

Criterion 5. *The structure and parameters of the model make practical sense. The structure of the model should be built based on reasonable logical relationships so that the inference process can be understood. The parameters in the BRB model are used to express expert knowledge, so they should have practical significance.*

Criterion 6. *Information is equivalent before and after transformation. Information is contained in the data, which are transformed in the model in different forms. To ensure that the inference process is reliable, reasonable transformations are needed to keep the information consistent.*

Criterion 7. *A transparent inference engine should be adopted. Confidence rules are the basis of inference, and how to reasonably utilize the confidence rules to obtain the prediction results is the key. The ER algorithm derives the results by probabilistic synthesis through rigorous mathematical calculations. Using ER as an inference engine allows BRB to have a transparent inference process.*

3.2. Method of Constructing Interpretable Model

In line with the above interpretability criteria, the student performance prediction model is constructed based on HBRB-I. HBRB was proposed to deal with complex systems with multiple attributes, offering a means to model all attributes without inducing a combination explosion [29]. HBRB uses a hierarchical structure, in which sub-BRBs at the base layer combine a limited set of attributes, channeling their outputs upward through the hierarchy until reaching the top layer for the ultimate result. However, there are two problems to be solved in this process. The foremost problem is how the attributes are grouped. The second problem is how to ensure that the process of HBRB can be understood.

The first problem addresses the sub-BRB. In the existing studies, it is stated that the correlation between attributes will affect the results of the BRB [20,30]. Because the pair of attribute inputs with strong correlation contain overlapping information [31,32], there is less valid information for BRB. If the sub-BRB inference is based on these attributes, the results passed to the upper layers would contain less valid information, which would affect the prediction results. Thus, the correlation between attributes in each sub-BRB should be reduced. Commonly used methods include principal component analysis (PCA) and linear discriminant analysis (LDA), but the results of these algorithms cannot be proven to replace the original data [33]. According to Criterion 4, the BRB has the simplest structure when there are only two attributes. Moreover, it is also easier to reduce correlation by dividing attributes into groups of two each. Therefore, an attribute grouping method based on Spearman's correlation coefficient is proposed in this paper. In this method, the correlation between each attribute and the result is firstly calculated and then ranked according to the size and thus divided into primary attributes and secondary attributes. After the division, the correlation between each pair of primary and secondary attributes is calculated, and the attribute pair that minimizes the correlation within the group as much as possible is selected as the final grouping. This process is represented as Equations (5) and (6):

$$X \xrightarrow{divide} \begin{cases} \{X_P, X_S\}, & if\ |X| = 2M \\ \{X_P, x_{M+1}, X_S\}, & if\ |X| = 2M+1 \end{cases} \tag{5}$$

$$Grouping(X_P, X_S) = \left\{(x_i, x_j) \middle| x_i \in X_P, x_j \in X_S\right\} \tag{6}$$

where X is the attribute set and $|X|$ denotes the number of attributes. X_P and X_S denote the primary and secondary attribute sets, respectively. If the number of attributes is even, there are M attributes in both X_P and X_S. Attributes can be classified into M groups. If the number of attributes is odd, the attributes cannot be grouped exactly. In this case, the $(M+1)$th attribute whose correlation size is in the middle position is taken as a separate group and directly participates in the inference of the top BRB. $Grouping(\cdot)$ denotes the process of selecting one attribute from the set of primary and secondary attributes, respectively, to form an attribute pair in the sub-BRB.

The second problem addresses the structure of HBRB. Following attribute grouping, each sub-BRB consists of only two attributes and the rule base is simple enough. However, this necessitates discussion whether insights derived from subsets of attributes yield comprehensible outcomes. Magdalena et al. argued that in a hierarchical structure, the whole system has interpretability only if each substructure can be interpreted independently [34]. Razak et al. proposed that the key to the interpretable framework of hierarchical systems

is to assess the interpretability of each subsystem [35]. Therefore, the entire system is interpretable when the intermediate variables of the HBRB are meaningful. Cao et al. argued that the hierarchical structure of HBRB responds to the process of moving from parts to the whole [36]. Integrating into the above analysis, the hierarchical structure can be understood as the process of "analyzing students' performance from different perspectives, and then synthesizing these dispersed viewpoints to form a comprehensive cognition". This way of understanding reflects the human thought process in the real world, which is conducive to the interpretability of the HBRB. Following this idea, the number of layers of HBRB should not be too high. The number of layers is not increased when the results can be derived from the bottom of a BRB aggregation. This is because one-sided perceptions based on one-sided perceptions are no longer easily accepted. The inference process for a two-layer HBRB is shown in Equation (7):

$$HBRB(X) = \begin{cases} BRB(BRB(x_1, x_{s(1)}), \ldots, BRB(x_M, x_{s(M)})), & if\ |X| = 2M \\ BRB(BRB(x_1, x_{s(1)}), \ldots, BRB(x_M, x_{s(M)}), x_{M+1}), & if\ |X| = 2M+1 \end{cases} \tag{7}$$

where $BRB(\cdot)$ denotes the inference process of a BRB. $x_{s(\cdot)}$ is a secondary attribute that has little correlation with the primary attribute. When $|x|$ is odd, x_{M+1} is used as an attribute of the top BRB. $HBRB(\cdot)$ denotes the complete inference process of a predictive model.

3.3. Interpretability Strategies of Model Optimization

Although the study in this paper concluded that expert knowledge is reliable, experts are unable to accurately describe all the working states of the system. To improve the prediction results of the model, the model parameters need to be adjusted. However, the original P-CMA-ES algorithm causes the optimized rules to deviate from the expert's judgment, thus destroying the interpretability of the HBRB-I model. For this reason, the following strategies are proposed for improving the optimization algorithm to ensure the interpretability of the optimized model.

Strategy 1: Make full use of expert knowledge. The initial parameters combining expert knowledge are the basis of model interpretability, and the optimization process of the algorithm should consider the setting of initial parameters. The original optimization algorithm searches randomly in the solution space and produces new solutions far from the initial value, thus destroying the interpretability. Therefore, using the initial parameters as the initial population of the algorithm and controlling the optimization process helps to ensure model interpretability.

$$mean^{(g)} = \begin{cases} \Omega_{EK}, & if\ g = 1 \\ mean^{(g)}, & if\ g \neq 1 \end{cases} \tag{8}$$

where $mean^{(g)}$ is the population of the gth generation.

Strategy 2: The optimized rules are consistent with the expert judgment. The initial rules are set based on the expert's knowledge of the actual working state of the system, and the semantics of the optimized rule description should be consistent with the initial rules. In addition, the meaning of belief degree is the possibility that a certain outcome occurs. If the optimized rule is "when the student's performance in the course is 'good' and the completion of homework is 'good', then the student's final performance index is {(excellent, 0.5), (medium, 0.2), (poor, 0.3)}", which is obviously not reasonable and does not conform to the expert's judgment. Therefore, the rule parameters should be subject to the following constraints:

$$\begin{aligned} R_1 &: \left\{ \theta_k^{lb} \leq \theta_k \leq \theta_k^{ub},\ (k = 1, 2, \ldots, L) \right\} \\ R_2 &: \left\{ \delta_i^{lb} \leq \delta_i \leq \delta_i^{ub},\ (i = 1, 2, \ldots, T_k) \right\} \end{aligned} \tag{9}$$

$$
\begin{aligned}
R_3 : \Big\{ & (\beta_{n,k}^{lb} \le \beta_{n,k} \le \beta_{n,k}^{ub})\ (n = 1, 2, \ldots, N) \\
& \wedge ((\beta_{1,k} \le \beta_{2,k} \le \cdots \le \beta_{N,k})\ (k = 1, 2, \ldots, L) \\
& \vee (\beta_{1,k} \ge \beta_{2,k} \ge \cdots \ge \beta_{N,k}) \\
& \vee (\beta_{1,k} \le \cdots \le \max(\beta_{1,k}, \ldots, \beta_{N,k}) \ge \cdots \ge \beta_{N,k})) \Big\}
\end{aligned} \tag{10}
$$

where R_1, R_2, R_3 denote the constraints of the parameters. lb and ub label the maximum and minimum values of the parameters that are desirable, and this range is determined by expert knowledge. The belief distribution satisfies at least the requirements of conforming to expert judgment and conforming to common sense, and the requirements vary from system to system. In the HBRB-I model of this paper, a reasonable distribution of confidence should be monotonic or convex. This is shown in Figure 3.

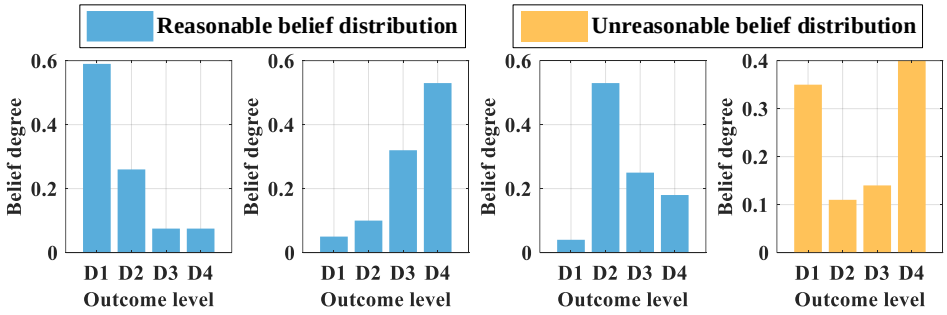


Figure 3. Distribution of belief degrees.

Strategy 3: The optimization process is centered on expert knowledge. The objective function of optimization algorithms is mostly MSE, accuracy, etc., aiming at the best result and not considering interpretability. Even though the above strategies limit the solution space to a certain range, the optimization process still searches towards the edges, thus affecting the interpretability. To solve this problem, the distance between the initial rule and the optimized rule is added to the objective function. Using this distance as a loss for model accuracy reduction allows the algorithm to take expert knowledge into account while achieving the best possible results. The new objective function enables the improved optimization algorithm to strike a balance between accuracy and interpretability. The distance between the rules before and after optimization is calculated by Equation (11):

$$
D(\beta^{in}, \beta^{op}) = \sqrt{\frac{1}{L_H} \sum_{l=1}^{L_H} \sum_{n=1}^{N_H} (\beta_{n,l}^{in} - \beta_{n,l}^{op})^2} \tag{11}
$$

where $D(\beta^{in}, \beta^{op})$ denotes the distance between the initial rule and the optimized rule. β^{in} and β^{op} denote the belief degrees before and after optimization, respectively. L_H is the number of rules in the HBRB. N_H is the number of outcome levels in the rule.

4. Inference and Optimization Process of HBRB-I Prediction Model

In this section, the method proposed in this paper for attribute grouping based on Spearman's correlation coefficient is first introduced. Then it describes the BRB inference process using the ER algorithm. Finally, the improved P-CMA-ES optimization algorithm is proposed. In Section 4.1, attributes are grouped to reasonably construct the HBRB. In Section 4.2, the inference process of the HBRB-I model is presented. In Section 4.3, the optimization process of the improved P-CMA-ES algorithm is illustrated.

4.1. Method of Attributes Grouping

The previous section proposed that the correlation between two attributes of each sub-BRB should be as small as possible. In statistics, the coefficients commonly used to measure attribute correlation are Pearson's coefficient, Kendall's coefficient, and Spear-

man's coefficient, respectively. Among them, Pearson's coefficient can only measure linear correlation and requires that the variables conform to a normal distribution, and Kendall's coefficient applies to ordered categorical variable, while Spearman's correlation coefficient does not require the attributes or distributional characteristics of the variables and is more widely applicable. Spearman's correlation coefficient is calculated based on the position of the data in the ranking and takes values between -1 and 1. The closer its absolute value is to 1, the greater the correlation between two attributes, and the closer it is to 0, the smaller the correlation. In this paper, attributes are grouped based on the absolute value of the Spearman correlation coefficient, which is calculated as shown in Equation (12):

$$\begin{gathered} \rho_x^y = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)},\ d_i = rg(x_i) - rg(y_i) \\ \Gamma_x^y = \left|\rho_x^y\right| \end{gathered} \tag{12}$$

where ρ_x^y denotes the Spearman correlation coefficient between the variables x and y. Γ_x^y is the absolute value of ρ_x^y, which describes the size of the correlation. n is the number of samples. d_i denotes the difference in the order of the ith set of data for the two variables. $rg(\cdot)$ denotes the rank of the variable's value, which is the position of the data in the descending order.

The first step in attribute grouping is to select the primary and secondary attribute sets from X. First, the correlation $\Gamma_{x_i}^y$ between each attribute x_i and the prediction result Y is calculated. The attributes are then sorted from largest to smallest and relabeled. The result of this process is represented as Equation (13):

$$\Gamma_{x_1}^y \geq \Gamma_{x_2}^y \geq \ldots \geq \Gamma_{x_{2M}}^y (\geq \Gamma_{x_{2M+1}}^y) \tag{13}$$

where $x_1, x_2, \ldots, x_{2M}(, x_{2M+1})$ are attributes relabeled by size, and x_1 has the highest correlation with the result.

If the number of attributes $|X|$ is even, then $x_1, \ldots, x_M$ are categorized as the primary attribute set and $x_{M+1}, \ldots, x_{2M}$ are categorized as the secondary attribute set. If the number of attributes is odd, the set of secondary attributes is $x_{M+2}, \ldots, x_{2M+1}$. At this point, x_{M+1} is not involved in the next step and stands alone as a set.

The second step is to calculate the correlation between each pair of primary and secondary attributes. One attribute is taken from each of the sets of primary and secondary attributes in turn, and the correlation between them is computed, and then a matrix of the form Equation (14) is formed:

$$[\Gamma] = \begin{cases} \begin{bmatrix} \Gamma_{x_1}^{x_{M+1}} & \cdots & \Gamma_{x_1}^{x_{2M}} \\ \vdots & \ddots & \vdots \\ \Gamma_{x_M}^{x_{M+1}} & \cdots & \Gamma_{x_M}^{x_{2M}} \end{bmatrix}, & if\ |X| = 2M \\ \begin{bmatrix} \Gamma_{x_1}^{x_{M+2}} & \cdots & \Gamma_{x_1}^{x_{2M+1}} \\ \vdots & \ddots & \vdots \\ \Gamma_{x_M}^{x_{M+2}} & \cdots & \Gamma_{x_M}^{x_{2M+1}} \end{bmatrix}, & if\ |X| = 2M+1 \end{cases} \tag{14}$$

where a row denotes the correlation between a primary attribute and each secondary attribute.

The final step is grouping to minimize the correlation between each pair of attributes as much as possible. The purpose of this step is to maximize the average valid information

contained in each set of attribute data so that the outputs of the sub-BRB are meaningful. The grouping method is based on Equation (15):

$$s(i) = \begin{cases} \min \sum_{i=1}^{M} \Gamma_{x_i}^{x_{s(i)}},\ s(i) \neq \ldots \neq s(M) \\ M+1, \ldots, 2M,\ if\ |X| = 2M \\ M+2, \ldots, 2M+1,\ if\ |X| = 2M+1 \end{cases} \tag{15}$$

The combination that minimizes the sum of correlations is the final grouping. It is denoted as $\left\{\left(x_1, x_{s(1)}\right), \ldots, \left(x_M, x_{s(M)}\right)\right\}$ or $\left\{\left(x_1, x_{s(1)}\right), \ldots, \left(x_M, x_{s(M)}\right), x_{2M+1}\right\}$.

4.2. Inference Process of HBRB-I Model

The structure of the HBRB-I model can be determined based on the result of attribute grouping. Then the final prediction results are obtained by inference layer by layer from bottom to top. Based on the interpretability criterion proposed in Section 3.1, the inference process for each BRB is as follows:

First, the matching degrees between the input data and the reference values are calculated. As in Equation (16),

$$S(x_i) = \left\{ \left(A_{i,j}^k, \alpha_{i,j}^k\right) \middle| i = 1, \ldots, T_k, j = 1, \ldots, J_i \right\}$$
$$\alpha_{i,j}^k = \begin{cases} \frac{A_{i,j+1}^k - x_i}{A_{i,j+1}^k - A_{i,j}^k} & , j = k \left(A_{i,j}^k \leq x_i \leq A_{i,j+1}^k\right) \\ \frac{x_i - A_{i,j}^k}{A_{i,j+1}^k - A_{i,j}^k} & , j = k+1 \\ 0 & , j = 1, \ldots, J_i (j \neq k, k+1) \end{cases} \tag{16}$$

where $S(x_i)$ denotes the process of information transformation on the input data x_i. $\alpha_{i,j}^k$ denotes the matching degree of x_i belonging to the jth reference value of the ith attribute. J_i denotes the number of reference values. $A_{i,j}^k$ and $A_{i,j+1}^k$ denote the two neighboring reference values.

After obtaining the matching degrees of the data in the rules, the activation weights are calculated. The calculation is shown in Equation (17):

$$\omega_k = \frac{\theta_k \prod_{i=1}^{T_k} \left(a_{i,j}^k\right)^{\overline{\delta_i}}}{\sum_{l=1}^{L} \theta_l \prod_{i=1}^{T_k} \left(a_{i,j}^l\right)^{\overline{\delta_i}}},\ \overline{\delta_i} = \frac{\delta_i}{\max\limits_{i=1,2,\ldots,T_k} \{\delta_i\}} \tag{17}$$

where ω_k denotes the activation weight of the kth rule. $\overline{\delta_i}$ denotes the relative weight of the attribute.

Then, the activated rules are used as a basis for inference, and the ER algorithm is used to reason about the belief degrees of the outcome levels. The inference process is shown in Equation (18):

$$\beta_n = \frac{\mu \left[\prod_{k=1}^{L} \left(\omega_k \beta_{n,k} + 1 - \omega_k \sum_{j=1}^{N} \beta_{j,k}\right) - \prod_{k=1}^{L} \left(1 - \omega_k \sum_{j=1}^{N} \beta_{j,k}\right) \right]}{1 - \mu \left[\prod_{k=1}^{L} \left(1 - \omega_k\right) \right]}$$
$$\mu = \left[\sum_{n=1}^{N} \prod_{k=1}^{L} \left(\omega_k \beta_{n,k} + 1 - \omega_k \sum_{j=1}^{N} \beta_{j,k}\right) - (N-1) \prod_{k=1}^{L} \left(1 - \omega_k \sum_{j=1}^{N} \beta_{j,k}\right) \right]^{-1} \tag{18}$$

where $\beta_n (n = 1, 2, \ldots, N)$ denotes the belief degree of the nth outcome level D_n.

The output of the BRB is represented in the form of Equation (19):

$$S(\widetilde{x}) = \{(D_n, \beta_n) | n = 1, 2, \ldots, N\} \tag{19}$$

where $S\left(\widetilde{x}\right)$ denotes the belief distribution obtained from a set of data $\widetilde{x}$ of the attributes.

Finally, the inference result of the BRB is calculated. As shown in Equation (20),

$$u(S(\widetilde{x})) = \sum_{n=1}^{N} u(D_n)\beta_n \tag{20}$$

where $u\left(S\left(\widetilde{x}\right)\right)$ denotes the final result of the BRB. $u(D_n)$ denotes the utility value of the nth outcome level.

Thus, the prediction result of the HBRB-I model is represented as $u(S(X))$.

4.3. Optimization Process of HBRB-I Model

The P-CMA-ES algorithm is characterized by its ability to adaptively adjust the covariance matrix and efficiently search the solution space. The P-CMA-ES algorithm is suitable for optimization problems with nonlinear, nonconvex functions in continuous domains. It has the advantages of adaptability, efficiency, and stability and is widely used in BRB optimization process [37,38]. However, due to the randomness of the search process, the new solution will deviate from expert knowledge, thus destroying the interpretability of HBRB-I. Therefore, based on the strategies proposed in Section 3.3, the improved P-CMA-ES algorithm is proposed as shown in Figure 4.

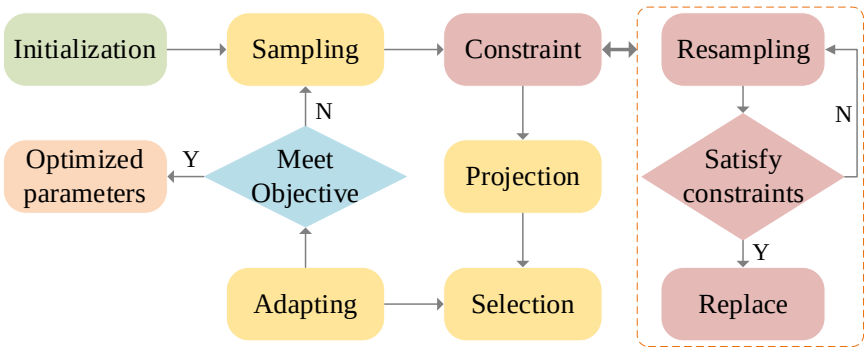


Figure 4. Process of the improved P-CMA-ES algorithm.

First, the objective function of the optimization algorithm is determined. According to strategy 3, the new objective function is

$$\min MES(\Omega) + \kappa D(\beta^{in}, \beta^{op}) \tag{21}$$

$$MES(\Omega) = \frac{1}{n}\sum_{i=1}^{n} (\hat{u}(X) - u(X))^2 \tag{22}$$

where $\Omega = \{\theta, \delta, \beta\}$ denotes that the parameters need to be adjusted. $D\left(\beta^{in}, \beta^{op}\right)$ serves as the loss of prediction accuracy. $\kappa (0 \leq \kappa \leq 1)$ is the loss intensity, which is determined by expert knowledge. The closer κ is to 1, the greater the loss, indicating that interpretability is more important. The closer κ is to 0, the smaller the loss, indicating that interpretability is hardly considered. $MSE(\cdot)$ denotes the mean-square error between predicted and true values. $\hat{u}(X)$ denotes the prediction result of HBRB-I, and $u(X)$ denotes the sample of actual values.

The steps of the improved optimization algorithm are as follows:

Step 1: Initialization operation. According to strategy 1, determine initial population $\varpi^0 = \Omega_{EK}$ and initial covariance matrix C^0, initial step size ϵ^0, population size λ.

Step 2: Sampling operation. The next generation of population is generated from Equation (23):

$$\Omega_i^{g+1} = \varpi^g + \varepsilon^g N(0, C^g),\ i = 1, 2, \ldots, \lambda \tag{23}$$

where Ω_i^{g+1} denotes the ith solution of the $(g+1)$th generation. ϖ^g denotes the mean of the population of the gth generation. ϵ^g denotes the step size. $N(\cdot)$ denotes the normal distribution. C^g denotes the covariance matrix.

Step 3: Constraint operation. According to strategy 2, resample the parameters that do not meet the constraints in Equation (24). Generate new solutions using the process of Equation (23) until all parameters satisfy the constraints. Then the new solution replaces any original solution that does not satisfy the constraints, represented as Equation (25).

$$\begin{gathered} \theta \sim R_1,\ \delta \sim R_2, \\ \textstyle\sum_{n=1}^{N} \beta_n \leq 1,\ \beta \sim R_3 \end{gathered} \tag{24}$$

$$\Omega_i^{g+1} \Leftarrow \Omega_i^{g+1} = \varpi^g + \varepsilon^g N(0, C^g) \tag{25}$$

where R_1, R_2, R_3 are interpretability constraints. $\Leftarrow$ denotes the replacement operation.

Step 4: Projection operation. The parameters of BRB satisfy the equational constraints, and the unsatisfied solutions need to be projected into the feasible domain hyperplane. The hyperplane of the equational constraints is represented by Equation (26), and the projection operation is represented by Equation (27):

$$I_e \Omega_i^{g+1}(1 + n_e \times (j-1) : n_e \times j) = 1 \tag{26}$$

$$\begin{gathered} \Omega_i^{g+1}(1 + n_e \times (j-1) : n_e \times j) = \Omega_i^{g+1}(1 + n_e \times (j-1) : n_e \times j) \\ -\frac{I_e^T}{I_e \times I_e^T} \times \Omega_i^{g+1}(1 + n_e \times (j-1) : n_e \times j) \times I_e \end{gathered} \tag{27}$$

where $I_e = [1 \ldots 1]_{1 \times n_e}$ denotes the parameter vector. j denotes the number of equation constraints.

Step 5: Selection operation. Sort the population according to Equation (21). Then select τ optimal solutions from them as the next generation and update their mean values. The process is as follows:

$$\varpi^{g+1} = \sum_{i=1}^{\tau} v_i \Omega_{i:\lambda}^{g+1} \tag{28}$$

where v_i denotes the coefficients used for weighted reorganization.

Step 6: Adapting operation. Update the covariance matrix and step size to control the direction and speed of population evolution. The updating process is shown in Equation (29) and Equation (30), respectively:

$$\begin{gathered} C^{g+1} = (1 - c_1 - c_2) C^g + c_1 p_c^{g+1} \left(p_c^{g+1}\right)^T \\ + c_2 \sum_{i=1}^{\tau} \xi_i \left(\frac{(\Omega_{i:\lambda}^{g+1} - \varpi^g)}{\varepsilon^g} \right) \left(\frac{(\Omega_{i:\lambda}^{g+1} - \varpi^g)}{\varepsilon^g} \right)^T \\ p_c^{g+1} = (1 - c_c) p_c^g + \sqrt{c_c(2 - c_c) \left(\sum_{i=1}^{\tau} \xi_i^2 \right)^{-1}} \frac{\varpi^{g+1} - \varpi^g}{\varepsilon^g} \end{gathered} \tag{29}$$

$$\begin{gathered} \varepsilon^{g+1} = \varepsilon^g \exp\left(\frac{c_\sigma}{d_\sigma} \left(\frac{\left\| p_\sigma^{g+1} \right\|}{E\|N(0,1)\|} - 1 \right) \right) \\ p_\sigma^{g+1} = (1 - c_c) p_\sigma^g + \sqrt{c_c(2 - c_\sigma) \left(\sum_{i=1}^{\tau} \xi_i^2 \right)^{-1}} (c^g)^{-\frac{1}{2}} \frac{\varpi^{g+1} - \varpi^g}{\varepsilon^g} \end{gathered} \tag{30}$$

where c_1 and c_2 denote the learning rate.

Step 7: Repeat the operation of sampling to adapting until the maximum number of iterations is reached.

5. Case Study

In this section, a dataset of student performance is used as case study to validate the effectiveness of the HBRB-I prediction model. In Section 5.1, attributes are grouped. In Section 5.2, the student performance prediction model based on HBRB-I is constructed and optimized. In Section 5.3, the interpretability of HBRB-I is validated. In Section 5.4, comparative experiments are designed.

5.1. Attributes Grouping

The dataset used in this paper has four attributes, which are (1) learning time x'_1, which represents the total number of hours of learning for each student; (2) previous scores x'_2, which represents the student's performance on the past test; (3) sleep time x'_3, which represents the average number of hours of sleep a student receives each day; (4) and number of practice x'_4, which represents the number of test papers the student usually uses as practice. In addition, students' performance is measured by the performance index Y, which is academic achievement.

The attributes are divided according to the method in Section 4.1. First, the correlations between the four attributes and the performance index are calculated. Then they are sorted by size in descending order and the attributes are relabeled in order. The correlations are shown in Table 1, and the sorting result is shown in Equation (31). The new labels of the attributes are shown in Table 2. Thus, the result of division is primary attribute set $X_P = \{x_1, x_2\}$ and secondary attribute set $X_S = \{x_3, x_4\}$.

$$\Gamma^Y_{x'_2} > \Gamma^Y_{x'_1} > \Gamma^Y_{x'_3} > \Gamma^Y_{x'_4} \tag{31}$$

Table 1. Correlation between attributes and student performance.

Attribute	x'_1	x'_2	x'_3	x'_4
Correlation	0.3854	0.9151	0.0856	0.0298

Table 2. New labels for attributes.

Attribute	**Previous Score**	**Learning Time**	**Sleep Time**	**Number of Practice**
Label	x_1	x_2	x_3	x_4

Then, the correlation between each pair of primary and secondary attributes is calculated in turn. The correlations are arranged in rows based on the labels of the primary attributes and constructed as a matrix as in Equation (32).

$$\begin{bmatrix} \Gamma^{x_3}_{x_1} & \Gamma^{x_4}_{x_1} \\ \Gamma^{x_3}_{x_2} & \Gamma^{x_4}_{x_2} \end{bmatrix} = \begin{bmatrix} 0.06 & 0.032 \\ 0.0213 & 0.0685 \end{bmatrix} \tag{32}$$

Finally, the grouping of attributes is selected according to Equation (15). The result of grouping is represented as

$$\{(x_1, x_4), (x_2, x_3)\} \tag{33}$$

5.2. Construction and Optimization of HBRB-I Model

Based on the results of attribute grouping, a two-layer HBRB-I model can be constructed as in Figure 5. The model has two sub-BRBs, called SBRB1 and SBRB2, which both aggregate two attributes with the output of y_1 and y_2, respectively. The top BRB, called

TBRB, is used to aggregate the outputs of the sub-BRBs and then reason about the final prediction result Y.

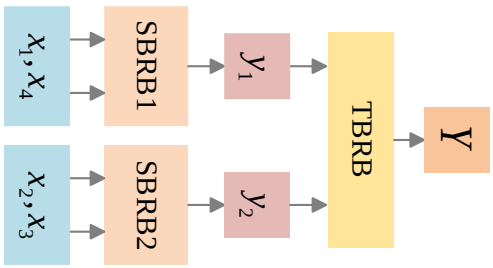


Figure 5. Structure of HBRB-I model.

After determining the structure, the student performance prediction model is constructed. In SBRB1, the previous score and the number of practice are set at four reference points, which are H (High), M (Medium), L (Low), and P (Poor). In SBRB2, learning time and sleep time are set at four reference points, which are L (Long), SL (Slightly Longer), SS (Slightly Shorter) and S (Short). Both y_1 and y_2 are set at E (Excellent), G (Good), A (Average) and P (Poor). Y is set the same as the previous score. Then the initial HBRB-I is constructed by combining the expert knowledge. The initial attribute weights and specific reference values are shown in Table 3. The initial rules of the TBRB are shown in Table 4. If a single-layer BRB structure is used, the number of rules is $4^4 = 256$, whereas the total number of rules in the HBRB proposed in this paper is $4^2 \times 3 = 48$. It can be seen that the number of rules is reduced dramatically, and the complexity of the system is greatly reduced.

Table 3. Reference values corresponding to reference points.

BRB	Input	Reference Point	Reference Value	δ	Output	Reference Point	Reference Value
Sub-BRB1	x_1 x_4	(H, M, L, P) (H, M, L, P)	(100, 85, 55, 30) (10, 7, 3, 0)	0.8 0.5	y_1	(E, G, A, P)	(4, 3, 2, 1)
Sub-BRB2	x_2 x_3	(L, SL, SS, S) (L, SL, SS, S)	(9, 7, 3, 0) (9, 7, 5, 3)	0.8 0.5	y_2	(E, G, A, P)	(4, 3, 2, 1)
TBRB	y_1 y_2	- -	- -	0.8 0.4	Y	(H, M, L, P)	(100, 80, 40, 10)

Table 4. Initial rules of TBRB.

Number	Initial Weight	Attributes	{H, M, L, P}
1	0.9	E∧E	{0.9, 0.1, 0, 0}
2	0.9	E∧G	{0.8, 0.2, 0, 0}
3	0.9	E∧A	{0.7, 0.3, 0, 0}
4	0.9	E∧P	{0.7, 0.2, 0.1, 0}
5	0.9	G∧E	{0.5, 0.4, 0.1, 0}
6	0.9	G∧G	{0.4, 0.5, 0.1, 0}
7	0.9	G∧A	{0.4, 0.3, 0.2, 0.1}
8	0.9	G∧P	{0.3, 0.3, 0.2, 0.2}
9	0.9	A∧E	{0.1, 0.2, 0.5, 0.2}
10	0.9	A∧G	{0, 0.2, 0.5, 0.3}
11	0.9	A∧A	{0, 0.2, 0.4, 0.4}
12	0.9	A∧P	{0, 0.1, 0.4, 0.5}
13	0.9	P∧E	{0, 0.1, 0.2, 0.7}
14	0.9	P∧G	{0, 0, 0.3, 0.7}
15	0.9	P∧A	{0, 0, 0.2, 0.8}
16	0.9	P∧P	{0, 0, 0.1, 0.9}

The initial model is optimized using the improved P-CMA-ES algorithm. Because the results produced by the sub-BRB are considered to be one-sided knowledge, the optimization process needs avoid falling into the local optimum, so the global optimization is used. The loss intensity κ in the objective function is set to 0.8, which indicates the high importance of interpretability. A total of 300 sets of data are selected from the student performance dataset, of which 210 sets are used for training and the remaining 90 sets are used for testing. The prediction results of the optimized model are shown in Figure 6, and the MSE between the predicted and actual values is 5.8609. It can be seen that HBRB-I has a good fitting effect, and the prediction results are reliable as far as accuracy is concerned. The optimized parameters are shown in Tables 5 and 6.

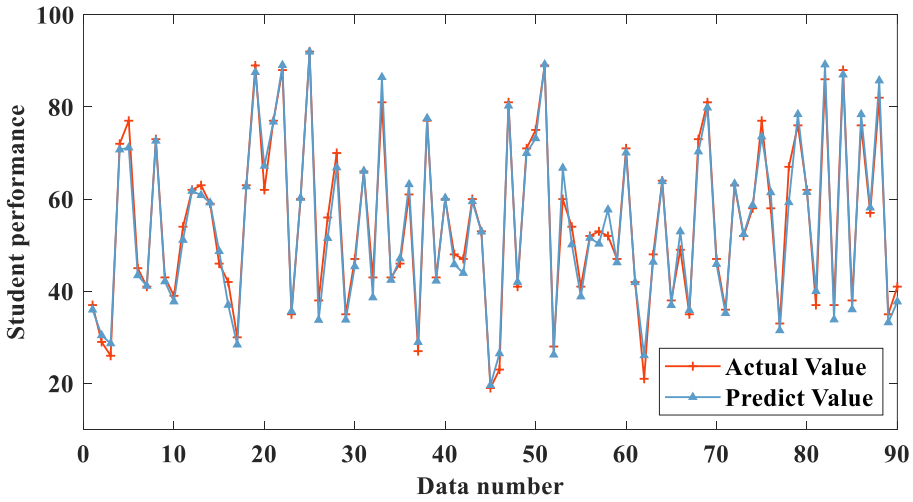


Figure 6. Effect of optimized HBRB-I.

Table 5. Optimized attribute weights.

x_1	x_4	x_2	x_3	y_1	y_2
0.761	0.3311	0.787	0.3521	0.8419	0.7081

Table 6. Optimized rules of TBRB.

Rule	Optimized Weight	Attributes	{H, M, L, P}
1	0.6018	E∧E	{0.9499, 0.053, 0, 0}
2	0.4632	E∧G	{0.8456, 0.1045, 0.0412, 0.0088}
3	0.292	E∧A	{0.6053, 0.2863, 0.0839, 0.0245}
4	0.483	E∧P	{0.6311, 0.2269, 0.1073, 0.0346}
5	0.7703	G∧E	{0.5926, 0.3751, 0.0316, 0.0007}
6	0.4321	G∧G	{0.31, 0.4148, 0.1959, 0.0793}
7	0.5567	G∧A	{0.3512, 0.2679, 0.2308, 0.1501}
8	0.8041	G∧P	{0.2297, 0.2374, 0.2847, 0.2482}
9	0.9704	A∧E	{0.1103, 0.2746, 0.4876, 0.1275}
10	0.4462	A∧G	{0.0711, 0.2574, 0.4296, 0.2419}
11	0.8063	A∧A	{0.054, 0.1648, 0.364, 0.4172}
12	0.9764	A∧P	{0.0243, 0.0816, 0.3750, 0.5191}
13	0.1468	P∧E	{0.0143, 0.0469, 0.2046, 0.7342}
14	0.46	P∧G	{0.0327, 0.0677, 0.2644, 0.6352}
15	0.7112	P∧A	{0.0015, 0.0109, 0.1091, 0.8785}
16	0.9203	P∧P	{0.0018, 0.0172, 0.0228, 0.9582}

5.3. Interpretability Analysis

The interpretability of the HBRR-I model derives primarily from expert knowledge. Based on their accumulated experience in teaching practice, experts abstract student behaviors and possible performances into the form of belief rules. In this process, the expert knowledge is transformed into the parameters in the rules, and at the same time, these parameters are given practical meanings. Therefore, the interpretability of HBRB-I can be guaranteed when the optimized rules and parameters do not deviate from the expert judgment.

The initial model is called HBRB0, the model using the original P-CMA-ES algorithm is called HBRB1, the model using the improved P-CMA-ES algorithm is called HBRB2, and the model proposed in this paper is HBRB-I. To analyze the interpretability of the model, the belief distributions of the initial rules and the optimized rules are compared, as shown in Figure 7. Among them, rules 1, 2, 3, 4, 9, 15, and 16 remain consistent with expert knowledge after optimization, and the belief degrees of the rest of the rules have a distribution similar to that of expert knowledge. This indicates that the optimized rules are still reliable and interpretability can be guaranteed. In contrast, in HBRB0, which does not consider the interpretability optimization strategy, the optimized rules completely deviate from the expert knowledge. Moreover, the prediction results of rules 3, 10, and 14 are "high" and "poor" at the same time, which is obviously unreasonable. This indicates that the interpretability of HBRB0 is destroyed in the optimization process.

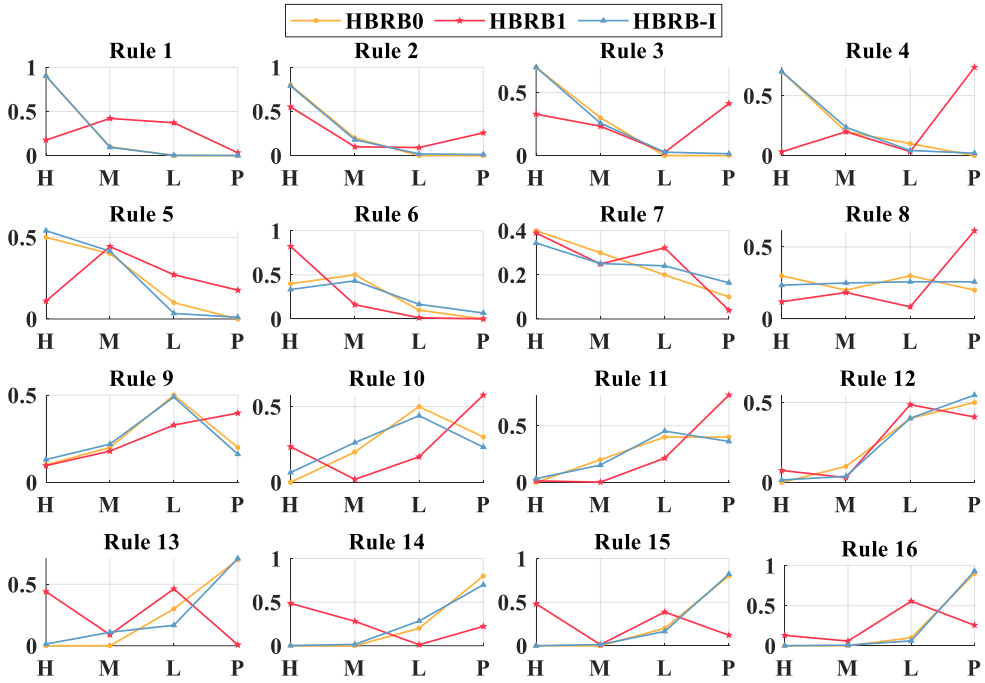


Figure 7. Comparison of optimized rules of HBRB0, HBRB1, and HBRB-I.

Euclidean distance represents the distance between two points in space as shown in Equation (34). Thus, interpretability can be measured using Euclidean distance. The initial rule and the optimized rule have corresponding points in the solution space, and the larger the distance between them, the more deviation from expert knowledge and the worse the interpretability. The Euclidean distances and errors for each HBRB model are shown in Table 7.

$$\rho = \sqrt{(\beta_1^{in} - \beta_1^{op}) + \ldots + (\beta_{L_H}^{in} - \beta_{L_H}^{op})} = \sqrt{\sum_{i=1}^{L_H} (\beta_i^{in} - \beta_i^{op})^2} \tag{34}$$

where ρ denotes the Euclidean distance between the initial rule and the optimized rule. L_H denotes the total number of rules in the HBRB.

Table 7. Euclidean distance and prediction error of different models.

Model	HBRB1	HBRB2	HBRB-I
Euclidean distance	4.5689	1.0065	0.739
MSE	5.6058	5.7408	5.8609

By comparison, HBRB1 has the smallest error, but its Euclidean distance is much larger than that of the rest of the models. This is because in parameter optimization, the optimization algorithm does not consider the effect of the initial parameters on the interpretability but searches in the solution space in a random manner. This leads to optimized rules which do not match expert knowledge, which affects the interpretability of the model. Although HBRB2 limits the algorithm's ability to search, it still achieves good prediction results. More importantly, the Euclidean distance is substantially reduced, suggesting that the rules of HBRB2 are close to expert judgment. The optimization process of HBRB-I is centered on expert knowledge, which has the smallest Euclidean distance. Compared with HBRB2, the error is only slightly increased, but the optimized rules are closer to the experts' knowledge, indicating that the HBRB-I model has the best interpretability.

5.4. Sensitivity Analysis

The traceability of the HBRB-I itself is also an aspect of interpretability [39]. The results of the sensitivity analysis are shown in Figure 8. In SBBR1, attribute x_1 has a greater impact on the prediction results than attribute x_4. In SBBR2, attribute x_2 has a greater impact on the prediction results than attribute x_3. The results of the analysis are consistent with the calculations in Table 1 and Equation (32) and with the experts' perceptions of student behavior. Using sensitivity analysis, teaching experts can discover the factors that have the greatest impact on students' academic performance. Based on the analysis results, they can adjust their teaching strategies in time to improve teaching quality.

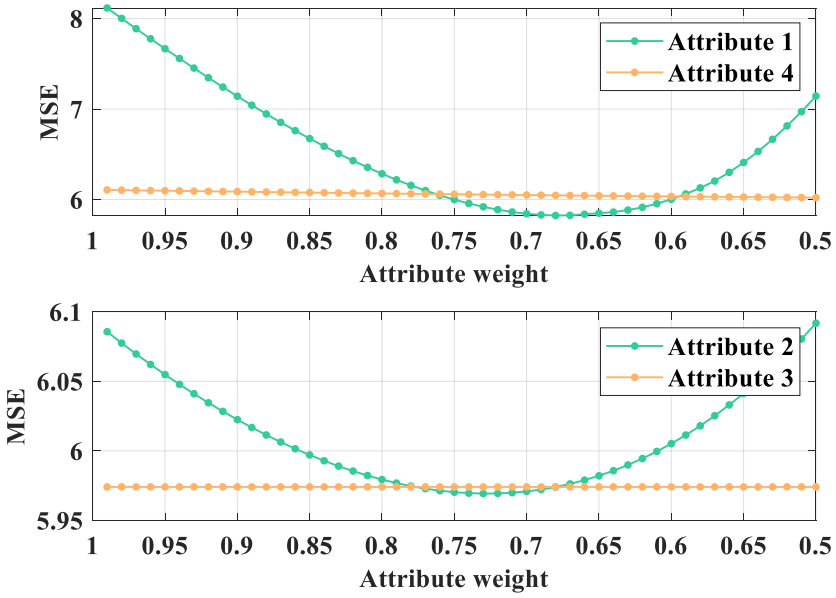


Figure 8. Sensitivity analysis of HBRB-I.

5.5. Comparison Experiment

In order to verify the effectiveness and reliability of the HBRB-I prediction model, three groups of comparison experiments are designed. In the first group, HBRB1, HBRB2, and HBRB-I are compared to verify whether the improved new method is indeed effective. In the second group, HBRB-I is compared with traditional machine learning algorithms, including back propagation neural networks (BPNNs), random forest (RF), and support vector machines (SVMs). Thirty repetitions are performed for all methods, and the experiment

results are shown in Table 8. In the third group, prediction models using different attribute groupings are constructed to explore the effect of attribute correlation on the results.

Table 8. Results of comparative experiments.

Comparison	Model	Max MSE	Min MSE	Average MSE	The Standard Deviation of MSE	Average ρ
Group 1	HBRB1	7.2072	4.7771	5.5675	0.6439	4.4065
	HBRB2	6.7221	4.7525	5.7916	0.4769	0.9899
	HBRB-I	6.2742	4.9807	5.8259	0.2948	0.7358
Group 2	BPNN	5.7598	3.5545	4.4335	0.5613	-
	RF	13.09	6.6558	8.9222	1.2808	-
	SVM	12.707	6.2551	8.8461	1.7048	-

Analyze the results of the first group of comparison experiments in Table 8. The average MSE is used as an index to evaluate the accuracy of the prediction model, and the average Euclidean distance is used as an index to evaluate the interpretability. HBRB2 decreased the accuracy of the prediction results by 4.03% and the average Euclidean distance decreased by 77.54% compared to HBRB1. HBRB-I decreased the accuracy of the prediction results by 4.64% and the average Euclidean distance decreased by 83.3% compared to HBRB1. All three prediction models show good accuracy, but the interpretability of HBRB1 is not guaranteed. With the addition of interpretability constraints, the accuracy of HBRB2 decreases slightly, but the interpretability is greatly improved. This indicates that the improved P-CMA-ES algorithm proposed in this paper ensures that the model remains interpretable while improving the prediction accuracy. The accuracy of HBRB-I is comparable to that of HBRB2, but the Euclidean distance is significantly smaller than that of HBRB2. This indicates that the new objective function can control the optimization process.

To further investigate the effects of constraints and new objective function on interpretability, the optimization processes of HBRB models with different optimization methods are compared, as shown in Figure 9. Under the condition that the initial populations are all expert knowledge, the Euclidean distance of HBRB1 using the original P-CMA-ES increases rapidly, and the offspring populations deviate far from expert knowledge. After adding interpretability constraints, the Euclidean distance of HBRB2 is effectively controlled in iterations. The Euclidean distance of HBRB-I with the new objective function is minimized, indicating that the interpretability is guaranteed to the greatest extent. The improvement of the optimization algorithm in this paper is proved to be effective.

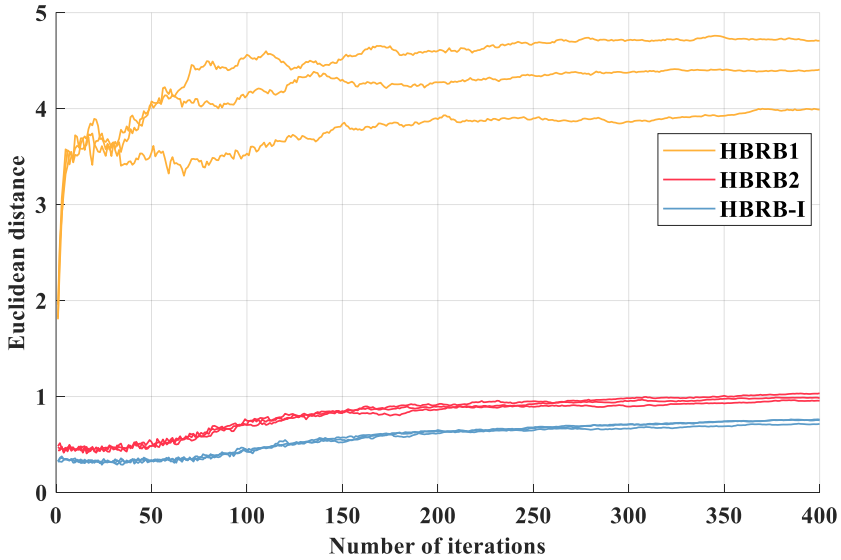


Figure 9. Euclidean distances of HBRB1, HBRB2 and HBRB-I in iterations.

In addition, the stability of the models needs to be verified. According to Table 8, the average MSE difference between the three prediction models is small, but the accuracy of HBRB1 fluctuates within a slightly larger range. Observing Figure 10, it can be seen that HBRB-I has the smallest change in accuracy in repeated experiments. Moreover, the standard deviation of the MSE of HBRB-I is also smaller than that of the other methods, so HBRB-I has the best performance in terms of stability. The results of one of the predictions of HBRB1, HBRB2, and HBRB-I are placed in Figure 11. As can be seen from the red markers in the figure, the individual prediction results of HBRB1 have a large difference from the actual values. These results may influence educational experts' judgment of student performance and lead them to make wrong decisions.

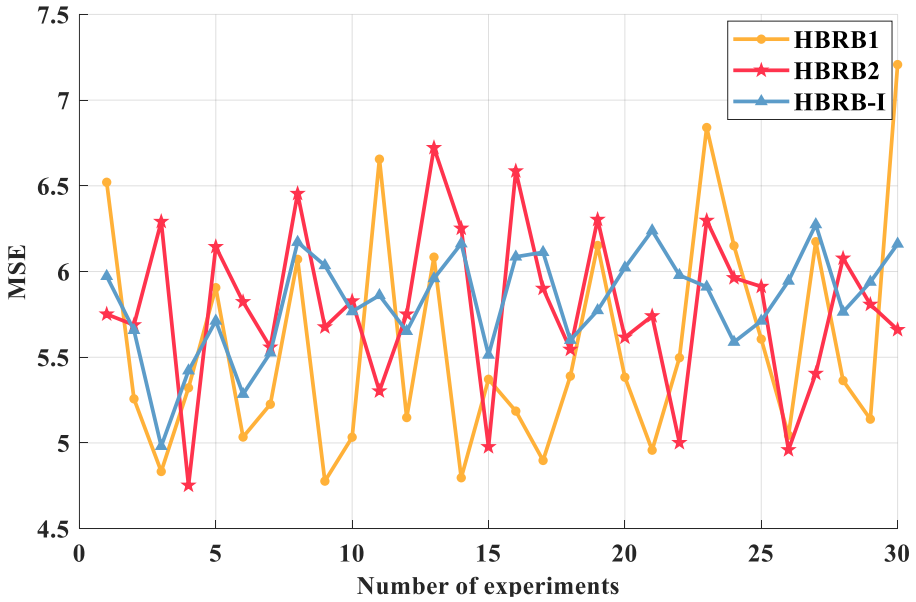


Figure 10. Prediction accuracy of HBRB1, HBRB2, and HBRB-I.

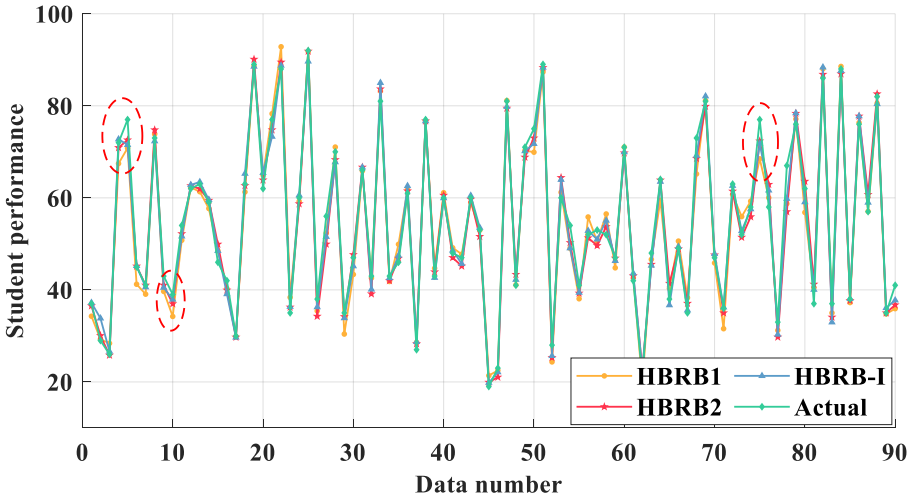


Figure 11. Prediction results of HBRB1, HBRB2 and HBRB-I.

Figures 12 and 13 exhibit the belief distribution of the prediction results for HBRB0 and HBRB-I, respectively. In the red region, student performance is the same, but the belief degrees have a large difference. According to the analysis of the figure, it can be seen that the belief distribution of HBRB-I clearly describes student performance because belief degrees can be explicitly assigned to outcome levels. In contrast, HBRB0 cannot handle uncertainty and it has difficulty differentiating between the levels. For example,

when student performance is 61, the belief distribution of HHBRB-I is {0.181, 0.3267, 0.3182, 0.1741}, which describes the semantics that the student performance is between M and L. Meanwhile, the confidence distribution of BRB0 is {0.3943, 0.1037, 0.1699, 0.332}, the semantics of which are incomprehensible.

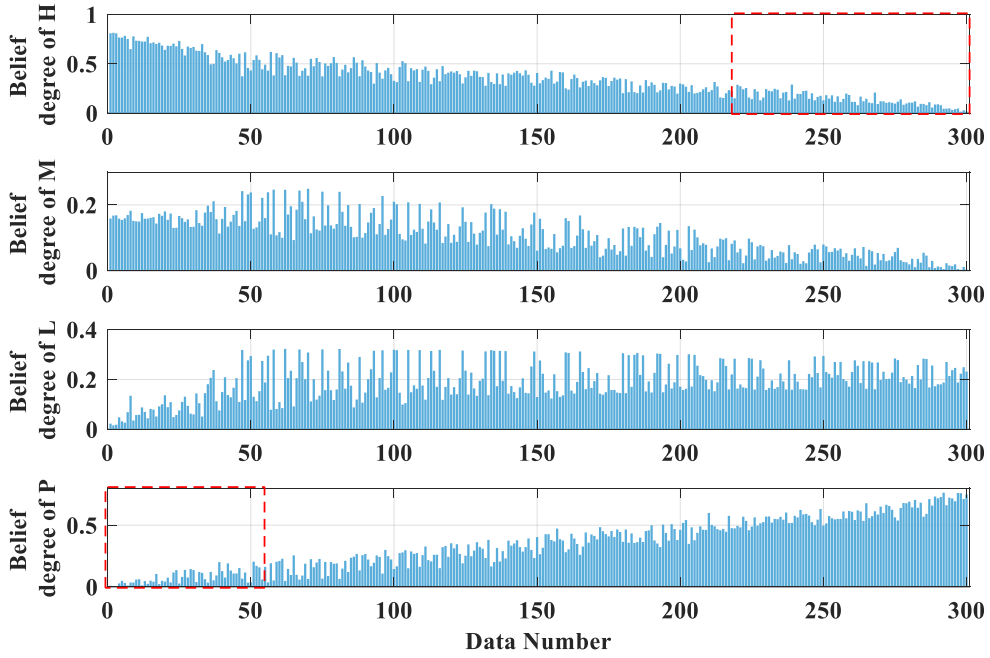


Figure 12. Belief distribution of HBRB0 prediction results.

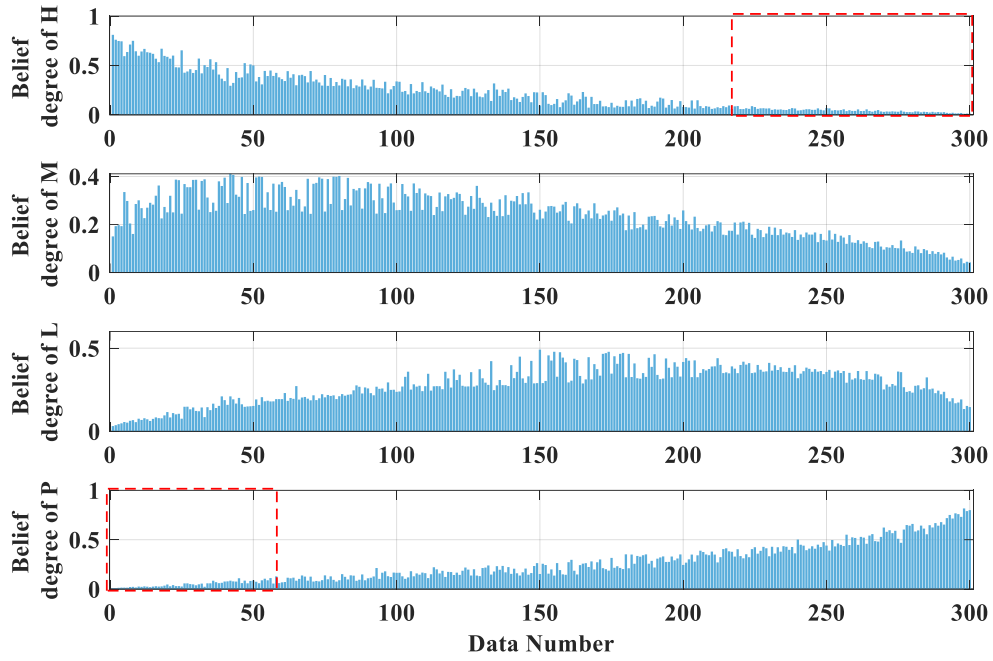


Figure 13. Belief distribution of HBRB-I prediction results.

The results of the second group of comparison experiments are shown in Figures 14 and 15, in which the methods are able to fit the results of student performance. It is clear that the accuracy and stability of RF and SVM are weaker than those of BPNN and HBRB-I. BPNN is more accurate compared to HBRB-I, but the stability of the predicted results is slightly lower. More critically, BPNN is a black-box model. Although BPNN can extract features from the data, these features are complex and abstract and are completely incomprehensible

to humans. This leads to users not understanding the internal mechanism and distrusting the prediction results. Unlike BPNN, HBRB-I has good interpretability and its process of producing prediction results can be understood by the user. Therefore, even though the accuracy is lower than that of BPNN, users still trust the results of HBRB-I.

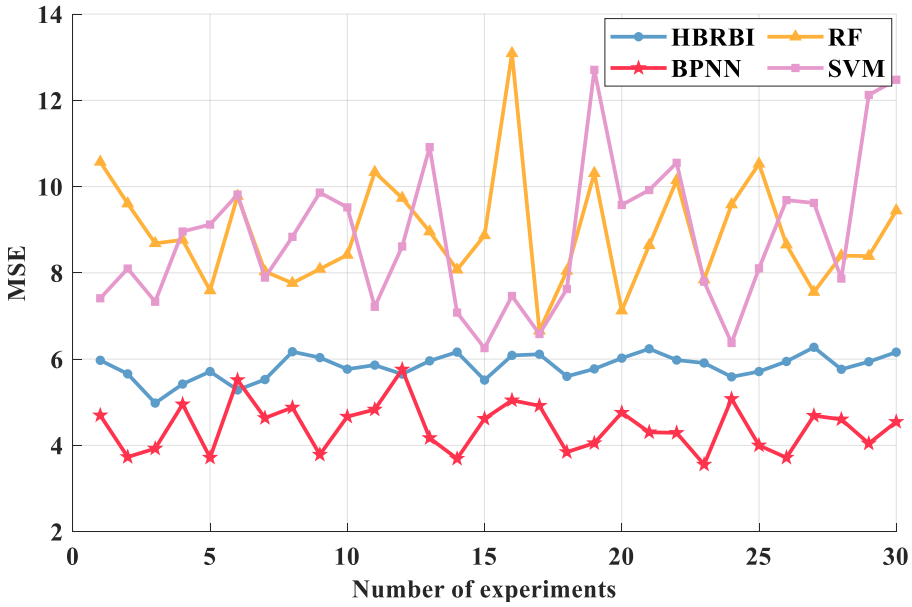


Figure 14. Prediction accuracy of HBRB1, BPNN, RF, and SVM.

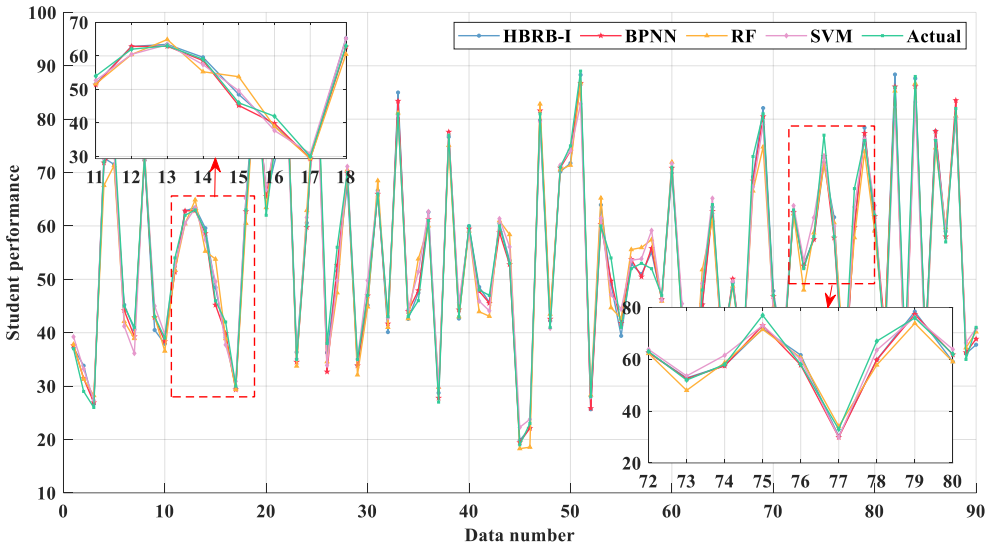


Figure 15. Prediction results of HBRB-I, BPNN, RF, and SVM.

In the third group of experiments, models with attributes grouped as $\{(x_1, x_2), (x_3, x_4)\}$, and $\{(x_1, x_3), (x_2, x_4)\}$ are constructed. In the former grouping, x_1 and x_2 have the highest correlation with the result. The mean MSE of the prediction results is 7.8109, which is much larger than that of the HBRB-I model. In the latter grouping, there is a high correlation between the same group of attributes. The average MSE of the prediction results is 6.2763, which is also larger than the HBRB-I model. The experimental results show that the correlation between attributes affects the valid information in the sub-BRB output, which in turn affects the prediction results of HBRB.

In conclusion, the effectiveness of methods to ensure model interpretability is verified in the comparison of various BRB models. Comparison with other algorithms through repeated experiments proves that the HBRB-I model performs well in terms of prediction

accuracy and interpretability, and is more stable than the rest of the models. The prediction results of taking different attribute groupings show that the grouping method of minimizing correlation ensures that the information is valid. Therefore, the reliability of the model can be proved.

Therefore, the key to the interpretability of HBRB-I can be summarized as follows: (1) Rules are used as the basis for inference, and a transparent inference engine is used to derive results. (2) The parameters in the model are given practical meanings, and the hierarchical structure reflects the cognitive process from one-sided to comprehensive. (3) The results of the model can be traced, and primary attributes have a strong influence on the result. (4) Expert knowledge is involved in the initial model construction process as priori knowledge to provide decision making for problems in specific domains such as education. (5) The optimization process revolves around expert knowledge to ensure that the rules do not deviate from expert judgment while improving accuracy.

6. Conclusions

The prediction results of student performance serve as a guide in teaching practice. Educators are able to identify in advance the risk of students failing academically so that they can change their teaching strategies in time. If the prediction results are reliable, the right teaching strategies can improve the quality of teaching. Conversely, unreliable results can lead to serious consequences. Therefore, the prediction model must be transparent and readily interpretable so that users can judge for themselves whether the prediction results are reasonable or not.

Within existing methods for predicting student performance, there are still problems such as poor interpretability, underutilized expert knowledge, and limited data processing. After clarifying these barriers, the use of the BRB that was used to construct the model is proposed to solve them. Therefore, a novel HBRB-I model with good interpretability is proposed in the paper. A hierarchical structure is constructed to simplify the rule base in response to the problem of excessive system complexity of BRB in education scenarios. Following a summary of the structure interpretability criterion for HBRB, the inference process of aggregating two attributes per sub-BRB is identified. It is also found that with only two attributes per group, the correlation between attributes in each group can be minimized using a simple method. For this reason, a new attribute grouping method based on Spearman's correlation coefficient is proposed. In this study, it is also explored how interpretability can be ensured in a system with a hierarchical structure, and it is proposed that the reasoning process of HBRB is a cognitive process from one-sided to comprehensive. In addition, three optimization strategies are proposed to address the fact that the randomness of the optimization algorithm will destroy the interpretability of the model. The purpose is to ensure that the optimized rules do not deviate from the expert judgment, and at the same time improve the accuracy of the prediction results. In the final experimental part, the HBRB-I model proposed in this paper is verified to be indeed effective from various aspects.

This study on interpretability presents certain limitations. First, expert knowledge is not always reliable. Both subjective perception and lack of experience affect the interpretability of BRB. Therefore, how to measure the credibility of expert knowledge needs to be investigated. Moreover, the rule base contains all the working states of the system, but the limited sample data may not match all the rules. The problem of parameters being over-optimized also needs to be considered in subsequent studies.

Author Contributions: Conceptualization, M.L.; methodology, M.L.; software, M.L.; validation, M.L. and H.C.; formal analysis, M.L. and G.Z.; investigation, W.H.; resources, G.Z.; data curation, H.C. and J.Q.; writing—original draft preparation, M.L.; writing—review and editing, W.H.; visualization, M.L.; supervision, G.Z.; project administration, G.Z. and W.H.; funding acquisition, G.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology Department, Heilongjiang Province, grant number GZ20220131.

Data Availability Statement: Data for this study were taken from https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression/data, accessed on 9 April 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wan, H.; Liu, K.; Yu, Q.; Gao, X. Pedagogical Intervention Practices: Improving Learning Engagement Based on Early Prediction. *IEEE Trans. Learn. Technol.* **2019**, *12*, 278–289. [CrossRef]
2. Yu, X.; Li, W.; Zhang, C.; Wang, J.; Zhao, Y.; Liu, F.; Pan, Q.; Liu, H.; Ding, J.; Chen, D. Time-Aware Multi-Behavior Graph Network Model for Complex Group Behavior Prediction. *Inf. Process. Manag.* **2024**, *61*, 103666. [CrossRef]
3. Mai, T.T.; Bezbradica, M.; Crane, M. Learning Behaviours Data in Programming Education: Community Analysis and Outcome Prediction with Cleaned Data. *Future Gener. Comput. Syst.* **2022**, *127*, 42–55. [CrossRef]
4. Xenos, M. Prediction and Assessment of Student Behaviour in Open and Distance Education in Computers Using Bayesian Networks. *Comput. Educ.* **2004**, *43*, 345–359. [CrossRef]
5. Qiu, F.; Zhang, G.; Sheng, X.; Jiang, L.; Zhu, L.; Xiang, Q.; Jiang, B.; Chen, P. Predicting Students' Performance in e-Learning Using Learning Process and Behaviour Data. *Sci. Rep.* **2022**, *12*, 453. [CrossRef]
6. Song, Q.; Luo, W. SFBKT: A Synthetically Forgetting Behavior Method for Knowledge Tracing. *Appl. Sci.* **2023**, *13*, 7704. [CrossRef]
7. Xiao, J.; Teng, H.; Wang, H.; Tan, J. Psychological Emotions-Based Online Learning Grade Prediction via BP Neural Network. *Front. Psychol.* **2022**, *13*, 981561. [CrossRef]
8. Imhof, C.; Comsa, I.-S.; Hlosta, M.; Parsaeifard, B.; Moser, I.; Bergamin, P. Prediction of Dilatory Behavior in eLearning: A Comparison of Multiple Machine Learning Models. *IEEE Trans. Learn. Technol.* **2023**, *16*, 648–663. [CrossRef]
9. Cui, C.; Zong, J.; Ma, Y.; Wang, X.; Guo, L.; Chen, M.; Yin, Y. Tri-Branch Convolutional Neural Networks for Top-k Focused Academic Performance Prediction. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 439–450. [CrossRef]
10. Mubarak, A.A.; Cao, H.; Zhang, W. Prediction of Students' Early Dropout Based on Their Interaction Logs in Online Learning Environment. *Interact. Learn. Environ.* **2022**, *30*, 1414–1433. [CrossRef]
11. Chen, C.-H.; Yang, S.J.H.; Weng, J.-X.; Ogata, H.; Su, C.-Y. Predicting At-Risk University Students Based on Their e-Book Reading Behaviours by Using Machine Learning Classifiers. *Australas. J. Educ. Technol.* **2021**, *37*, 130–144. [CrossRef]
12. Hao, J.; Gan, J.; Zhu, L. MOOC Performance Prediction and Personal Performance Improvement via Bayesian Network. *Educ. Inf. Technol.* **2022**, *27*, 7303–7326. [CrossRef]
13. Li, X.; Zhang, Y.; Cheng, H.; Li, M.; Yin, B. Student Achievement Prediction Using Deep Neural Network from Multi-Source Campus Data. *Complex Intell. Syst.* **2022**, *8*, 5143–5156. [CrossRef]
14. Shou, Z.; Chen, P.; Wen, H.; Liu, J.; Zhang, H. MOOC Dropout Prediction Based on Multidimensional Time-Series Data. *Math. Probl. Eng.* **2022**, *2022*, 2213292. [CrossRef]
15. Mumtaz, F.; Jehangiri, A.I.; Ishaq, W.; Ahmad, Z.; Alramli, O.I.; Ala'anzy, M.A.; Ghoniem, R.M. Quality of Interaction-Based Predictive Model for Support of Online Learning in Pandemic Situations. *Knowl. Inf. Syst.* **2024**, *66*, 1777–1805. [CrossRef]
16. Shou, Z.; Xie, M.; Mo, J.; Zhang, H. Predicting Student Performance in Online Learning: A Multidimensional Time-Series Data Analysis Approach. *Appl. Sci.* **2024**, *14*, 2522. [CrossRef]
17. Sohail, S.; Alvi, A.; Khanum, A. Interpretable and Adaptable Early Warning Learning Analytics Model. *CMC-Comput. Mat. Contin.* **2022**, *71*, 3211–3225. [CrossRef]
18. Azcona, D.; Hsiao, I.-H.; Smeaton, A.F. Detecting Students-at-Risk in Computer Programming Classes with Learning Analytics from Students' Digital Footprints. User Model. *User-Adapt. Interact.* **2019**, *29*, 759–788. [CrossRef]
19. Yang, J.B.; Liu, J.; Wang, J.; Sii, H.S.; Wang, H.W. Belief Rule-Base Inference Methodology Using the Evidential Reasoning Approach-RIMER. *IEEE Trans. Syst. Man Cybern. A* **2006**, *36*, 266–285. [CrossRef]
20. Zhou, Z.-J.; Hu, G.-Y.; Hu, C.-H.; Wen, C.-L.; Chang, L.-L. A Survey of Belief Rule-Base Expert System. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *51*, 4944–4958. [CrossRef]
21. Cao, Y.; Zhou, Z.; Hu, C.; He, W.; Tang, S. On the Interpretability of Belief Rule-Based Expert Systems. *IEEE Trans. Fuzzy Syst.* **2021**, *29*, 3489–3503. [CrossRef]
22. Gao, F.; Zhang, A.; Bi, W.; Ma, J. A Greedy Belief Rule Base Generation and Learning Method for Classification Problem. *Appl. Soft. Comput.* **2021**, *98*, 106856. [CrossRef]
23. Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.* **2019**, *51*, 1–42. [CrossRef]
24. Zhou, Z.; Cao, Y.; Hu, G.; Zhang, Y.; Tang, S.; Chen, Y. New Health-State Assessment Model Based on Belief Rule Base with Interpretability. *Sci. China Inf. Sci.* **2021**, *64*, 172214. [CrossRef]
25. Han, P.; He, W.; Cao, Y.; Li, Y.; Zhang, Y. Deep Belief Rule Based Photovoltaic Power Forecasting Method with Interpretability. *Sci Rep* **2022**, *12*, 14467. [CrossRef] [PubMed]
26. Xie, Y.; He, W.; Zhu, H.; Yang, R.; Mu, Q. A New Unmanned Aerial Vehicle Intrusion Detection Method Based on Belief Rule Base with Evidential Reasoning. *Heliyon* **2022**, *8*, e10481. [CrossRef]

27. Yin, X.; Xu, B.; Hu, L.; Li, H.; He, W. A New Interpretable Belief Rule Base Model with Step-Length Convergence Strategy for Aerospace Relay Health State Assessment. *Sci. Rep.* **2023**, *13*, 14066. [CrossRef]
28. Yang, L.-H.; Ren, T.-Y.; Ye, F.-F.; Nicholl, P.; Wang, Y.-M.; Lu, H. An Ensemble Extended Belief Rule Base Decision Model for Imbalanced Classification Problems. *Knowl.-Based Syst.* **2022**, *242*, 108410. [CrossRef]
29. Zhou, Z.-G.; Liu, F.; Jiao, L.-C.; Zhou, Z.-J.; Yang, J.-B.; Gong, M.-G.; Zhang, X.-P. A Bi-Level Belief Rule Based Decision Support System for Diagnosis of Lymph Node Metastasis in Gastric Cancer. *Knowl.-Based Syst.* **2013**, *54*, 128–136. [CrossRef]
30. Han, W.; Kang, X.; He, W.; Jiang, L.; Li, H.; Xu, B. A New Method for Disease Diagnosis Based on Hierarchical BRB with Power Set. *Heliyon* **2023**, *9*, e13619. [CrossRef]
31. Oreski, S.; Oreski, D.; Oreski, G. Hybrid System with Genetic Algorithm and Artificial Neural Networks and Its Application to Retail Credit Risk Assessment. *Expert Syst. Appl.* **2012**, *39*, 12605–12617. [CrossRef]
32. Arora, N.; Kaur, P.D. A Bolasso Based Consistent Feature Selection Enabled Random Forest Classification Algorithm: An Application to Credit Risk Assessment. *Appl. Soft Comput.* **2020**, *86*, 105936. [CrossRef]
33. Hu, G.; He, W.; Sun, C.; Zhu, H.; Li, K.; Jiang, L. Hierarchical Belief Rule-Based Model for Imbalanced Multi-Classification. *Expert Syst. Appl.* **2023**, *216*, 119451. [CrossRef]
34. Magdalena, L. Semantic Interpretability in Hierarchical Fuzzy Systems: Creating Semantically Decouplable Hierarchies. *Inf. Sci.* **2019**, *496*, 109–123. [CrossRef]
35. Razak, T.R.; Garibaldi, J.M.; Wagner, C.; Pourabdollah, A.; Soria, D. Toward a Framework for Capturing Interpretability of Hierarchical Fuzzy Systems-A Participatory Design Approach. *IEEE Trans. Fuzzy Syst.* **2021**, *29*, 1160–1172. [CrossRef]
36. Cao, Y.; Tang, S.; Yao, R.; Chang, L.; Yin, X. Interpretable Hierarchical Belief Rule Base Expert System for Complex System Modeling. *Measurement* **2024**, *226*, 114033. [CrossRef]
37. Ming, Z.; Zhou, Z.; Cao, Y.; Tang, S.; Chen, Y.; Han, X.; He, W. A New Interpretable Fault Diagnosis Method Based on Belief Rule Base and Probability Table. *Chin. J. Aeronaut.* **2023**, *36*, 184–201. [CrossRef]
38. Zhou, Z.; Feng, Z.; Hu, C.; Hu, G.; He, W.; Han, X. Aeronautical Relay Health State Assessment Model Based on Belief Rule Base with Attribute Reliability. *Knowl.-Based Syst.* **2020**, *197*, 105869. [CrossRef]
39. Han, P.; He, W.; Cao, Y.; Li, Y.; Mu, Q.; Wang, Y. Lithium-Ion Battery Health Assessment Method Based on Belief Rule Base with Interpretability. *Appl. Soft Comput.* **2023**, *138*, 110160. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.