A Study of Population Physical Health Rating Model and Its Correlation Based on Multilevel Constructs

Chao Liang †, Chen Zhang †

School of Safety Engineering, Heilongjiang University of Science and Technology, Harbin, China

† These authors also contributed equally to this work

Abstract. This paper conducts a detailed study on the correlation between residents' dietary and living habits and common chronic diseases. Firstly, the paper constructs and analyzes a multilevel evaluation model for the data in the annex; secondly, it uses the MRMR feature selection model to explore the relationship between age, gender and other factors and residents' living habits and dietary habits; and lastly, it puts forward a series of recommendations for the residents' physical health.

Keywords: Multilevel evaluation modeling, MRMR feature selection, population health, chronic diseases.

1. Introduction

Urban lifestyles have a significant impact on the health of the population, including increasing the risk of chronic diseases and affecting mental health. Although cities provide abundant medical resources and health education, unhealthy lifestyles and environmental stress remain major problems. This study aims to analyze the rationality of urban residents' dietary habits and their major problems; and explore whether living and dietary habits are related to factors such as age, gender, marital status, education level, and occupation. Through these analyses, this paper hopes to provide targeted recommendations for improving the health of urban residents.

2. Modeling principles

2.1. Urban lifestyle and health associations under multifactorial influences

A multilevel, multiconstructed evaluation model was used as an infrastructure to comprehensively analyze the eating habits of urban residents. The model uses three main levels of evaluation: individual, social, and environmental. Within each level, constructs are identified that may influence eating habits. For example, at the individual level, considerations include age, gender, and health status; at the social level, factors such as socio-cultural, family background, and education level are addressed; and at the environmental level, factors may include the food supply chain, food prices, and advertising strategies. These levels and elements are interconnected to form a comprehensive evaluation framework. The model utilizes a variety of data sources, including questionnaires, statistical databases, and academic literature, for rigorous analysis. The goal is to identify factors that have a key influence on the rationality or irrationality of eating habits and to assess their level of influence at each level [1].

2.2. Statistical analysis and data interpretation

The Maximum Relevance Minimum Redundancy (MRMR) algorithm provides a robust approach to feature selection, which is essential for analyzing associations between dietary habits and multiple variables such as age, gender, marital status, education level, and occupation. The algorithm isolates the most relevant and least redundant features in the dataset, thereby increasing the accuracy and reliability of the statistical analyses. A detailed implementation of the MRMR algorithm is provided in a subsequent section, demonstrating its effectiveness in identifying interrelated factors that influence dietary and lifestyle choices of urban residents [2].

3. Modeling

3.1. Dietary habits and health risk model based on multivariate statistical analysis

3.1.1 Data preprocessing

The study covered several key dimensions, including baseline conditions, smoking, alcohol consumption, diet, physical activity, and health status, before being synthesized. These data were conducted in compliance with the A3 Eight Guidelines for a Balanced Diet of the Chinese Residents Dietary Guidelines (2022) [3]. For data preprocessing, the presence of missing values or outliers may lead to statistical bias and model instability. To optimize the model performance, detailed data cleaning and noise reduction were performed. First, segmented cubic Elmit interpolation was used to fill in the missing values, reducing the model error and optimizing its stability. In addition, a polynomial regression method was applied to reduce noise in the data. The residual test and white noise sequence analysis demonstrated that the accuracy and reliability of the model were significantly improved after data cleaning. In summary, this series of preprocessing activities not only enriched the integrity of the data, but also improved the accuracy and credibility of the subsequent model analysis. This is of great significance for the accurate interpretation and application of the research results as shown in **Fig. 1**.

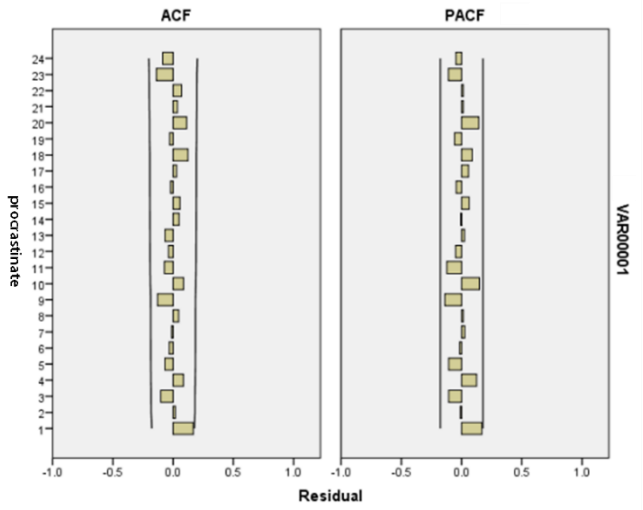


Fig. 1 ACF detection

3.1.2 Construction of a multilevel evaluation model for residents' diet

The construction of a multilevel evaluation model of the population's diet is a complex task that integrates multiple aspects of diet, including nutrition, health, food diversity, and environmental impact. In this paper, the construction of the evaluation model is based on the existing research data given in the title.

1) Define evaluation objectives: The application scenarios and purposes of the model are clarified, such as public nutrition policy guidance or food industry development.

2) Classify evaluation dimensions: Based on the evaluation objectives, identify the main dimensions such as nutritional health, environmental friendliness, and food diversity.

3) Setting evaluation indicators: Under the selected dimensions, specific evaluation indicators such as vitamin intake, mineral intake, protein intake, etc., are further specified.

4) Determination of weights: Assigning weights to different dimensions and indicators through expert assessment or statistical analysis.

5) Data processing: Perform pre-processing steps such as data cleaning and normalization to ensure data quality.

6) Model construction: Select appropriate mathematical methods, such as hierarchical analysis or linear weighting method, to construct the evaluation model.

7) Model validation: Validate the validity and reliability of the model through historical data or field tests.

8) Model optimization: Based on the validation results, the model is refined and optimized.

9) Model application: The optimized model is applied to real-world scenarios, such as for the development of public nutrition policy or food production.

This series of steps forms a comprehensive and systematic evaluation model construction process, which requires interdisciplinary cooperation and a large amount of data support. In particular, the different levels of the dietary pyramid can be refined into sub-dimensions of evaluation to increase the accuracy and application scope of the model.

In this paper, different dietary intakes are structured separately: this paper decomposes the dietary pyramid in Annex III as shown in **Fig. 2**:



Fig. 2 Dietary mapping

When constructing dietary evaluation models based on the Dietary Guidelines for Chinese Residents, multiple dimensions and indicators should be considered for different populations and needs. For example, the dietary evaluation model for the general population focuses on nutritional balance, food diversity, dietary fiber, and water intake. The dietary model for children focuses on protein, carbohydrate, fat, and vitamin and mineral intake. The elderly model focuses on calcium, vitamin D and vitamin B12, as well as protein and dietary fiber intake. The maternal model emphasizes folic acid, calcium, iron, and fruit and vegetable intake.

In constructing these models, assessment methods such as dietary nutrient content calculations, food diversity indices and nutritional balance assessment indices should be used in an integrated manner. The process needs to consider the actual situation and data sources to ensure the accuracy and practicability of the models. Ultimately, these models aim to scientifically guide the public to improve their dietary habits and provide a basis for relevant policy formulation.

Therefore, this paper establishes dietary standards for different population groups as shown in **Table 1**:

Table 1. Dietary standards for different population groups

Name	Nutrient	Intake	Notes
Adults	Protein	0.8-1.2 g/kg/day	Depends on weight
Fat	20-35% energy	Mostly good fats	
Carbs	45-65% energy	Mostly complex carbs	
Fiber	25-30 g/day	From veggies, fruits, whole grains	
Calcium	800-1000 mg/day		
Vitamin C	100-200 mg/day		
Children	Protein	1.0-2.0 g/kg/day	Depends on age, growth
Calcium	700-1000 mg/day	Needed more when growing	
Iron	7-10 mg/day	Depends on growth, gender	
Elderly	Protein	1.0-1.2 g/kg/day	To avoid muscle loss
Calcium	800-1200 mg/day	For bone health	
Vitamin B12	2.0-3.0 mcg/day	Absorbs less easily	
Pregnant Women	Folic Acid	400-600 mcg	For baby's neural health
Calcium	1000-1300 mg	For bone health	
Iron	27-30 mg/day	To avoid low iron	

Please note that the values in the above tables are for illustrative purposes and the actual dietary standards should be personalized according to specific factors such as age, gender, physiological conditions, and activity level. Meanwhile, the intake ranges mentioned above are only general reference values, and specific nutritional requirements may vary according to individual differences, physical conditions, and health goals. It is best to consult a professional doctor or dietitian for personalized advice on diet.

3.1.3 Determination of the weights of the indicators for the evaluation of the population's diet

The weights of the indicators for the evaluation of the population's diet are set to reflect the importance and degree of contribution of different indicators in the overall evaluation. In dietary evaluation models, there are usually several indicators used to measure different aspects of dietary quality, such as nutritional balance, food diversity, environmental friendliness, etc. These indicators are not equal in the overall evaluation. These indicators are not equal in the overall evaluation; some indicators are more important in terms of their impact on diet quality, while others have a relatively small impact.

By setting weights, different indicators can be considered and processed in a quantitative manner, and each indicator can be assigned a corresponding weight according to its importance. The benefits of doing so include: 1) highlighting important indicators to accurately reflect dietary quality; 2) reducing the impact of secondary indicators to stabilize the evaluation results; 3) enabling multi-dimensional considerations, such as nutrition versus environment; and 4) providing model flexibility to adapt to different needs and scenarios. The setting of weights provides flexibility in the model, allowing it to be adapted to specific needs or populations.

CRITIC (Criteria Importance Through Intercriteria Correlation) is a commonly used method for assigning weights to indicators used to evaluate residents' diets. The CRITIC weighting method is an objective approach. The method utilizes two indicators: contrast intensity and conflictiveness. Contrast intensity is measured using the standard deviation of the data. A higher standard deviation indicates greater fluctuation and, consequently, a higher weight. Conflictiveness is measured using

the correlation coefficient. A larger correlation coefficient value suggests less conflictiveness, resulting in a smaller weight. The CRITIC method determines the weights of evaluation indicators by analyzing their correlations in the overall evaluation process. The following steps outline the procedure for determining the weights of evaluation indicators for residential diets using the CRITIC method [4]:

In the construction of the population diet evaluation model, the key steps include: 1) defining the evaluation indicators, including nutrition, food types and other factors; 2) creating a judgment matrix based on expert judgment to quantify the relative importance of the indicators; and 3) mathematically calculating a vector of weights to represent the degree of importance of the indicators in the overall evaluation. The computational process of the CRITIC method mainly consists of the following steps:

Column normalization judgment matrix: ensure that the column sum is 1.

Calculate the positive correlation matrix: normalize the rows by adding them.

Calculate the negative correlation matrix: normalize after adding the columns.

Calculate correlation degree: multiply positive and negative correlation matrix.

Calculate the weight: normalize the correlation.

For the CRITIC method, when the standard deviation is certain, the smaller the conflict between the indicators, the smaller the weight; the larger the conflict, the larger the weight; in addition, when the degree of positive correlation between the two indicators is larger, (the correlation coefficient is closer to 1), the smaller the conflict, which suggests that there is a greater similarity in the information reflected in the evaluation of the program's strengths and weaknesses of the two indicators, and this paper's evaluation of the weights to be solved in **Fig. 3**:

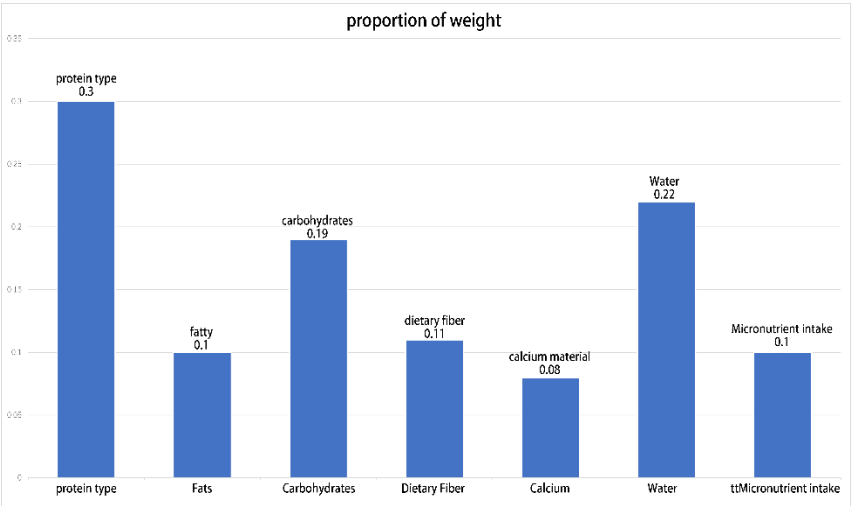


Fig. 3 Weight distribution

3.2. Modeling the association between socio-demographic variables and lifestyle

The MRMR algorithm (Minimum Redundancy Maximum Relevance) is a commonly used feature selection algorithm for selecting the most informative features in each dataset and eliminating redundant features. In this problem, this paper can use the MRMR algorithm to analyze the correlation between the living and eating habits of residents and factors such as age, gender, marital status, education, and occupation [5].

The following steps are based on the MRMR algorithm:

1) Data preparation: collect and organize data sets covering factors such as residents' living habits, dietary habits, age, gender, marital status, literacy, occupation, and so on. Ensure that the data have been cleaned and pre-processed to deal with missing values, outliers, etc.

2) Feature selection: divide the dataset into feature set and target variables. In this problem, living and eating habits are the features and age, gender, marital status, education, and occupation are the target variables.

3) MRMR algorithm calculation: the MRMR algorithm is applied to rank the features in the feature set and select the features that have the greatest correlation with the target variables and are not redundant with each other.

4) Correlation analysis: the top-ranked characteristics and target variables are analyzed for correlation, and correlation coefficients, chi-square tests and other methods can be used to measure the degree of association between them.

5) Statistical analysis: use appropriate statistical methods, such as analysis of variance (ANOVA), t-tests, etc., to verify whether the differences between the characteristics and the target variables are significant.

6) Interpretation of results: Based on the results of the analysis, a conclusion is drawn as to whether there is a correlation between living and eating habits and factors such as age, gender, marital status, literacy, and occupation. If correlated, the degree of influence of the characteristics on the target variables and the relationship between them can be further explored.

4. Model solving

4.1. Strategies for optimizing eating habits and health risks

Based on the establishment of the above weights, this paper can establish a resident diet evaluation model. The model can help assess the diet quality of a resident, and at the same time can guide the improvement of diet structure to improve diet quality.

Assume that there are N indicators for evaluating residents' diets in this paper, and the weight vectors of these indicators have been obtained W = [w1, w2, ..., wN], where wi denotes the weight of the ith indicator.

The steps are as follows:

1) Data collection: collect data on N indicators, which can include information such as population dietary surveys, frequency of food intake, and nutrient content.

2) Data standardization: The data collected are standardized to ensure that they are in the same scale and range. This can be done using methods such as normalization or standardization.

3) Calculate the resident diet score: For each resident's diet data, the following steps were followed to calculate the score:

(1) Multiply the data of each indicator with the corresponding weights to get the weighted score.

(2) Sum the weighted scores of all indicators to get the residents' diet score.

A higher diet score indicates a better-quality diet for the population.

4) Giving dietary advice: Based on the residents' dietary scores, corresponding dietary advice can be provided to help improve the dietary structure. For example, if the score of an indicator is low, it can be recommended to increase the intake of the relevant food or supplement the corresponding nutrients.

5) Model evaluation and optimization: The established dietary evaluation model needs to be continuously evaluated and optimized. The accuracy and usefulness of the model can be improved by collecting more data, introducing more indicators, or adjusting the weights.

In this paper, the following evaluation model is based on the distance between superior and inferior solutions (TOPSIS) algorithm [6].

In this paper, the original data is first normalized as well as standardized, and since all the data in this paper are normalized, only standardization is required.

The methods to standardize the data are Min-max standardization, z-score standardization, and Decimal scaling. In this problem, z-score standardization is used in this paper. The mathematical expectation and standard deviation of each variable were first found, then the data were standardized using the formula $Z_{ij} = (X_{ij} - X_i)/S_i$ Normalize the data and finally swap the plus and minus signs before the inverse indicator.

In this paper, the urban data indicator matrix Q, which is obtained after the normalization process of the above steps, contains all very large data, from which the ideal optimal solution and the worst solution can be taken out.

$$Q = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix} \tag{1}$$

The optimal solution and the worst solution are:

$$q_{min} = [\ (q_{12} \ \cdots \ q_{m1}), (q_{1n} \ \cdots \ q_{mn})] \tag{2}$$

The distance between the ith component and the optimal solution and the worst solution can then be calculated.

$$d_i^+ = \sqrt{\sum_{j=1}^{m} [w_j (q_{ij}{}^+ - q_{ij})]^2} \tag{3}$$

In turn, the score of the ith component can be calculated:

$$d_i{}^- = \sqrt{\sum_{j=1}^{m} [w_j (q_{ij}{}^- - q_{ij})]^2} \tag{4}$$

0≤S,≤1, When d, is smaller, it means that the distance from the optimal solution is smaller, and the score is higher; When *d* is smaller, it means that the distance from the worst solution is smaller, and the score is lower. In this way, this paper can analyze the reasonableness of different residents' diets under different categories according to the height of the scores, and this paper visualizes the relevant scores according to the model in **Fig. 4**:

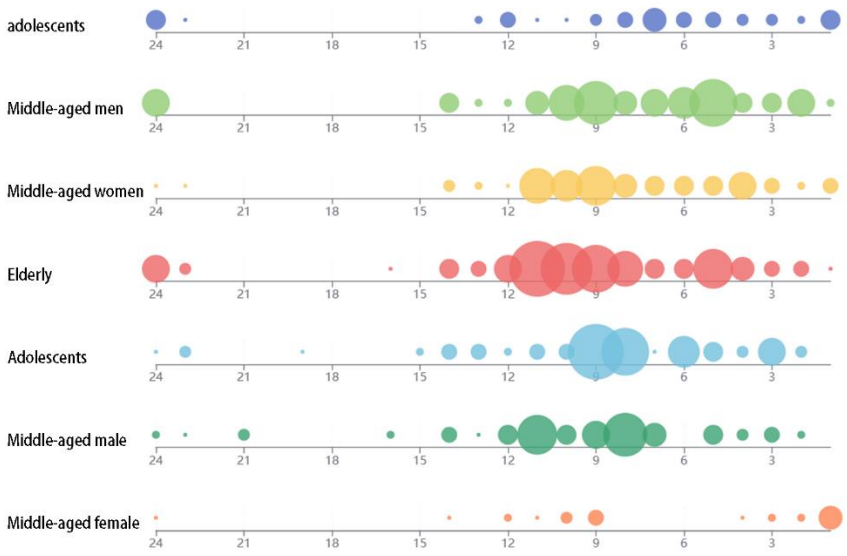


Fig. 4 Visualization Score

In the visualization results of the above figure, the following conclusions can be drawn in this paper:

1. Summarizing the above results, this paper concludes that for the intake of protein, carbohydrates, and minerals, etc., young people obviously score the highest and have a more reasonable intake; followed by middle-aged people, and then the elderly.

2. The sampled population's intake of dietary fiber, calcium, water, and micronutrients was reasonable, while their intake of protein and fat did not meet the criteria for a healthy diet.

3. Some of the people have a low intake in terms of variety and quantity, which is not up to the standard.

4.2. Validation of the validity of the model and sensitivity analysis

In this paper, mRMR Maximum Correlation Minimum Redundancy is performed on the available data and the process is shown below:

Based on the characteristic variable Xi and the dependent variable value Yi in the data, their probability density function (corresponding to continuous variables) is $P^{(x)}, P_x^{(y)} P^{(x,y)}$, Then the mutual information is:

$$I_{(x;y)} = \int\int_{p(x,y)} log \frac{p(x,y)}{p(x)P(y)} dx\, dy \tag{5}$$

The correlation between the feature set S and the function value c of the dependent variable is defined by the average of all mutual information values between the individual features fi and the function value c as follows:

$$D(S,c) = \frac{1}{|S|} \sum_{f_i \epsilon S} I\ (f_i;c) \tag{6}$$

The redundancy of all features in a set S is the average of all mutual information values between feature f_i and feature f_j force:

$$R(S) = \frac{1}{|S|^2} \sum_{f_i, f_j \epsilon S} I\ (f_i;f_j) \tag{7}$$

The MRMR criterion is a combination of the two measures given above, i.e., relevance and redundancy are defined below:

$$mRMR = \underset{s}{Max} \left[\frac{1}{|S|} \sum_{f_i \in S} I\ (f_i;C) - \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I\ (f;f_j) \right] \tag{8}$$

The optimization problem can be reduced to a maximum value using the ISM method:

$$\underset{x_j \in X - S_{m-1}}{Max} \left[I(x_j;c) - \frac{1}{m-1} \sum_{x_j \in S_{m-1}}^{1} I(x_j;x_i) \right] \tag{9}$$

The results of the final test are shown in **Fig. 5**:

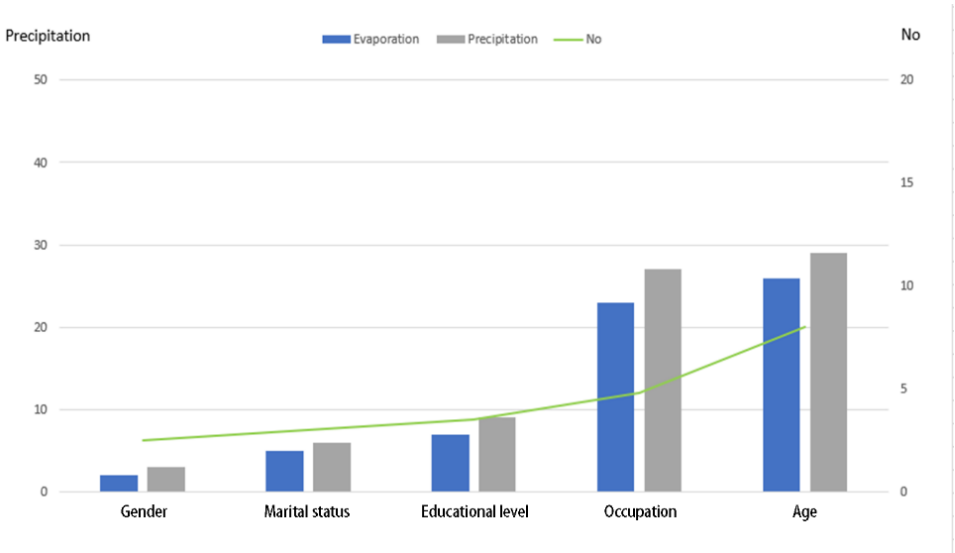


Fig. 5 Inspection results

Based on the above results, the paper can be concluded that:

1) Age: Age has a strong correlation, indicating that the older you are the less likely you are to participate in sports, but you need good habits and diet to maintain a healthier body.

2) Occupation: the strong correlation of occupation may indicate that certain types of occupations are more inclined to healthy eating.

3) Marital status: the correlation between marital status and all variables is weak, indicating that marital status has less influence on these living and eating habits.

4) Literacy: Its weak correlation suggests that literacy may have some influence on these factors.

5. Summary

The model proposed in this paper has multiple advantages, especially in the combination of subjective and objective weights using the CRITIC weighting method, which in turn improves the accuracy of the comprehensive evaluation. Through the integrated application of MATLAB and SPSSPRO software, the model realizes easy, intuitive, and fast data processing. However, the model suffers from certain errors and poor matching effects in data processing and modeling of attachment data. To improve these shortcomings, the research team optimized the feature combinations and algorithms and made continuous improvements based on medical experts' recommendations, especially in the field of chronic disease prevention and treatment. It is worth mentioning that the model is also highly generalizable and can be applied not only to the analysis of healthy diets, but also extended to a variety of fields such as corporate advantage and regional influence.

References

[1] Liu Fei. Promoting healthy and civilized lifestyles to raise the level of public health literacy[N]. Anyang Daily,2023-09-04(002).

[2] SUN Lin,XU Feng,LI Shuo et al. Multi-marker feature selection based on ReliefF and maximum correlation minimum redundancy[J/OL]. Journal of Henan Normal University (Natural Science Edition):1-9[2023-09-24]. http://kns.cnki.net/kcms/detail/41.1109.n.20230625.1558.002.html.

[3] . Dietary Guidelines for Chinese Residents (2022) released in Beijing[J]. Journal of Nutrition,2022,44(06): 521-522.DOI: 10.13325/j.cnki.acta.nutr.sin.2022.06.019.

[4] Yanyan Wang,Wenjie Zhao,Lijun Qi. Precast supplier partner selection based on CRITIC and TOPSIS[J]. Project Management Technology,2023,21(09):86-91.

[5] BUGATA P, DROTAR P. On some aspects of minimum redundancy maximum relevance feature selection[J]. Science China (Information Sciences),2020,63(01):89-103.

[6] Li Xiaolu,Chen Shaoqing. Comprehensive evaluation and spatial analysis of new urbanization development level in Jiangsu Province based on entropy weight TOPSIS[J]. Popular Standardization,2023(18):93-96.