J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 5615--5631

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Multidimensional Regression Analysis and Dynamic Adjustment Modeling of Students' Physical Fitness Data in Physical Education Teaching in Colleges and Universities

Zhirong Zhao[1,✉]

[1] *Physical Education College, Luoyang Normal University, Luoyang, Henan, 471934, China*

ABSTRACT

College students' physical fitness is an important part of national health, and analyzing physical fitness data in college physical education teaching helps to dig out the factors affecting students' physical fitness and adjust the teaching plan in time. The article reviews some basic regression tools and selects variables such as BMI dietary habits for logistic regression analysis to analyze the factors affecting students' physical fitness. The similarity, uncertainty and dissimilarity between students and their friends are calculated by Top-N recommendation set algorithm, and the physical education teaching program is dynamically adjusted with the new SFD recommendation algorithm. Finally, values were assigned to different movement banks and risk factors, and the experts' agreement with the new adjusted program was examined. The intensity of physical activity had the greatest relationship with passing or failing physical fitness among all factors (regression coefficient = 0.927, $p<0.05$), while balanced meat and vegetarian meals helped to enhance students' physical fitness. Most of the professional physical fitness coaches agreed with the teaching adjustment strategies given by this methodology and the level of agreement was high (>70%), reflecting the rationality and feasibility of this study.

Keywords: multidimensional regression analysis, logistic regression, instructional adjustment, physical fitness data, SFD recommendation algorithm

1. Introduction

At present, the physical education work of Chinese colleges and universities has yet to be improved, in order to better improve the teaching level of colleges and universities in a comprehensive manner, it is imperative to carry out in-depth physical training in physical education teaching [1-2]. Basic physical fitness training is a key link for college students to enhance their physical fitness and improve their athletic performance [3]. Basic physical training in physical education in colleges and universities is of

✉ Corresponding author.
E-mail address: luoyangzzr@163.com (Z. Zhao).
Received 12 January 2024; Revised 05 March 2024; Accepted 25 December 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-313

great significance for improving the physical fitness and athletic performance of college students as well as cultivating the spirit of teamwork [4-5]. Colleges and universities in the process of teaching physical education courses, physical training content into it, in this time requires physical education teachers in the choice of methods, not only to combine the daily teaching experience, but also need to be based on the performance of the students, the development of effective training strategies, so that the students can make a deep understanding of physical training, and ultimately obtain the ideal training effect [6-8]. By following the principles of safety, progressivity, pertinence and comprehensiveness, coaches can analyze students' physical fitness data and then develop personalized training plans, combine college students' interests with training, focus on the diversity of training methods, and conduct regular physical fitness tests [9-12]. In this way, the level of basic physical fitness training in sports teaching in Chinese colleges and universities will be significantly improved. However, in the actual teaching process, it is not difficult to find that many teachers still use traditional teaching methods, and do not pay much attention to students' physical training [13-14]. In order to promote the healthy development of sports in Chinese colleges and universities, it is necessary to strengthen the physical training of students, so that students can have a better physical quality to join the fierce social competition [15-17].

This paper briefly introduces the basic principles of the multiple linear regression algorithm and various commonly used Logistic regression models. 800 students' physical fitness data were extracted from a university, and the physical measurement data were categorized according to the different levels of BMI, physical exercise intensity, and dietary habits. Logistic single-factor regression analysis and multiple regression analysis were successively conducted to test the significance factors leading to college students' physical fitness failure. The dynamic adjustment model of physical education was designed from the perspective of personalized recommendation of physical exercise program. The similarity, uncertainty and dissimilarity between students were calculated from the physical test data, and the recommendation of physical education program was completed according to the Top-N recommendation algorithm with reference to the performance of students' friends. Based on different movements and risk factors, we construct a movement library to increase the accuracy of recommendation. Finally, professionals are invited to evaluate the effectiveness of the recommendation program.

2. Multidimensional regression analysis of student fitness data

2.1 Multiple linear regression algorithm

Multiple linear regression prediction method is a more general and intuitive method among many prediction methods, and the key to prediction using multiple linear regression prediction method is to determine the multiple linear regression equation [18]. Assuming that Y is the dependent variable and $X_j(j=1,2,\cdots,m)$ is the independent variable, the multiple linear regression prediction equation can be expressed as:

$$Y_t = b_0 + b_1x_1 + b_2x_2 + \cdots + b_mx_m \tag{1}$$

where $b_0, b_1, b_2, \cdots, b_m$ is the coefficient to be determined. It can be seen that in the multiple linear regression equation, once the coefficient to be determined, the equation is also determined.

In order to establish multiple linear regression prediction equations, it is necessary to grasp the actual data of the dependent variable and the independent variable in a certain period, if you know the actual data of the dependent variable in period n and the actual data of the independent variable in period $y(y_1, y_2, \cdots, y_n)$ and n:

$$\begin{cases} x_1(x_{11}, x_{21}, \cdots, x_{n1}) \\ x_2(x_{12}, x_{22}, \cdots, x_{n2}) \\ \vdots \\ x_m(x_{1m}, x_{2m}, \cdots, x_{nm}) \end{cases} \tag{2}$$

Bringing the actual data into equation (1) yields the following system of simultaneous equations:

$$\begin{cases} y_1 = b_0 + b_1 x_{11} + b_2 x_{12} + \cdots + b_m x_{1m} \\ y_2 = b_0 + b_1 x_{21} + b_2 x_{22} + \cdots + b_m x_{2m} \\ \vdots \\ y_n = b_0 + b_1 x_{n1} + b_2 x_{n2} + \cdots + b_m x_{nm} \end{cases} \tag{3}$$

$$Y = XB \tag{4}$$

Where, Y is the matrix of actual values of the dependent variable, X is the matrix of actual values of the independent variable, and B is the matrix of coefficients to be determined in the following equations:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} \tag{5}$$

$$B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \tag{6}$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ 1 & x_{31} & x_{32} & \cdots & x_{3m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{7}$$

When multiple linear regression modeling, the coefficient to be determined, $b_0, b_1, b_2, \ldots, b_m$, is usually obtained by using the least square method to build a system of simultaneous equations and then solving the system of simultaneous equations. The system of simultaneous equations obtained by the least square method is as follows:

$$\begin{cases} \sum y = nb_0 + b_1 \sum x_1 + \ldots + b_m \sum x_m \\ \sum x_1 y = b_0 \sum x_1 + b_1 \sum x_1^2 + \ldots + b_m \sum x_1 x_m \\ \sum x_2 y = b_0 \sum x_2 + b_1 \sum x_1 x_2 + \ldots + b_m \sum x_2 x_m \\ \vdots \\ \sum x_m y = b_0 \sum x_m + b_1 \sum x_1 x_m + \ldots + b_m \sum x_m^2 \end{cases} \tag{8}$$

The above system of equations (8) can be expressed in matrix form as:

$$X^T Y = X^T XB \tag{9}$$

Multiplying both sides of Eq. (9) by $(X^T X)^{-1}$ simultaneously yields the following matrix expression:

$$B = (X^T X)^{-1} X^T Y \tag{10}$$

The result of the matrix obtained from equation (10) is the value of the coefficient to be determined $b_0, b_1, b_2, \ldots, b_m$, After the coefficient to be determined, the prediction equation (3) can be determined.

2.2 *Logistic regression analysis*

In order to explore what relationship exists between the dependent variable and the independent variable, people usually use the method of regression analysis to study it, and according to the different characteristics of the dependent variable, independent variable and the relationship between them, the regression analysis can be divided into linear regression, curvilinear regression and non-linear regression and other types. When the relationship between the independent variable and the dependent variable is linear, the regression analysis is linear regression. When the relationship between the independent variable and the dependent variable is nonlinear, the regression analysis is nonlinear regression. Logistic regression is a kind of non-linear regression, Logistic regression is a kind of regression analysis method that analyzes the dependent variable as a fixed class variable. According to the different number of independent variables, Logistic regression can also be divided into univariate Logistic regression, multivariate Logistic regression. According to the different classification of the dependent variable, Logistic regression can also be divided into binary Logistic regression, multi-category Logistic regression. Among them, the simplest and most common is binary one-dimensional logistic regression [19].

2.2.1. Dichotomous one-dimensional logistic regression models. Dichotomous one-way Logistic regression is a regression analysis where the dependent variable is a dichotomous variable and there is only one independent variable.

Let the dependent variable be y, when $y = 0$ indicates that the event did not occur and $y = 1$ indicates that the event occurred. The independent variable is x, which represents an influencing factor. Recall that the conditional probability of the event occurring is $P(y = 1 | x) = P$. The regression model for the one-dimensional logistic is:

$$P = \frac{\exp(\alpha + \beta \mathrm{x})}{1 + \exp(\alpha + \beta \mathrm{x})} \tag{11}$$

$$1 - P = 1 - \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} = \frac{1}{1 + \exp(\alpha + \beta x)} \tag{12}$$

The occurrence ratio of events, referred to as Odds, is the ratio of the probability of an event occurring to the probability of it not occurring $P / 1 - P$, i.e., the ratio of the probability of an event occurring to the probability of it not occurring. Odds has no upper bound and is necessarily positive, and a logarithmic transformation of Odds yields a one-element Logistic linear regression model:

$$\mathrm{Logit}(P) = \ln\left(\frac{P}{1 - P}\right) = \alpha + \beta x \tag{13}$$

That is, instead of directly describing the relationship between the dependent variable traffic misbehavior and the independent variable influencing factors, the Logistic regression model is described in the profile by describing the relationship between the probability of traffic misbehavior occurring and the influencing factors.

The ratio of the two Odds is known as the dominance ratio, abbreviated as OR, and $OR \approx e^{\beta}$. Approximates that the probability of an event occurring increases or decreases by (OR-1)% for each unit increase in the independent variable x.

2.2.2. Binary multivariate logistic regression models. Dichotomous multiple logistic regression is a regression analysis where the dependent variable is a dichotomous variable and there are multiple independent variables [20]. Let the dependent variable be y, when $y=0$ means that the event did not occur, and $y=1$ means that the event occurred. The m independent variables affecting y are $x_1, x_2, \dots, x_m$. Denote the conditional probability of the event occurring as $P(y=1 \mid x_1, x_2, \dots, x_m) = P$, and the binary multivariate logistic regression model is:

$$P = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)} \tag{14}$$

$$\begin{aligned} 1 - P &= 1 - \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)} \\ &= \frac{1}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)} \end{aligned} \tag{15}$$

Then the linear regression model of multivariate logistic is:

$$\text{Logit}(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m \tag{16}$$

At this point $OR_i = e^{\beta_i}$ approximates that the likelihood of an event occurring increases or decreases by (OR-1)% for each unit increase in x_i i.e., the i rd influencing factor, holding other influencing factors (i.e., the independent variables) constant.

2.2.3. Multiclassified multivariate logistic regression models. Multicategorical multivariate Logistic regression is a regression analysis in which the dependent variable is three and more than three categorical variables and the independent variables are more than one [21]. As an example, if the dependent variable has three categorical categories, assume that the dependent variable is y. y=0 indicates category A, which is set as the reference category. $y=1$ denotes class B. $y=2$ denotes class C. Then the logistic regression model when the dependent variable is $y=i$ is:

$$\text{Logit}P_{i/_0} = \ln\left[\frac{P(y=i \mid x)}{P(y=0 \mid x)}\right] = \beta_{i0} + \beta_{i1} x_1 + \cdots + \beta_{im} x_m \quad (i=1,2) \tag{17}$$

Two sets of non-zero regression coefficients are thus obtained. If the coefficient on one of the independent variables x is significantly positive, the probability of the dependent variable being in the current category is greater than the probability of it being in the reference category, and vice versa, holding other independent variables constant.

2.2.4. Ordered Logistic Regression Models. An ordered Logistic regression model is a regression model in which the dependent variable is three or more categorical variables and there are hierarchical categories between the dependent variables. Let the dependent variable have j hierarchical category and assigned a value of $y=1, y=2, \dots, y=j-1, y=j$, and the relationship between each y-value is $(y=1) < (y=2) < \cdots < (y=j)$. There are $j-1$ cut-offs separating neighboring categories: if $y' \le \mu_1$, then $y=1$; if $\mu_1 < y' \le \mu_2$, then $y=2$. If $\mu_j - 1 < y'$, then $y=j$. μ is the cut-off point separating the categories, and $\mu_1 < \mu_2 < \cdots < \mu_j$. The probability estimation model for categorization is constructed using an ordered logistic regression model, and the logistic regression model corresponds to a regression model that contains $j-1$ Logit functions:

$$\ln\left(\frac{p_1}{1-p_1}\right)=\beta_{01}-\sum_{k=1}^{k}\beta_k x_k \tag{18}$$

$$\ln\left(\frac{p_1+p_2}{1-p_1-p_2}\right)=\beta_{02}-\sum_{k=1}^{k}\beta_k x_k \tag{19}$$

$$\ln\left(\frac{p_1+p_2+\cdots+p_{j-2}}{1-p_1-p_2-\cdots-p_{j-2}}\right)=\beta_{0j-2}-\sum_{k=1}^{k}\beta_k x_k \tag{20}$$

$$\ln\left(\frac{p_1+p_2+\cdots+p_{j-2}+p_{j-1}}{p_j}\right)=\beta_{0j-1}-\sum_{k=1}^{k}\beta_k x_k \tag{21}$$

Where: $p_1, p_2, \ldots, p_j$ is the probability of occurrence of different classifications, the base level used for comparison is category j, and the sum of the probabilities of the j categories is 1, i.e., $p_1+p_2+\cdots+p_{j-1}+p_j=1; x_k(k=1,2,\ldots,k)$ denotes the k th independent variable. $\beta_{0j}(j=1,2,\ldots,j)$ denotes the intercept term of the regression. $\beta_k(k=1,2,\ldots,k)$ denotes the regression coefficient.

In a logistic regression model, each logit function corresponds to a β_{oj} estimate and each $\beta_o j$ estimate varies, but each x_k corresponds to an identical β_k regression coefficient across all logit functions, i.e., the regression lines for each ordered logit are parallel.

Similar to the dichotomous dependent variable, the cumulative probability for each classification can be calculated by the following equation:

$$P(y\le j\mid x)=P(y'\le \mu_j\mid x)=\frac{e^{\beta_{0j}-\Sigma_{k=1}^{k}\beta_k x_k}}{1+e^{\beta_{0j}-\Sigma_{k=1}^{k}\beta_k x_k}} \tag{22}$$

From this, the probability that the dependent variable is in a certain classification can be obtained and is calculated as shown in Equation (23):

$$\begin{gathered}P(y=1)=P(y\le 1)\\ P(y=2)=P(y\le 2)-P(y\le 1)\\ \vdots\\ P(y=j)=1-P(y\le j-1)\end{gathered} \tag{23}$$

And the sum of the probability values for each category is 1, i.e., $P(y=1)+P(y=2)+\cdots+P(y=j)=1$.

2.3 Data processing and regression statistical analysis

2.3.1. Basic information on survey respondents. 800 students were randomly selected from a university as survey respondents, and the basic status distribution of grade and gender of this survey respondents are shown in Table 1 and Table 2.As can be seen from the table, the distribution of students' grades is almost the same, and there is no statistically significant difference by the chi-square test, with the chi-square value=0.512, P=0.769 (P>0.05). The gender ratio is also close, with a chi-square value = 0.015, P = 0.9 (P>0.05), allowing for comparison.

Table 1. The grade distribution of student physical ability

	Freshman	Sophomore	Junior	Sum.
Qualified physical ability	133	132	135	400
Percentage %	33.25%	33%	33.75%	10%
Unqualified physical ability	129	140	131	400

Percentage %	32.25%	35%	32.75%	10%

Table 2. The gender distribution of student physical ability

	Female	Male	Sum.
Qualified physical ability	189	211	400
Percentage %	47.25%	52.75%	10%
Unqualified physical ability	205	195	400
Percentage %	51.25%	48.75%	10%

2.3.2. Definition of Variables and Code Assignment. The specific codes and assigned values are shown in Table 3. Among them, BMI is the body mass index of the survey respondents, body mass index (BMI) = weight (kg) ÷ height2 (m), and the normal range is 18.5 $\leqslant$ BMI < 24. Intensity of physical activity is the number of times of physical activity in a week multiplied by the length of the exercise, calculated in minutes. Dietary habits are the meat and vegetable preferences of the usual meals. Electronic product use habit is the length of time spent watching videos and playing games on the Internet on computers and cell phones every day. Self-perception factor is whether or not one understands one's physical fitness status and the significance of physical fitness test. Psychological tendency factor refers to the love of physical education class and whether to actively participate in school-organized activities. External factors refer to whether the teaching level of physical education teachers and the equipment of sports grounds have an impact on their physical fitness development.

Table 3. Name and code evaluation of the variable in the physical ability survey

Code	Name	Evaluation			
B1	BMI	1=normal		2=abnormal	
B2	Physical exercise intensity	1= >180mins	2= 90-180mins	3= 20-90mins	4= <20mins
B3	Eating habits	1= moderate meat and vegetable eating	2= mainly meat	3= mainly vegetable	
B4	Use of electronic products	1= 1 hour	2= 2 hours	3= 3 hours	4= 4 hours
B5	Self-cognitive factor	1= yes		2= no	
B6	Psychological tendency	1= yes		2= no	
B7	External factor	1= yes		2= no	

2.3.3. Statistical description of the sample of college students' physical fitness influencing factors. According to the questionnaire, the statistical description of the survey data of BMI index, physical exercise intensity, dietary habits, electronic product use habits, self-perception factors, psychological tendency factors, and external factors of college students' physical fitness unqualified group and qualified group is given by Table 4. It was initially found that among the unqualified students, the distribution of the group with normal and abnormal BMI index was uneven (64%:36%), and the distribution of physical exercise intensity was closer to grades 1 and 2. The above phenomenon implies that there may be a significant difference between students with different physical fitness statuses in each of the above mentioned indexes.

Table 4. Statistical description of physical ability factors

Name	Value	Qualified	Percentage%	Unqualified	Percentage%
B1	1	189	47.25%	256	64%
	2	211	52.75%	144	36%

B2	1	8	2%	129	32.25%
	2	59	14.75%	184	46%
	3	171	42.75%	48	12%
	4	162	40.5%	39	9.75%
B3	1	70	17.5%	107	26.75%
	2	255	63.75%	96	24%
	3	75	18.75%	197	49.25%
B4	1	10	2.5%	73	18.25%
	2	114	28.5%	196	49%
	3	109	27.25%	94	23.5%
	4	167	41.75%	37	9.25%
B5	1	261	65.25%	244	61%
	2	139	34.75%	156	39%
B6	1	206	51.5%	296	74%
	2	194	48.5%	104	26%
B7	1	210	52.5%	242	60.5%
	2	190	47.5%	158	39.5%

2.3.4. Binary Logistic Regression Analysis. Logistic regression one-way analysis was conducted with whether college students failed in physical fitness as the dependent variable, and BMI, physical exercise intensity, dietary habits, electronic product use habits, self-perception factors, psychological tendency factors, and external factors as the independent variables, and the results of the analysis were obtained as in Table 5 at the P=0.05 significance level.

The data from the binary logistic regression analysis of SPSS software showed that self-perception (P=0.165>0.05) was not a risk factor, and that in both groups of students investigated for physical fitness failing and passing, there were almost the same proportion of students with deficiencies in this aspect of cognition, and the role of the influence on the students' physical fitness results was not significant. The risk factors (P<0.05) screened by the study to induce college students to fail in physical fitness are: BMI, intensity of physical exercise, dietary habits, habits of using electronic devices, psychological predisposition factors, and external factors.

Table 5. Analysis result

	B	S.E.	Wals	df	P	OR
B1	0.692	0.056	29.006	1	0	2.001
Cons_	-1.031	0.259	26.056	1	0	0.380
B2	1.507	0.025	262.047	1	0	4.723
Cons_	-4.047	0.273	237.998	1	0	0.018
B3	-0.411	0.079	19.450	1	0	0.715
Cons_	0.861	0.210	17.376	1	0	2.202
B4	1.030	0.150	176.401	1	0	2.967
Cons_	-2.888	0.255	161.413	1	0	0.039
B5	-0.164	0.142	1.949	1	0.165	0.840
Cons_	0.321	0.194	1.661	1	0.194	1.253
B6	0.911	0.197	49.838	1	0	2.599
Cons_	-1.229	0.183	44.903	1	0	0.324
B7	0.375	0.145	8.535	1	0.002	1.441
Cons_	-0.544	0.206	7.655	1	0.005	0.506

2.3.5. Logistic multiple regression analysis based on physical fitness data. The risk factors for inducing physical disqualification of college students screened out by the above logistic univariate regression analysis may have a direct relationship with the physical disqualification of college students, but the role of each factor is not independent, and the BMI index, physical exercise intensity, eating habits, electronic product use habits, psychological predisposition factors, and external factors are used as independent variables for further research by logistic multiple regression analysis, and the results are shown in Table 6.

As can be seen from Table 6, BIM index, physical exercise intensity, electronic product use habits, psychological tendency factors, external factors and other factors have a certain degree of correlation with whether college students' physical fitness is unqualified or not, while the balance of dietary meat and vegetarian food is a protective factor for whether college students in higher vocational education are unqualified or not (B = -0.361), probably because of the present information-based society, college students from many channels are Can be to pay attention to the balance of diet meat and vegetarian, diet meat and vegetarian balance in order to maintain the body's most basic nutritional needs in the long term.

The dependent variable is denoted by y, and the probability of college students failing physical fitness is denoted by p, then $y = Ln\left(\frac{p}{1-p}\right)$. The independent variables BMI index, physical exercise intensity, dietary habits, electronic product use habits, psychological predisposition factors, and external factors are denoted by x_1, x_2, x_3, x_4, x_5, and x_6, respectively, and the following regression equation is obtained as follows: $y = -8.952 + 0.927x_1 + 1.447x_2 - 0.361x_3 + 1.217x_4 + 0.744x_5 + 0.44x_6$. It can be seen that the physical exercise intensity has the greatest influence on inducing the college students' physical fitness failure, and the remaining factors are, in the order of daily behaviors, Internet time, BMI index, psychological predisposition factors, and external factors. The influence of physical exercise intensity has the greatest effect, and the remaining ones are, in order, the online time of daily behavior, BMI index, psychological tendency factors and external factors.

Table 6. The multivariate regression analysis of unqualified physical ability

Variable	B	S.E.	Wals	df	P	OR
B1	0.927	0.164	22.507	1	0	2.402
B2	1.447	0.081	199.504	1	0	4.371
B3	-0.361	0.059	11.874	1	0.002	0.682
B4	1.217	0.097	132.43	1	0	3.261
B6	0.744	0.242	12.796	1	0	2.008
B7	0.44	0.142	4.922	1	0.027	1.488
Cons_	-8.952	0.664	161.314	1	0	0.000

3. Dynamic Adjustment Model for Physical Education

3.1 Dynamic Adjustment Model of Physical Education Teaching Based on Recommendation Algorithm

In the process of physical education teaching in colleges and universities, timely adjustment of teaching strategies according to changes in physical fitness data can help to improve teaching effectiveness, help to meet the teaching needs of different students, and enhance students' motivation for physical exercise.

3.1.1. General design of the algorithm. The SFD recommendation algorithm for college students' physical test is an algorithm designed specifically based on college students' physical test big data, and the algorithm flow is shown in Fig. 1.The concept of set-pair analysis is introduced in this study to transform the traditional similarity. In the recommendation process, we first need to calculate the similarity (S), uncertainty (F) and dissimilarity (D) between two objects, and then calculate the set-pair recommendation degree rec(A, B) according to the theory of set-pair analysis, and finally filter the suitable recommendation objects and recommend them by set-pair recommendation degree.

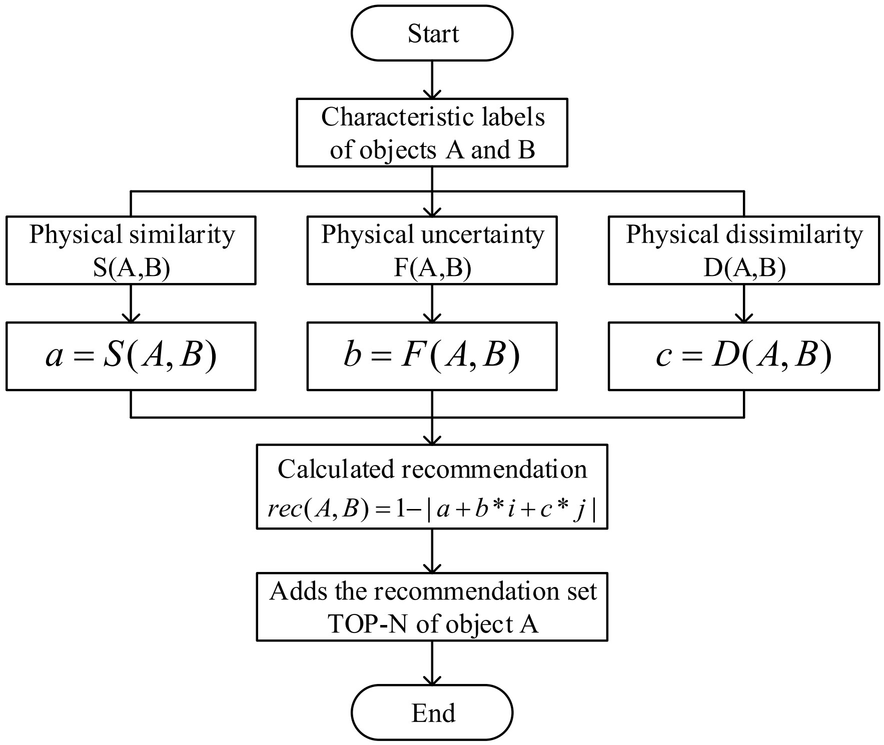

Fig. 1. Algorithm flow of physical test for college students

3.1.2. Data pre-processing. Before the design of the recommendation algorithm, in order to calculate the degree of recommendation of physical testing between the 2 objects, it is necessary to score college students' physical testing programs according to the "National Physical Fitness Standard", the students' physical testing program data percentile, and the classification of students' physical condition: explosive, endurance, flexibility, strength, different categories of scores by the students' physical testing program scores according to the calculation of different weights, and then through the threshold grading to determine the groups of strong, medium and weak students in each category, and finally the literal motor characteristics need to be fitted to give the definition: {"strong": 1, "medium": 2, "weak": 3}.

Assuming the existence of objects A and B, their motion characteristics are represented as feature vectors, i.e., the difference between the features of $A = < a_1, a_2, a_3, a_4 >, B = < b_1, b_2, b_3, b_4 >$, A and B can be represented by $|a_k - b_k|$, $k \in \{1,2,3,4\}$, and

$|a_k - b_k| = 0$, denotes A has similarity with B.

$|a_k - b_k| = 1$, denotes that A & B have uncertainty.

$|a_k - b_k| = 2$, denotes that A & B have degree of dissimilarity.

The study will be designed with similarity, dissimilarity, and uncertainty of body measurements in the context described above.

3.1.3. Set Pair Recommendation Degree. Unlike traditional similarity calculation recommendation algorithms, it is necessary to consider the degree of similarity between users and users, the degree of difference between users and users, and the uncertainty between users and users. Therefore, the concept of set pairs is introduced in the similarity calculation of recommendation algorithms and the following definition exists:

$$rel(A,B) = a + b*i + c*j \tag{24}$$

where $rel(A,B)$ denotes the correlation between A and B, $rel \in [-1,1]$, and the larger rel indicates higher similarity and vice versa lower dissimilarity. a represents the similarity between A and B, i.e., $a = S(A,B)$; b represents the uncertainty between A and B, i.e., $b = D(A,B)$; c represents the dissimilarity between A and B, i.e., $c = F(A,B)$, and satisfies $a+b+c=1$. iv9 is the marking of the uncertainty, j is the marking of the dissimilarity, and in the operation, i and j participate in the operation as coefficients at the same time, and it is stipulated that j takes the value of -1, and i takes the value in the interval [-1, 1] as the case may be. interval as appropriate. Set pair recommendation degree is deformed on the basis of correlation rel, comprehensively considering the similarity, difference, uncertainty three factors, to avoid the error of high recommendation degree due to too much similarity or dissimilarity, adopting the following definition:

$$rec(A,B) = 1\left|a + b*i + c*j\right| \tag{25}$$

$rec(A,B)$ denotes the pairwise recommendation between A and B, and the more $rec \in [0,1]$ and rec converge to 0, the less likely they are to be recommended. The more it converges to 1, the more likely it is to be recommended.

1) Physical similarity. Similarity is a numerical measure of the degree of similarity between two objects, which is an important reference index in the personalized recommendation system, and the traditional similarity is calculated based on the user's rating of the item and the recommendation using the relevant formula for similarity. The method of calculating the similarity of physical measurement in this study is different from the traditional method, which refers to obtaining students' physical measurement data, analyzing and processing the data in order to extract the sports characteristics describing the students, and calculating the similarity by comparing the characteristic values between different students through certain methods. Suppose there are objects A, B in the recommender system, then the following similarity exists for the recommended objects A, B:

$$S(A,B) = \frac{N(a_k = b_k)}{n} \tag{26}$$

2) Body Measurement Similarity. Where $S(A,B)$ is the similarity of body measurement between object A and B, n is the sum of the kinds of characteristic attributes, and a_k and b_k represent the characteristic values of the 2 objects respectively. Through certain rules, students' performance is divided into different levels, and when 2 students' physical performance levels are equal, they are considered to have certain similarity, i.e., there exists $N(a_k = b_k)$ is the number of features with the same eigenvalues of the 2 objects.

The degree of dissimilarity is a numerical measure that describes the degree of difference between 2 objects, in the problem of recommendation based on physical fitness test scores, it is often unreasonable to recommend only based on the degree of similarity, if the recommended two sides only exist similarity, it represents that the two sides can learn very little from each other, so the two sides need to exist a certain degree of dissimilarity. Only the existence of dissimilarity between the two sides, there is the possibility of learning from each other. Let there are 2 objects A and B with dissimilarity in the recommender system, then the following dissimilarity exists in the recommended objects A and B:

$$D(A,B) = \frac{\sum_{k=1}^{n} |a_k - b_k - 1| - N(a_k = b_k)}{n} \tag{27}$$

where $D(A,B)$ is the degree of dissimilarity between objects A and B, n is the sum of the types of feature attributes, and a_k and b_k denote the feature values of the 2 objects, respectively. $\sum_{k=1}^{n}|a_k - b_k - 1| - N(a_k = b_k)$ is the number of dissimilar features of 2 objects, where $\sum_{k=1}^{n}|a_k - b_k - 1|$ is the sum of the number of similar and dissimilar features of 2 objects, $N(a_k = b_k)$ is the sum of similar features of 2 objects, and n is the type of feature attributes.

3) Body measurement uncertainty. If there are objects A and B in the recommender system, when there exists between A and B that one party's sports performance is moderate and the other party's performance is strong or weak, this cannot be used to determine whether the gap between the two parties can really be large enough to teach the other party, and therefore there is uncertainty. In set-pair theory, similarity, dissimilarity, and uncertainty sum to 1, which is $a+b+c=1$, so there exists for the recommended object A, B there is uncertainty as follows:

$$F(A,B) = 1 - \frac{\sum_{k=1}^{n}|a_k - b_k - 1|}{n} \tag{28}$$

4) Determination of linkage degree i. The value of the difference uncertainty coefficient i corresponding to objects A and B in the set pair recommendation degree corresponding to them in the recommender system is a key point to be determined. As the value of i tends to be closer to 1, the similarity between the 2 objects is higher, therefore, using the cosine similarity, a method of determining the value of i using the computational value method is proposed, which exists in the following definition:

$$i = \frac{1}{1 + d/s} \tag{29}$$

where s is the cosine similarity of object A, B with the following formula:

$$s = \frac{\sum_{k=1}^{n}(a_k \times b_k)}{\sqrt{a_k}\sqrt{b_k}} \tag{30}$$

d is the cosine dissimilarity of object A, B. Under certain conditions, the cosine dissimilarity can be transformed from the cosine similarity. The following formula is used to transform the similarity:

$$d = e^{-s} \tag{31}$$

d indicates the degree of dissimilarity between 2 objects, s indicates the degree of similarity between 2 objects, when s takes a larger value, the smaller the value of d, and the more i converges to 1. Conversely, i is farther away from l.

5) Calculation of Top-N recommendation set and friend recommendation. The Top-N recommendation algorithm is to sort the data according to certain rules and select the largest or smallest N data from the sorted list for recommendation [22]. Different rules can be formulated for different social environments to screen the data in the recommendation set, based on the university campus student population for research, in the friend recommendation at the same time need to consider a variety of factors, according to the user set obtained on the set of the degree of recommendation, calculating the average of the degree of recommendation of the user and the recommended user as a threshold value r, will be greater than the threshold value of the degree of recommendation is stored in the user's Top-N recommendation set.

Because in the process of friend recommendation, not only need to consider the user and the user's recommendation degree, there are also some practical issues need to be considered, such as the user's class, gender, distance from the living area, etc., recommended according to the user's needs for screening.

3.2 Analysis of the algorithm assignment process

3.2.1. Assignment of different training movement libraries. According to the logic of the recommendation algorithm, when assigning values, the key point is the size of the difference between the value and the value rather than the absolute size of the value itself. By assisting national coaches and professional physical fitness coaches and discussing with them, quantitative values were assigned to the movement libraries of resistance class, super isometrics, time class, and specialized training. Figure 2 shows the average contribution of different movement banks to physical fitness.

The resistance category itself mainly increases resistance by the students themselves or with the help of equipment, and mainly focuses on increasing the students' strength, so the overall value of strength will be higher than others in terms of the value of applicability compared to the other physical fitness sub-competencies. The movement library contains 4 resistance training movements, and the average value of the overall applicability is 0.18 for strength, while it is 0.17 for explosive strength, 0.12 for speed, and 0.15 for endurance.

For example, the weighted squat has an overall suitability of 0.25 for strength, which is greater than the suitability of 0.21 for explosiveness and 0.11 for endurance, and the suitability of this movement for sensitivity is 0 because the weighted squat does not require special control by the trainer. This movement does not require the trainer to specifically control balance, change direction, or accelerate or decelerate. The single-leg pull has an applicability of 0.32 for strength, which is greater than the 0.21 for explosive strength, and greater than the 0.12 and 0.22 for speed and endurance, and has a value of 0.06 for agility. The single-leg pull is a movement operation itself, in the practice students mainly single-leg force, and need to control the body not in the frontal plane, sagittal plane, horizontal plane tilt to ensure the stability of the center of gravity, so the single-leg pull has the degree of applicability for the students' sensitivity.

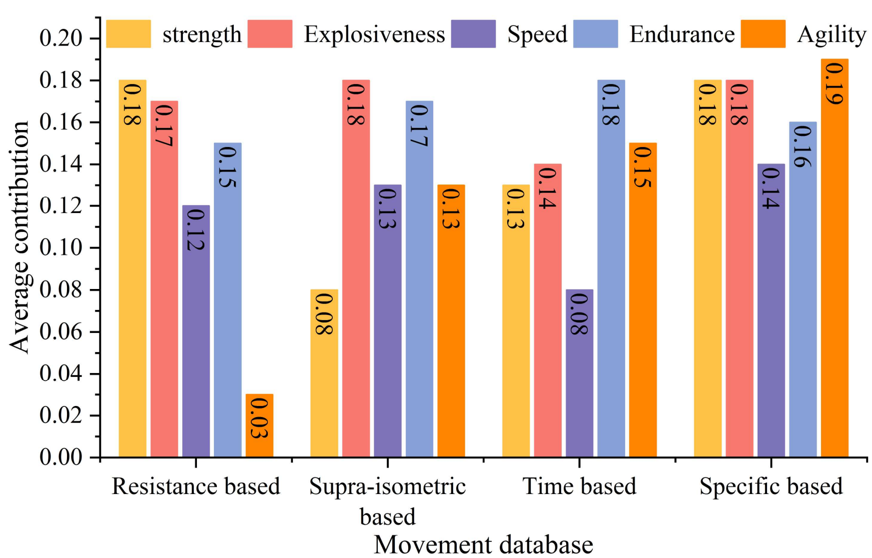

Fig. 2. Average contribution of different movement database to physical ability

3.2.2. Assignment of different risk factors. In this study, we take actions as the object of assignment, by assigning values to 4 action sub-banks with 40 actions, totaling 1,443 assignment units, in order to complete the regulation of risk factors for the SFD algorithm's recommended exercise program, so as to achieve a more accurate and more realistic recommended program. The principle when assigning values is the same as that of applicability assignment, which is to consider only the relationship between risk factors and specific compliant movements, and still take some impedance class movements as an example for assignment. Table 7 shows the action assignments considering different risk factors, with vi, vI, Vi and VI to indicate the 4 action types respectively.

Based on the relationship between students and competitions, the training cycle was divided into post-exercise conditioning, normal training, and competition preparation phases. Each phase has a training bias. Based on the experience of national coaches, values are assigned to the different phases. In the post-exercise adjustment, students should do some low-intensity training to help students' energy recovery, and it is less suitable for high-intensity or large amount of training. Therefore, the post-exercise conditioning is positively adjusted for the training loads of vi and Vi, which are 1 to 1.4, and negatively adjusted for the training of VI and VI, which are 0.2 to 0.5, so as to avoid the risk of injuries that may be brought about by the students' continuing to train with high loads after performing high-intensity sports. The normal training phase has no positive or negative regulation for training, so the value is assigned to 1. Students in the preparation phase are more suitable for training loads with higher intensity and smaller volume, which can help students without excessive fatigue, and better prepare for the competition. Therefore, the preparation phase has a positive conditioning effect for vi and VI, assigned a value of 1.2 to 1.3, and a negative conditioning effect for Vi and VI, assigned a value of 0.3 to 0.6 to simulate the competition load.

In the given action cases, all the training applicability involving high intensity is 0. This is because students should avoid high intensity exercise after injury and should train with VI loads to promote recovery from injury, so in the four case actions, the rehabilitation phase has a positive regulatory effect on the VI loads, with an assigned value of 1.2 to 1.4. And in the super-equal-length actions in the library of the actions, because the action itself involves the contraction cycle of muscle elongation one by one, such as continuous box jumping exercises, standing jump sprint and CM jump, etc., are not suitable for training during the rehabilitation period, so the value of risk factor is assigned as 0. As for the movements involving static exercises in the library of time-based movements, such as the balance pad one-legged support and four-point plank support, the value of the rehabilitation stage for the training loads Vi and vi of small intensity is 1.2 and 1.4, and the other value is assigned as 0.

Table 7. The evaluation of different movement considering risk factors

Factor	State	Bulgarian squat				Hard pull			
		vi	vI	Vi	VI	vi	vI	Vi	VI
Work out cycle	Adjustment	1.3	0.4	1.0	0.4	1.3	0.4	1.1	0.2
	Exercise	1.1	1.1	1.1	1.0	1.1	1.1	1.0	1.1
	Preparation	1.2	1.3	0.5	0.3	1.2	1.4	0.9	0.2
Injury	Recovery	1.2	0.0	1.0	0.0	1.4	0.0	1.4	0.0
	Nonrecovery	1	1	1	1	1	1	1	1
Factor	State	Heavy squat				Single leg hard pull			
		vi	vI	Vi	VI	vi	vI	Vi	VI
Work out cycle	Adjustment	1.4	0.6	1.2	0.3	1.4	0.5	1.2	0.2
	Exercise	1.0	1.1	1.0	1.0	1.2	1.0	1.0	1.1
	Preparation	1.3	1.3	0.9	0.2	1.2	1.4	1.0	0.2
Injury	Recovery	1.2	0.0	1.4	0.0	0.0	0.0	0.0	0.0
	Nonrecovery	0	1	1	1	1	1	1	1

At this point, the recommendation system of physical training program by SFD algorithm is constructed. The physical program consists of two parts: the applicability of the movement itself with different loads to the student's physical sub-capability, and the positive and negative moderating effects of the risk factors on the training loads.

3.3 Training program recommendation results and application analysis

After adjusting the numerical assignments of the movement library, the adjusted results were presented to the coaches and physical trainers, totaling 30 people. In this study, a questionnaire survey was conducted on 30 novice coaches by presenting the results recommended by the algorithm, including the training movements and training loads, and asking the coaches whether they approved of their training program. The number of people who recognized the recommended program and the number of cases of recognition are shown in Figure 3, respectively.

It can be found that 50% (15 coaches in total) of the coaches agree with most (70%~80%) of the dynamic adjustment strategies given by the recommended algorithm, and about 66% of the coaches strongly agree with (agree with more than 80% of the recommended content of) the dynamic adjustment program provided by the recommended algorithm in this paper. It can be seen that the SFD algorithm designed in this paper can propose a good dynamic adjustment program for physical education based on students' physical fitness data.

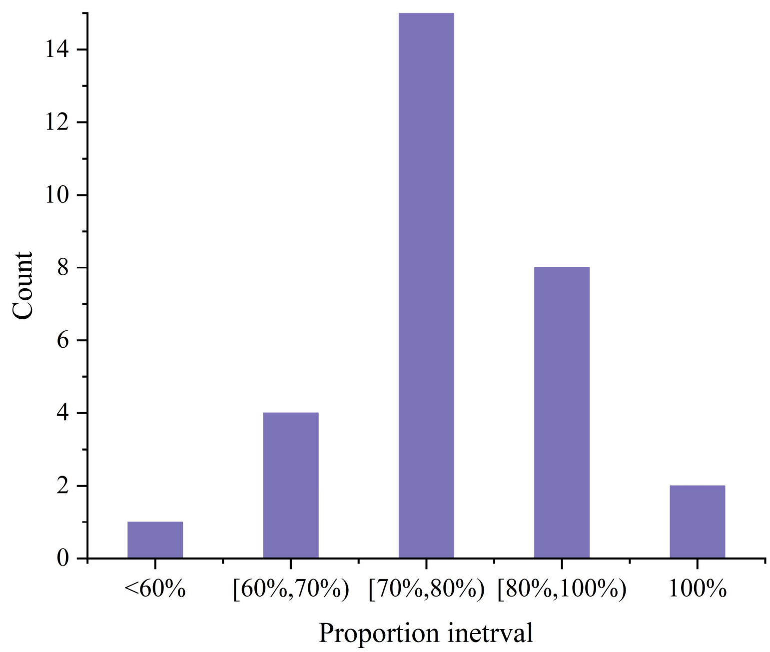

Fig. 3. The proportion of people who agree on the recommended plan

4. Conclusion

Through the multidimensional regression analysis of college students' physical fitness data and teaching dynamic adjustment model construction, this paper provides a method to optimize college physical education teaching through physical fitness data.

In the regression results, BIM index, physical exercise intensity, electronic product use habits, psychological tendency factors, external factors and other factors have correlation with college students' physical fitness pass or fail ($p<0.05$), and physical exercise intensity has the greatest effect on inducing college students' physical fitness failure (regression coefficient=0.927). And the balance of dietary meat and vegetables is a protective factor for physical fitness pass (regression coefficient = -0.411). Professional physical fitness coaches were invited to evaluate the adjustment program given by the dynamic adjustment model of physical education in this paper, and more than half of the coaches strongly agreed (agreed with more than 70% of the entries) with the new adjustment program.

This shows that this paper improves the scientificity of college physical education teaching by using students' physical fitness data through multidimensional regression analysis and dynamic teaching adjustment model.

References

[1] Opstoel, K., Chapelle, L., Prins, F. J., De Meester, A., Haerens, L., Van Tartwijk, J., & De Martelaer, K. (2020). Personal and social development in physical education and sports: A review study. European Physical Education Review, 26(4), 797-813.

[2] Beni, S., Fletcher, T., & Ní Chróinín, D. (2017). Meaningful experiences in physical education and youth sport: A review of the literature. Quest, 69(3), 291-312.

[3] Li, L., Liu, L., & Zou, S. (2019). An Experimental Study on the Training Method of Athletes' Physical Fitness in College Physical Education. DEStech Transactions on Social Science, Education and Human Science.

[4] Zhang, H. (2022). STUDY ON THE INFLUENCE OF PHYSICAL TRAINING ON ALLEVIATING COLLEGE STUDENTS'ANXIETY AND EMOTIONAL PRESSURE. Psychiatria Danubina, 34(suppl 2), 35-35.

[5] Wang, J. (2019). The association between physical fitness and physical activity among Chinese college students. Journal of American College Health, 67(6), 602-609.

[6] Sun, J., Lin, C., Wang, L., Bi, C., & Qiao, B. (2024). A comparative study on the physical fitness of college students from different grades and majors in Jiangxi province. Plos one, 19(10), e0308483.

[7] Gao, X., Rosas, M. F., Zhang, Y., Gatpandan, M. P., Gatpandan, P. H., & Dasig, D. (2021, July). The Research and Analysis of Student Health Status on the Basis of Physic Fitness Test'Data. In 2021 International Conference on Public Health and Data Science (ICPHDS) (pp. 110-115). IEEE.

[8] Liu, K. T., Kueh, Y. C., Arifin, W. N., Kim, Y., & Kuan, G. (2018). Application of transtheoretical model on behavioral changes, and amount of physical activity among university's students. Frontiers in psychology, 9, 2402.

[9] Mc Guire, P., Elsey, S., Barrett, M., & Ruhala, S. (2019). Getting freshmen "off and running" to success in college. The International Journal of Health, Wellness and Society, 10(1), 33.

[10] Krabak, B. J., Tenforde, A. S., Davis, I. S., Fredericson, M., Harrast, M. A., d'Hemecourt, P., ... & Roberts, W. O. (2019). Youth distance running: strategies for training and injury reduction. Current sports medicine reports, 18(2), 53-59.

[11] Wang, L., Wang, R., Ying, Y. Q., & Xu, W. Y. (2021, June). Research on Prediction Model of Endurance Running Ability of College Students Based on BP Neural Network. In 2021 IEEE 3rd International Conference on Computer Science and Educational Informatization (CSEI) (pp. 328-332). IEEE.

[12] Räisänen, A. M., van den Berg, C., Owoeye, O. B., McKay, C. D., & Emery, C. A. (2022). Running every time, planking sometimes: Youth adherence to a neuromuscular training program. Translational Journal of the American College of Sports Medicine, 7(4), e000209.

[13] Dubey, A. K., & Yadav, S. K. S. (2023). Effect of four weeks plyometric training program on vertical jump and speed on male university players of Chhatrapati Sahu ji Maharaj University, Kanpur. Journal of Sports Science and Nutrition, 4(1), 48-51.

[14] Gómez-Molina, J., Ogueta-Alday, A., Camara, J., Stickley, C., & Garcia-Lopez, J. (2018). Effect of 8 weeks of concurrent plyometric and running training on spatiotemporal and physiological variables of novice runners. European Journal of Sport Science, 18(2), 162-169.

[15] Chu, T. (2022). Research on college students' physique testing platform based on big data analysis. Mathematical Problems in Engineering, 2022(1), 4615020.

[16] Tang, P., Wang, Y., & Shen, N. (2020). Prediction of college students' physical fitness based on K-means clustering and SVR. Computer Systems Science and Engineering, 35(4), 237-246.

[17] Liu, Y. (2020, September). College students' physical fitness test data analysis, visualization and prediction

using data mining techniques. In Journal of Physics: Conference Series (Vol. 1631, No. 1, p. 012121). IOP Publishing.

[18] Hinpetch Daungsupawong & Viroj Wiwanitkit. (2024). A comment on manuscript Comparison of machine learning algorithms and multiple linear regression for live weight estimation of Akkaraman lambs.Tropical Animal Health and Production(8),338-338.

[19] Yuto Aizawa,Takeshi Emura & Hirofumi Michimae. (2025). Bayesian ridge estimators based on copula-based joint prior distributions for logistic regression parameters. Communications in Statistics - Simulation and Computation(1),252-266.

[20] Onu, Obineke Henry,Amakuro, Okuata Avula & Alabge, Samson Adekola. (2022). Study of Binary Logistic and Poisson Regression Models of Diabetic Patients in Nigeria using Dichotomous and Non- Dichotomous Predictors. Asian Journal of Probability and Statistics37-48.

[21] Wang Chunyang,Wang Shuangting,Guo Zengzhang,Wang Liping & Ma Chao. (2014). A New Semi-supervised Classification Method of Hyperspectral Image based on Combining Renyi Entropy and Multinomial Logistic Regression Algorithm. International Journal of Signal Processing, Image Processing and Pattern Recognition(5),59-70.

[22] Sharon Moses J. & Dhinesh Babu L.D.. (2020). Genetic Algorithm Influenced Top-N Recommender System to Alleviate New User Cold Start Problem. International Journal of Swarm Intelligence Research (IJSIR)(2),62-79.