Research on the development trend of intelligent leisure sports based on big data analysis

Feng Liang¹,²

¹School of Leisure and Health Guilin Tourism University, Guilin, 541006, China
²East China University of Science and Technology, Shanghai 200237, China

Corresponding author and e-mail: Feng Liang, lf@gltu.edu.cn

Abstract. Big data analysis often faces huge data samples. Its characteristics can be summarized as large amount of data, fast speed, multiple types, value, and authenticity. It is often used in data warehouse, data security, data analysis, data mining, etc. PLS regression analysis is a predictive modeling technique in big data analysis, which integrates principal component analysis, canonical correlation, and multiple linear regression algorithms. In a complex case analysis with multiple dependent variables, the main idea of components simplifies the model. The means of deriving the expressions of independent variables and dependent variables from the mapping relationship between components has a great advantage. This paper collects the attributes and market data of China's smart leisure sports equipment in 2020, establishing a PLS regression analysis model to analyze the development trend of China's smart leisure sports.

1. Introduction
"Artificial Intelligence" was included in the government’s work report for the first time at the Fifth Session of the Twelfth National People’s Congress on March 5, 2017. The research and development and transformation of technologies such as intelligence, integrated circuits, biopharmaceuticals, and fifth-generation mobile communications will expand and strengthen industrial clusters. It provides strategic goals and safeguards for the development of artificial intelligence in my country. With the advent of the Internet age, people’s lives are gradually digitized, the Internet is combined with big data, and emerging information technologies are constantly updated, providing a technical foundation for the combination of smart industry and leisure sports industry, and the smart leisure sports industry is gradually emerging. In 2021, Sun Dehao and others developed a LoRa-based sports data collection system design, through which the system monitors the athlete's exercise coefficient, so as to achieve the purpose of improving technical movements, and combines the measurement results of the VM8 ECG monitor and the IoT multi-physiological parameter monitoring system. For comparison, the error does not exceed 3.3 times/min, and the result is accurate [1]. In 2019, Jia Jianping studied the intensity control method of the treadmill based on the linear regression algorithm. The experiment tested 10 experimenters of the same age but different physical fitness and physical state, and monitored the heart rate fluctuation records of the testers after the experiment started. During the experiment the maximum deviation between the actual heart rate and the target heart rate is 8 and the minimum is -10; the root mean square error (RMSE) from the target heart rate is 4.54, 4.61, 4.01, 4.23, 2.21, 4.47, 4.79, 4.34, 4.72, 2.89, 3.87, respectively, the overall is relatively smooth [2]. In 2020, based on Jia Jianping’s experiments, Yue Haifeng and others designed a smart treadmill based on fuzzy control of heart rate...
and blood oxygen. Based on the Android application platform, the system can detect the user’s exercise coefficient and display the real-time condition of the body, using a fuzzy control algorithm to control the heart rate error at [-90, 90] [3]. In 2020, Zhang Ruiquan conducted big data analysis on the physical activity of Chinese adults from the perspective of smart sports bracelets. The survey found that smart bracelets of different brands have different data from the same tester. The error is greater than 0.02 and less than 0.4. The conclusion that the accuracy of smart bracelets developed by some brands is relatively low indicates that there is still a large gap in the effect of current smart sports products [4]. Through research, it can be seen that the current sports equipment combined with algorithm technology is quite common, but there are few articles that use big data to statistically analyze the development trend of smart leisure sports. Therefore, this article analyzes the current status of the development of the smart sports industry based on Optimization Strategy of PLS regression analysis and big data research.

2. Algorithm

PLS regression is a statistical method that uses projections to project predictor variables and observed variables into a new space to find a linear regression model. It is used to find the basic relationship between two matrices (X and Y), find the multi-dimensional direction in X space to explain the multi-dimensional direction with the largest variance in Y space, that is, an implicit variable method for modeling the covariance structure in these two spaces.

PLS regression analysis requires the dependent variable Y to be normal. Combined with the principle of principal component analysis, multiple independent variables X and multiple dependent variables Y are condensed into components U, V (X corresponds to the principal component U, Y corresponds to the principal component V), then use the canonical correlation principle to analyze the relationship between X and U, and the relationship between Y and V; and combine the principle of multiple linear regression to analyze the relationship between X and V, so as to study the relationship between X and Y.

The general multiple underlying model of PLS regression analysis is shown below.

\[
X = TP^T + E \tag{1}
\]

\[
Y = UQ^T + F \tag{2}
\]

Where X is a prediction matrix of \(n \times m\), Y is a response matrix of \(n \times p\), T and U are a matrix of \(n \times l\), which are the projection of X and Y respectively, and P and Q are \(m \times l\) respectively. The orthogonal load matrix of, and the matrices E and F are error terms, which are independent and identically distributed normally distributed random variables. Decompose X and Y to maximize the covariance between T and U.

PLS regression is mainly divided into two steps, as shown below, PLS regression is mainly divided into two steps, as shown below.

(1) Confirm the extraction of the number of principal components through the VIP analysis table of cross-validity and projection importance, that is, when the number of principal components increases, the VIP index of X changes little, or perform cross-validity analysis on different numbers of principal components. When Qh is greater than 0.0975, it is considered the principal component contributes to the model, and then the number of principal components satisfying Qh greater than 0.0975 is calculated.

(2) Perform regression analysis based on the number of principal components in the first step. Include:

a) Establish the relational expressions between the principal components and X and Y.

b) Describe the information interpretation rate (concentration rate) between the principal component and the research item, that is, the accuracy analysis.

c) Establish the regression equation of the original independent variable X to the dependent variable Y.

d) Analyze the strength of X's interpretation of Y.

The flow chart of PLS regression analysis is shown in Figure 1.
3. Experimental results and discussion

In order to discuss the development trend of smart leisure sports in the current market, this article investigates the platform channels for consumers to purchase smart sports equipment in 2020, the types of smart sports equipment purchased, and the occasions where smart sports equipment is used more frequently. Investigated the use area of the smart sports equipment, the wearability, the fun of the smart sports equipment is measured by whether it has social ability, the portability of the smart sports equipment, the purchase price and other performance as input independent variables. The performance of smart sports equipment is quantified as (0,1), “1” means that the attribute is strong, and “0” means that it does not have this attribute. In the statistical data, public places are set to 1, and personal devices are represented by 0. The same purchase method is set to “1” for online purchase, and “0” for offline orders placed through physical store shopping platforms, etc. If the two purchase methods coexist, the value is between (0,1) The same applies to the quantification of equipment use occasions. Select representative equipment among smart leisure sports equipment:

Wearables are represented by smart bracelets/watches, and non-wearable devices are divided into large-scale devices, such as treadmills, smart bicycles, and portable types, such as smart sensing dumbbells, and grip meters. To measure the popularity of smart leisure sports equipment in the consumer market, select the total number of sports equipment sold in 2020 and the frequency of user usage (times/day). See Table 1 for the establishment of a variable naming table.

Figure 1. Flow chart of PLS algorithm.
Table 1. Variable naming table.

| Variable name | Significance               |
|---------------|---------------------------|
| X1            | Land use area(m²)         |
| X2            | Device wearability        |
| X3            | Social skills             |
| X4            | Device portability        |
| X5            | Price(Ten thousand yuan)  |
| X6            | Use occasion              |
| X7            | Ways to purchase          |
| Y1            | Total annual sales of equipment(station) |
| Y2            | User frequency(Times/day) |

According to the survey, the attributes of smart leisure sports equipment and consumer market surveys in 2020 are shown in Table 2.

Table 2. Attributes of Smart Leisure Sports Equipment and Consumer Market Survey in 2020.

| Equipment number | Y1  | Y2  | X1       | X2   | X3   | X4   | X5  | X6   | X7   |
|------------------|-----|-----|----------|------|------|------|-----|------|------|
| 1                | 2439| 13  | 0.0001   | 1    | 1    | 1    | 149 | 0    | 1    |
| 2                | 4728| 6   | 0.1368   | 0    | 0.5  | 1    | 48  | 0.5  | 0.5  |
| 3                | 3781| 14  | 0.5      | 0    | 0    | 0    | 69  | 0.5  | 0.5  |
| 4                | 3851| 8   | 0.81     | 0    | 0.5  | 0    | 137 | 0.5  | 1    |
| 5                | 2572| 4   | 0.903    | 0    | 0.5  | 0    | 268 | 0.5  | 0.5  |
| 6                | 4168| 2   | 1.2      | 0    | 0.5  | 0    | 357 | 0.5  | 1    |
| 7                | 6832| 2   | 2.628    | 0    | 0.5  | 0    | 160 | 1    | 0    |
| 8                | 5383| 5   | 5.25     | 0    | 0    | 0    | 468 | 1    | 0.5  |
| 9                | 6372| 6   | 7.27     | 0    | 1    | 0    | 843 | 1    | 0.5  |
| 10               | 5974| 3   | 10.24    | 0    | 0    | 0    | 928 | 1    | 0    |
| 11               | 6853| 2   | 238.5    | 0    | 0    | 0    | 1199| 1    | 0    |

First, determine the optimal number of components of the sample data, and perform cross-validation analysis on the data to obtain a data table shown in Table 3.
Table 3. Cross-validation analysis data table.

| Component h | SS       | PRESS    | Qh²  |
|-------------|----------|----------|------|
| 1           | 3655993.252 | 9146077.836 | 1.000 |
| 2           | 2133708.253 | 8087242.962 | -1.212 |
| 3           | 1940880.722 | 12717028.522 | -4.960 |
| 4           | 1807090.937 | 15000178.679 | -6.729 |
| 5           | 1747596.034 | 24663359.973 | -12.648 |
| 6           | 1714708.810 | 80943357.986 | -45.317 |
| 7           | 1639571.545 | 97689970.168 | -55.972 |

The criterion for judging the number of principal components is that Qh is greater than 0.0975, indicating that the principal component contributes to the model. If Qh is less than 0.0975, it means that the component does not contribute to the model. It can be seen from the above table that only when the principal component is 1, the Qh value conforms to the judgment to be greater than 0.0975, so one principal component is finally selected as the conclusion.

Then the PLS regression analysis is performed with the number of principal components as 1, firstly, the relationship expressions between principal components and independent variable X and dependent variable Y are obtained.

Table 4. Expressions of mathematical relationship between principal components and research items.

| Independent variable | Principal component U1 |
|----------------------|-------------------------|
| X1                   | 0.472                   |
| X2                   | -0.475                  |
| X3                   | 0.421                   |
| X4                   | 0.265                   |
| X5                   | 0.195                   |
| X6                   | 0.202                   |
| X7                   | 0.475                   |

Dependent variable

| Principal component V1 |
|-------------------------|
| Y1                      | 1.059                   |
| Y2                      | 1.074                   |

Which is

\[ U_1 = 0.472 \times X_1 - 0.475 \times X_2 + 0.421 \times X_3 + 0.265 \times X_4 + 0.195 \times X_5 + 0.202 \times X_6 + 0.475 \times X_7 \]  

\[ V_1 = 1.059 \times Y_1 + 1.074 \times Y_2 \]  

Therefore, after the independent variables and the principal components of the dependent variables are obtained, the correlation between the original independent variable X and the dependent variable Y can be analyzed, and the standardized regression coefficients are shown in Figure 2.
From this get the relation
\[ Y_1 = 0.220 \times X_1 - 0.221 \times X_2 + 0.196 \times X_3 + 0.123 \times X_4 + 0.091 \times X_5 + 0.094 \times X_6 + 0.221 \times X_7 \]  
\[ (5) \]
\[ Y_2 = 0.223 \times X_1 - 0.224 \times X_2 + 0.199 \times X_3 + 0.125 \times X_4 + 0.092 \times X_5 + 0.096 \times X_6 + 0.224 \times X_7 \]  
\[ (6) \]

For the two dependent variables, look for the independent variable with the best explanatory importance. The higher the value, the higher the explanatory degree.

Figure 2. Standardized regression coefficients.

Figure 3. Projected importance index VIP of each variable.

It can be seen from the table that the independent variable X7 has the highest explanatory power for all Y (VIP value is 1.257), followed by X2 (VIP value is 1.022), and then X1 (VIP value is 1.248). Indicates that the independent variables that have the greatest impact on all output variables are X1, X2, and X7.

Then the precision analysis of the extracted principal components is used to study the ratio of information extraction of the principal components to the analysis items. The results are shown in Table 5.
Table 5. Accuracy analysis of principal components U, V and independent variables X, Y.

|       | Principal component U1 | Comprehensive |
|-------|------------------------|---------------|
| X1    | 0.959                  | 0.959         |
| X2    | 0.940                  | 0.940         |
| X3    | 0.738                  | 0.738         |
| X4    | 0.369                  | 0.369         |
| X5    | 0.208                  | 0.208         |
| X6    | 0.232                  | 0.232         |
| X7    | 0.947                  | 0.947         |
|       | 0.628                  | 0.628         |
| Y1    | 0.996                  | 0.996         |
| Y2    | 0.997                  | 0.997         |
|       | 0.997                  | 0.997         |

The information with the main component U being X is concentrated, and the information with the main component V being Y is concentrated. It can be seen from the analysis table that the extraction ratio of the principal component U1 to all three pieces of X information is 0.628, which means that the extraction information is relatively high. The information extraction ratio of X1 and X2 is very high (0.959 and 0.940 respectively), but the information extraction ratio of the principal component U1 to X5 and X6 is lower, 0.208, 0.232 means that the principal component U1 is difficult to extract the information of X5, X6. The principal component V1 has an extraction ratio of 0.735 for all three Y information, and the ratio of extracted information is also relatively high. For all dependent variables Y, there is high information extraction, which proves that the results obtained in this experiment are reliable.

Through the analysis of the projection importance index diagram of the respective variables in Figure 3, it can be seen that the independent variables that have the greatest impact on all output variables are X1, X2, and X7, and the VIP values are 1.257, 1.022, and 1.248 respectively, that is, the footprint of the device and the wearability of the device. The purchase method will have a greater impact on the market sales of smart leisure sports equipment. From Figure 2 and equations (5) and (6), it can be seen that X1 and X7 are positively correlated with all Y, while X2 and Y are negatively correlated, which means that sports leisure equipment is more inclined to occupy a large area and be wearable. In the direction of weaker sex, and consumers’ purchase methods are more inclined to online purchases, it can be explained that with the development of smart technology, consumers prefer large-scale smart leisure sports equipment. Then the current data can analyze smart leisure Sports equipment tends to be large and weakly wearable, and the sales network is platform-oriented.

Therefore, suggestions can be made: When developing smart leisure sports equipment, businesses can prefer to introduce this type of sports equipment into classrooms and gyms and other indoor large-scale equipment, and at the same time, combined with consumers’ buying habits, adopt online-based sales channels and offline sales channels. As a supplement, try webcast sales and other methods to meet the needs of consumers.

4. Conclusions
From a statistical point of view, this article analyzes the development trend of smart sports and leisure equipment, and establishes a PLS regression analysis model to find the factors that affect the needs and development of its shopping malls. The independent variable principal components and the dependent variable principal components generated are all relevant to the research item. The high extraction degree
and high reliability of the results prove that the analysis results of PLS applied to the trend research of smart leisure sports equipment have high reference value, and provide a new idea for data analysis in this field.

Acknowledgement
This paper is based on the following projects “Supported by 1000 Middle-aged and Young Backbone Teachers Cultivation Plan of Guangxi Colleges and Universities” Program Research funding project. Author: Feng Liang was born in Zibo, Shandong, P.R. China, in 1983. He received the PH.D from Korea National Sport University. Now, he is working at Guilin Tourism University, Associate professor. His research interest in Research direction: Leisure sports industry, sports tourism.

References
[1] Sun Dehao, Cui Guoqiang, Wang Xin, Liu Kaiyang, Li Zhengping. LoRa-based sports data acquisition system design [J]. Internet of Things Technology, 2021, 11(03): 8-9+13.
[2] Jia Jianping, Wang Wenbo, Liu Guoping. Treadmill intensity control method based on linear regression [J]. Journal of Nanchang University (Engineering Science Edition), 2019, 41(01): 76-79+84.
[3] Yue Haifeng, Liu Yinchi, Qian Nini, Sui Xiuwu. Research on intelligent treadmill based on fuzzy control of heart rate and blood oxygen [J]. Henan Science and Technology, 2020, 39(28): 42-44.
[4] Zhang Ruiquan. Big data analysis of physical activity of Chinese adults based on smart sports bracelet [J]. Journal of Shangqiu Teachers College, 2020, 36(09): 78-81.
[5] F Liang, R Li, L Mu. Research on the spread effect of data recording sports app on outdoor running group [J]. Microprocessors and Microsystems, 2021, 82.
[6] Feng Liang. Promoting the Integrated Development and Path Optimization of Sports Industry and Tourism Industry [J]. International Journal of Frontiers in Sociology, 2020, 2(9):114-122.