Big data applied in secondary education student's achievement by using principal component analysis

Rui Li ¹, Meiyan Tian ²

¹. HongHe University, Honghe, Yunnan, PR China
². First class teacher of information technology in middle school, Meng Qiao Middle School of Jinping County in Yunnan Province, Honghe, Yunnan, PR China

Keywords: Education information; Big data technologies; Student performance; Analysis

Abstract. With the advent of the era of big data, data analysis, penetrated into all walks of life among the analytical method has become more abundant. And inside the field of education, especially for high school inside analysis of the data it is relatively simple, which most high school's data is students' test scores. Principal component analysis method to middle school students mainly use the results were dimensionality reduction is calculated for each principal component scores and principal component composite score, then score cluster analysis method subjects isolated partial science students, science and migraine students were characteristic analysis, finally using a regression analysis of student test results were estimated, analyzed in the comprehensive examinations can play well and play mad two-part student discipline characteristics. Finally got partial science students tend to either science or the main component of integrated ranked among test exam has good performance.

1. Introduction

Big data is a huge amount of data, typically PB or EB data storage magnitude [1]. Byte, KB, MB, GB, TB, PB, EB, ZB, YB, DB, NB, in accordance with the feed rate of 1024 (2^10) increase [2, 3]. According to the current mainstream hard drive TB per unit, a PB storage concept is not an ordinary room can hold, and perhaps only EB store search engine company [4]. Education information currently developing "two-way platform" big data applications for the construction of education has laid a solid foundation [5]. Big Data management involves education is very extensive. The daily management of educational activities generated by the teachers, students and parents, and other acts data, such as teachers and students of basic information, education and attendance management, school credits process, teaching information, teaching and research education, management of equipment and instruments, all types of communications and information, education can be a big data.

In this information age, resource sharing big data applications play an increasingly important role of information technology in education, and to a large extent helped teachers innovative teaching models and teaching methods [7]. How to make good use of large data-depth features is an important issue we face. Second, data analysis applications on education some people say: We are using big data, downloading of text, pictures, video, courseware, are using internet and internet resources to carry out a series of educational activities. Indeed, this is the application, but not analytical applications, not large data-level applications [8, 9]. Big data in the most primitive way into the "cloud", simply put, is every individual, every department, every family, every industry, different individuals, units summarize various actors operation. It records not only the input data itself, but extends implanted themselves enter all the associated information.

Primary and secondary education teaching quality monitoring has "all the elements" "the entire process," "full" of "full" and other four characteristics. "Total factor" refers to the teaching process, students learn to master all the factors managers work quality; "the entire process" means the entire process of education and teaching, teaching plans include teaching process, teaching implementation, teaching reflection phase, students school, learning, assessment, graduation and so on; "full" means...
that teachers, students, workers and service personnel management of all relevant content related to primary and secondary schools; "comprehensive" means not only teaching, teaching effect, including on the basis of scientific, effective control of the teaching process. With the support of modern teaching environment of big data platform, four "all" cited problems solved.

2. Education big data analysis

2.1 Timely and accurate feedback student information

Usually center site provides various courses of learning courses in various disciplines of knowledge function, and the popular Mu class (MOOC) platform, participants can ask questions online. Others can communicate answer, there may be interspersed with some small tests can feedback learning happening. In this teaching process, a lot of real-time data to be recorded class platform, and after appropriate analysis process, after the original data accumulated clutter of a single student's behavior can present a regularity, which reflect the students throughout the science teaching process in which the state and problems. This can have a base to adjust the learning content or targeted for study, a substantial increase in the quality of teaching.

According to the overall results of the analysis of large data, management can also be found under the traditional model can not find the problem, and to take specific solutions, a more favorable course design, to better promote the development of education. In the information environment, the teacher in designing a course, you can ask questions and gradually by the feedback from students the answers to the mobile terminal. In the teaching process, time node by the teacher control, but this time students only need the appropriate knowledge click here "A" or "B", yes" or "no", as left and right mouse button clicks are generally simple, until after the lecture ended, the teacher can immediately know how many students have when an error occurs.

2.2 Scientific compare different types of data

We can imagine, in the education sector organizations to carry out the evaluation of teachers or students or the game when, in the case of organizational difficulties would not necessarily need to be grouped sub disciplines. Primary and secondary schools in various disciplines call big data past performance statistics as the basis for a large sample, primary group players score big data comparison and secondary school and secondary school players score big data comparison, the respective relative ranking, to determine the outcome. Even as rigorous college entrance examination, candidates can also choose the examination time in batches.

College entrance examination of the volume random by the computer, automatic printing, and automatic recovery after the candidates to answer, marking intelligent achieve a series of full-automatic mode, the absolute guarantee fair. The most important point is that the candidates different questions and performance criterion. We can compare according to the city, the country's big achievement data, regardless of the candidate's questions the degree of difficulty, the difficulty factor can automatically analyze assessment scores. To be based on time to answer, answer mode answer as to master some auxiliary evaluation indicators.

3. Results and discussion

3.1 Mathematical model analysis

There is provided sample n records, each record observation have p-dimensional variables: X1, X2, Xp. n × p matrix by order of the original data as follows:

\[
X = \begin{pmatrix}
X_{11} & X_{12} & \cdots & X_{1p} \\
X_{21} & X_{22} & \cdots & X_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
X_{n1} & X_{n2} & \cdots & X_{np}
\end{pmatrix}
\]  
(1)
In it: \( X_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}, i = 1, 2, \ldots, p \) \( (2) \)

The vector \( p \) data matrix \( X \), the linear transformation can be combined to form a new variable, namely:

\[
\begin{align*}
    f_{i1} &= a_{i1}x_{i1} + a_{i2}x_{i2} + \cdots + a_{ip}x_{ip} \\
    f_{i2} &= a_{21}x_{i1} + a_{22}x_{i2} + \cdots + a_{2p}x_{ip} \\
    &\vdots \\
    f_{ip} &= a_{p1}x_{i1} + a_{p2}x_{i2} + \cdots + a_{pp}x_{ip} \\
\end{align*}
\]

\( F = \begin{pmatrix} f_{i1} & f_{i2} & \cdots & f_{ip} \\ f_{21} & f_{22} & \cdots & f_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{np} \end{pmatrix} = (F_1, F_2, \ldots, F_p) \) \( (3) \)

\( A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix} = (A_1, A_2, \ldots, A_p) \) \( (4) \)

If not bound by a linear transformation are endless. According to the definition of a mathematical model of principal component analysis, to make use of principal component analysis, it is required according to the original data, and the model of the three conditions require calculating the principal component coefficient, whereby the principal component model.

3.2 Calculating a correlation matrix

Import the data into the software, in order to understand the degree of correlation between the subjects and the subjects, first calculated the correlation matrix \( R \) each subject variables, the results was shown in Tab. 1.

<table>
<thead>
<tr>
<th>correlation coefficient</th>
<th>Moral</th>
<th>Chinese</th>
<th>Math</th>
<th>English</th>
<th>History</th>
<th>Geography</th>
<th>Biology</th>
<th>Physics</th>
<th>Chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moral</td>
<td>1.000</td>
<td>.828</td>
<td>.666</td>
<td>.723</td>
<td>.815</td>
<td>.659</td>
<td>.777</td>
<td>.766</td>
<td>.655</td>
</tr>
<tr>
<td>Chinese</td>
<td>.828</td>
<td>1.000</td>
<td>.622</td>
<td>.755</td>
<td>.754</td>
<td>.599</td>
<td>.668</td>
<td>.652</td>
<td>.588</td>
</tr>
<tr>
<td>Math</td>
<td>.666</td>
<td>.622</td>
<td>1.000</td>
<td>.648</td>
<td>.666</td>
<td>.677</td>
<td>.752</td>
<td>.889</td>
<td>.756</td>
</tr>
<tr>
<td>English</td>
<td>.723</td>
<td>.755</td>
<td>.648</td>
<td>1.000</td>
<td>.632</td>
<td>.600</td>
<td>.633</td>
<td>.658</td>
<td>.613</td>
</tr>
<tr>
<td>History</td>
<td>.815</td>
<td>.754</td>
<td>.666</td>
<td>.632</td>
<td>1.000</td>
<td>.688</td>
<td>.799</td>
<td>.702</td>
<td>.725</td>
</tr>
<tr>
<td>Geography</td>
<td>.659</td>
<td>.599</td>
<td>.677</td>
<td>.600</td>
<td>.688</td>
<td>1.000</td>
<td>.768</td>
<td>.735</td>
<td>.721</td>
</tr>
<tr>
<td>Biology</td>
<td>.777</td>
<td>.668</td>
<td>.752</td>
<td>.633</td>
<td>.799</td>
<td>.768</td>
<td>1.000</td>
<td>.788</td>
<td>.768</td>
</tr>
<tr>
<td>Physics</td>
<td>.766</td>
<td>.652</td>
<td>.889</td>
<td>.658</td>
<td>.702</td>
<td>.735</td>
<td>.788</td>
<td>1.000</td>
<td>.886</td>
</tr>
<tr>
<td>Chemistry</td>
<td>.655</td>
<td>.588</td>
<td>.756</td>
<td>.613</td>
<td>.725</td>
<td>.721</td>
<td>.768</td>
<td>.886</td>
<td>1.000</td>
</tr>
</tbody>
</table>

From Tab. 1, the section "correlation coefficient" is easily seen, this is a symmetric matrix, and the correlation coefficient between the variables subject larger than, for example: the correlation coefficient between the moral and the language was 0.828, indicating that subjects a high degree of correlation between variables.
As can be seen from Fig. 1, the graphical presentation types two very dense more and more middle loose. That is, the top-notch academic achievement of students and poor students, or whether it is out of the ranking are ranked PCA- differ significantly; while ranking in the middle of some of the students there is a clear difference, in the middle of the greater difference, maximum the difference was more than more than 50, indicating that some students are in the middle of making some reason for this discrepancy, it is necessary for the existence of differences of these students for further study.

4. Summary

In this paper, the use of principal component analysis of the main high school subjects: ethics, Chinese, math, English, history, geography, biology, physics and chemistry were analyzed, the dimensionality reduction of nine dimensions represent the three main components. Several different types of disciplines, subjects out of the use of Z and F, respectively principal component score ranking and the ranking of the two made the difference analysis. As used herein, the method and analysis process can be repeated use, enter the appropriate student achievement data, step by step implementation of the algorithm article listed out, you can get the desired results accordingly.

References


