

Health data analysis based on multi-calculation of big data during COVID-19 pandemic

Yang Bo and Wang Chunli*

Nanjing Medical University, Nanjing, Jiangsu, China

Abstract. Under the influence of the COVID-19, the analysis of physical health data is helpful to grasp the physical condition in time and promote the level of prevention and control of the epidemic. Especially for novel corona virus asymptomatic infections, the initial analysis of physical health data can help to detect the possibility of virus infection to some extent. The digital information system of traditional hospitals and other medical institutions is not perfect. For a large number of health data generated by smart medical technology, there is a lack of an effective storage, management, query and analysis platform. Especially, it lacks the ability of mining valuable information from big data. Aiming at the above problems, the idea of combining Struts 2 and Hadoop in the system architecture of the platform is proposed in this paper. Data mining association algorithm is adopted and improved based on MapReduce. A service platform for college students' physical health is designed to solve the storage, processing and mining of health big data. The experiment result shows that the system can effectively complete the processing and analysis of the big data of College students' physical health, which has a certain reference value for college students' physical health monitoring during the COVID-19 epidemic.

Keywords: Physical health data, parallel processing, health monitoring, multi calculation big data, COVID-19

1. Introduction

Since December 2019, the surveillance of influenza and related diseases has been continuously carried out in Wuhan, Hubei Province. Multiple cases of viral pneumonia were found, all of which were diagnosed as viral pneumonia/pulmonary infection. On January 12, 2020, the World Health Organization officially named it 2019-nCov. On February 17, 2020, the Chinese Center for Disease Control and Prevention (CDC) released the largest epidemiological analysis of new corona virus pneumonia to date. It is found that:

- (1) Most novel corona virus pneumonia patients showed mild symptoms, and mild/moderate cases reached 80.9%.

- (2) The crude mortality was 2.3%. The death rate of patients aged 60 and over accounted for 81%, and the mortality rate of patients with complications was much higher.
- (3) The novel corona virus pneumonia presents an epidemic pattern. The cases in December 2019 may be a small-scale exposure transmission mode. January 2020 is likely to be the diffusion mode. This is in line with the conclusion of previous investigations: the new corona virus is transmitted from a still unknown wild animal to humans, and then spreads from person to person.

Under the influence of the COVID-19, the analysis of physical health data is helpful to grasp the physical condition in time and promote the level of prevention and control of the epidemic. Especially for novel corona virus asymptomatic infections, the initial analysis of physical health data can help to

*Corresponding author. Wang Chunli, Nanjing Medical University, Nanjing 211166, Jiangsu, China. E-mail: 6207011@qq.com.

detect the possibility of virus infection to some extent.

For the digital acquisition of medical and health services, mainly through intelligent acquisition sensors [1-7]. Common medical sensors include piezoelectric pulse, three-point pulse sensor, heart rate sensor, ECG sensor and skin contact sensor. They can be used to collect the four vital sign parameters of human body, such as breath, body temperature, pulse and blood pressure [8]. The collected physiological information is transformed into electrical signals for computer identification by Internet of things technology. The electric signal is used to represent the relevant information for the future transmission and analysis. For the digital storage and processing of medical and health services, when the data collected and transmitted by sensors, they need to be transformed into valuable information [9-11]. Medical and health data has the characteristics of continuous, high growth and complex structure, which also contains rich and diverse information value. By mining and analyzing these data, we can improve the diagnosis and treatment and service level of health service institutions. How to mine and analyze these data is mainly through data mining technology.

In 2019, the development process of China's student physique monitoring pointed out that: the physique of college students in China still shows a downward trend, but the downward speed is slowing down. At the same time, body shape is changing. In particular, the obesity rate continues to rise, increasing by 2% to 3% every five years [12]. In addition, the 2019 national physique monitoring report shows that the physique state of students aged 7 to 19 is the most serious decline. The unhealthy life style and the indifference of health consciousness make the physical health of college students seriously threatened. Therefore, it is a very important task to monitor and analyze the physical health data of college students. The digital information system of traditional hospitals and other medical institutions is not perfect. For a large number of health data generated by intelligent medical technology, there is a lack of effective storage, management, and query and analysis platform [13]. Especially, it lacks the ability of mining valuable information from big data. To solve these problems, this paper proposes the idea of combining Struts 2 and Hadoop in the platform architecture. The association algorithm of data mining based on MapReduce is adopted and improved. In order to solve the storage, processing and mining of health big data, a health

service platform for college students is researched in this paper.

2. Platform design and data mining algorithm optimization

2.1. Overall system architecture

This paper is based on the user health service system. The system collects the vital signs of users through wireless sensor network and information service node to realize the health application for users. This paper focuses on the intelligent terminal technology and information perception technology under the complex application environment and multiple service requirements. Then through the research and analysis of home service terminal adaptive access technology, the health service network is designed. The data is transferred to the health service platform, and the big data system and intelligent information analysis and processing of software definition service are carried out on the virtual information platform. Through the analysis and processing of big data, valuable information can be obtained. Through the user's health data, the user is given the guidance of life habits such as exercise, diet, etc. The accuracy of user diagnosis can be improved by medical data mining. Through this method, we can find out the hidden correlation between diseases and prevent the occurrence of related diseases. The architecture of the overall system is shown in Fig. 1.

As can be seen from Fig. 1, user oriented health service is mainly divided into three parts: intelligent node data collection technology, adaptive multi network access technology and big data analysis and processing technology. The big data analysis and processing platform for user health services studied in this paper is based on regional intelligent data center and high-speed Internet. Based on the Internet service system, supported by key technologies such as big data storage, processing, mining and interactive visual analysis. It can provide users with data management, data storage and data mining analysis services through a variety of mobile terminals and the Internet.

The big data analysis and processing platform consists of three layers, namely service layer, function layer and platform layer. The service layer provides users with a series of operations and related interfaces for Web-based big data analysis and processing, and allows users to manage user information. The

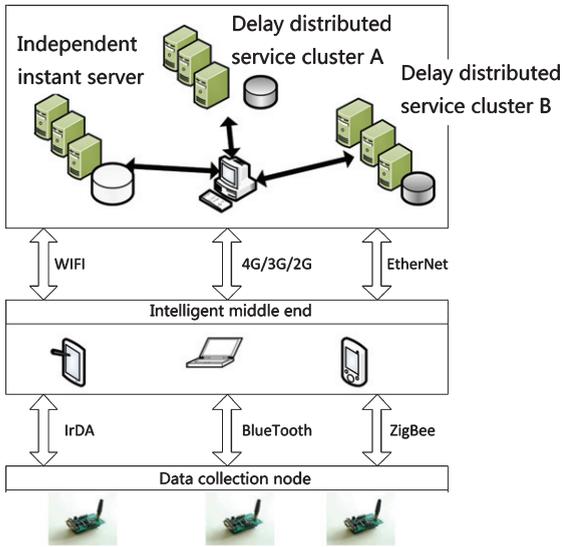


Fig. 1. Network architecture of user oriented health services.

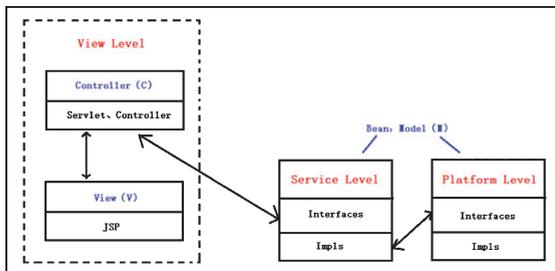


Fig. 2. Logical architecture of big data analysis and processing platform.

function layer provides some functions of big data storage and mining algorithm. The platform layer provides the basic distributed environment support for the whole big data analysis platform. Its logical architecture is shown in Fig. 2 above. According to the above logical architecture, in order to realize the service layer, function layer and platform layer respectively. The technical architecture used in this article is shown in Fig. 3.

The service layer is the top layer of the system, showing the user the most intuitive web interface operation. This facilitates the analysis and processing of health big data. Through invoking the service interface of the functional layer, diversified services are formed for different objects such as medical institutions and management departments, including disease early warning, patient behavior analysis, health guidance and other services. It mainly uses HTML, JSP, JavaScript to write the front page, and

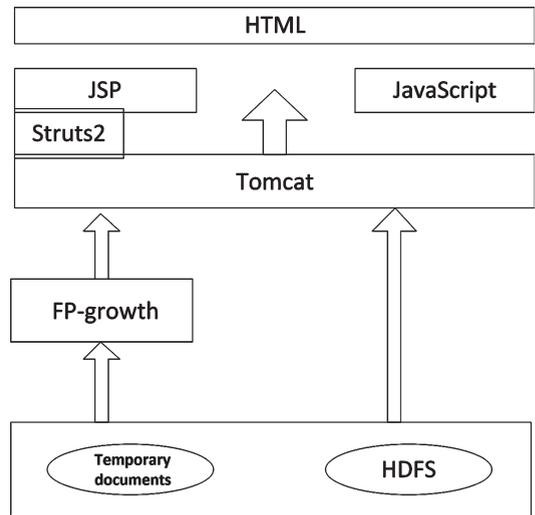


Fig. 3. Logical architecture of big data analysis and processing platform.

uses the currently very popular web development framework Struts 2. This architecture deploys web pages on Tomcat web server. The function layer is located under the service layer, which mainly provides users with data management, data storage, data mining and other functions in the background. It can be called by the service layer, including the association rule algorithm FP growth algorithm involved in the platform. The platform layer is located below the function layer, providing basic support for big data, computing and processing various medical and health data. This system mainly uses Hadoop big data processing framework as the technical support of platform layer. The three-layer structure has the following advantages:

- (1) To increase the reusability and encapsulation of code, developers only need to pay attention to the specific implementation of each layer, instead of paying extra attention to all layers when implementing the functions of one layer.
- (2) It can reduce the coupling between layers and make it convenient for developers to decouple.
- (3) It can easily expand or modify the original functions.
- (4) The logic is clear and easy to reuse.
- (5) It is convenient for later maintenance and reduces the manpower and material resources needed in the maintenance process.

2.2. Association rules mining algorithm

The core of the design, analysis and processing of the big data platform for user health services is to find the rules contained in the data through the knowledge mining of the big data and show them in a form that ordinary users can understand. The algorithms used in this paper are as follows:

- (1) Association Rules: association rules are used to find out whether multiple things will appear at the same time, as well as the number and possible probability of occurrence, find out the relationship between valuable items from a large number of data sets, analyze the association of various diseases in various complications in the field of medical and health care, and study the probability of another complication under the condition of a certain complication To explore the relationship between complications and causality.
- (2) Clustering idea: birds of a feather flock together, which explains the connotation of clustering algorithm. Because birds of a feather flock together, clusters or subclasses can be formed according to the differences between different kinds and the similarities between the same kinds. The cluster generated by clustering is a set of objects, which are similar to the objects in the same cluster, but different from the objects not in the same cluster. There are several applications in the field of medical and health care, such as clustering analysis on the age and gender characteristics of various diseases, finding different age and gender characteristics of various diseases, clustering analysis on the related physiological indicators of various diseases, and studying the distribution characteristics of the related physiological indicators of various diseases.
- (3) Classification idea: classification idea refers to mapping data to predefined groups or classes, and dividing transactions or abstract objects into multiple classes based on a certain supervised learning. Classification is a continuous learning process. We can illustrate it with examples in the field of health care. In some cases, when we have a cold, we may summarize the characteristics of the cold, such as nasal congestion, headache, and general weakness. Then we can judge whether we have a cold according to these characteristics next time.

The possibility of accurate judgment of the cold will be great, thus forming a certain rule of experience which belongs to the classification process.

2.3. Definition of association rule algorithm

Association rules found a relationship between things and other transactions or interdependence. Assuming that $I=\{i_1,i_2,\dots i_m\}$ is a collection and the related data task D is a collection of database transactions, in which each transaction T is a collection, and making $T \subseteq I$. Every transaction has an identifier TD . Assuming that A is a set of items, $A \subseteq T$. Association rules are the containing type of $A \Rightarrow B$, among them, $A \subset I, B \subset I$ and $A \cap B = \Phi$. The rules $A \Rightarrow B$ is in the transaction which sets up with support s , s is the percentage for the transaction contains $A \cup B$ in the D .

The LIPI algorithm through scanning data sets of frequency, then finding relevant data and finishing dig, its principle as follows:

Assuming that $I=\{i_1,i_2,\dots i_n\}$ is a collection, which composed of different characteristics, the characteristics of each item as a constituted set of items. And the item set is not an empty set, but is a subset of the set of I , which can be expressed as $(x_1x_2 \dots x_m)$, every x_k is a term.

The sample variance and sample proportion of variance have established the following relationships [14–16]:

$$S^{*2} = \frac{S^2}{\mu^2} \tag{1}$$

Proof: by the definition, we have:

$$\begin{aligned} S^{*2} &= \frac{1}{n-1} \sum_{i=1}^n \left(\frac{\alpha_i - \bar{\mu}}{\bar{\mu}^2} \right)^2 \\ &= \frac{1}{(n-1)\bar{\mu}^2} \sum_{i=1}^n (\alpha_i - \bar{\mu})^2 \\ &= \frac{1}{\bar{\mu}^2} S^2 \end{aligned} \tag{2}$$

Consider delay, the L can be expressed as:

$$L^0 = \left\| \begin{matrix} C_{ijkl}^0 & e_{kij}^0 \\ e_{ikl}^{0T} & -\eta_{ik}^0 \end{matrix} \right\| \tag{3}$$

These functions can be expressed in the following form:

$$C(x) = C^0 + C^1(x), e(x) = e^0 + e^1(x),$$

$$\eta(x) = \eta^0 + \eta^1(x), \rho(x) = \rho_0 + \rho_1(x) \quad (4)$$

The value with superscript of 1 represents the difference below:

$$C^1 = C - C^0, e^1 = e - e^0, \eta^1 = \eta - \eta^0,$$

$$\rho_1 = \rho - \rho_0, e^1 = e - e^0 \quad (5)$$

The whole function can be simplified into the following integral equation set:

$$f(x, \omega) = f^0(x, \omega) + \int_V S(x - x')(L^1 F(y') + \rho_1 \omega^2 g(R)T_1 f(y'))S(y')dy' \quad (6)$$

In addition, we can introduce the abbreviated formula:

$$g(x, \omega) = \left\| \begin{matrix} G_{ik}(x, \omega) & \gamma_i(x, \omega) \\ \gamma_k(x, \omega) & g(x, \omega) \end{matrix} \right\|,$$

$$s(x, \omega) = \left\| \begin{matrix} G_{ik,l}(x, \omega) & \gamma_{i,k}(x, \omega) \\ \gamma_{k,l}(x, \omega) & g_{,k}(x, \omega) \end{matrix} \right\|,$$

$$L^1(x, \omega) = \left\| \begin{matrix} C_{ijkl}^1 & e_{kij}^1 \\ e_{kij}^{1T} & -\eta_{ik}^1 \end{matrix} \right\|,$$

$$F(x, \omega) = \left\| \begin{matrix} u_{(i,j)}(x, \omega) \\ \varphi, i(x, \omega) \end{matrix} \right\| \quad (7)$$

In these expression, $G_{ik}(x, \omega), \gamma_i(x, \omega), g(x, \omega)$ can be represented as:

$$g(x, \omega) = \frac{1}{(2\pi)^3} \int g(k, \omega) \exp(-ikx)dk \quad (8)$$

$$g(k, \omega) = \left\| \begin{matrix} G_{ik}(k, \omega) & \gamma_i(k, \omega) \\ \gamma_k^T(k, \omega) & g(k, \omega) \end{matrix} \right\| \quad (9)$$

Where $G_{ik} = (\Lambda_{ik} + \frac{1}{\lambda} h_i h_k^T)^{-1}$,

$$g = -(\lambda + h_i^T \Lambda_{ij}^{-1} h_j)^{-1}, \gamma_i = \frac{1}{\lambda} h_k^T G_{ki}. \quad (10)$$

The algorithm is to propose the original data processing for many times, then use the effective information contained in the original data.

3. FP-growth algorithm

3.1. Algorithmic thinking

FP growth algorithm can be divided into two stages:

The first stage: the establishment of FP Tree, the second stage: frequent pattern mining for FP Tree. The establishment of FP Tree needs to scan the transaction database twice. We use the data in Table 1 for algorithm analysis, as follows in Fig. 1.

The first scan accumulates the support of each item and selects the items that meet the minimum support, that is, one-dimensional frequent item set. These transaction items are arranged in descending order according to the frequency number to form an f-list. We assume that the minimum support of association rules is 3.

The second scan is used to build the FP Tree. First, according to the transactions recorded in F-list, non-frequent items are eliminated. The main operation is to sort the data in Table 1 according to the one-dimensional frequent item set, and delete the items whose occurrence times are lower than the minimum support degree. After the above operations, we can get the updated data set of FP growth algorithm. I si shown in the Table 2 below.

According to the updated data set in Table 2, you can start to build FP Tree. The establishment rules of FP Tree are as follows:

Rule 1: The root of the tree is the root node, which is defined as null.

Rule 2: A node is inserted into a transaction set. A node is a transaction item. Each node contains the name and occurrence times of the transaction.

Table 1
Initial data set of parallel FP-growth algorithm

Transaction ID	Transaction set
1	a, c, d, f, g, i, m, p
2	a, b, c, f, l, m, o
3	b, f, h, j, o
4	b, c, k, s, p
5	a, c, e, f, l, m, n, p

Table 2
Data set updated by FP-growth algorithm

Transaction ID	Transaction set
1	f, c, a, m, p
2	f, c, a, b, m
3	f, b
4	c, b, p
5	f, c, a, m, p

Rule 3: If the first n nodes of the inserted transaction have the same path in the tree, no new node needs to be created. Directly add 1 to the number of nodes in the original path. The path here refers to the path from the root node to the leaf node.

Rule 4: If the inserted transaction has no project path in the tree, create a new node.

Rule 5: Attaching a header table during tree building. This table is used to represent one-dimensional frequent item set, and the pointer of the elements in the header table can be used to point to the node with the same name in the tree for the first time.

Rule 6: Each node in the tree contains a pointer to the next node with the same name. If there is no node with the same name, the pointer points to null.

3.2. FP Tree parallel mining

The FPMiningReducer class is used to implement the reducer process in MapReduce and the received transaction list is merged. Then a new TransactionTree is constructed. Finally, the list of projects localFlist from cTree is exported. The key codes are as follows:

```

FPGrowthIds.generateTopKFreqentPatterns
(cTree.iterator(),
freqList,
minSupport,
maxHeapSize,
PFPGrowth.getGroupMembers(key.get(),
maxPerGroup,
numFeatures),
IntegerStringOutputConverter,
ContextStatusUpdater)
    
```

First, truncating freqList and only keep items that are more frequent than minSupport. This list of items is used for the subsequent export of frequent patterns. Then construct the FP-tree and traverse all transactions in cTree, that is, all branches of the tree. These transactions are used to construct frequent trees. The fpGrowth() method is called to export FrequentPatternMaxHeap for each item that meets the minimum support level.

For each project, its frequent pattern growth() method is derived from the FP tree. For the growth() method, if the frequency count of the current item is not large enough, an empty FrequentPatternMaxHeap is returned directly. Otherwise, the creatMoreFreqConditionalTree method of the FP tree is called. All the locations are found where the current node appears and frequent nodes on the path is

exported. A condTree is constructed, and then the splitSinglePrefix() method is called. The shared items are put in pTree while the bottom items are put in qTree. Then pTree and qTree can save the results obtained by mining frequent patterns up and down, respectively.

3.3. Algorithm performance analysis and test

In order to verify the performance of the parallel FP-growth algorithm, this paper uses the speedup ratio as the evaluation index of the performance experiment. The speedup ratio in this paper refers to the ratio of time consumed by a single task running in a stand-alone processing environment and a parallel processing environment. In the field of communication and computer, speedup ratio is often used to measure the performance of parallel processing. In this paper, we test the speedup ratio of the parallel FP growth algorithm under different slave node numbers and different test data set sizes (1MB, 1GB and 2GB respectively).

The experimental results are shown in Fig. 4. The experimental results show that the acceleration ratio is less than 1 when the amount of data is small. This shows that the performance of single FP growth algorithm is better than that of parallel FP growth algorithm when the amount of data is small. This situation is mainly because it takes a certain amount of time to start and call the algorithm based on MapReduce. When the amount of data is small, the running time of the algorithm is not long. However, the running time of Hadoop platform in parallel environment

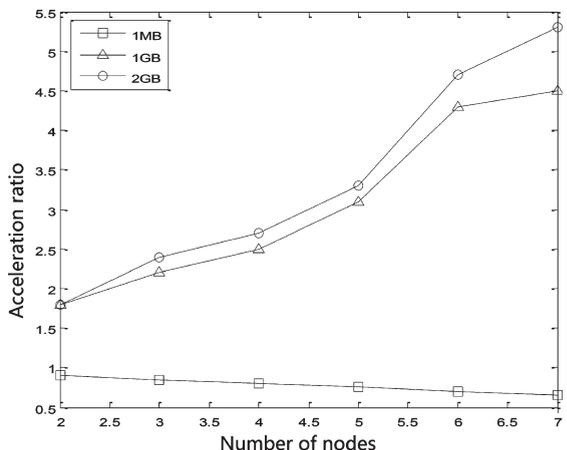


Fig. 4. Acceleration ratio test results.

causes the speedup ratio to be lower than that in single machine environment.

When the amount of data increases, the proportion of time consumed by the Hadoop platform decreases, and the advantages of the algorithm are reflected. It can be seen that the speedup ratio of the improved parallel FP growth algorithm for association mining gradually increases, which shows its superiority in the distributed environment. Therefore, the improved parallel FP growth algorithm based on MapReduce can greatly improve the search speed of frequent item sets in the case of large amount of data. We use this algorithm to apply to the data mining module of the big data platform for user health services, mining association rules for massive health big data. It can be used in the research and analysis of medical results and provide health guidance for users.

4. Implementation of health service platform

The platform is divided into three technical framework layers: platform layer, function layer and service layer.

(1) Platform layer

The platform layer build Hadoop cluster distribution is mainly used for data storage and reading. The database used is MySQL, and the file system includes HDFS and Linux local files. Upload local files to the HDFS file system. The files uploaded to the HDFS file system are calculated by the parallel FP-growth algorithm engine based on MapReduce to obtain the relevant data, and then copied to the local Linux system file. Then parse the local Linux system file and import it into the database persistent storage.

(2) Functional layer

The functional layer is mainly composed of two aspects, including Tomcat website server and parallel FP-growth algorithm engine. The Tomcat server mainly faces the web application layer. Struts2 intercepts requests from the web application layer, and then responds differently to different requests. The parallel FP-growth algorithm engine based on MapReduce is mainly responsible for the Tomcat server. For the Hadoop cluster built on the platform layer, enter the directory where the data files are located in the HDFS file system. Then call the parallel FP-growth algorithm based on MapReduce for calculation. Finally, the MapReduce process written by Hadoop is called to sort and analyze the HDFS output directory, and the results of the strong associ-

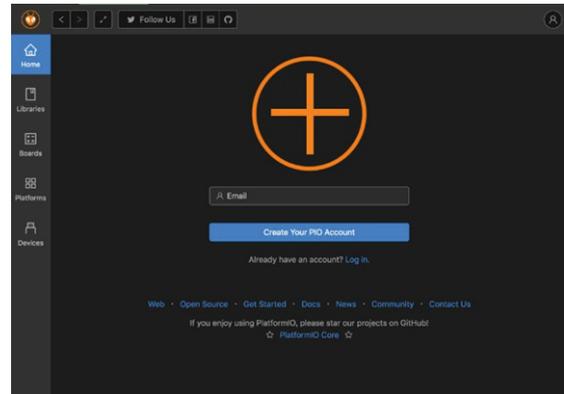


Fig. 5. The main interface.

ation rules obtained by the final mining are stored in the MySQL database.

(3) Service layer

The service layer mainly receives the user's request through JSP and HTML performance, and sends the corresponding request to the Struts2 interceptor, that is, the corresponding function layer module. Then the functional layer module performs corresponding calculations and data interaction with the platform layer to obtain corresponding result data. Then display in HTML or JSP on the front page.

After successful login, the user can enter the main interface of the big data analysis and processing platform, as shown in Fig. 5 below.

5. Conclusions

In the novel corona virus pneumonia epidemic situation, the analysis of health data helps to grasp the physical condition in time, and promote the improvement of epidemic prevention and control level. Especially for the new asymptomatic corona virus infection, the preliminary analysis of health data is helpful to find the possibility of virus infection to a certain extent. Through data correlation analysis on the physical health data mining of college students, it can help the immunity of college students to a certain extent to prevent the invasion of new coronary pneumonia virus. In the post-epidemic period, according to physical health big data, relevant parts of the school can adopt targeted training to improve the physical fitness of college students, so as to better prevent and treat new coronary pneumonia.

This paper mainly describes the implementation of big data platform for user health services. Firstly,

the development environment of the platform is introduced, including hardware environment and software environment. Then the Distributed installation of Hadoop big data processing platform is introduced in detail, including the steps of JDK installation, Hadoop cluster network setting, SSH password free login, Hadoop installation on name node and data node. After the above steps are completed, the background big data processing and analysis function is completed. Then the foreground is mainly developed by using Struts 2 framework. The innovation lies in the introduction of Hadoop technology to complete the construction and deployment of Hadoop distributed cluster in the engineering center. Struts 2 is used to deploy MVC design pattern and to define data management, data storage, data mining and other interfaces. This makes the control layer, model layer and view layer separate and reduces the coupling and at the same time. It is convenient to expand the function. At the same time, it is convenient for the interaction between the back-end Hadoop cluster and the front-end user health service platform, and the Hadoop cluster is connected to the platform for big data analysis and processing. The platform is implemented in a hierarchical structure. The specific content of each level is described in this paper. The experiment result shows that the system can effectively complete the processing and analysis of the big data of College students' physical health, which has a certain reference value for college students' physical health monitoring during the COVID-19 epidemic.

References

- [1] M. Nie, L. Yang, J. Sun, et al. Advanced forecasting of career choices for college students based on campus big data, *Frontiers of Computer Science* **12**(3) (2018), 26–35.
- [2] L. Dingkun, P.H. Woo B. Erdenebileg, et al. Application of a mobile chronic disease health-care system for hypertension based on big data platforms, *Journal of Sensors* **1** (2018), 1–13.
- [3] J. Huang, D. Zhu and Y. Tang, Health diagnosis robot based on healthcare big data and fuzzy matching, *Journal of Intelligent & Fuzzy Systems* **33**(5) (2017), 2961–2970.
- [4] M. Chen and S. Jiang, Analysis and research on mental health of college students based on cognitive computing, *Cognitive Systems Research* **56**(AUG.) (2019), 151–158.
- [5] M.J.Y. Suh, H.J. Yi, H.J. Kim, et al. Is asymmetric hearing loss a risk factor for vestibular dysfunction? lesson from big data analysis based on the Korean National Health and Nutrition Survey, *Otology & Neurotology* (2019), 1.
- [6] B. Xue and T. Liu, Physical health data mining of college students based on DRF algorithm, *Wireless Personal Communications* **102**(2) (2018), 1–11.
- [7] Chen, Hung-Ming, Chang, Kai-Chuan, Lin and Tsung-His, A cloud-based system framework for performing online viewing, storage, and analysis on big data of massive BIMs. *Automation in Construction* **71**(pt.1) (2016), 34–48.
- [8] W. Zhang, B.Y. Wang, X.Y. Du, et al. Big-data analysis: A clinical pathway on endoscopic retrograde cholangiopancreatography for common bile duct stones, *World Journal of Gastroenterology* **25**(8) (2019), 1002–1011.
- [9] Y. Cai, D. Li and Y. Wang, Intelligent crime prevention and control big data analysis system based on imaging and capsule network model, *Neural Processing Letters* (2020) 1–15.
- [10] A. Hanani, A.M. Rahmani and A. Sahafi, A multi-parameter scheduling method of dynamic workloads for big data calculation in cloud computing, *Journal of Supercomputing* **73**(11) (2017), 4796–4822.
- [11] S. Jiang, M. Lian, C. Lu, et al. Ensemble prediction algorithm of anomaly monitoring based on big data analysis platform of open-pit mine slope, *Complexity* **2018** (2018), 1–13.
- [12] W. Song, M. Xu and Y. Dolma, Design and implementation of beach sports big data analysis system based on computer technology, *Journal of Coastal Research* **94**(sp1) (2019), 327.
- [13] J.C.M. Tarazon, K. Suzuki and B. Hu, Big data analysis for improved mental healthcare: A population-based perspective, *European Psychiatry* **33** (2016), S606–S607.
- [14] D. Zhang, High-speed train control system big data analysis based on the fuzzy RDF model and uncertain reasoning, *International Journal of Computers Communications & Control* **12**(4) (2017), 577–591.
- [15] Y. Zhang, M. Chen, S. Mao et al. CAP: community activity prediction based on big data analysis, *IEEE Network* **28**(4) (2014), 52–57.
- [16] Q. Zhang, Z. Chen and L.T. Yang, A nodes scheduling model based on Markov chain prediction for big streaming data analysis, *International Journal of Communication Systems* **28**(9) (2015), 1610–1619.