Study on the intelligent evaluation model after stoke based on random forest

Yaru Ge$^*$

$^*$School of Software, Zhengzhou University, Zhengzhou, Henan, 450000, China
$^*$Corresponding author’s e-mail: geyaru@stu.zzu.edu.cn

Abstract: According to the summary of "China cardiovascular report 2018", in recent years, there are 2 million new stroke patients in China every year, and nearly 70% of them have lost hand function due to stroke. Therefore, the rehabilitation evaluation of postoperative hand function of stroke patients is complicated with low intelligence. In this paper, data collection methods are introduced firstly, and the original data are artificially classified. Then, starting from the decision tree method, the motion data of stroke patients are intelligently evaluated by establishing a random forest, and the evaluation results are returned to the patients for review in the form of floating point number based on their probability proportion. Finally, the results of the model experiment and its advantages and disadvantages are analyzed, and the future development direction and prospect are put forward.

1. Introduction

The summary of China cardiovascular report 2018 points out that in recent years, there were 2 million new stroke patients in China every year, and nearly 70% of them suffered from hand function loss due to stroke. Such patients have a high degree of living dependence and a significant decline in quality of life. In our country, the rehabilitation medical industry is at the early stage of development with few resources, and the medical personnel distribution is not balanced. So the present study of stroke rehabilitation of hand function is not perfect, and there is also the current situation of the shortage of corresponding rehabilitation equipment. It is a huge challenge for the health of the increasing stroke patients.

In this context, various new intelligent, miniaturized and portable devices have become the focus of research, which can be broadly divided into the following categories: exoskeleton based upper limb rehabilitation robot [1], air pump based upper limb rehabilitation equipment [2], and emg based rehabilitation equipment. Based on the above rehabilitation methods and considering the advantages and disadvantages of the former rehabilitation devices, we developed a new rehabilitation device based on mirror rehabilitation therapy [3] and based on the Internet of things -- the home intelligent rehabilitation system. The system consists of smart wearable devices, applications, servers and other parts.

After the innovative mirroring rehabilitation exercise, the patient can choose the rehabilitation evaluation function of the application program and conduct intelligent evaluation. The intelligent system will feedback the evaluation results in the form of floating point quantization.

2. Data collection

The sensor device will generate binary data when the stroke patient is wearing the smart wearable device for evaluation, which will specify the data frame encapsulated in hexadecimal 18-byte frame
format, and the first two bytes and the last two bytes of the data frame are the check bytes. The data frame is transmitted to the mobile application via Bluetooth 4.0, which automatically parses the MAC frame and verifies the data frame with hamming code algorithm [4].

The data will be stored in an array, and the mobile application will parse a series of discrete motion data. In order to parse the motion data, classification method is adopted to plan and process the data. The planned data will be calculated based on the difference between the current data and the previous.

\[ \Delta_2 = \text{Array}[\text{End}] - \text{Array}[\text{Begin}] \]  
\[ \Delta \begin{cases} 
\Delta_1 + \Delta_2, & \Delta_1 \cdot \Delta_2 > 0 \\
0, & \Delta_1 \cdot \Delta_2 \leq 0 
\end{cases} \]  

\[ \Delta_1 \] is for the previous one of \( \Delta \), \( \Delta_2 \) is the data difference between this data frame and the last data frame. And the value is updated as shown in formula (2), which will be encapsulated as an array of length 10, encapsulated [5] as a Jason string format, and uploaded to the cloud server.

3. Model construction
After obtaining the movement data of stroke patients on the server side, the data is stored in the cloud database.

3.1. Data classification and decision tree construction
The doctors divided the patients' exercise data into two categories: qualified and unqualified. The classified data constituted the primary data set. We extract its features and divide it into dexterity, myodynamia and duration, so as to construct the table structure as shown in figure 1:

| Dexterity | Myodynamia | Duration | Decision |
|-----------|------------|----------|----------|

Figure 1. Decision tree table structure.

The input part of the decision tree was constructed by randomly sampling the movement data of hundreds of groups of patients. After preprocessing the data, the information entropy, association entropy and trust entropy of each attribute were calculated according to the following formulas [6].

\[ \text{Ent}(D) = - \sum_{k=1}^{n} p_k \log_2 p_k \]  
\[ H(D|A) = \sum_{i=1}^{n} \frac{|D_i|H(D_i)}{|D|} = - \sum_{i=1}^{n} \frac{|D_i|}{|D|} \sum_{k=1}^{K} \frac{|D_k|}{|D_i|} \log_2 \frac{|D_k|}{|D_i|} \]  
\[ g(D, A) = H(D) - H(D | A) \]

Ent(D) is the information entropy, where pk is the proportion of sample k in the sample set d. The smaller the value of Ent(D) is, the higher the purity of the sample set D is. Finally, it can be found that the attribute value of flexibility has the greatest influence on the evaluation of patients' movement, among which the index finger is the most important reference index. Finally, the ID3 decision tree model as shown in figure 2 is constructed.
3.2. Establishment of random forest model

Based on the decision tree model constructed in 3.2, in order to further improve the accuracy of judgment, 10 samples were randomly put back for a total of 20 times. 20 training sets were generated, and for the 20 training sets, we trained 20 decision tree models respectively. For the single decision tree model, pruning was not required during the splitting process of the decision tree, and multiple decision trees were generated to form a random forest.

For the classification problem, the final classification result is decided by voting according to several tree classifiers. This builds the final model to complete the evaluation as shown in figure 3.
3.3. Intelligent evaluation

Based on the random forest model established in 3.3, the rehabilitation index of patients can be accurately evaluated. The rehabilitation index of patients' hand function can be obtained by the following formula:

\[
\text{grade} = \frac{N_1}{M_1} \times 100 \times P_1 + \frac{N_2}{M_2} \times 100 \times P_2
\] (6)

- \(N_1\): Quantity of flexibility qualified data
- \(M_1\): Total flexibility data
- \(N_2\): Quantity of muscle strength qualified data
- \(M_2\): Total muscle strength data
- \(P_1\): Flexibility weighting factor
- \(P_2\): Muscle strength weighting factor

The above results are stored as an array, and then encapsulated as JSON data and fed back to the mobile application, as shown in figure 4:

![Figure 4. Quantitative evaluation index.](image)

4. Experiment and comparison

After the clinical trial tracking of 5 stroke patients, a total of 217 groups of exercise data were collected. The data of the 217 groups were evaluated by using the intelligent evaluation algorithm based on random forest, and it was found that the qualified rate of data was 34.37%, the rejected rate was 65.63%, and the error was controlled within 5 times. The model has an accuracy of 97.7%. The experimental results show that the intelligent evaluation model based on random forest is accurate and reliable.

5. Summary and outlook

The intelligent evaluation algorithm based on random forest model can accurately diagnose the functional recovery of patients. The correctness and reliability of the model can be verified experimentally and theoretically.

5.1. Deficiencies of the data set

Problems following the model, the model adopted by the training data set was derived from the movement of different patients, so the random forest model on this basis to build is not good for individual patients, which is the intelligent evaluation model based on random forest biggest shortage,
but if for each patient to establish its unique random forest model will cost huge financial and material resources. Therefore, it is difficult to realize personalized intelligent evaluation.

5.2. Future expectation
Believing that with the improvement of the patients with intelligent evaluation system data set, the degree of refinement of data will be improved. Random forest will be combined with BP neural network [7] to improve the intelligent evaluation system, and the feedback mechanism of BP neural network will be used to continuously improve the fitness of the model, making it a personalized, high-precision and intelligent evaluation system.

In the future, popular intelligent analysis system can not meet the needs of users, then personalized and intelligent development will become the development trend of intelligent algorithms, artificial intelligence will be more advanced.

6. Conclusion
In this paper, the current rehabilitation methods and evaluation methods for post-stroke patients were reviewed and studied. Based on the traditional machine learning methods and deep learning methods, an intelligent evaluation system was established for the deficiency of existing technologies. At the same time, this paper also compares and analyzes the different experimental results of the intelligent evaluation system based on random forest. Finally, the future research direction and potential development trend are discussed.

References
[1] Yi Li, Suiran Yu. Design of an upper extremity exoskeleton active strength trainer for convalescent stroke patients [J]. Mechanical design and research. 2019, 35(04):5-10.
[2] Deng Yang, Wenhui Liu, Congxiao Wang, Jie Wang, XiShu Yan, Yanping Peng, Ying wang. Upper limb rehabilitation robot adjuvant therapy on upper extremity function in patients with cerebral apoplexy hemiplegia rehabilitation [J]. The application of nerve damage and function reconstruction, 2019, 14(09):447-449.
[3] Fucai Huang, Tianjia Liang. Effects of mirror therapy on lower limb balance and motor function in stroke patients with hemiplegia [J]. Journal of Guangxi Medical University, 2019, 36(09):1506-1509.
[4] Fengfu Du. Automatic error checking and correction of hamming code in PIC MCU serial communicaton [J]. Application of MCU and embedded system, 2004(08):49-50.
[5] Zhenfan Ding. Jason message encapsulation and parsing of Spring rest-style Web services [J]. Intelligent computer and application, 2012, 2(02):9-10+17.
[6] LunaJoséMarcio, Gennatas Elstathios D, Ungar Lyle H, Eaton Eric, Diffenderfer Eric S, Jensen Shane T, Simone Charles B, Friedman Jerome H, Solberg Timothy D, Valdes Gilmer. Building more accurate decision trees with the additive tree.[J]. Proceedings of the National Academy of Sciences of the United States of America, 2019, 116(40).
[7] Chun Yan, Meixuan Li, Xiao Zhou. Optimization of BP neural network based on improved genetic algorithm for vehicle insurance fraud identification model [J/OL]. Journal of Shandong University of Science and Technology (natural science edition):1-9[2019-10-22]. https://doi.org/10.16452/j.cnki.sdkjzk. 2019.05.009.