Balanced risk assessment system for the elderly based on factor analysis and random forest

Xinlin Ying

Abstract

For the elderly, their balance ability is greatly reduced due to age, leading to accidents such as falls. In this paper, a balance ability assessment system is established to help elderly people to perform balance ability assessment, thus helping them to correct their movement posture and reduce the occurrence of accidents.

First, the data were pre-processed to remove individuals with serious data loss and to fill in individual data for individuals with missing data. Then, according to the 19 node model in the literature, the corresponding 25 links were selected from 42 monitoring points, and the center of gravity of the human body was solved by the multiplicative coefficient method to calculate the number of risks at each moment.

Second, stepwise regression was performed with the number of falls as the dependent variable and the amplitude of left and right sway as the independent variable. The correlation coefficient was used to express the effect of left and right sway amplitude on falls. Then random forest was used to analyze right and left sway amplitude, BMI and age to derive their influence factors about falls.

Finally, evaluate system were assessed based on the entropy weighting method.
We found that swing range and BMI have similar effects on falls, but age is not really the most important consideration in the process. Therefore, we should actively correct the walking posture of the elderly in order to play a preventive role.

Keywords: Multiplication coefficient method, random forest, Entropy weight method

1. Restatement of the Problem

Based on the data provided, establish a feature extraction model through analyzing steps, center of gravity and motion. Meanwhile, 25 body balance features coming from 42 monitoring points should be extracted to evaluate the overall body balance of the elderly. On the basis of the 25 indicators in question I and the influencing factors in annex I, establish a balanced risk assessment system. Moreover, propose improvement suggestions or precautions for the elderly with different abilities.

2. Assumptions of the Model

1. Ignore individual differences in innate balance

3. Notations

Symbol	Definition
x	Abscissa of each monitoring point
y	Ordinates of monitoring points
z	Vertical coordinates of monitoring points

4. Establishment and Solution of the Model

4.1 Modeling based on Coefficient Method

4.1.1 Concepts of the Correlation

Balance
When the human body begins to move, it is often subject to sudden internal or external disturbances, and the elderly often fail to respond quickly and appropriately, resulting in an accidental fall. When disturbed, the position of the body's center of gravity will change, and the body's balance ability is to ensure that the body's center of gravity always remains on the standing support surface through the change of the spatial position of the limbs. In addition, the height of the center of gravity is also related, the smaller the support surface, the higher the center of gravity, the more difficult it is to maintain balance. The greater the left and right or up and down swing is also the more likely to fall

4.1.2 Modeling based on the Extraction of center of gravity

From the 42 points, 25 points that are distributed over most parts of the human body and can represent the walking characteristics of the elderly were selected. The body balance of the elderly was evaluated from the two aspects of the walking height and shaking degree of 80 elderly people.

4.1.3 Coefficient Method

The coordinate value of each joint's center of the body multiplies by the corresponding coefficient, then their sum is the coordinate value of the center of gravity of the body. It comes from the analysis method with high accuracy (namely torque synthesis method).[1]

4.1.4 Selection of the Node and Extraction of the Coordinate

At present, the joint coefficient provided by the German barycenter data is widely used to construct the calculation model of the barycenter of the human body, which contains 19 human nodes.[1] However, the 42 nodes given in the problem cannot correspond one to one with the nodes in the coefficient synthesis method, which is specifically shown in the following aspects:

• In the data provided, there are no coordinates of the center of gravity of the head, the midpoint of the two shoulder lines and the center of the two hip lines. Therefore, the 3d coordinates of feature points 41 and 42 are averaged to obtain the coordinates of the center of gravity of the head. The coordinates of feature points 16 and 17 were averaged to obtain the coordinates of the midpoint of the two shoulder lines. The coordinates of feature points 13 and 14 were averaged to obtain the central coordinates of the two hip lines.

• The center point of each human body module point after using the provided feature points and constructing the torque synthesis model is shown in figure 1.1. The corresponding coefficients and coordinate sources of each human body module point in the model are shown in table 1.1. where the red point is the selected point and the blue and yellow points are the points to be verified

$$X_e(t) = \sum_{i=1}^{19} k_i * X_i(t)$$

$$Y_e(t) = \sum_{i=1}^{19} k_i * Y_i(t)$$

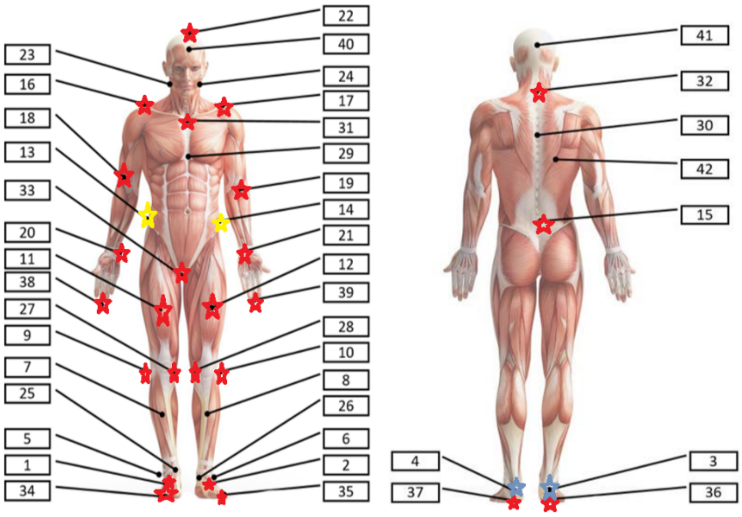


Fig. 1: The layout of the monitoring points

Figure 1.1

Table 1.1 Human joints, coefficients and coordinate source

The body parts	Coefficient of link	Origin of coordinates （right; left）
the center of gravity of the Head	0.0706	The average of part 41 and part 42
shoulders	0.0356	16; 17
The midpoint of the shoulder line	0.2391	The average of part 16 and part 17

elbows	0.0580	18；19
wrists	0.0192	20；21
the center of gravity of the hands	0.0180	The average of part 20 and part 38；The average of part 21 and part 39
hip	0.1297	15
The midpoint of the hip line	0.1879	The average of part 13 and part 14
knees	0.1630	The average of part 9 and part 27；The average of part 10 and part 28
ankles	0.0643	The average of part 5 and part 25；The average of part 6 and part 26
heels	0.0158	36；37

In the third column of the table, the coordinates of the left and right sides of the human body, such as left shoulder and right shoulder, so ‘;’ are used to separate them.

Table a total of 11 human links, using 25 of the 42 feature points given in the problem

4.1.5 Determination of chance to fall down

Once we know the body xyz coordinates, the key to a balance is that the height of the body cannot change to a great extent during a continuous period of time

$$\begin{cases} z > \dfrac{\max(z)-\min(z)}{2}+\min(z) \\ \left|z(i+1)-z(i)\right| < 5 \end{cases}$$

$$\begin{cases} z < \dfrac{\max(z)-\min(z)}{2}+\min(z) \\ \left|z(i+1)-z(i)\right| < 5 \end{cases}$$

4.2 Classification of indicators

4.2.1 Factor analysis

Factor analysis is to decompose the original variable into a linear combination of several common factors, so as to better understand the internal relationship of the original variable.

Set $X_i(i=1,2,...p)$ be p variables, it can be denoted as:

$$\begin{cases} X_1 = L_{11}F_1 + L_{12}F_2 + L_{13}F_3 + \ldots + L_{1m}F_m + \varepsilon_1 \\ X_2 = L_{21}F_1 + L_{22}F_2 + L_{23}F_3 + \ldots + L_{2m}F_m + \varepsilon_2 \\ \ldots \\ X_p = L_{p1}F_1 + L_{p2}F_2 + L_{p3}F_3 + \ldots + L_{pm}F_m + \varepsilon_p \end{cases}$$

$$X_i = L_{i1}F_1 + L_{i2}F_2 + L_{i3}F_3 + \ldots + L_{im}F_m + \varepsilon_i$$

$$X = LF + \varepsilon$$

$F_1, F_2, \cdots, F_m$ are called common factors, which are unobservable variables. $L = (L_{ij})_{p \times m}$ is called a factor loading matrix. ε_i is a special factor, which cannot be included by the first m common factors. And it satisfies $\operatorname{cov}(F, \varepsilon) = 0$, meanwhile, F, ε are irrelevant.

Steps of the factor analysis:
1. Find the potential common factors $F_1, F_2, \ldots, F_m$;
2. Calculate factor loading;
3. Explain the relationship between variables.

Age, BMI, swing range were used for factor analysis, so as to classify the four indicators and establish a risk assessment system.

4.2.2 Risk assessment system based on Random Forest

Machine Learning (ML) is cool. It's what AI is all about. It's a bit of reverse Science: Instead of building a theoretical model with traditional Science, and then use that theoretical model to predict the future, we use data and from that data we build a statistical model of the data, and then use that model to predict the future. It's so cool, we need a preview of this.

Random forests (RF) are cool, invented in the 1950s[5] (as a decision tree algorithm) and optimized using in modern libraries like Scikit-Learn

4.3 Evaluation based on entropy weight method

Information entropy borrows the concept of dryness in thermodynamics to describe, on average, the amount of information about an event. According to the definition of information flag, entropy value can be used to judge the high dispersion degree of an index for an index. The smaller the entropy value is, the higher the high dispersion degree of the index is. The greater the influence of the index on the comprehensive evaluation (that is, the weight) is.

The procedure of objective weight determination by entropy weight method:
Step 1 Data Standardization

Firstly, the data is preprocessed to remove the data dimension. The specific method is as follows:

$$x_{ij}' = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}$$

x_{ij} is the jth index of the ith elderly, the normalized version is x_{ij}'; $\max(x_j)$ is the maximum value of the jth index; $\min(x_j)$ is the smallest value of the jth index.

Step 2 Calculate the entropy of each index

$$E_j = -\frac{\sum_{i=1}^{n} y_{ij} \ln y_{ij}}{\ln n}$$

E_j is the information entropy of the ith index; y_{ij} is the proportion of the jth index of the ith elderly.

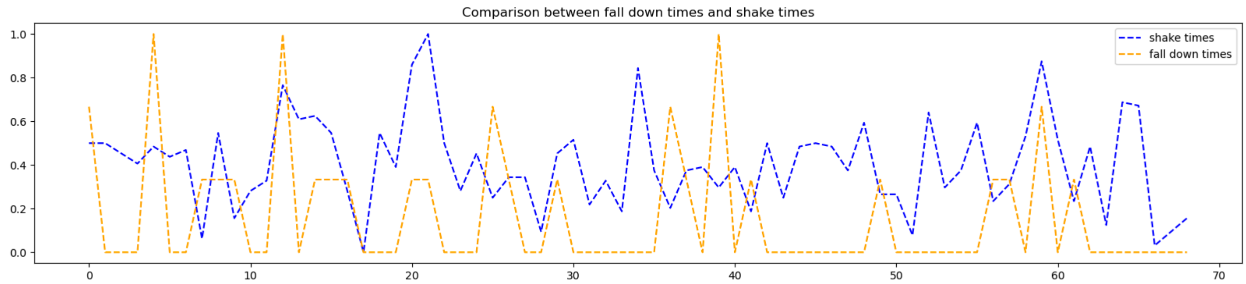
Step 3 Calculate the weight of each index

$$w_j = \frac{1 - E_j}{m - \sum_{j=1}^{m} E_j}$$

w_j is the weight of the jth index.

5. Result and Evaluation

5.1 The chance that old people might fall down VS times of old people fall down in a year



Use pearson correlation coefficient to analysis the the connection between swing range and fall down times. p value is small which means they have strong connections. PearsonRResult(statistic=0.12478894933391799, pvalue=0.3069522680264505)

5.2 Factors' weight of some specific old people

Name(age): wen yanfang（81）

table 2.1 weight based on RF and EWM for wen yanking

	BMI	Age	walking posture
Weight in random forest	0.43	0.29	0.28
Weight in EWM	0.36	0.34	0.30

More attention should be paid to the adjustment of overall body situation rather than the walking posture. Reduce the hunchback to change BMI.

Name(age): wang deqin（67）

table 2.2 weight based on RF and EWM for wang deqin

	BMI	Age	walking posture
Weight in random forest	0.30	0.12	0.58
Weight in EWM	0.45	0.15	0.40

Walking posture should be given priority to and internal feedback and regulation mechanism should be improved.

Name(age): liu zaoheng（76）

table 2.3 weight based on RF and EWM for lie zaoheng

	BMI	Age	walking posture
Weight in random forest	0.33	0.30	0.37
Weight in EWM	0.38	0.29	0.33

Treatment of the health situation and adjustment of posture are both needed.

table 3.4 weight based on Random Forest over all samples

index	BMI	Age	walking posture (swing range)
weight	0.46	0.26	0.28

table 3.5 weight based on EWM over all samples

index	BMI	Age	walking posture (swing range)
weight	0.38	0.24	0.38

We can see the value of age, swing range and BMI is around 0.26, 0.28, 0.46, which is close to the random forest algorithm result, 0.34, 0.38 and 0.28. Means the Random forest give a good analysis. This result also indicates that age, are not that important when consider whether an old people will fall down or not. What's most important is the gesture and the situation of the old people.

6. Discussion

Advantages and Disadvantages of the model

- Model advantages: Based on the study of human center of gravity 25 points are taken to calculate the center of gravity of the body, and the concept of human balance is reasonably quantified to determine the likelihood of fall of the elderly. The influence factors between the indicators were given intuitively according to the random forest algorithm. The entropy weighting method was used to evaluate the weights and to test the significance of the objective indicator assignment according to the actual meaning.

- Model disadvantages: Didn't make full use of the information in the attachment, didn't consider the situation of going up stairs and climbing slope. Some elderly people may go upstairs to help them, which will reduce the probability of falling down.

Reference

[1] 王蔚,冯亚琴,杨再兴,王晓燕.基于体感交互设备的人体重心计算方法.[J].南京师范教育科学学院.2018,33(4):596-599.
[2]Jin-Zhuang Xiao, Zhi-Fang Yang, Hong-Rui Wang, Xin-Cai Yang.[J]. Detection Method of Human Three-Dimensional Body Center of Gravity Based on Inclinometer Network. 2017,29(7): 1081-1087.
[3]Duan Zengwu. Capturing and Application of Human Motion Information Based on Kinect[D]. Hebei: Hebei university.
[4]Jiao shanshan. Influence of aging on body dynamic posture stability at the start of walk-ing [D]. Tianjin institute of physical education, 2017.
[5]https://en.wikipedia.org/wiki/Ensemble_learning