Fitness Movements Recognition and Evaluation Based on LSTM

Yi Wang, Xiaowen Zhu and Chengzhang Qu*

College of Information Engineering, Wuhan Business University, No. 816 Wuhan Economic and Technological Development Zone, China
Email: quchengzhang@foxmail.com
*Corresponding Author

Abstract. In this paper, we use Kinect to obtain human skeleton information, then extract 24-dimensional eigenvalues to represent the dynamic movements. A lighted refined LSTM model is employed to recognition the movement serious, finally we propose a movement evaluation model to mainly avoid sports injured. For experiment, we collect 5 kinds of movements as 250 movement samples totally. Using our model, the accuracy can get 80% as the accuracy we test on MSR dataset is 82%. At last, the vulnerable judgment for fitness movement is calculated based on characteristic angles of each dynamic action which can also enlarge the effect of fitness.

1. Introduction
Nowadays, information technology has been fast developed all over the world, and it makes our daily life more convenient. In recent years, extracting the features of 3D human skeleton informations by Kinect to recognize the movements is becoming more and more popular.

On the other hand, the new technology about the Artificial Intelligence has tremendous progresses in pattern recognition. The most popular technology is deep learning. RNN(Recurrent Neural Network) and LSTM(Long Short-Term Memory) are famous deep learning models to deal with the sequence data.

To classify the different actions, as deep neural network shows the powerful capability in various classification work we employ a neural network model to deal with the dynamic actions. Deep learning has more neural net layers and more complex processing to achieve better performance on classification or recognition tasks. Neural network is widely used, including industrial parameter optimization, modification of parameters in medical image Analysis, risk prediction in industrial industry, etc. There are two methods to used widely, RNN and LSTM, which have better effect in time serious classification problem. RNN is suitable for handwritten recognition or voice recognition. But RNN has Long-term dependency, it can’t identify which is importance in time serious. LSTM model can fit the problem based on its particular structure, thus this paper chooses LSTM neural network model as the classification model.

Based on the weight of eigenvalue, the score of action is obtained by synthesizing penalty factor. Perframe in the dynamic action is scored by vulnerable angle, We get the average value of all frames during the action, average value of the score is the score of the action.

In this paper, we employ a light refined LSTM model to recognize the movement with 3D features, then evaluate the movement by calculating the score to both prevent the sports injury and enlarge the fitness effect.
2. Model of Classify

The neural network algorithms include RNN, LSRM and so on. RNN is used in prediction and classification based on a large of data. LSTM fits to analyse data which time series has important events which have long interval and delay, such as speech recognition, emotional analysis and gesture recognition[9][10]. In cells of LSTM, the outputs are adjusted by adding some gates, including forget gate, input gate, and output.

![Figure 1. Model of LSTM](image)

The forget gate reads the output value of the previous cell and input values for this cell, it outputs a figure in 0 to 1 whether the previous results need to be abandoned. $y_{t-1}$ is an output of previous cell, $x_t$ is an input of this cell. The limitations of the forget gate $f_t$ followed.

$$f_t = \delta(W_f \cdot [y_{t-1}, x_t] + b_f) \quad (1)$$

The sigmoid and tanh layers of the input gate need to be constantly debugged, and the cell status is updated to set the appropriate values. The state of the former cell is calculated with the forget gate and the input gate of this cell to decide the change of state of the cell. $\tilde{c}_t$ is status of the previous cell, $c_t$ is the status of this cell, $i_t$ represents the input threshold.

$$i_t = \delta(W_i \cdot [y_{t-1}, x_t] + b_i) \quad (2)$$
$$\tilde{c}_t = \tanh(W_c \cdot [y_{t-1}, x_t] + b_c) \quad (3)$$
$$c_t = f_t \cdot \tilde{c}_{t-1} + i_t \cdot \tilde{c}_t \quad (4)$$

The output gate needs the sigmoid layer to determine the output part, through processing with the tanh layer. $o_t$ represents the output threshold, $y_t$ represents the current cell output.

$$o_t = \delta(W_o \cdot [y_{t-1}, x_t] + b_o) \quad (5)$$
$$y_t = o_t \cdot \tanh(c_t) \quad (6)$$

3. Experiment

3.1. The Type of Actions

Five movements were used in the experiment, there are basketball shooting, serve of badminton, horse walking of martial arts, big tree of yoga and sidekick of taekwondo. In the experiment, there are a total of 10 experimenters. They do 5 kinds of actions, which each experimenter does 5 times. Altogether, we collects 250 data. We judge the injury-prone angle of the data and get the result of the score.
3.2. Parameter of LSTM

The paper uses the keras in LSTM to classify the dynamic motion data. We convert the skeleton data which is collected by Kinect into a format and we set the maximum frame number of 200 frames. We have five types of actions, and then we divide 250 sets of data into train data and test data, according to a scale.

The acquired data is randomly selected from each class as test data, and the rest is training data. The list of the data is converted into array. Because there are five types of actions, the label for each action of test and train is expressed as a list of 5D. For example, the mark of the first action is [1,0,0,0,0]. Label also needs to be converted to array. The train data’ shape is (200,200) and test data’ shape is (50,200).

For 250 groups of data, some data are extracted from each class as test data and the others as training data. Through the test and training data are divided by 1:4 and 2:3 to carry on the contrast test. And the classification effect of 60-dimensional data obtained by Kinect is compared with that 24-dimensional eigenvalues are extracted. We obtain a model with an average accuracy of 80%. We also use the model to classify MSR's skeleton model and get 82% accuracy.

We use the KNN model to classify movements. We extract feature data of perframe, and compare the length of test data with the length of train data, the short data should padding the last frame data to the same length. The sum of perframe of feature data, we add them to get a final distance of two movements. And we compare the difference of 5 template data, the classification of least value is result. We get 67.2% accuracy.

4. Our Method

4.1. Extract Feature Data

We extract the valuable joints to get angle and distance from the 20 skeleton points that has 60 dimensions. According to the coordinates of the node, feature distances is calculated. And the angle between the two vectors of the adjacent joints is calculated by the feature angle. Finally, 12 dimensional distance features and 12 dimensional angle features are extracted for classification. Blue angles and red distances are shown below.
4.2. Movement Recognition
We use LSTM model to classify the data by extract feature data. More specifically, we normalize the length of each action, we use feature data of perframe and LSTM model to classify. We get accuracy of 80%, and we compare the result with KNN model, LSTM model is better. The more discussion see section 3.

4.3. Method of Evaluation
Deduction points are appropriate for the angular displacement that may cause injury and the angle of operation that is not up to standard. The angles of easy injury include a range of angles, such as the shoulder, neck, elbow, heel, knee, spine and standing center of gravity. Normalize the scoring range of feature angles that are extracted from motion, then we give the segmented functions $f^{[14]}$

$$f = \begin{cases} 
0 & \theta \leq \beta_1 \\
\frac{1}{\cos \alpha - \cos \beta} \cdot \cos \theta \cdot \frac{\cos \beta}{\cos \beta - \cos \alpha} & \beta_1 < \theta < \alpha_1 \\
1 & \alpha_1 < \theta < \alpha_2 \\
\frac{1}{\cos \alpha_2 - \cos \beta} \cdot \cos \theta \cdot \frac{\cos \beta}{\cos \beta - \cos \alpha_2} & \alpha_2 < \theta < \beta_2 
\end{cases} \quad (7)$$

The vulnerable condition is variable for each joint angle. Here are the criteria for the function of the vulnerability angle. As shown below.
Figure 4. Vulnerable angle display diagram

Figure 4 draws the gray plane as the body plane, using the left shoulder joint, right shoulder joint and spine joint. Calculating normal vector $\mathbf{n}$ of body plane, which is used to describe the orientation of the body in three-dimensional space.

The angle named $\alpha_6$ between the vector of “shoulder center-spine” and the “hip-knee”, it indicates the angle that the leg is raised. $\alpha_7$ is the angle of the “hip center-left knee” and “hip center-right knee” vectors used to describe hip joint stretching. $\alpha_{11}$ is the tilt of the standing foot. The detailed calculation steps of the angle are followed.

1. Take the shoulder joint as a point, make the vertical line of the shoulder joint.
2. Across the elbow joint as a plane, intersecting the vertical line to the center r1 of the circle.
3. At the center of the circle, the xyz axis which is relative to the Kinect camera is established.
4. Over the center making a vector $\vec{n}$ which is parallel to normal vector of body plane.
5. Join the vector “center of circle-elbow joint”, we calculate the angle $\alpha_6$ or $\alpha_7$ by the vector and the normal vector $\vec{n}$.

5. Conclusion
This paper collects the dynamic data of five actions, there are basketball shooting, serve of badminton, horse walking of martial arts, big tree of yoga and sidekick of taekwondo. To classify these data by
LSTM neural network. We get 80% accuracy of LSTM model, and we get 67.2% accuracy of KNN according to the 250 data. In the dynamic motion score, we can increase the score of the standard degree of movement, and combine the score of vulnerable angle to evaluate comprehensively.

6. Acknowledgement
Fund by National College Students' innovation and entrepreneurship training program (No.201711654009)

7. Reference
[1] Mike Schuster and Kuldip K. Paliwal, Bidirectional Recurrent Neural Networks, Trans. on Signal Processing 1997
[2] GERS, Felix A, SCHMIDHUBER, et al. Learning to forget : Continual prediction with LSTM[J]. Neural Computation, 2000, 12(10):2451-2471.
[3] Aleksendrić, Bellini, Carlone, Ćirović, Rubino, Sorrentino. Neural-fuzzy optimization of thick composites curing process[J]. Materials and Manufacturing Processes, 2019, 34(3).
[4] Chuyang Ye, Xiuli Li, Jingnan Chen. A deep network for tissue microstructure estimation using modified LSTM units[J]. Medical Image Analysis, 2019, 55.
[5] Peng Xu, Rui Du, Zhongbao Zhang. Predicting pipeline leakage in petrochemical system through GAN and LSTM[J]. Knowledge-Based Systems, 2019, 175.
[6] Wang J, Yang Y, Mao J, et al. Cnn-rnn: A unified framework for multi-label image classification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 2285-2294.
[7] Ordóñez F, Roggen D. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition[J]. Sensors, 2016, 16(1): 115.
[8] Lipton Z C, Kale D C, Elkan C, et al. Learning to diagnose with LSTM recurrent neural networks[J]. arXiv preprint arXiv:1511.03677, 2015.
[9] Chang Su, Junchao Li, Ying Peng, Yijiang Chen. Chinese metaphor sentiment computing via considering culture[J]. Neurocomputing, 2019.
[10] Fitness Movement Recognition and Evaluation Based on Kinect[J]. Computer Science and Application, 2018, 08(07): 1134-1145.
[11] Liu L, Wu X, Wu L, et al. Static human gesture grading based on Kinect[C]//5th International Congress on Image and Signal Processing. IEEE, 2012: 1390-1393.