Quantitative Taxonomy of Hand Kinematics Based on Long Short-Term Memory Neural Network

Chenguang Li[1,2], Hongjun Yang[1], Shiying Sun[1], and Long Cheng[1,2](✉)

[1] State Key Laboratory for Control and Management of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China
{lichenguang2017,hongjun.yang,shiying.sun,long.cheng}@ia.ac.cn

Abstract. Assessment of hand motor function is crucial to stroke patients. However, the commonly used Fugl-Meyer assessment (FMA) scale requires 11 hand and wrist movements. To simplify these movements, this study proposes a hand motion classification framework based on deep learning to achieve the quantitative taxonomy of hand kinematics. To the best of our knowledge, this is the first study to use deep learning for the quantitative taxonomy of the hand kinematics. First, we use the Long Short-Term Memory (LSTM) neural network to extract deep features of 20 hand movements (including 11 FMA movements) from 37 healthy subjects, and rank the LSTM neural network output value (predicted probability) of each sample. The similarity between the movements obtained by the nonlinear transformation can be used to draw the confusion matrix. Then the confusion matrix is taken as the category feature to obtain the clustering dendrogram to show the similarity between different hand movements intuitively. Next, the 20 hand movements are divided into four groups by hierarchical clustering. The silhouette coefficient of the clustering results is 0.81, which is close to the ideal value of 1, indicating the validity of the clustering result. Finally, the clustering center is calculated to find the corresponding movement as the representative movement for motor function assessment. As a result, we reduced the 20 movements to 5 movements, allowing for a faster quantitative assessment of hand motor function than the FMA scale. This also lays the foundation of the assessment paradigm for our follow-up research on evaluation algorithm.

Keywords: Hand kinematics · Hand motor function assessment · Deep learning · LSTM · Hierarchical clustering

Project supported by the Beijing Municipal Natural Science Foundation under Grant JQ19020, National Natural Science Foundation of China under Grant 62025307, Grant U1913209, Grant62073319, Grant62103412 and China Postdoctoral Science Foundation under Grant 2021M693403.

T. Mantoro et al. (Eds.): ICONIP 2021, CCIS 1517, pp. 221–229, 2021.
https://doi.org/10.1007/978-3-030-92310-5_26

1 Introduction

Hand motor function is usually evaluated and judged by the hand's performance in fine grasp and strong grasp. However, these two aspects are obviously not enough to fully assess the complex functions of the hand. Fugl-Meyer assessment (FMA) scale is one of the most common subjective assessment methods, which can comprehensively evaluate the hand motor function of stroke patients. However, it is time consuming to evaluate so many movements. Can we simplify and optimize these movements? In order to better understand the law of hand movement, it is necessary to analyze the hand kinematics.

Santello et al. proposed an early method by applying principal component analysis (PCA) to the finger joint angles under a set of significant grasping movements [1]. Many other works that followed [2–4], were inspired by the grasp of taxonomy to select the correct hand movements. Until the past two years, the analysis of hand kinematics still continued. In [5], the hand kinematics in activities of daily living was characterized by PCA, and five synergistic effects were obtained. The 20 human hand grasping movements are classified, and the Mahalanobis distance between different movements is quantitatively analyzed in [6]. After PCA dimensionality reduction and hierarchical clustering of the finger joint movement data of 20 hand grasping movements, three synergistic effect modes were obtained [7]. There are also some studies introducing methods for analyzing hand kinematics based on sEMG signals [6,8].

Although there have been some studies on the analysis of hand kinematics, the movement of the wrist joints was ignored. The wide range of wrist movement enhances the functions of the hands and fingers, while providing sufficient stability. Therefore, the analysis of the hand kinematics cannot be separated from the wrist joint.

In [6], the use of manually extracted features for kinematics analysis is inferior in the ability of generalization and robustness. It is well known that deep learning can automatically extract appropriate deep features based on the tasks by using neural networks, which has been widely confirmed in convolutional neural networks. Therefore, long short-term memory (LSTM) is applied in this study. In addition, although the clustering result of hand kinematics was provided in [6], it did not give the evaluation metrics of the result, we thus do not know how effective the clustering was. This concern is addressed in this study. The contributions of this paper are summarized as follows:

1) Currently, most researchers have only studied the kinematics of the hand grasping movements, but we also analyze four wrist movements and two other finger movements, including wrist flexion, wrist extension, wrist supination, wrist pronation, thumb adduction and abduction of all fingers, facilitating a comprehensive quantitative taxonomy of hand kinematics.
2) A novel framework for the quantitative taxonomy of hand kinematics is proposed. It mainly includes the LSTM neural network to automatically extract the hand motion features and the nonlinear transformation to calculate the hand motion similarity. LSTM neural network can effectively extract deep

features with significant differences between categories. To the best of our knowledge, this is the first study using deep learning to perform quantitative taxonomy of hand kinematics.

3) The 20 hand movements are divided into four groups by the method of hierarchical clustering, and then the hand movements corresponding to the cluster centers are found. The silhouette coefficient of the clustering results is 0.81, which demonstrates the more effective performance than the result in [6]. As a result, we reduced the 20 movements to 5 movements, allowing for a faster quantitative assessment of hand motor function than the FMA scale.

2 Methodology

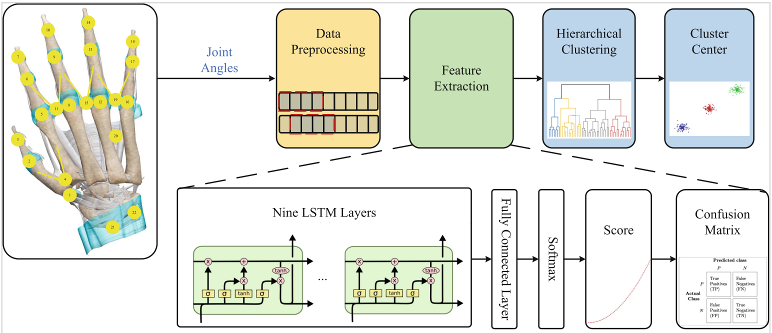

Fig. 1. The overall framework of the study

The proposed method mainly includes three parts: data capturing and preprocessing, feature extraction and hierarchical clustering. The overall framework of the study is shown in Fig. 1.

2.1 Data Capturing and Preprocessing

Data Capturing. The data were recorded from 22 finger joint angles of 37 intact subjects. There are more details about the experiment in [9]. In our study, the selected hand movements refer to the movements in the FMA scale, and are appropriately extended on this basis. All the 20 hand movements analyzed in this study are shown in Fig. 2. It should be noted that wrists and finger movements are crucial to the assessment of hand motor function, so they are added in the 20 hand movements.

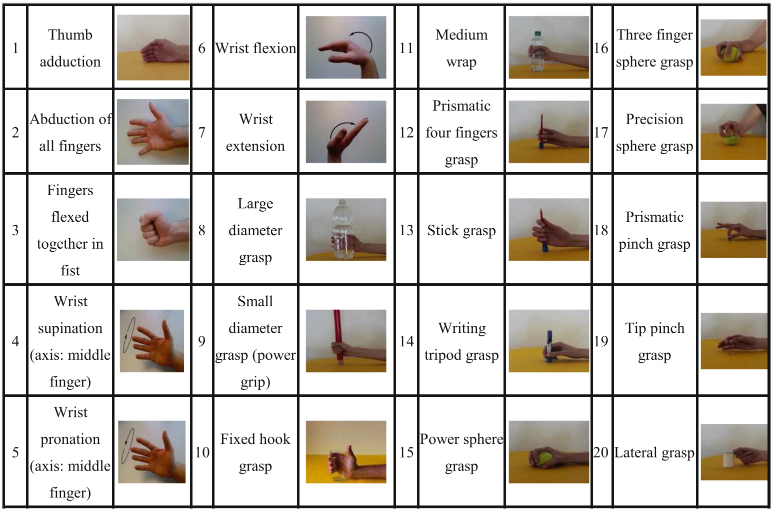

Fig. 2. Twenty different hand movements

Data Preprocessing. Since the data in two data sets (DB1 and DB5) are measured at different sampling rates (100 Hz 200 Hz respectively), a uniform sampling rate is required before data analysis. In this study, the resampling method [10] is used to unify the sampling rate. As the joint angles of the hand movements does not change very quickly, the sampling rate of the two data sets is unified 100 Hz. DB1 contains 20 types of hand movement data from 27 subjects, each of which is repeated 10 times; DB5 contains 20 types of hand movement data from 10 subjects, each of which is repeated 6 times. So the total number of movement repetitions is 27 * 20 * 10 + 10 * 20 * 6 = 6600, which make up the entire data set. The data set is randomly shuffled, 70% of which is used for training and 30% for testing. To increase the sample size, this study performs the sliding window method, and the window size and sliding distance are consistent with those in [6]. So each movement repetition has a window of 200 ms (20 sampling points), with an overlap of 100 ms (10 sampling points). The whole data preprocessing is performed with MATLAB R2019a, and the followed work is achieved by the Python language (version 3.6) based on the Tensorflow framework.

2.2 Feature Extraction

LSTM Neural Network. In this study, multi-layer LSTM network is introduced to extract the deep features of hand motion, which improves the generalization and robustness of the model, compared with manually feature extraction [6]. The network includes nine LSTM layers and one fully connected layer. The input size of the first LSTM layer is 22. The data of each time window is input to the LSTM in batches. The output size of each LSTM layer is 32, 64, 128, 256, 256, 128, 64, 32, 16 respectively and the output size of the fully connected layer is 20, which is activated by the algorithm of softmax. Sparse categorical cross entropy is used as the loss function. The Adams algorithm is used as an optimization method for network training.

Category Feature. The LSTM neural network is used to calculate the probability value of each sample, and the value is used to obtain the category feature. All data (including the training set and the test set) are input to the neural network, and the full connection layer outputs N (the total number of samples) vectors of 20 dimensions, where each movement contains $N/20$ vectors. The n-th sample vector of the m-th movement is represented by $p_{m,n} \in R^{20}$, where $m = 1, ..., 20; n = 1, ..., N/20$. The 20 elements in the vector represent the probabilities that this movement is respectively classified as the 20 movements. The elements in each vector are sorted from smallest to largest, and the sequence numbers are used to replace the corresponding elements in $p_{m,n}$ to form a new vector $p'_{m,n} \in R^{20}$. To maximize the distinction between the previous categories, each element in $p'_{m,n}$ is squared to get $p''_{m,n} \in R^{20}$, and then $N/20$ vectors $p''_{m,n}$ of each movement are summed to get the final score vector $S_m \in R^{20}$. Therefore, the score for 20 movements can be written as $Score = (S_1^T, S_2^T ..., S_{20}^T) \in R^{20*20}$. $Score$ is actually an initial confusion matrix, and the value of each element in the matrix describes the distance or similarity between the two movements represented by the abscissa and ordinate. A smaller value indicates that the two movements are more similar. It requires further normalization and diagonalization to obtain the final confusion matrix A as follows:

$$A = \left| \left(\frac{(Norm(Score) + Norm(Score)^T) * 100}{2}\right)^{\star} \right|, \tag{1}$$

where $Norm$ is the normalization operation, $\star$ is an operation that the rows and columns of the matrix are subtracted from the diagonal value, $||$ means take the absolute value. Confusion matrix A obtained by the above nonlinear transformation is used as the category feature of hierarchical clustering.

2.3 Hierarchical Clustering

Each column of the confusion matrix A is taken as the category feature, and the clustering dendrogram is obtained by the hierarchical clustering method. To judge the quality of the clustering results, the silhouette coefficient (SC) is used as a metric calculated by:

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \tag{2}$$

$$SC = \frac{1}{M} \sum_{i=1}^{M} s(i), \tag{3}$$

where $a(i)$ is the average distance between a movement i and other movements in the same cluster, $b(i)$ is the average distance between movement i and the movements of the other cluster, and $M = 20$ means the total number of movements, $SC \in [-1, 1]$. The clustering effect is better if SC is closer to 1.

Finally, we need to find the movement corresponding to the cluster center that has the minimum distance sum between the other movements in the same cluster, which could be calculated as follows:

$$SN(j) = \underset{i}{arg\,min}(a_j(i)), \tag{4}$$

where $SN(j)$ represents the sequence number of the movement of the j-th cluster center, i is the movement in the j-th cluster, $a_j(i)$ is the average distance between a movement i and other movements in the j-th cluster. The cluster center movements will be used to evaluate the hand motor function of stroke patients.

3 Results and Discussion

3.1 LSTM Neural Network

After 15 epochs of training, the classification accuracy of LSTM neural network can reach 94.43%. If all the samples (train set and test set) are put into the network to obtain the category probability for each movement, we get that the accuracy of Top-1 classification is 94.69%, and the accuracy of Top-5 classification is 99.90%. The results show the deep feature extracted by the LSTM neural network leads to a high classification accuracy through the basic softmax classifier, indicating it can appropriately represent the characteristics of the corresponding category.

3.2 Confusion Matrix

The confusion matrix including 20 hand movements is shown in Fig. 3(a). The horizontal and vertical coordinates represent 20 movements, respectively. The labels 1–20 are consistent with the movement labels in Fig. 2. A smaller value (darker background color) of elements in the matrix indicates a greater similarity between the two movements represented by the abscissa and the ordinate.

3.3 Hierarchical Tree

The confusion matrix is transformed into a hierarchical tree by the single algorithm, as shown in Fig. 3(b). The abscissa in the Fig. 3(b) shows 20 hand movements, and the ordinate designates the distance between the movements, which builds the relationship between different hand movements intuitively. For the taxonomy of hand movements and decreasing the number of them as much as possible, three dashed lines of different colors are drawn. Their ordinates are approximately 34, 33, and 30, respectively.

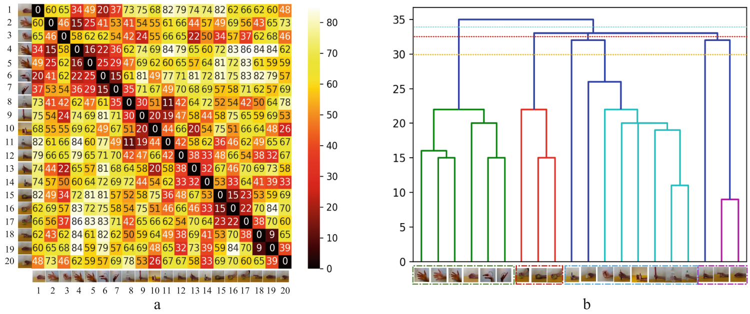

Fig. 3. Clustering results. (a) Cluster Confusion matrix of 20 movements. (b) Hierarchical tree of 20 movements (Color figure online)

- If the blue dashed line is used as the threshold line, the 20 movements can be divided into two categories. The movements on the left part are wrist movements and finger extension movements, and the ones on the right part are grabbing movements. The result of this classification is rough.
- If the red dashed line is used as the threshold line, the 20 movements can be divided into four groups. Counting from left to right, the icons below the abscissa are respectively framed by green, red, blue and purple dot-dash line. The first two groups represent movements of wrist and finger extension and movements of sphere grasp. The characteristic of the third group is that, except the thumb, the other four fingers all present the columnar grip posture, but the bending angles of the four fingers are different. In the forth group, the first movement (prismatic four fingers grasp) is the pinch between the thumb and the other four fingers, and the next two movements (prismatic pinch grasp and tip pinch grasp) are the pinch of thumb and index finger.
- If the yellow dashed line is used as the threshold line, the 20 movements can be divided into six groups. The reason is that on the basis of the above four groups, the writing tripod grasping in the third group and the lateral grasping in the fourth group are separately classified to form two new groups.

Considering the rapidity and accuracy of the assessment, we choose the second grouping way (divided into four groups) for the further research.

3.4 Metric of Clustering Result

We calculate that the SC of the confusion matrix obtained by the artificial feature method in [6] was 0.22. And if the framework of our study was used in [6], movements would be divided into six categories and the SC is $0.77 > 0.22$. It indicates that the features extracted by the LSTM neural network are more effective. In addition, the LSTM neural network is used to cluster 20 movements in this study, and the result of SC is 0.81, which also verified the validity of clustering results. More detailed information on the comparison of clustering is summarized in Table 1.

Table 1. Comparison of clustering results

Characteristic	Method in [6]	Method of our study in[6]	This study
Number of movements	20	20	20
Characteristic	Hand grasp	Hand grasp	Wrist movement finger extension and grasp
Data sources	DB2	DB2	DB1 and DB5
Number of subjects	40	40	37
Feature extraction	Artificial features	LSTM	LSTM
Clustering method	Hierarchical clustering	Hierarchical clustering	Hierarchical clustering
Number of clusters	five	six	four
SC	0.22	0.77	0.81

3.5 Cluster Center

Although we have classified the 20 movements into 4 groups, we still need to select 1 movement from each group as the representative one of this group to simplify and optimize the process for hand motor function assessment. After calculation, there are five cluster centers in the four groups, of which there are two in the first group (two points having the same minimum distance sum) and one in each of the remaining groups. The five movements are shown in Fig. 4.

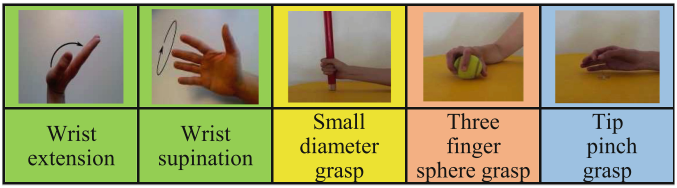

Fig. 4. Five representative hand movements

4 Conclusion

In this study, 20 hand movements involving wrist and finger extension are selected to comprehensively analyze the hand kinematics and serve the assessment of hand motor function of stroke patients. A novel framework for quantitative taxonomy of hand kinematics is proposed to extract the feature using LSTM neural network. A good classification effect is achieved to verify this feature has the common characteristics of this category. Then, the 20 behaviors are divided into four categories by using the nonlinear transformation, with a SC of 0.81. Finally, the cluster center of each group movement is calculated, and the hand movement corresponding to each cluster center is obtained. In the future work, we will collect the information of the 5 hand movements of a large number of stroke patients to replace the FMA scale, and propose an assessment algorithm to evaluate the hand motor function of stroke patients.

References

1. Santello, M., Flanders, M., Soechting, J.F.: Postural hand synergies for tool use. J. Neurosci. **18**(23), 10105–10115 (1998)
2. Charlotte, H.R., Schieber, M.H.: Quantifying the independence of human finger movements: comparisons of digits, hands, and movement frequencies. J. Neurosci. **20**(22), 8542–8550 (2000)
3. Veber, M., Bajd, T.: Assessment of human hand kinematics. In: 2006 IEEE International Conference on Robotics and Automation (ICRA), pp. 2966–2971 (2006)
4. Jarrass, N., Ribeiro, A.T., Sahbani, A., Bachta, W., Roby-Brami, A.: Analysis of hand synergies in healthy subjects during bimanual manipulation of various objects. J. Neuroeng. Rehabil. **11**(113), 1–11 (2014)
5. Jarque-Bou, N.J., Vergara, M., Sancho-Bru, J.L., Gracia-Ibanez, V., Roda-Sales, A.: Hand kinematics characterization while performing activities of daily living through kinematics reduction. IEEE Trans. Neural Syst. Rehabil. Eng. **28**(7), 1556–1565 (2020)
6. Stival, F., Michieletto, S., Cognolato, M., Pagello, E., Mller, H., Atzori, M.: A quantitative taxonomy of human hand grasps. J. Neuroeng. Rehabil. **16**(28), 1–17 (2019)
7. Jarque-Bou, N.J., Scano, A., Atzori, M., Mller, H.: Kinematic synergies of hand grasps: a comprehensive study on a large publicly available dataset. J. Neuroeng. Rehabil. **16**(63), 1–14 (2019)
8. Bao, T., Zaidi, S., Xie, S., Yang, P., Zhang, Z.: A deep kalman filter network for hand kinematics estimation using semg. Pattern Recogn. Lett. **143**(1), 88–94 (2021)
9. Jarque-Bou, N.J., Atzori, M., Mller, H.: A large calibrated database of hand movements and grasps kinematics. Sci. Data **7**(12), 1–10 (2020)
10. Richard, G.L.: Streamlining Digital Signal Processing. John Wiley & Sons, Hoboken (2012)