Anthropometric Features Based Gait Pattern Prediction Using Random Forest for Patient-Specific Gait Training

Shixin Ren[1,2], Weiqun Wang[1,2(✉)], Zeng-Guang Hou[1,3], Xu Liang[1,2], Jiaxing Wang[1,2], and Liang Peng[1,2]

[1] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
{renshixin2015,weiqun.wang,zengguang.hou,liangxu2013,wangjiaxing2016, liang.peng}@ia.ac.cn

[2] University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

[3] The CAS Center for Excellence in Brain Science and Intelligence Technology, Beijing 100190, China

Abstract. Using lower limb rehabilitation robots to help stroke patients recover their walking ability is becoming more and more popular presently. The natural and personalized gait trajectories designed for robot assisted gait training are very important for improving the therapeutic results. Meanwhile, it has been proved that human gaits are closely related to anthropometric features, which however has not been well researched. Therefore, a method based on anthropometric features for prediction of patient-specific gait trajectories is proposed in this paper. Firstly, Fourier series are used to fit gait trajectories, hence, gait patterns can be represented by the obtained Fourier coefficients. Then, human age, gender and 12 body parameters are used to design the gait prediction model. For the purpose of easy application on lower limb rehabilitation robots, the anthropometric features are simplified by an optimization method based on the minimal-redundancy-maximal-relevance criterion. Moreover, the relationship between the simplified features and human gaits is modeled by using a random forest algorithm, based on which the patient-specific gait trajectories can be predicted. Finally, the performance of the designed gait prediction method is validated on a dataset.

Keywords: Patient-specific gait · Anthropometric features
Random forest · Gait prediction

1 Introduction

Stroke is one of the common diseases that cause nervous system damage and even lead to death. Fortunately, due to timely treatment after stroke,

This research is supported by the National Natural Science Foundation of China (Grants 91648208) and Beijing Natural Science Foundation (Grant3171001).

L. Cheng et al. (Eds.): ICONIP 2018, LNCS 11304, pp. 15–26, 2018.
https://doi.org/10.1007/978-3-030-04212-7_2

the relative death rate has dropped rapidly [1]. However, the damages caused by stroke usually have long-term negative effects on patients' mobility, muscle control ability and gait patterns, and almost a half of stroke patients cannot walk without assistance [2].

Lower limb rehabilitation robots (LLRRs) have been developed to assist stroke patients to recover their walking ability in the last 20 years. Lokomat [3], ALEX [4] and Rewalk [5] are typical examples of the LLRR. Since gait training robots will be used by different patients, it is crucial to design personalized gait patterns for this kind of robot, which can be predicted by using the anthropometric parameters. However, the accurate relationship between anthropometric parameters and the gait pattern has not been well researched.

Luu et al. [8] proposed a gait trajectory generation method based on finite Fourier series (FFS) and modeled the relationship between the Fourier coefficients and gait feature, i.e., cadence and stride length. It can be seen from [8], human gait patterns can be represented by the associated Fourier coefficients. Koopman et al. [7] selected six key events to describe an individual's gait pattern in a gait cycle. Then a linear model was used to describe the relationship among the key events and human heights and walking speeds.

Luu et al. also adopted the multi-layer perceptron neural networks (MLPNN) model [6] and the general regression neural network (GRNN) model [9] for the gait pattern prediction, which are based on the gait parameters and four anthropometric features of the human legs. However, human gait trajectories are related to more factors, as is shown in [10]. Fourteen anthropometric features were used to estimate the hip, knee, and ankle joint angles, and the Gaussian process regression (GPR) method was used to design the estimation model [10]. The shortcoming of this method is that the estimation time is longer and the estimated gait patterns are inconvenient to be implemented on the platform of LLRRs.

This paper mainly focuses on developing a machine learning approach to predict personalized gait patterns. A Random Forest (RF) algorithm is designed to learn the relationship between the anthropometric features and the gait trajectories. To reduce calculation load, the gait trajectories are represented by Fourier coefficients which are used as the outputs of the RF algorithm. Fourteen anthropometric features which are same as [10] are used as the inputs of the RF algorithm. For the purpose of easy application on LLRRs, the minimal-redundancy-maximal-relevance (mRMR) criterion is adopted for the feature optimization, which is implemented based on the mutual information of inter-anthropometric features and the mutual information between the anthropometric features and Fourier coefficients. It can be found in the experiment that the modeling efficiency of the RF algorithm is higher than that of GPR. It can seen from the result of the anthropometric feature optimization of this paper and the comparison experiment that, the features used in [8,9] is insufficient for accurate prediction of personalized gait patterns, and the features used in [10] are too redundant for human gait prediction.

The remainder of this paper is organized as follows. Section 2 illustrates an LLRR developed at Institute of Automation, Chinese Academy of Sciences and the gait fitting method. The gait pattern prediction and the feature optimization are also given in Sect. 2. The results and discussion are presented in Sect. 3. This paper is summarized in Sect. 4.

2 Method

2.1 A Lower Limb Rehabilitation Robot

An LLRR has been developed at Institute of Automation, Chinese Academy of Sciences recently, based on which multiple training modes for patients at different rehabilitation phases can be provided.

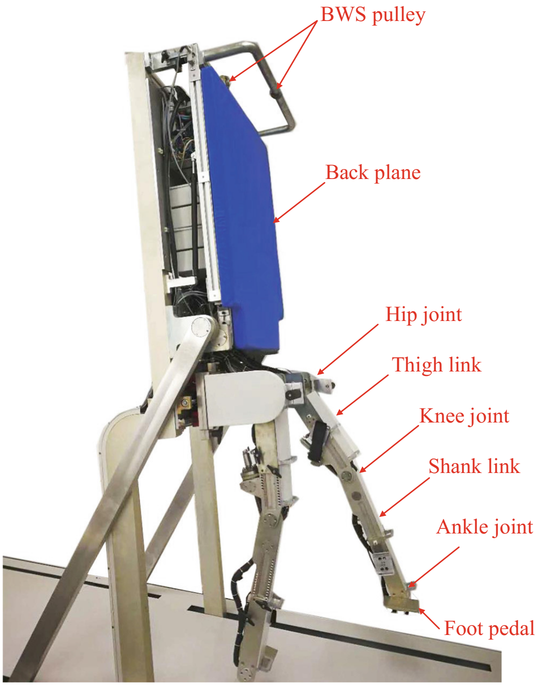

Fig. 1. Lower limb rehabilitation robot developed by Institute of Automation, Chinese Academy of Sciences.

As shown in Fig. 1, the LLRR has two leg exoskeletons, each of which has thigh and shank links and hip, knee and ankle joints. Each joint has one rotational degree of freedom in the sagittal plane, i.e., flexion/extension. The training effect can be improved by using a natural and personalized gait trajectory. On one hand, the lengths of thigh and shank links can be adjusted according to patients' shapes, which is one of the typical properties of the robot. On the other hand, how to design a personalized gait trajectory based on patients' features is one of the key points, which is investigated in the following text.

2.2 Anthropometric Features and Gait Trajectory Fitting

The human gait data which describes healthy subjects' gait kinematics and anthropometric features are obtained from [10]. Human age, gender and 12 body parameters are considered in this paper and given in Table 1.

Table 1. The associated human anthropometric features

Features	Ranges	Features	Ranges
Age (years old)	20–69	Bi-iliac width (cm)	26.1–35.8
Height (cm)	149–185	ASIS breath (cm)	20–30.6
Mass (kg)	43.3–99	Knee diameter (cm)	8.2–13
Gender	F/M	Foot length (cm)	20.5–28
Thigh length (cm)	27.5–41.6	Malleolus height (cm)	5.2–9
Calf length (cm)	30.5–46.3	Malleolus width (cm)	5.5–8
Bi-trochanteric width (cm)	28.8–38.6	Foot breath (cm)	6.4–11

The joint trajectories of human legs are continuous, smooth and periodic during walking. So, when patients participate in training on the LLRR, the output angle of LLRR must be smooth and compliant. Using Fourier series to fit the gait trajectories can generate the smooth and compliant output values. It has been proved that five terms of the Fourier series are enough for accurate fitting joint trajectories of human legs, as follows:

$$f(t_n) = a_0 + \sum_{n=1}^{5}(a_n \cos(n\omega t) + b_n \sin(n\omega t)), n = 1, \cdots, 5 \tag{1}$$

where $\omega = \frac{2\pi}{T}$ (T is the period of the gait pattern). a_n and b_n are Fourier coefficients. Hence, the human leg joint trajectory can be represented by a vector consisting of Fourier coefficients, as follows:

$$\mathbf{Y}_{(i,j)} = (a^0_{i,j}, a^1_{i,j}, b^1_{i,j}, a^2_{i,j}, b^2_{i,j}, a^3_{i,j}, b^3_{i,j}, a^4_{i,j}, b^4_{i,j}, a^5_{i,j}, b^5_{i,j}) \tag{2}$$

where $i = 1, 2, ..., P$; P is the number of subjects; $j = 1, 2, 3$, corresponding to the hip, knee and ankle joints, respectively.

2.3 Feature Selection Based on mRMR

Selecting the proper number of anthropometric features, which are strongly correlated with human gaits, is helpful for the performance improvement of the human gait prediction model. The mRMR method [11] is adopted to rank the importance of anthropometric features. It bases on the mutual information (MI) between the anthropometric features and the Fourier coefficients given in (3), which is a measure of the mutual dependence between the two variables.

By using the mRMR, an anthropometric feature subset can be obtained. The features in subset are not only highly related to Fourier coefficients, but also less redundancy among themselves. The mutual information, $I(X, Y)$, of two variables X and Y can be calculated in terms of their marginal probability functions $p(x)$, $p(y)$ and joint probability distribution function $p(x, y)$, as follows:

$$I(X, Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right). \tag{3}$$

A feature subset, where the redundancy of inter-features is minimum, can be obtained by:

$$\min W = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} I(s_i, s_j), \tag{4}$$

where s_i and s_j are features in set S that contains N features.

Similarly, the features of subset having max-relevance with Fourier coefficients can be obtained by:

$$\max S = \frac{1}{N} \sum_{i=1}^{N} I(s_i, r_t), \tag{5}$$

where r_t is one of the Fourier coefficients.

The mRMR method combining the above two conditions can be described by:

$$\max_{s_i \in S - D_m} \left[I(s_i, r_t) - [\frac{1}{N-M} \sum_{j=1}^{N-M} I(s_i, s_j)] \right], \tag{6}$$

where D_m is an already selected feature subset with M features. Then one feature can be selected from the set $\{S - D_m\}$ into D_m by implement of (6) for one time. It can be seen that the features selected early into set D_m are more closely related to the Fourier coefficients.

2.4 The Random Forest Model

RF is a machine learning algorithm that combines the advantages of Bagging and Decision trees for classification or prediction [12]. The performance of RF has been verified in plenty of applications in the last ten years. Therefore, the RF algorithm is adopted to develop a model for describing the relationship between the anthropometric features and the Fourier coefficients. The RF algorithm designed in this paper is given by:

– Give a dataset containing X samples, where each sample contains L anthropometric features. X_{temp} ($X_{temp} < X$) samples are randomly chosen from the original dataset to construct a subset X_{sub}^i. There will be T sample subsets, i.e. X_{sub}^i ($i = 1, ..., T$), after T bootstrap iterations.
– Build T regression trees based on T sample subsets. A feature subset Θ_{sub} is formed by using K ($K < L$) features which are randomly selected from the L anthropometric features. The best feature of the subset Θ_{sub} is used for division at each node of the tree. This is very effective to avoid the correlation of inter-trees. T trees can be built by this method to form a RF model.
– Let the response value of a tree to an input sample x is $f^t(x)$, the output value of the RF model can be given as follows:

$$Y(x) = \frac{1}{T}\sum_{t=1}^{T} f^t(x). \tag{7}$$

In this paper, two parameters of the RF model, which are highly related to the performance, are needed to be optimized: $ntree$, number of the trees, and $mfeature$, number of the features in the subset Θ_{sub}. The grid search method is used to find the optimal values of the two parameters. And the mean square error (MSE) is chosen to evaluate the predictive accuracy of the RF model.

3 The Results and Discussion

3.1 The Result of Fourier Series Fitting

The comparison between the reconstructed trajectories and actual trajectories of a sample are shown in Fig. 2. It can be seen that the actual trajectories are fitted very well by the Fourier coefficients.

3.2 Feature Selection and Optimization of the RF Model

The importance rankings of an anthropometric feature to each of the associated 11 Fourier coefficients can be obtained by the mRMR. Since a joint trajectory is represented by the 11 Fourier coefficients, the importance of each feature to a joint trajectory can be represented by the mean value of its 11 importance rankings corresponding to 11 Fourier coefficients. The importance rankings of 14 features are given in Table 2.

By adding one feature each time, 14 feature subsets can be obtained from the 14 features. Then the final optimal feature subset can be selected by using the optimal RF model. The specific steps are given as follows: Firstly, the first subset was obtained by using only the first feature in Table 2. Secondly, new subsets can be formed by adding one by one in the order of rankings until 14 subsets were obtained. Thirdly, the RF model was optimized for each feature subset. The grid search method was adopted to optimize the two parameters of each RF model: $ntree$ and $mfeature$ with ranges of 100–550 (in our experiments, the

MSE couldn't be decreased above 550 and below 100) and 1–12, respectively. The performance of each RF model was verified by using the 5-fold cross validation method [13]. For example, for the feature subset containing 12 features, the optimal $ntree(150)$ and $mfeature(3)$ can yield the lowest MSE for the ankle joint, as is shown in Fig. 3. Finally, the optimal feature subset can be obtained by selecting that with the lowest MSE of the RF model. It can be seen from Fig. 3 that, the best performance of the RF model for the ankle joint can be

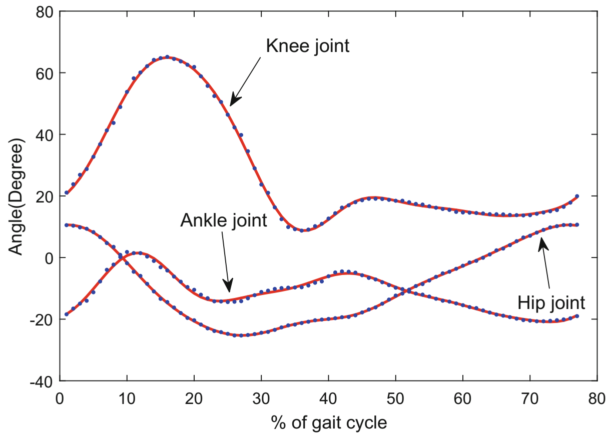

Fig. 2. Reconstructed joint trajectories by the Fourier coefficients and actual joint trajectories.

Table 2. The importance ranking of anthropometric features based on mRMR.

Features	Hip joint	Knee joint	Ankle joint	Mean
Mass	2.75	1.33	1.08	2.24
Age	4.17	2.58	2.67	3.73
Height	4.42	3.33	2.92	3.96
Calf length	6.33	5.42	4.08	5.56
Thigh length	6.25	5.83	6.33	6.28
ASIS breath	6.42	5.92	7.42	6.73
Bi-trochanteric width	7.58	8.17	7.5	7.63
Malleolus height	7.00	10.33	9.25	7.64
Bi-iliac width	8.17	7.25	8.00	8.21
Foot length	8.75	9.92	9.25	8.91
Malleolus width	9.25	12.25	12.08	10.09
Knee diameter	9.75	9.50	11.00	10.16
Foot breath	11.75	11	10.75	11.41
Gender	12.42	12.17	12.67	12.46

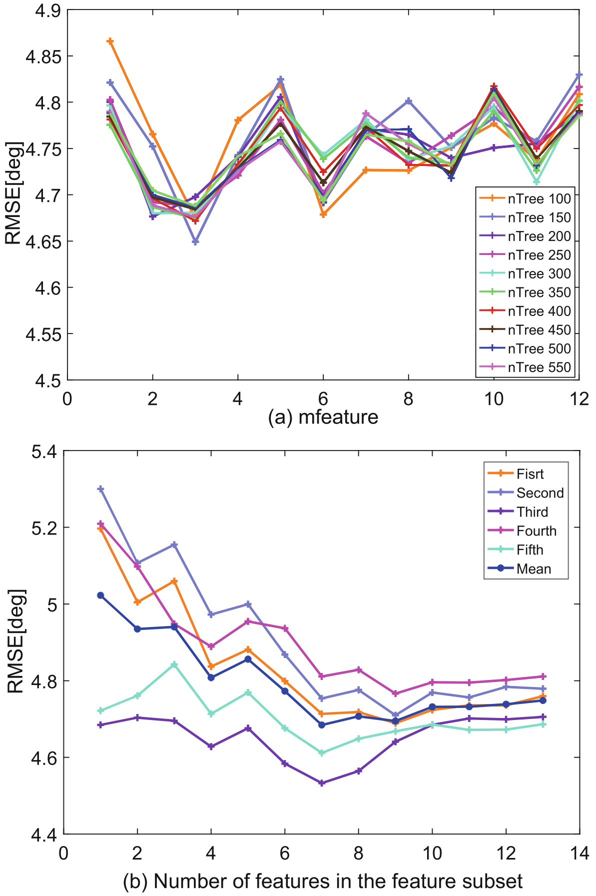

Fig. 3. (a) Optimaization of the RF models and (b) Feature selection for ankle joint.

obtained by using the feature subset, namely S_{OP}, containing the top seven features. Therefore, four anthropometric features are not enough for prediction, as is presented in the introduction section.

3.3 Predicted Performance of the RF Models

The predicted Fourier coefficients, obtained by using the optimized feature subset and the RF models, were used to reconstruct the joint trajectories. The reconstructed trajectories matched the actual trajectories well, which can be seen from Fig. 4. In order to validate the performance of the RF models, the 5-fold cross validation were used. The root mean squared error (RMSE) was adopted as the evaluation criterion. Meanwhile, the Pearson correlation coefficient was used to evaluate the similarity between the predicted and actual trajectories.

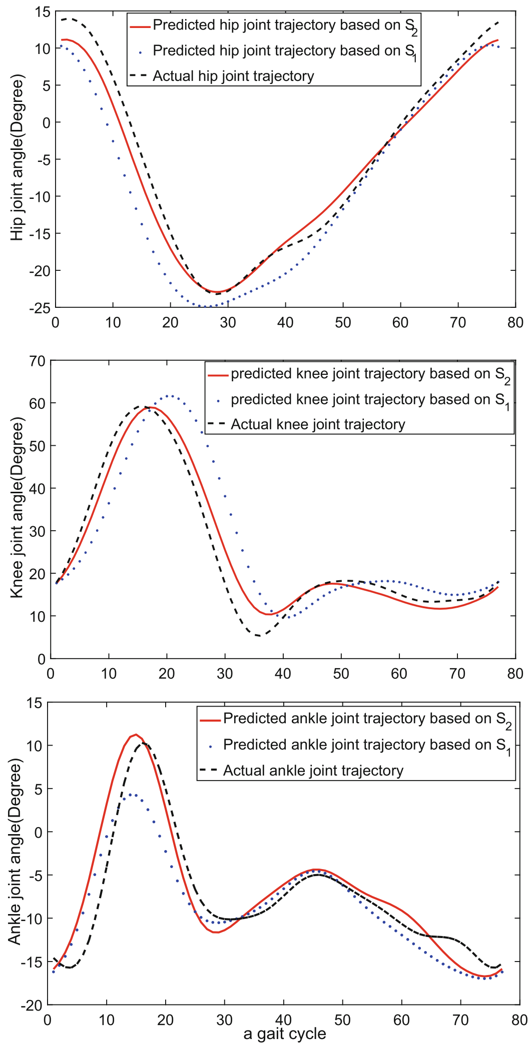

Fig. 4. Three joint trajectories comparison

Additionally, the feature dataset of [9], refered as S_2 in this paper, were also used to predict the gait trajectories, where the prediction model is the RF model as well. In addition to four anthropometric parameters, S_1 includes two gait parameters, namely stride length and cadence. They were not included in the gait dataset of this paper; however, they can be calculated by the gait period and walking speed as follows:

$$C = 2\frac{1}{T_{gait}}, \tag{8}$$

$$L_{stride} = \frac{V_{walking}}{T_{gait}}, \tag{9}$$

where C is the cadence; T_{gait} is the period of a gait; L_{stride} is the stride length; $V_{walking}$ is the walking speed.

Table 3. The performance comparison by using the feature dataset S_1 and S_2

Joint	S_1		S_2	
	e	ρ	e	ρ
Hip	5.06	0.91	4.72	0.93
Knee	8.02	0.881	7.70	0.897
Ankle	5.15	0.74	4.92	0.75

The performance comparison by using the feature dataset S_1 and S_2 are given in Table 3, where the mean error, e, is defined by:

$$e = \frac{1}{5}\sum_{m=1}^{5}\left(\frac{1}{5}\sum_{n=1}^{5} e_{m,n}\right), \tag{10}$$

$$e_{m,n} = \frac{1}{P}\sum_{i=1}^{P}\sqrt{\frac{1}{Q}\sum_{j=1}^{Q}|v_i^*(j) - v_i(j)|}, \tag{11}$$

where i is the sample index; j is the time index; Q is the max time index; P is the total number of the test samples; v_i^* and v_i is the predicted and actual sample joint trajectory, respectively.

The mean value of Pearson correlation coefficients ρ is defined by:

$$\rho = \frac{1}{5}\sum_{m=1}^{5}\left(\frac{1}{5}\sum_{n=1}^{5} \rho_{m,n}\right), \tag{12}$$

$$\rho_{m,n} = \sum_{i=1}^{P}\frac{cov(v_i^*, v_i)}{\sigma_{v_i^*}\sigma_{v_i}}, \tag{13}$$

where $cov(v_i^*, v_i)$ is the covariance; $\sigma_{v_i^*}$ and σ_{v_i} are the standard deviations.

It can be found from Table 3 that the performance of the gait prediction model based on S_2 is better than that based on S_1 in the RMSE and correlation. Meanwhile, as shown in Fig. 5, the volatility of the RMSE and correlation of the gait prediction model based on S_2 is much smaller than that based on S_1. It indicates that the gait prediction model based on the optimized feature subset can generate relatively stable prediction results. Moreover, the performance comparison experiment between the RF and GRNN models was also carried out, from which it was found that the performance of these two models were similar to each other.

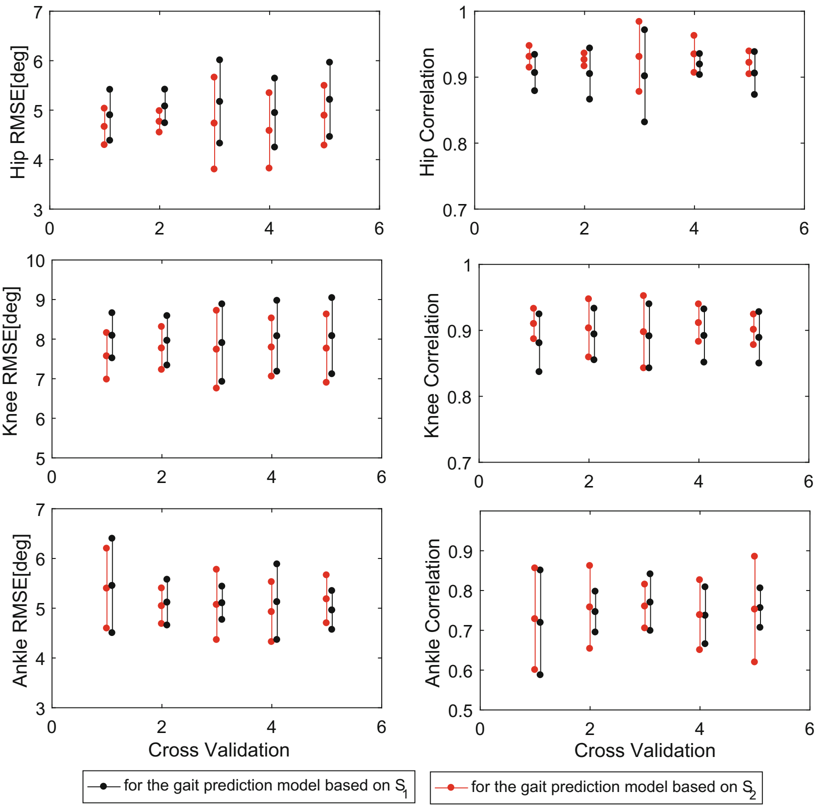

Fig. 5. RMSE and correlation coefficients of the predicted trajectories based on S_1 and S_2. Each bar is obtained from a 5-fold cross validation. The middle point in bar is the average, and two end points of the bar indicate the standard deviation.

Besides, the MLPNN model of [6] was also used to predict the gait trajectories of this paper. Lavenberg-Marquardt algorithm used in [6] was applied. It was found that, the volatility of RMSE for MLPNN was relatively large. It can be explained by that the performance of the MLPNN is unstable for gait dataset with small number of samples.

4 Conclusion

To generate the patient-specific gait trajectories for the LLRR, a RF algorithm is designed to learn the relationship between the anthropometric features and the Fourier coefficients, which are used to represent the gait trajectories. The anthropometric features are simplified by using an optimization method based on the criterion of mRMR. The tree and feature numbers of the RF model are optimized by the grid search method. The experiment results show that the performance of the proposed method based on the optimized feature subset is satisfactory. Patients will be included in experiments to further prove and improve the gait prediction method of this paper in the future.

References

1. Winstein, C.J., et al.: Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the american heart association/american stroke association. Stroke **47**(6), e98 (2016)
2. Swinnen, E., Beckwe, D., Meeusen, R., Baeyens, J.P., Kerckhofs, E.: Does robot-assisted gait rehabilitation improve balance in stroke patients? a systematic review. Top. Stroke Rehabil. **21**(2), 87–100 (2014)
3. Jezernik, S., Colombo, G., Morari, M.: Automatic gait-pattern adaptation algorithms for rehabilitation with a 4-dof robotic orthosis. IEEE Trans. Robot. Autom. **20**(3), 574–582 (2004)
4. Banala, S.K., Kim, S.H., Agrawal, S.K., Scholz, J.P.: Robot assisted gait training with active leg exoskeleton (alex). IEEE Trans. Neural Syst. Rehabil. Eng. **17**(1), 2–8 (2009)
5. Esquenazi, A., Talaty, M., Packel, A., Saulino, M.: The rewalk powered exoskeleton to restore ambulatory function to individuals with thoracic-level motor-complete spinal cord injury. Am. J. Phys. Med. Rehabil. **91**(11), 911–921 (2012)
6. Luu, T.P., Lim, H., Qu, X., Hoon, K., Low, K.: Subject-specific lower limb waveforms planning via artificial neural network. In: 2011 IEEE International Conference on Rehabilitation Robotics, pp. 1–6. IEEE (2011)
7. Koopman, B., van Asseldonk, E.H., Van der Kooij, H.: Speed-dependent reference joint trajectory generation for robotic gait support. J. Biomech. **47**(6), 1447–1458 (2014)
8. Luu, T.P., Lim, H.B., Qu, X., Low, K.: Subject tailored gait pattern planning for robotic gait rehabilitation. In: 2010 IEEE International Conference on Robotics and Biomimetics, pp. 259–264, December 2010
9. Luu, T., Low, K.H., Qu, X., Lim, H.B., Hoon, K.H.: An individual-specific gait pattern prediction model based on generalized regression neural networks. Gait Posture **39**(1), 443–448 (2014)
10. Yun, Y., Kim, H.C., Shin, S.Y., Lee, J., Deshpande, A.D., Kim, C.: Statistical method for prediction of gait kinematics with gaussian process regression. J. Biomech. **47**(1), 186–192 (2014)
11. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 1226–1238 (2005)
12. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
13. Refaeilzadeh, P., Tang, L., Liu, H.: Cross-validation. In: Encyclopedia of Database Systems, pp. 532–538. Springer, New York (2009)