Human action recognition based on a single acceleration sensor

Yi Wu

Signal and Information Processing, Communication University of China, Beijing 100024, China
wuyi511001@126.com

Abstract. In this paper, based on the acceleration signal acquired by a single triaxial acceleration sensor worn at the bottom of the finger, an algorithm for recognizing the daily movement behavior of the human body is proposed. The algorithm converts the three-axis acceleration into the resultant acceleration, horizontal acceleration and vertical acceleration, and extracts the features from them. The 37 features are sorted by the F-score as a measure of the importance of the feature, and a subset of features that meet the application requirements is found. Model training is performed on the feature subset using multi-class SVM to obtain the final classification model. Experiments show that the model has high recognition accuracy and the accuracy can reach 93.17%.

1. Introduction

With the rapid development of the economy, the pace of urban life in China is accelerating, and more and more people are caught in the sub-health state. Therefore, health has become a life issue that everyone is paying more and more attention to. The definition of health is not limited to the body, but also includes psychological and social adaptation. As the main way to improve physical and mental health, exercise has received more and more attention. In this sense, it is a meaningful research direction to give relevant opinions and suggestions accordingly by analyzing and recording the movement data of the human body.

Human action recognition is mainly divided into two types: visual-based human action recognition and sensor-based human action recognition. The latter is widely used in daily human action recognition because of its strong mobility and less redundant information. The most direct manifestation of human behavior is acceleration, so in recent years, human action recognition based on wearable acceleration signal sensors has developed rapidly. The sports ring is worn at the bottom of the finger, which can obtain the acceleration signal of the human hand through the three-axis acceleration sensor carried by the finger. This paper analyzes the daily behavior of the human body by analyzing the acceleration signal of the hand.

2. Pattern Recognition

Pattern recognition is also known as pattern classification. It is divided into supervised classification and unsupervised classification based on whether the categories to which each experimental sample belongs are known in advance. The specific steps of pattern recognition are: obtaining information, preprocessing the information, processing the recognition sample into a feature vector for analysis, designing the classifier to generate a discriminant function according to the feature vector, and
classifying the information to be identified. This article will also focus on the steps of pattern recognition described above.

3. Data Processing and Analysis

3.1. Date collection

The data used in the experiments herein was derived from a three-axis acceleration sensor in a motion ring worn at the base of the finger. Acceleration data for six daily movements of 6 people (3 men and 3 women) were collected, including sitting, bicycling, walking, going upstairs, going downstairs and running. The motion ring collects data at a sampling frequency of 25 Hz, and the collected data is transmitted to the PC through the USB interface, and the existing dynamic link library is called and converted into csv files for storage and analysis.

3.2. Data preprocessing

The raw data returned by the triaxial acceleration sensor includes acceleration signals in three axial directions of X, Y, and Z. Fig.1 shows the three-axis acceleration signal of the same person in different motion states. From the raw data, the data of walking, going upstairs and moving downstairs have a high degree of similarity, which is difficult to distinguish. Therefore, signal preprocessing is necessary.

(1) Resultant acceleration

Since the three-axis acceleration sensor is worn on the hand and the hand is quite flexible in motion, the orientation of the acceleration sensor is always changing. So the three axes of the sensor and the three axes in the motion space have little correlation. In order to reduce the impact of the sensor orientation on the raw data, the summation of the three-axis acceleration signals is recorded as the resultant acceleration. The formula is as follows. Where \(i\) denotes the sampling point number, \(x_i, y_i, z_i\) represent the \(i\)th acceleration value on the three axes, respectively, and \(a_i\) represents the \(i\)th value of the resultant acceleration.

\[
a_i = \sqrt{x_i^2 + y_i^2 + z_i^2}
\]

(2) Horizontal acceleration and vertical acceleration

In the paper [1], a device orientation-independent algorithm is mentioned. It can be assumed that the orientation of the three-axis acceleration sensor does not change in a short time. The algorithm works as follows: for a chosen sampling interval, typically a few seconds, obtain an estimate of the gravity component on each axis by averaging all the readings in the interval on that axis. The difference between the original acceleration and the gravity component can be denoted as dynamic acceleration. The projection of dynamic acceleration on the gravitational acceleration is the vertical acceleration. And we can get the horizontal acceleration by subtracting the vertical acceleration from the dynamic acceleration.

Through the above two steps, the three sets of signals of resultant acceleration, horizontal acceleration and vertical acceleration can be used instead of the original three-axis acceleration signal for analysis.

(3) Normalized processing

In order to reduce the difference between the data collected by different acceleration sensors, the acceleration signal needs to be normalized. The acceleration signal caused by gravity is consistent for all sensors. Therefore, it is more beneficial for model training to change the dimensional data into dimensionless data in the form of multiples of gravity acceleration.

(4) Data windowing

The acceleration signal is time series. In order to classify the motion, we need to window the data. The choice of window length is very important for the recognition of human motion state. In order to ensure the integrity of human motion, the window length should not be too short. However, the state of motion in daily life does not guarantee the same state of motion for a long time, so the window...
length selection should not be too long. Through the relevant experimental results, the window length is set to 5s.

The signal obtained after the above pretreatment process is shown in Figure 2.

![Fig.1 Raw data](image1)

![Fig.2 Data after preprocessing](image2)

### 3.3. Feature extraction

Feature extraction is a very important part of machine learning, and it has a great impact on the performance of machine learning. How to extract more concise and more efficient features has always been an important topic for people to study. This paper integrates several papers on the characteristics of acceleration signal extraction, including mean, standard deviation, median, over-average rate, skewness, kurtosis, energy, which are commonly used. We also get mean of power spectral density, standard of deviation power spectral density\(^2\), and the peak position of it, and the average maximum and minimum values based on the step recognition proposed in the literature\(^3\). We extracted the above 12 features from the resultant acceleration, the horizontal acceleration and the vertical acceleration respectively. In addition, the correlation coefficients of the horizontal acceleration and the vertical acceleration are calculated. The above features together constitute a 37-dimensional feature vector, which can be used by later feature selection.

### 3.4. Feature selection

F-score is a simple technique which measures the discrimination of two sets of real numbers. Given training vectors \(x_{k}, k = 1, ..., m\), if the number of positive and negative instances are \(n_+\) and \(n_-\), respectively, then the F-score of the \(i\)th feature is defined as:

\[
F(i) = \frac{(\bar{x}_{i}^{(+)\prime} - \bar{x}_{i}^{(-)\prime})^2 + (\bar{x}_{i}^{(-)\prime} - \bar{x}_{i}^{(-)\prime})^2}{\frac{1}{n_+ + 1} \sum_{k=1}^{n_+} (x_{k, i}^{(+)\prime} - \bar{x}_{i}^{(+)\prime})^2 + \frac{1}{n_- + 1} \sum_{k=1}^{n_-} (x_{k, i}^{(-)\prime} - \bar{x}_{i}^{(-)\prime})^2}
\]

(2)

where \(\bar{x}_{i}, \bar{x}_{i}^{(+)\prime}, \bar{x}_{i}^{(-)\prime}\) are the average of the \(i\)th feature of the whole, positive, and negative data sets, respectively; \(\bar{x}_{k, i}^{(+)}\) is the \(i\)th feature of the \(k\)th positive instance, and \(\bar{x}_{k, i}^{(-)}\) is the \(i\)th feature of the \(k\)th negative instance. The larger the F-score is, the more likely this feature is more discriminative\(^4\). Therefore, we use the F-score as a standard to evaluate the importance of features. We sort the F-scores of features and the results are shown in Tab.1. As can be seen from the sorting results, the characteristics of the vertical acceleration and the resultant acceleration have higher F-scores than the characteristics of the horizontal acceleration.

| Tab.1 The sort result of feature's F-score value |
|-----------------------------------------------|
| **Feature F-score Ranking Number** | **Feature Name**                  |
| 1–5                              | R-Std, R-AveMax, V-AveMax, V-Mean, V-Median |
| 6–10                             | R-PsdMean, R-PsdStd, V-E, V-AveMin, R-E    |
According to the above sorting result, the feature subsets are sequentially selected for training, and compared according to the correct rate results of the cross-validation of the training results, as shown in Fig.3. When the number of features is 29, with the best parameters found, the highest accuracy is obtained, which is 96.01%. When the number of features is greater than 14, the accuracy rate line graph tends to be stable, which meets the practical requirements of the classifier. In order to balance work efficiency and accuracy, we chose the top 14 features as a feature subset of the training model.

| 11–15      | R-Mean, R-PsdMaxPos, V-PsdMean, H-AveMax, H-Median |
|-------------|---------------------------------------------------|
| 16–20       | H-Mean, R-Median, V-Psd Std, H-Std, V-Std        |
| 21–25       | H-E, H-AveMin, R-AveMin, R-OverMean, V-PsdMaxPos |
| 26–30       | H-PsdMean, H-PsdMaxPos, H-Psd Std, HV-Corr, H-Skew |
| 31–35       | H-Skew, C-Skew, V-Kurt, V-OverMean, C-kurt      |
| 36,37       | H-Kurt, H-OverMean                               |

![Fig.3 Feature subset training model cross-validation results line graph](image)

3.5. Feature extraction
Support Vector Machine (SVM) is a generalized linear classifier that classifies data in a supervised learning manner. Its basic idea is to map data into high-dimensional space and find a separate hyperplane with the largest boundary. SVM shows many unique advantages in solving small sample, nonlinear and high dimensional pattern recognition [5]. Professor Lin Chih-Jen from Taiwan University has designed an easy-to-use, fast and efficient software package for SVM pattern recognition and regression, namely LIBSVM. In this paper, C-SVC with RBF radial basis as the kernel function is adopted. The optimal parameters \( C=512, \gamma=2.0 \) are selected by grid traversal method. Model training is performed on the features selected in Section 3.4, and the training model is tested. The specific results are in Chapter 4.

4. Test Results
The 2082 sets of data obtained by the test are subjected to the above processing to obtain corresponding 14-dimensional feature vectors. The 600 of them were randomly selected as test data, and the rest were training data. The number of test data and the test results are shown in Fig.4. Among them, the recognition accuracy of the running state is up to 100%, followed by sitting, going downstairs, walking and riding, and the accuracy rate is above 90%. The recognition accuracy of the upstairs state is only 82%. The overall recognition accuracy rate was 93.17%.
5. Conclusion

In this paper, we use the triaxial acceleration sensor in the moving ring worn at the bottom of the finger to collect the sample data of the tester in various motion states, and perform a series of processing on the sample data to extract the feature. By analyzing the F-score of the feature, more important features can be selected to form a feature subset. The model is trained using the selected features, and the best parameter values are selected by grid analysis. The accuracy of the model obtained in this way can reach 93.17%, which can meet the application requirements. However, the testers in this experiment are all between 20 and 30 years old, and the height is between 155 and 180 cm. Whether the accuracy of this model will be affected by other ages, height, weight and other aspects needs further verification. This is also our research direction, and we hopes to propose more universal algorithms through subsequent research.

References

[1] Mizell D. Using gravity to estimate accelerometer orientation[C]//Proc of the 7th IEEE Int Symp on Wearable Computers. Los Alamitos. CA: IEEE Computer Society. 2003: 252-252
[2] XU Bing-xue, SHI Jian-hua, et al. Design and implementation of an acceleration sensor system based on human behavior recognition[J]. Computer Development and Application, 2014,27(12):55-57(in Chinese).
[3] Li Feng, Pan Jingkui. Human motion recognition based on triaxial accelerometer[J]. Journal of Computer Research and Development,2016,53(3):621-631(in Chinese).
[4] Chen Y W, Lin C J. Combining SVMs with various feature selection strategies[J]. 2006.
[5] XU Chuan-Long, GU Qin-Long, et al. Activity recognition method based on three-dimensional accelerometer[J]. Computer System Application,2013,22(06):132-135(in Chinese).