Real-time Dynamic Data Analysis Model Based on Wearable Smartband

Xiangyu Li, Chixiang Wang, Haodi Wang, and Junqi Guo(✉)

Beijing Normal University, Beijing, China
guojunqi@bnu.edu.cn

Abstract. Since the traditional annual physical fitness test for adolescents in schools lacks of real-time dynamic data collection and deep analysis, we propose a data analysis model mapping from raw dynamic data to physical fitness evaluation, based on a self-developed wearable smartband by which we collect dynamic data from middle school students in Beijing. Firstly, the model presents a preprocessing algorithm which consists of a smoothness priors approach (SPA) and a median filter (MF), aiming for preprocessing of both photo plethysmo graphy (PPG) signals and three-axis acceleration data collected from the wearable smartband. Secondly, the model implements physiological and physical index estimation to acquire heart rate (HR), blood oxygen saturation ($\boldsymbol{SpO_2}$) and exercise amount estimates from the preprocessed data. Thirdly, the model extracts several key features closely related to physical fitness evaluation from the estimated HR, SpO_2 and acceleration data. Finally, a support vector machine (SVM) algorithm is employed for classification of physical fitness level of different smartband users. An application testing of our self-developed wearable smartband has been implemented in Tongzhou No. 6 middle school of Beijing. Experimental evaluation results demonstrate feasibility and effectiveness of both the smartband hardware/software and the proposed data analysis model.

Keywords: Dynamic data · Physical fitness evaluation · PPG signal · Smoothness prior approach · Median filter · Support vector machine

1 Introduction

Nowadays studies generally think that health-related physical fitness evaluation is an important part of physical education [1]. Traditional physical fitness test for adolescents, carrying out annually, cannot reflect young people's daily physical health status. Otherwise, its ability for data analysis and evaluation is also poor, lacking effective feedback and tracking mechanism. According to C.J.Caspersen's research, physical fitness is a set of attributes that are either health-related or skill-related and the degree of these attributes can be measured with specific test [2]. To improve the traditional physical fitness test, we consider wearable devices, which can monitor physical fitness condition, have set physiological information collection, storage, display and other functions in one. However, it is still basically stay in the stage of simple visualization of physiological data, but lack a model of analysis from raw data to physical health evaluation.

F. Xhafa et al. (eds.), *Advances in Intelligent Systems and Interactive Applications*, Advances in Intelligent Systems and Computing 686,
https://doi.org/10.1007/978-3-319-69096-4_61

After collecting real-time dynamic physiological data: real-time heart rate (HR), blood oxygen saturation (SpO_2)and three-axis acceleration through self-developed wearable smartband. The model proposed in this paper does preprocessing and analyzing, then extracts eigenvector and synthesized feature matrix with the static features obtained by traditional physical test. Finally, the result of the physical fitness evaluation is attained by Support Vector Machine (SVM) algorithm.

2 Existing Solutions

The dynamic data is a series of data ordered by time. It is available to extract feature index efficiently from the dynamic data to evaluate teenagers' health condition. Yang Z, Wang G. and others have done researches basing on the annual "dynamic" data of adolescents' health status by manual testing [3, 4]. However, this method can not reflect the real-time conditions reasonablely. So reseachers consider using wearable smartband to obtain real-time data. Photo Plethysmo Graphy (PPG), the most widely used method to extract dynamic data in wearable smartband, uses basic principle of photoelectric signal conversion, utilizing the character that the arterial blood has a different level of absorbing light from the muscles, bones, veins and other connective tissue, and extracting the AC signal from the light signal to reflect the trait of blood fluxion [5, 6].

The electro cardio signal (ECG) preprocess is divided into denoising and removing baseline drift. The former includes one-dimensional digital filtering, wavelet denoising, median filter denoising, and so on. Meanwhile, the most common way to deal with baseline drift is filter. H.S.Niranjana and others did researches on ECG denoising and proposed an optimized wavelet function [7]. B. Sharma and J. Suji analyzed the effects of different window sizes in ECG denoising [8]. As for evaluating the physical fitness, A.R. Lipu and others proposed a clustering algorithm using FPGA [9]. Y.Omard applied the double-hashing to sort [10]. Y. Dai, X. Yang utilized the fuzzy mathematics and established a sorting model applied to the evaluation of students' physical fitness [11].

3 A New Dynamic Data Analysis Model

Our model's main goal is establishing a mapping relation from raw physical data to physical fitness evaluation. This process contains five steps: raw data acquisition, data preprocessing, index estimation, extracting eigenvector and training SVM classifier (Fig. 1).

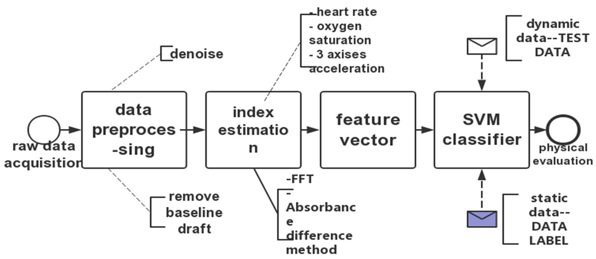

Fig. 1. Model framework

3.1 Data Acquisition

Basing on previous researches, a series of typical raw data are selected. This series includes static data from conventional physical fitness test: BMI, vital capacity, physical quality (strength, endurance, flexibility, speed, sensitivity) and dynamic data obtained from smartband: realtime PPG signal (HR, SpO2) and three-axis acceleration.

3.2 Data Preprocessing

3.2.1 SPA (Smoothness Priors Approach) Denoising

SPA [12] is effective as a non-linear approach to detrend. Raw PPG signal data contain periodic term and aperiodic term (trend term). SPA filters out the aperiodic term which has high frequency, and remains the periodic term as result.

$$z = z_{stat} + z_{trend} \tag{1}$$

where: z_{stat} is periodic term, z_{trend} is aperiodic term (trendterm).

We use single parameter to filter the target term. Calculate parameters by the optimal estimation method. In this way,we solve the equation below:

$$\lambda^2 \times \ddot{g} + g = f \tag{2}$$

If $f = \cos \omega t$, then $g = \frac{1}{(1 + \lambda\omega^4)} \cos \omega t$. It is a typical lowpass filter for discrete time series. Tune the parameters λ to isolate terms and take out the high-frequency noise.

3.2.2 Median Filter

Intercept a data window from the raw data, reorder and replace the previous median with new median. In this way, we can filter the frequency component of baseline.

- Take k as midpoint, and define the data window which possesses N points. If N is even, the window is: $\left[k - \frac{N}{2}, k + \frac{N}{2} - 1\right]$, else the window is: $\left[k - \frac{N-1}{2}, k + \frac{N-1}{2}\right]$
- For every k, reorder the data in the window and replace x with the new midpoint.
- Tune the size N of the window till the frequency component of baseline is elected.

Get rid of it and then the PPG signal data become horizontal.

3.3 Index Estimation

3.3.1 Heart Rate Estimation

Discrete Fourier Transform (DFT) converts time-domain to frequency-domain so as to acquire the signal period. Heart rate can be attained from the signal period. Discrete Fast Fourier transform (FFT) is a fast algorithm of DFT, which uses low point' DFT to count high point' DFT repeatedly. In other words, FFT decomposes Fourier transform to a smaller one iteratively in order to reduce computing complexity. $F = \frac{\max|\mathrm{fft}(\mathrm{x})| * \mathrm{fs} * \frac{60\mathrm{s}}{\min}}{N}$, N is the length of the data, fs is sampling frequency.

3.3.2 SpO₂

In human blood deoxyhemoglobin and oxyhemoglobin have different unique absorption spectrum in red and near infrared regions. Arteriopalmus can cause change of blood flow, so that the absorption spectrum will change meanwhile. On the other hand, non-blood tissue's absorption spectrum is constant, such as skin, muscle, skeleton. So detect the change of absorption spectrum caused by fluctuation in blood volume, eliminate effects of non-blood tissue, and then calculate SpO_2 [13]:

$$SpO_2 = A\frac{I_{AC}^{\lambda_1}/I_{DC}^{\lambda_1}}{I_{AC}^{\lambda_2}/I_{DC}^{\lambda_2}} - B,\ \mathrm{A} = \frac{\varepsilon_{Hb}^{\lambda 2}}{\varepsilon_{HbO_2}^{\lambda 1} - \varepsilon_{Hb}^{\lambda 1}},\ \mathrm{B} = \frac{\varepsilon_{Hb}^{\lambda 1}}{\varepsilon_{HbO_2}^{\lambda 2} - \varepsilon_{Hb}^{\lambda 2}} \tag{3}$$

light λ_i's absorptivity: $\varepsilon_{HbO_2}^{\lambda i}, \varepsilon_{Hb}^{\lambda i}$, intensity's DC and AC component is $I_{AC}^{\lambda_i}, I_{DC}^{\lambda_i}$

3.3.3 Three-axis Acceleration Estimation

- Speed estimate

i. Known three-axis acceleration x, y, z, $A = \sqrt{x^2 + y^2 + z^2}$, Compute A every 0.19 s, recognize pace by $A_{max} - A_{min}$. When the difference $\leq 0.13g$ and $\geq 0.07g$ ($g = 9.8\mathrm{ms}^{-2}$), record it as a single step.

ii. The distance is the sum of stride, which is computed by Based Stride Length (*BSL*)

$$BSL = Height \times GenderFactor \times 1.1 \tag{4}$$

$$Stride = BSL \times StepRateFactor \tag{5}$$

where: *GenderFactor* is 0.415(male)/0.413(female). *StepRateFactor* is based on speed.

iii. Define time window: $Speed = \frac{distance}{time}$, time is the length of the window.

- Energy expended per step

$$\text{Calories} = \frac{MetabolicFactor \times 0.00029}{StepRate} \times Weight \tag{6}$$

where: *weight* is measured in kilograms; MetabolicFactor depends on speed.

3.4 The Feature Extraction

After data preprocessing and estimation, attain HR, SpO_2 and three-axis acceleration ordered by time, which are still original and their features are not obvious. So extract features from them and build eigenvector to make evaluations on the physical conditions.

3.4.1 Define Eigenvector

A: Resting Heart Rate (**RHR**), Heart Rate Reserve (**HRR**) [14], Recovery Heart Rate (**RHR**) [15], Immediate Heart Rate after Exercise (**IHRE**)
B: Resting Blood Oxygen Saturation (**R-*SpO₂***) [16], Falling Time of SpO_2<high-intensity exercise>(**FT- *SpO*₂**) [16]
C: Average Speed (**AS**), Maximum Speed (**MS**), Energy Expended Per Step (**EEPS**)

The eigenvector is consist of three parts: $S = \{A, B, C\}$. A is related to HR, B is related to SpO_2, C is related to three-axis acceleration.

3.4.2 Experimental Paradigm

Wearing the activated smartband, fitstly subjects sit for 3 minutes on the playground. Then they are requested to run a 2000-meter dash. After running, all subjects sit and rest for at least 2 minutes and then trun off the smartband. During the test, subjects are not allowed to drink and stop.

3.5 Classification

3.5.1 Classification Model

We regard the core from static data as data label for every sample, and use it to train our classifier. That is, input the eigenvector defined above, adjust the parameter till the output closes to its data label. Repeat this step to train the classifier. There will be 4 categories after classifying, standing for excellent, above-average, medium-low and bad.

3.5.2 SVM

It is a typical multi-classification problem, as we should give each sample a evaluation result. Because conventional SVM can only apply to binary classification problems, we choose one-against-one method to train the classifier. That is, build $k(k-1)/2$ binary classifier to solve k classification problem. Furthermore, we use C-SVC algorithm and radial basis function (RBF) to realize Multilinear map [17].

Further discussion—*Adding static data to the model*

Above, we mainly discuss dynamic data attained from smartband. Static data from conventional physical fitness test only provides data label and does not take part in the model. More valid index will improve the accuracy of out classifier. So consider to extract features indexes from static data, add the dimension of eigenvector and train the new SVM classifier. The new eigenvector is $S' = \{S, S_1\}^T$ S_1 is extracted from static data.

4 Experimental Evaluation

We conducted extensive experimental studies to verify the validity of this model, using MATLAB and MySQL. 302 students took part in the test who had done traditional physical fitness test recently. They weared smartband and tested in accordance with the experimental paradigm determined in advance. The sampling frequency of PPG signal and three-axis acceleration was 25 times/sec.

4.1 Data Preprocessing and Index Estimation

In the data preprocessing, frequently use Root-Mean-Square Error (RMSE), Signal-Noise Ratio (SNR) to assess preprocessing's effect. Larger RMSE means more effective preprocessing. Smaller SNR means more effective preprocessing. Meanwhile, in order to attain real-time physical fitness evaluation, we require superior time requirement.

- RMSE is the square roots of the variance between raw data and the preprocessing result. RMSE $= \left\{\frac{[f(n)-f_1(n)]^2}{n}\right\}^{\frac{1}{2}}, f(n)$ is raw signal data, $f_1(n)$ is preprocessing result.
- SNR is also a traditional approach: SNR $= 10\log_{10}\left(\frac{p_s}{p_z}\right), p_s = \frac{\left[\sum_n f^2(n)\right]}{n}$ is the power of raw data, $p_z = RMSE^2$ is the power of noise.
- Compare our solution with wavelet transform, median filter and butterworth low pass filter. Then record the average RMSE, SNR and TIME of each algorithm in Table 1. Shown in which, our solution is better in RMSE and SNR, especially in TIME.

Table.1 Comparation of preprocessing

Denoising	SPA+ MF	Wavelet theory	Median filter	Butterworth
RMSE	4.79e + 05	1.17e + 06	8.81e + 05	7.99e + 05
SNR	112.0021	94.2014	99.8440	103.2775
TIME	0.04603	0.34622	0.12426	0.24355

To test accuracy of the model's index estimation, we compared HR and SpO_2 measured by professional medical equipment between our model based on the test. The accuracy reached 90%, which was enough for our evaluation.

4.2 Classification Evaluation

All samples were divided into two parts, 60% of samples were set as training set and 40% were set as test set. There were 302 subjects in the sample data set and sample's dimension was 9. All samples were divided into 4 categories. Compared the data label and classification result, accuracy in training set reached 93%, and in test set accuracy reached 90%.

Table 2 indicates the result of classification, all points are divided into 4 parts, which matches the 4 categories. Points in grey means the sample's data label, and points in blue means the sample's result of classification. Based on the Fig. 2, the clear majority of points is coincident (in grey), which means the result is up to standard. Some dominant blue points stand for the sample whose result is not accord with the data label.

Table.2 Result of experiment

Category	Category 1 (%)	Category 2 (%)	Category 3 (%)	Category 4 (%)
Training set-data label	34	39	20	7
Training set-result	34	42	17	7
Test set-data label	34	37	19	10
Test set-result	34	40	18	8

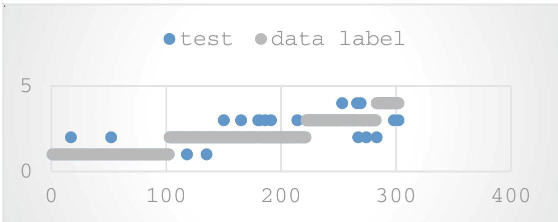

Fig. 2. Accuracy of experiment

5 Conclusion

A model is proposed to evaluate physical fitness in this paper. Based on the designed experimental paradigm, we obtain real-time dynamic data from smartband. Preprocess data using SPA to de-noises and MF to remove baseline drift. Then FFT is adopted to obtain the real-time sequence of HR. Make use of the difference of DC component between AC component of PPG signal to obtain the real-time sequence of SpO_2. Through three-axis acceleration we can obtain subjects' speed, etc. Based on the preprocessed dynamic data, we extract eigenvector and train SVM with known data label from traditional physical fitness test. Finally, we classify the eigenvector by trained SVM. Our work established a complete model from raw dynamic phyciological data to physical fitness evaluation. This model remedys the traditional physical fitness test's weakness and takes advantage of wearable smartband whose data is real-time. The actual test we conducted validate the usability of the model.

Acknowledgments. This research is sponsored by National Natural Science Foundation of China (No.61401029), and Beijing Advanced Innovation Center for Future Education (BJAICFE2016IR-004)

References

1. Corbin, C.B., Welk, G.J., Richardson, C., Vowell, C.: Youth physical fitness: ten key concepts. J. Phys. Educ., Recreat. Dance, 24–31 (30 Jan 2014)
2. Caspersen, C.J., Powell, K.E., Christenson, G.M.: Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep. **100**(2), 126–131 (Mar–Apr 1985)
3. Zhang, Y., He, L.: Dynamic analysis of the physical health of Chinese adolescents—Four national health monitoring data in 2014. Beijing University of Aeronautics and Astronautics
4. Wang, G.: Students Constitution of He Nan Province in 1995-2010 to study the dynamic characteristics analysis and trends of research data. Zhengzhou University
5. Heerlein, J., Rueheimer, T.: LED Based Sensors for Fitness Tracking Wearables (2014)
6. Application document "SFH 7050—Photoplethysmography Sensor". Osram Opto Semiconductors (2014)
7. Niranjana Murthy, H.S., Meenakshi, M.: Optimum choice of wavelet function and thresholding rule for ECG signal denoising. In: 2015 International Conference on Smart Sensors and Systems (IC-SSS)
8. Sharma, B., Suji, J.: Analysis of various window techniques used for denoising ECG signal. In: 2016 Symposium on Colossal Data Analysis and Network (CDAN)
9. Lipu, A.R., Amin, R., et al.: Exploiting parallelism for faster implementation of Bubble sort algorithm using FPGA. Telecommunication Engineering (ICECTE)
10. Omar, Y.M.K., Osama, H., Badr, A.: Double hashing sort algorithm. Comput. Sci. Eng. (2017)
11. Dai, Y., Xu, Y.: The Fuzzy Clustering Analyzes in the Forecasting of College. College of Science, Liaoning Technical University
12. Decomposition forecasting of displacement of landslide based on SPA (Smoothness Priors Approach)
13. Lv, C.: Study on Non-invasive Pulse Monitor of Oxygen Saturation of Blood. Liaoning Technical University
14. http://www.shapesense.com/fitness-exercise/calculators/heart-rate-reserve-calculator.aspx
15. http://www.med-health.net/Recovery-Heart-Rate.html
16. https://en.wikipedia.org/wiki/Oxygen_saturation
17. Ma, Y., Guo, G.: Support Vector Machine Application. ISBN: 978-3-319-02299-4. Springer International Publishing