An enhanced learning framework for the classification of college student physical fitness

Jiajun Wang[2], Zixin Xie[1], Jingwen Yan[1,3], Dazhi Jiang[1,3*] and Teng Zhou[1,3,*]

[1]College of Engineering, Shantou University, Shantou 515063, China;

[2]Ministry of Sports, Shantou University, Shantou 515063, China;

[3] Key Laboratory of Intelligent Manufacturing Technology (Shantou University), Ministry of Education, Shantou, 515063 China

*Corresponding author: Dazhi Jiang (dzjiang@stu.edu.cn) and Teng Zhou (zhouteng@stu.edu.cn)

Keywords: Physical test, Collage student, Selection strategy, Enhanced learning.

Abstract. Physical test (PF) is demonstrated by a variety of factors including body weight status, cardiorespiratory fitness, musculoskeletal fitness (muscular strength and endurance) and flexibility, which are related to college student's physical condition. The test helps them require an understanding of behavioral attributes and causative mechanisms that promote their physical fitness. Judging the level of college student's physical fitness quickly is very necessary. In this paper, we present an enhanced learning framework for the classification of college student physical fitness. To achieve this, we first design a selection strategy for the students by considering the features that have the greatest impact on. Then, we present an enhanced learning framework equipped with our custom loss function and boosting strategy to classify the level of students' physical fitness. Extensive experiments have demonstrated the outpeformance of our framework.

1. Introduction

Body weight status, cardiorespiratory fitness, musculoskletal fitness, the main health-related fitness components in youth, are the leading makers of health. Physical test is to evaluate the students' health level [1,2]. Then each student will obtain corresponding result, which has been recognized a vital method for the college students. It has been consistently reported that lack of exercise makes students more susceptible to disease and overweight. As the backbone of society, the health of college students has been paid close attention [3]. According to the physical test, it is urgent to find a method to judge the level of college student's physical fitness quickly [4]. So, the students could clearly know how to improve it [5].

Actually, judging students' physical health level can be essentially converted to a multi-classification task. It is necessary to find the most effective to classify and apply it on this problem. In the current, machine learning and data-driven approaches are becoming very important in many areas [6]. We could use the effective (statistical) models that capture the complex data and learn the model of interest from large datasets [7].

The machine learning is the main method of classification, such as Support Vector Machines (SVMs), K-nearest Neighbor (KNN) and Decision Tree (DT). SVMs also can be solved in small sample situations and improves generalization performance [8], which avoids the problem of selecting the neural network structure and local minimums. As for KNN and DT, the model of KNN is simple and effective [9]. It is easy to understand and interpret the structure of the Decision trees. DT are able to handle multi-output problems [10]. However, the effects of these methods are not ideal, especially for high-dimensional data. Even they need hundreds of small classifiers, which leads to a significant increase in the solution time. In short, they require a lot of computation and exist some drawbacks.

To address the issues of decision trees above, ensemble learning is defined by building multiple weak classifiers and combining them into a strong classifier to complete the classification task [11]. The key to ensemble learning is how to build differentiated classifiers and integrate the predicted results of these classifiers. In this paper, to improve the understanding of such challenging dataset, we

present an enhanced learning framework for the classification of college student physical fitness. We first design a feature selection strategy for the dataset of physical test [12]. Then we use the improved XGBoost as an end-to-end framework to learn and classify.

2. Methodology

2.1 Problem description

We formulate the problem of classification of college student physical fitness as follow. We first denote $x^d \in R^D$ as the physical test result of every student respectively. In this study, physical fitness battery comprises the following tests: weight and height to assess anthropometry, lung capacity and running to assess cardiorespiratory fitness, body flexion to assess flexibility, 50-meters running to assess speed-agility, standing long jump and leaping up (sitting up for girl) to assess muscular strength and endurance. Then, we list 8 features denoted as f_i for each student. We use $y_i \in 0,1,2$ as the classification of the *ith* student's physical fitness, where $y_i = 2$ means the students have good physical fitness, $y_i = 1$ means the students have general physical fitness and $y_i = 0$ means the students have unqualified physical fitness. Thus, the feature and outcome for a student is denoted as (x_i, y_i). Our task is to train a model for judging the level of student's physical fitness quickly.

2.2 Feature selection strategy

For different each college student, the feature vector of a physical test is unknown beforehand. The factors affecting the physical fitness of students are different. We present the status of a student by selecting the feature importance of the threshold to carry out the weighting combination. For example, the status $\hat{x}^d$ of each student can be denoted as:

$$\hat{x}^d = \sum_{i=0}^{I} \beta_i f_i, \quad subject\ to \quad Threshold(f_i) > l \tag{1}$$

where i is the index of the features? $Threshold(\cdot)$ is a shrinkage function that can be calculated every feature importance? β_i is a weighted factor and l is the value of threshold setting in advance?

2.3 Enhanced learning framework for physical fitness classification

XGBoost is based on an end-to-end tree boosting system. It provides a faster and more accurate way to solve classification and regression-type problems. The key feature in XGBoost is that it weights the predictors and tries to keep the new decision tree away from the errors made by previous decision trees, and it strengths its accuracy. This idea of reweighting predictors where errors occurs is the key idea behind all boosting algorithms. Here we will use the XGBoost as an enhanced learning framework for the classification of college student physical fitness, combined with the feature selection strategy above.

3. Experiments and discussions

3.1 Data description

A total of 6,249 students, including 3,079 boys and 3,170 girls, aged 18 to 22 were recruited from 17 provinces of south China. They are mainly freshmen and seniors. Informed consents were obtained from all subjects previously. The study protocol was designed in accordance with the guidelines outlined in the Declaration of Helsinki and approved by the Ethics Committee of Shantou University. Test data and researchers will be kept strictly confidential.

3.2 Performance evaluation

In the first experiment, we also compared our enhanced learning framework with four state-of-art models for the classification of college student physical fitness, which includes Naive Bayes, Logistic regression, XGBClassifier and artificial neural network. The accuracy and $F_\alpha - score$ of our

framework and comparisons are listed in Table.1 and Table.2. In boys and girls, the results of the accuracy or $F_\alpha - score$ obtained by the comparisons are in the range of 80% to 95%, while our enhanced learning framework achieve the accuracy rate of 91.54%, 96.37%, and $F_\alpha - score$ achieves 94.50%, 97.80% respectively.

Table 1. The accuracy and $F_\alpha - score$ of our framework and comparisons in boys.

Models	Accuracy	$F_\alpha - score$
Our framework	91.54%	94.50%
Naive Bayes	82.69%	89.27%
Logistic regression	83.09%	89.93%
XGBClassifier	90.36%	93.80%
Artificial neural network	82.20%	89.75%

Table 2. The accuracy and $F_\alpha - score$ of our framework and comparisons in girls.

Models	Accuracy	$F_\alpha - score$
Our framework	96.37%	97.80%
Naive Bayes	91.88%	95.35%
Logistic regression	92.93%	96.00%
XGBClassifier	95.22%	97.17%
Artificial neural network	92.55%	95.85%

4. Conclusion

The physical fitness is critically important college student, but how to judge the level of the student's physical fitness is still a difficult issue. In this study, we present an enhanced learning framework for the classification of college student physical fitness. The experimental results demonstrate that our framework is suitable for this task and achieve the better results. Our framework is beneficial to college student, which helps them to improve the physical fitness. In future, we plan to extend this method to the applications of other domains, such as basketball game prediction or time series analysis [13-18].

Acknowledgement

This research was financially supported by the National Natural Science Foundation of China (Grant NO. 61902232, 61672335, 61902232), the Natural Science Foundation of Guangdong Province (No. 2018A030313291, 2018A030313889), the Education Science Planning Project of Guangdong Province (2018GXJK048), and the STU Scientific Research Foundation for Talents (NTF18006) and Major Provincial Scientific Research Projects of Universities in Guangdong Province (Grant NO.2017KCXTD015).

References

[1] A. Packham and B. Street, The effects of physical education on student fitness, achievement, and behavior, *Economics of Production Education Review,* vol.72, pp. 1-18, 2019.

[2] J. K. Ward, P. A. Hastie, D. D. Wadsworth et al. A Sport education fitness season's impact on students' fitness levels, knowledge, and in-class physical activity, *Research quarterly for exercise and sport,* vol.88, pp. 346-351, 2017.

[3] O. Yarmak, Y. Galan, A. Hakman et al. The use of modern means of health improving fitness during the process of physical education of student youth, *Journal of Physical Education and Sport*, vol.17, pp. 1935-1940, 2017.

[4] E. Jianwei, J. Ye, and H. Jin, A novel hybrid model on the prediction of time series and its application for the gold price analysis and forecasting, *Physica A: Statistical Mechanics and its Applications*, vol. 527, p. 121454, 2017.

[5] A. Osipov, V. Vonog, O. Prokhorova, and T. Zhavner, Student learning in physical education in Russia (problems and development perspectives), Journal of Physical Education and Sport, vol. 16, p. 688, 2016.

[6] L. Bottou, F. E. Curtis, and J. Nocedal, Optimization methods for large-scale machine learning, *Siam Review*, vol. 60, no. 2, pp. 223–311, 2018.

[7] S. Suthaharan, Machine learning models and algorithms for big data classification, *Integr. Ser. Inf. Syst*, vol. 36, pp. 1–12, 2016.

[8] S. Maldonado and J. L´opez, Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for svm classification, Applied *Soft Computing*, vol. 67, pp. 94–105, 2018.

[9] Z. Deng, X. Zhu, D. Cheng, M. Zong, and S. Zhang, Efficient knn classification algorithm for big data, *Neurocomputing*, vol. 195, pp. 143–148, 2016.

[10] Y.-Y. Song and L. Ying, Decision tree methods: applications for classification and prediction, *Shanghai archives of psychiatry*, vol. 27, no. 2, p. 130, 2015.

[11] Z.-H. Zhou, Ensemble learning, *Encyclopedia of biometrics*, pp. 411–416, 2015.

[12] M. El Fatini, M. El Khalifi, R. Gerlach, A. Laaribi, and R. Taki, Stationary distribution and threshold dynamics of a stochastic sirs model with a general incidence, *Physica A: Statistical Mechanics and its Applications*, p. 120696, 2019.

[13] W. Cai, D. Yu, Z. Wu, X. Du, and T. Zhou, A hybrid ensemble learning framework for basketball outcomes prediction, *Physica A: Statistical Mechanics and its Applications*, vol. 528, p. 121461, 2019.

[14] L. Cai, Z. Zhang, J. Yang, Y. Yu, T. Zhou, and J. Qin, A noise-immune kalman filter for short-term traffic flow forecasting, *Physica A: Statistical Mechanics and its Applications*, p. 122601, 2019.

[15] L. Cai, Q. Chen, W. Cai, X. Xu, T. Zhou, and J. Qin, Svrgsa: a hybrid learning based model for short-term traffic flow forecasting, *IET Intelligent Transport Systems*, pp. 1–10, 2019.

[16] T. Zhou, G. Han, X. Xu, Z. Lin, C. Han, Y. Huang, and J. Qin, δ-agree adaboost stacked autoencoder for short-term traffic flow forecasting, *Neurocomputing*, vol. 247, pp. 31–38, 2017.

[17] T. Zhou, D. Jiang, Z. Lin, G. Han, X. Xu, and J. Qin, Hybrid dual kalman filtering model for short-term traffic flow forecasting, *IET Intelligent Transport Systems*, vol. 13, no. 6, pp. 1023–1032, 2019.

[18] T. Zhou, G. Han, X. Xu, C. Han, Y. Huang, and J. Qin, A learning-based multi-model integrated framework for dynamic traffic flow forecasting, *Neural Processing Letters*, vol. 49, no. 1, pp. 407–430, 2019.