

College Student Classification in Swimming Teaching and Training

Yang Jinyan, Xu Guanlei*

Student Brigade and Ocean department of Dalian Navy
Academy
Dalian, China

*Corresponding author: xgl_86@163.com

Li Yu, Zou Qiyang

Student Brigade of Dalian Navy Academy
Dalian, China
Liyu@163.com

Abstract—The swimming teaching and training play an important role in the university education. In order to improve the efficiency of swimming teaching and training for college students, the classification of college students is of much of signification before teaching and training. In this paper, we propose a classification method which is based on body shape such as body length, body width, body perimeter, body corporeity, body harmony that can be employed for training and classification. We establish a classification decision tree according to the distance between the different categories in the process of training and building support vector machine (SVM) classifier for the decision tree. It can classify the college students automatically and intelligently. Through testing a huge amount of college students sampled from many years, the experimental results show that the correct classification rate is over 90%.

Keywords—Swimming Teaching; Swimming Training; Support vector Machine (SVM); Classification; Tested Samples

I. INTRODUCTION

In sports and athletics, the body shape plays an important role. In many cases, before the selection of athletes or candidates in sports, the body shape is often analyzed in great details to show the potential ability in sports. Now there have been lots of researches on this field such as [1-5] and so on, e.g., in many applications, automatic dress size measurement and virtual try-ons, but also for virtual stunt men in movie productions, virtual models of real persons have to be created that are as detailed as possible and faithfully represent the true body skin surface. However, in this paper, we mainly focus on these fields such as body length, body width, body perimeter, body corporeity, body harmony and so on because they are related with swimming capability.

With the development of society and the improvement of human life, more and more people will take swimming. In the university education, swimming training becomes an interesting sport for college students. So, the swimming teaching and training will play a more important role in the university education. In order to improve the efficiency of swimming teaching and training for college students, the classification of college students is of much of signification before teaching and training. In this paper, we will propose a classification method which is based on body shape such as weight, height, thinness and fatness, body harmony that can be employed for training and classification. We establish a

classification decision tree according to the distance between the different categories in the process of training and building support vector machine (SVM) classifier for the decision tree. It can classify the college students automatically and intelligently. Finally, the experiments via testing a huge amount of college students sampled from many years will be made to show the efficiency of our proposed idea and method.

II. CLASSIFICATION AND SUPPORT VECTOR MACHINE

A. Methods of Classification

The classification methods based on different features are to classify the students for better swimming training and teaching. The traditional classification methods are as follows:

1) Artificial neural network [6] of classification

An artificial neural network (ANN) is a network inspired by biological neural networks (i.e., the central nervous systems of animals, in particular the brain) which are used to estimate functions that can depend on a large number of inputs that are generally unknown. In an ANN, the architecture specifies what variables are involved in the network and their topological relationships. The activity rule most neural network models have short time-scale dynamics: local rules define how the activities of the neurons change in response to each other. Typically the activity rule depends on the weights (i.e., the parameters) in the network. The learning rule specifies the way in which the neural network's weights change with time. This learning is usually viewed as taking place on a longer time scale than the time scale of the dynamics under the activity rule. Usually the learning rule will depend on the activities of the neurons. It may also depend on the values of the target values supplied by a teacher and on the current value of the weights. Generally, there are some main kinds of ANN such as BP, Hopfield and RBF are widely used[6].

2) Bayesian classifier[6]

In pattern recognition and statistics, classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. Unlike other procedures, Bayesian classification procedures provide a natural way of taking into account any available information about the relative sizes of the sub-populations associated with

This work is sponsored by the NSFCs (61002052, 61471412).

the different groups within the overall population. Bayesian procedures tend to be computationally expensive and, in the days before Markov chain Monte Carlo computations were developed, approximations for Bayesian clustering rules were devised. A Bayesian network model is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

3) Support vector machine (SVM) [6]

In pattern recognition, SVMs are supervised learning models with associated learning algorithms that analyze input data used for classification and furthermore regression analysis. Given a set of training samples, each marked as belonging to one or the other, an SVM training algorithm builds a model that assigns new samples to one category or the other to make it a non-probabilistic binary linear classifier. An SVM model is a representation of the samples as points in space. The samples of the separate categories are divided by a clear gap that is as wide as possible. New samples are then mapped into that same space and predicted to belong to a category based on which side of the gap they will fall on. SVMs can efficiently perform a non-linear classification using the kernel function mapping their inputs (data) into high-dimensional feature (data) spaces. After the comparison of these three methods for classification, we will employ SVM for our classification in this paper.

B. Support Vector Clustering for Student Classification

The clustering algorithm which provides an improvement to the support vector machines is called support vector clustering [6] and is often used in industrial applications either when data is not labeled or when only some data is labeled as a preprocessing for a classification pass.

1) The main features of body shape for swimming

In this paper, different from the work in [7,8] mainly focused on the main types of body shape such as Fig.1, this paper will mainly employ the features such as weight, height, thinness and fatness, body harmony.

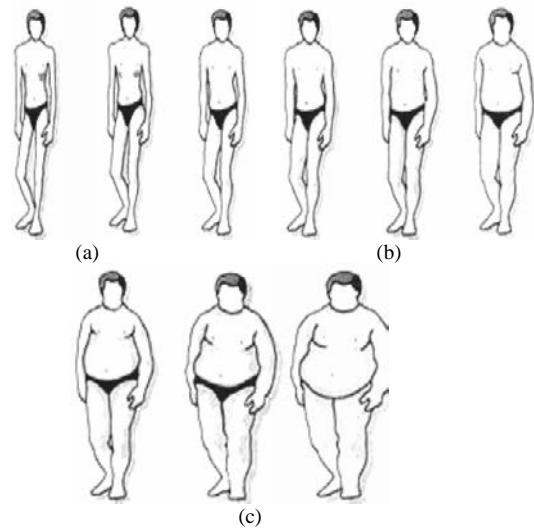


Fig. 1. Swimmers with different body shapes. (a) thin-long swimmers; (b) stout swimmers; (c) motile swimmers.

The features of bodies can be defined as four aspects: body length (including stature or height at standing, height at seating, the length of arms, the length of legs and so on), body width (including width of head, width of shoulders, width of waist, width of antrum and so on), body perimeter (including the perimeter of head, perimeter of chest, perimeter of arms, perimeter of legs and so on), body corporeity (including the substantiation of muscle and so on).

Features of thin-long: thin-long is a relatively concept because if one is thin then he/she will be relatively long. Therefore, thin is the main feature of the thin-long type. Thin-long will mainly include the height at standing, height at seating, the length of arms and the length of legs. In fact, in application, if one is thin-long in whole, or his/her ratio of length and width is high, then he/she will be considered as thin-long. Features of stout: stout is a relatively concept because if one is thin (fat) and short (long), then he or she will be relatively stout. Stout will mainly include the ratio between height at standing, height at seating, the length of arms, the length of legs and including width of head, width of shoulders, width of waist, width of antrum and so on. In fact, in application, if one is stout in whole, or his/her ratio of length and width is middle, then he/she will be considered as stout. Features of motile: motile is a relatively concept because if one is thin (fat) and very short, then he or she will be relatively motile. Motile will mainly include the ratio between height at standing, height at seating, the length of arms, the length of legs and including width of head, width of shoulders, width of waist, width of antrum and so on. In fact, in application, if one is motile in whole, or his/her ratio of length and width is low, then he/she will be considered as motile.

2) The design and realization of SVM classifier

The elementary principle of SVM is to minimize the structure risk and it has the perfect mathematical base and generalization ability. This paper will divide the students to two classes: the training set and the tested set, and through extracting the features of training set we gradually build up the SVM classifier. Then we will test the accuracy of the SVM

classifier. For the linear inseparable problem, there are two ways. The first one is the linearization of this linear inseparable case, which will result in the introduction of relaxed vectors. The other one is the trick of kernel function that is employed to project/map the data to the higher dimensional space. By mapping functions, the linear inseparable problem becomes the linear separable problem. In the new space, the dot product corresponds with the kernel function, as greatly improve the ability of nonlinear processing. Up to now, there have such kernel functions: linear kernel function, P-rank kernel function, multi-level sensing kernel function and RBF kernel function. After the introduction of kernel function, the inner product of vectors will be taken place by the kernel function:

$$\min Q(a) = \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j k(x_i, x_j) - \sum_{i=1}^n a_i, \quad (1)$$

$$\text{s.t. } a_i \geq 0, i=1, 2, \dots, n \quad \sum_{i=1}^n a_i y_i = 0. \quad (2)$$

The classifier is as follows:

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n a_i^* y_i k(x_i \cdot x) + b^* \right\}. \quad (3)$$

By selecting a SVM, we can get b^* :

$$y_i \left[\sum_{i=1}^n a_i^* y_i k(x_i \cdot x) + b^* \right] = 1. \quad (4)$$

In the new space, the dot product corresponds with the kernel function, and SVM greatly improves the ability of nonlinear processing. In the higher dimensional space, the SVM avoids the problem of "The Curse of Dimensionality". SVM has many kernel functions [6] such as linear kernel function, P-rank kernel function, multi-level sensing kernel function and RBF kernel function and so on. In this paper, we will employ the polynomial kernel function.

In fig.1, this is a linear separable problem of two classes. The segmentation lines (solid lines) can separate the two classes of data without any risk. The number of this kind of lines has countless many. However, if this line has the max distance away the two classes, it is the optimal. We call such line as the optimal classification line.

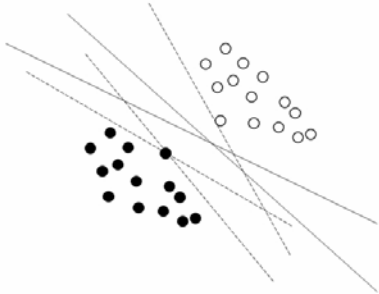


Fig. 2. An example of SVM for two classes

Set the linear separable sample set $(x_i, y_i)(i=1, 2, 3, \dots, n; x \in \mathbb{R}^d, y \in \{-1, 1\})$ is the label of class). In d -dimensional space, the general function has the form of $g(x) = w \cdot x + b$ with the separation function $w \cdot x + b = 0$. The decision function $f(x) = \text{sgn} \{w \cdot x + b\}$. After normalization

and the modulation of parameters W and b , finally our task is to minimize $\|w\|$.

So, we will have

$$\min \frac{1}{2} \|w\|^2, \text{ s.t. } y_i (w \cdot x_i + b) \geq 1 \text{ where } i=1, 2, \dots, n.$$

Then we define the Lagrange function:

$$L(w, a, b) = \frac{1}{2} \|w\|^2 - \frac{dy}{dx} \sum_{i=1}^n a_i [y_i (w \cdot x_i + b) - 1], \quad (5)$$

where, $a_i > 0$ is the parameter of the Lagrange function. This is a convex programming under the restrictive condition of $a_i \geq 0, i=1, 2, \dots, n \quad \sum_{i=1}^n a_i y_i = 0$. Here we note that the new signal transform can be employed such as EMD and other methods [9,10] for mapping function instead of the kernel function. But in this paper, we only use the traditional kernel functions.

The main steps of SVM for classification are defined as follows:

- ① Select the correct kernel function;
- ② Solve the optimal equation and the Lagrange function;
- ③ Give the optimal separation line equation;
- ④ Get the output of $\text{sgnf}(x)$.

III. THE EXTRACTION OF THE MAIN FEATURES OF STUDENTS

A. Body Length

The body length mainly includes the stature or height at standing, height at seating, the length of arms, the length of legs and so on. In most cases, the fast swimmer will have long body length including the long height, the long length of arms and the long length of legs. If the height, the length of arms and the long length of legs are short, it is nearly impossible that the swimming ability is better. However, if the height, the length of arms and the long length of legs are too long, it is also nearly impossible that the swimming ability is better. We can find that most of the perfect swimmers have the long height, the long length of arms and the long length of legs such as Sun Yang and Michael Phelps.

B. Body Width

The body width mainly includes the width of head, width of shoulders, width of waist, width of antrum and so on. In many cases, the fast swimmer will have thin width of head, wide width of shoulders, median width of waist, thin width of antrum and so on. If the width of head is big, then the water resistance will be big and the speed of swimming will be slow. The wide width of shoulders is the good guarantee of physical power for fast swimming. So generally, the fast swimmer will have wide width of shoulders. The median width of waist is an important element for perfect swimmer. The too wide or too thin width of waist is impossible for better swimming. There is also clear relation between perfect swimmers and the thin width of antrum for test.

C. Body Perimeter

The body perimeter mainly includes the perimeter of head, perimeter of chest, perimeter of arms, perimeter of legs and so on. In most cases, the fast swimmer will have relatively small perimeter of head because the water resistance will be big and the speed of swimming will be slow if the perimeter of head is big. On the other hand, if the swimmers want to get fast speed of swimming, the perimeter of arms and the perimeter of legs should be big so that the swimmers have enough physical power for fast swimming.

D. Body Corporeity

The body corporeity mainly includes the substantiation of muscle and so on. It is well known that the good swimmers must have the better body corporeity. However, if he/she has bad/poor body corporeity, it is impossible that he/she swims well. In many cases, the good body corporeity includes the good body harmony. So in this paper, we will combine the two elements. If we say the good body corporeity, we will mean good substantiation of muscle and good body harmony.

In table 1, we give the statistical variance values of over 1500 students that are tested in swimming and body shape. All these data are normalized to [0,1]. For example, if the body length is equal to 1, it means that this person has the height over 2.0m. If the body length is equal to 0, it means that this person has the height less than 1.5m. Other data implies the same physical sense. This means that the better swimmers have the similar body shape features such as the body length, the body width, the body perimeter and the body corporeity. Instead, if the swimmers have very different body shape features, this means most people won't swim very well.

TABLE I. THE VARIANCE VALUES OF FOUR ELEMENTS FOR SAMPLED STUDENTS

Terms	Better swimmer	General swimmer	Poor swimmer
Body Length	0.2534	0.3874	0.5174
Body Width	0.1789	0.2766	0.4034
Body Perimeter	0.3244	0.4252	0.5987
Body Corporeity	0.1034	0.2453	0.3345

IV. SVM BASED STUDENT CLASSIFICATION

Like the classification of texture [11], some main features must be employed. As shown in above, the main four elements (body length, body width, and body perimeter and body corporeity) are used in this paper as the main features. In this paper, the SVM used here is for multi-class. The inputs are the four parameters: body length, body width, and body perimeter and body corporeity. The outputs are the three classes: better swimmers, general swimmers and poor swimmers. By the SVM classifier, once the inputs are given, the output will be one of the three classes: better swimmers, general swimmers or poor swimmers. Finally, the accuracy defined as:

Classification accuracy = the correct classification samples/all the tested samples. (6)

In this paper, we test 1000 students in a few years in all. Before this, we select 1000 students for our training samples. Table 2 is the final classification result.

TABLE II. THE TESTED RESULTS OF SVM CLASSIFICATION FOR SWIMMING STUDENTS

Swimming classes	Number of students	Classification accuracy
Better swimmer	419	93.1%
General swimmer	276	92.3%
Poor swimmer	305	90.8%

According to the result of table 2, the experimental results show that the correct classification rate is over 90%. This implies that in the future we can use this method for student swimming classification. In such way, we can make different teaching and training plans.

V. CONCLUSIONS

In sports and athletics, the body shape plays an important role. In many cases, before the selection of athletes or candidates in sports, the body shape is often analyzed in great details to show the potential ability in sports. Similarly, in order to make different swimming teaching and training plans, the method for student swimming classification is very important. In this paper, we propose a classification method which is based on body shape such as body length, body width, body perimeter, body corporeity, body harmony that can be employed for training and classification. We establish a classification decision tree according to the distance between the different categories in the process of training and building support vector machine (SVM) classifier for the decision tree. It can classify the college students automatically and intelligently. Through testing a 1000 college students sampled from a few years, the experimental results show that the correct classification rate is over 90%. That is to say, in the future, we can use this method for student swimming classification. In such way, we can make different teaching and training plans.

REFERENCES

- [1] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, J. Davis, Scape: shape completion and animation of people, *ACM Trans. on Graphics* 24 (3) (2005) 408–416.
- [2] O. Weber, O. Sorkine, Y. Lipman, C. Gotsman, Context-aware skeletal shape deformation, *Computer Graphics Forum* 26 (3) (2007) 265–274.
- [3] Li Shiming. Research on theory and experiment of bodies' sports via balanced board[D].Beijing: Beijing Tiyu University, 2003.
- [4] A. O. Balan, M. J. Black, The naked truth: Estimating body shape under clothing, in: D. A. Forsyth, P. H. S. Torr, A. Zisserman (Eds.), *Proc. ECCV*, Vol. 5303 of Lecture Notes in Computer Science, Springer-Verlag, Marseille, France, 2008, pp. 15–29.
- [5] Yang Jinyan, Li Yu, Xu Guanlei, General Running Pose Analysis via Sparseness, 2nd International Conference on Electric Information and Control Engineering(ICEICE 2012): 4648-4651.
- [6] Richard O. Duda, Peter E. Hart, David G. Stork. *Pattern Classification* [M]. John Wiley Press., 2004(2nd Edition)
- [7] Yang Jinyan, Li Yu, Xu Guanlei, General Running Pose Analysis via Sparseness, 2nd International Conference on Electric Information and Control Engineering(ICEICE 2012): 4648-4651.

- [8] Jinyan Yang, Yu Li, Tao Jiang, Yu Wei, Guanlei Xu. Body Shape Analysis via Image Processing, International Conference on Educational Research and Sports Education (ERSE 2013): 101-104.
- [9] Xu Guanlei, Wang Xiaotong, Xu Xiaogang. Improved bi-dimensional EMD and Hilbert spectrum for the analysis of textures[J].Pattern Recognition,2009,42(5):718 – 734
- [10] Xu Guanlei, Wang Xiaotong, Xu Xiaogang. On Analysis of Bi-dimensional Component Decomposition via BEMD[J]. Pattern Recognition,2012,45(4):1617-1626.
- [11] R. Chellappa and S. Chatterjee, “Classification of textures using Gaussian Markov random fields,” IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-33, no. 4, pp. 959–963, Apr. 1985.