Children Abnormal Gait Analysis Based on SVM

Zhelong Wang, Ming Jiang, Yun Zhang

***Abstract*—Support Vector Machine (SVM) has become a hotspot of machine learning because of its rigorous theory background and remarkable generalization performance. Recent researches on SVM mainly concentrate on the property of SVM and variety of its applications. In this paper, an improved SVM algorithm is proposed for children abnormal gait analysis. The algorithm combines SVM with fuzzy clustering in order to improve the accuracy of SVM. Only samples that have weak relationships with all clusters are involved in SVM. Simulation experiment has been carried out to show that the algorithm based on the improved SVM may obtain better effectiveness than the normal SVM when it is applied for children abnormal gait analysis.**

***Index Terms*—Children Abnormal Gait Analysis; SVM**

I. Introduction

Gain analysis is a very important technique for early diagnosis of gait disease and rehabilitation assessment. It is inexact that most doctors diagnose gait disease based on their own judgments of comparing patients' gait with many curves created by a certain gait analysis system. In recent years, machine learning technology has gained much regard in gait analysis field, which may help doctors gain a more objective and exact disease assessment [1].

Artificial neural network (ANN) has been applied to distinguish normal gait and diseased gait due to its strong non-linear learning ability [2]. But, ANN often gets into a local minimum and overstrains training samples which may reduce the accuracy of classifier. In addition, the collection of gait data is time-consuming, so the number of gait samples used for its analysis is usually small and then it is not suitable to apply ANN. Support Vector Machine (SVM) is widely applied in pattern recognition because of its remarkable learning ability. Begg [3] applied SVM for the gait classification of the children and the elder. Kamruzzaman [4] used SVM to distinguish the gait of children with cerebral palsy.

In this paper, an algorithm named as F-SVM has been proposed for children gait analysis by combining SVM with fuzzy clustering [5],[6]. The algorithm first clusters training samples into several clusters by using FCM. Only samples that have a weak relationship with each cluster are chosen to be trained in SVM. By removing those un-support vectors, F-SVM may acquire much higher accuracy than standard SVM. The validity of F-SVM has been shown by simulation experiment conducted in this study.

Manuscript received July 6, 2009. This work was supported in part by the National Natural Science Foundation of China under Grant 60605022, in part by the China Earthquake Research Funds (200808075).

Z. Wang, M. Jiang and Y. Zhang are with School of Electronic and Information Engineering, Dalian University of Technology, Dalian, Liaoning, 116024, CHINA. (Corresponding author Z. Wang, e-mail: wangzl@dlut.edu.cn)

Z. Wang is also with Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, Liaoning, 110016, CHINA.

II. Description of F-SVM

A. Short introduction of SVM

SVM introduced by Vapnik [7] is usually used for classification tasks. In a task of binary classification, SVM aims to find an optimal separating hyperplane (OSH) which generates a maximum margin between two categories of data. To construct an OSH, SVM maps data into a higher dimensional feature space. SVM performs this nonlinear mapping by using a kernel function. Then, SVM constructs a linear OSH between two categories of data in the higher feature space. Data vectors which are nearest to the OSH in the higher feature space are called support vectors (SVs) and contain all information required for classification. In brief, the theory of SVM is as follows [7].

Consider a training set $D=\left\{\left(x_i, y_i\right)\right\}_{i=1}^{L}$ with each input $x_i \in R^n$ and an associated output $y_i \in \{-1,+1\}$. Each input x is firstly mapped into a higher dimension feature space F, by $z=\phi(x)$ via a nonlinear mapping $\phi: R^n \to F$. When data are linearly non-separable in F, there exists a vector $w \in F$ and a scalar b which define the separating hyperplane as:

$$y_i\left(w' \cdot z_i + b\right) \ge 1-\xi_i,\ \ \forall i \tag{1}$$

where $\xi_i(\ge 0)$ are called slack variable. The hyperplane that optimally separates the data in F is one that

$$\begin{aligned} &\mathit{minimize}\ \ \frac{1}{2}\cdot w' \cdot w + C \cdot \sum_{i=1}^{L}\xi_i \\ &\mathit{subject\ to}\ \ \ y_i\left(w' \cdot z_i + b\right) \ge 1-\xi_i,\ \ \xi_i \ge 0, \forall i \end{aligned} \tag{2}$$

where C is called regularization parameter that determines the tradeoff between a maximum margin and a minimum classification error. By constructing a Lagrangian, the optimal hyperplane according to (2) may be shown as the solution of

$$\begin{aligned} &\mathit{maximize}\ \ W(\alpha)=\sum_{i=1}^{L}\alpha_i - \frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{L}\alpha_i\alpha_j y_i y_j K\left(x_i, x_j\right) \\ &\mathit{subject\ to}\ \ \sum_{i=1}^{L} y_i\alpha_i = 0,\ 0 \le \alpha_i \le C,\ \forall i \end{aligned} \tag{3}$$

where $\alpha_1, \cdots, \alpha_L$ are the nonnegative Lagrangian multipliers. The data points x_i that correspond to $\alpha_i > 0$ are SVs. The weight vector w is then given by

$$w = \sum_{i \in SVs} \alpha_i y_i z_i \tag{4}$$

For any test vector $x \in R^n$, the classification output is then given by

$$y = sign(w \cdot z + b) = sign\left(\sum_{i \in SVs} \alpha_i y_i K(x_i, x) + b \right) \tag{5}$$

To build an SVM classifier, a kernel function and its parameters need to be chosen. So far, no analytical or empirical studies have established the superiority of one kernel over another conclusively. In this study, the following three kernel functions have been applied to build SVM classifiers:

1) Linear kernel function, $K(x,z) = \langle x, z \rangle$;
2) Polynomial kernel function, $K(x,z) = (\langle x, z \rangle + 1)^d$, d is the degree of polynomial;
3) Radial basis function, $K(x,z) = \exp\left\{ -\|x - z\|^2 / 2\sigma^2 \right\}$, σ is the width of the function.

B. Introduction of FCM algorithm

FCM algorithm proposed by Dunn [8] and extended by Bezdek [9] is one of the most well-known methods in clustering analysis. FCM partitions a set of *s*-dimensional vectors $X = \{X_1, X_1, \cdots, X_n\}$ into c clusters, where $X_j = (X_{j1}, X_{j2}, \cdots, X_{js})'$ represents the *j*th sample for $j = 1, \cdots, n$. The *i*th cluster is supposed to have a center vector $v_i = (v_{i1}, \cdots, v_{is})$ and FCM aims to determine cluster centers v_i , where $1 \le i \le c$. For the *j*th sample X_j and the *i*th cluster center v_i, there is a membership degree $u_{ij}\,(\in [0,1])$ indicating in what degree the sample X_j belongs to the cluster center vector v_i, which results in a fuzzy partition matrix $U = (u_{ij})_{n \times c}$. The objective function J is defined as follows:

$$J(U, v_1, \cdots, v_c, X) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2 \tag{6}$$

$$d_{ij} = \left(\sum_{k=1}^{s} (v_{ik} - x_{jk})^2 \right)^{\frac{1}{2}} \tag{7}$$

$$v_i = \sum\nolimits_{j=1}^{n} u_{ij}^m X_j \Big/ \sum\nolimits_{j=1}^{n} u_{ij}^m \tag{8}$$

$$u_{ij} = \left(\sum\nolimits_{k=1}^{c} (d_{ij} / d_{kj})^{\frac{2}{(m-1)}} \right)^{-1}, \; m \ne 1 \tag{9}$$

subject to $\sum_{i=1}^{c} u_{ij} = 1$, $\forall j = 1, \cdots, n$, where m (usually set to be 2) in (8) and (9) is used to adjust the weight effect of membership values. The FCM algorithm [10] may be described as follows:

1) Choose an integer c and a threshold value ε. Initialize the fuzzy partition matrix U by $c \times n$ random numbers in the interval $[0,1]$;
2) Compute $v_i\,(i = 1, \cdots, c)$ according to (8);
3) Compute all d_{ij} and u_{ij} according to (7) and (9) respectively. Thus update the fuzzy partition matrix U by the new computed u_{ij}.
4) Compute the objective function J by using (6). If the difference between two adjacent values of J is less than the given threshold ε, then stop. Otherwise go to step 2.

C. F-SVM algorithm

SV plays a decisive role in SVM, but non-SVs are usually inoperative. In F-SVM, treatment is not the same for each sample. Training samples that are affirmatively not SVs are not involved in SVM, and only samples that have a weak relationship with each cluster are chosen to be trained in SVM. By using this method, the classification accuracy and the training time may be effectively improved [5],[6].

The selection of training data is executed by using FCM which may cluster samples non-compulsively. The membership degree u_i in FCM indicates with what degree a sample belongs to the cluster center vector v_i. If there exists one $u_{i_0}\,(i_0 \in \{1, \cdots, c\})$ which is bigger than a threshold λ (80% in this study), it is clearly that the sample is far from OSH and its probability to be a SV is small. So, this sample is not involved in SVM. The block diagram of F-SVM is shown in Figure 1.

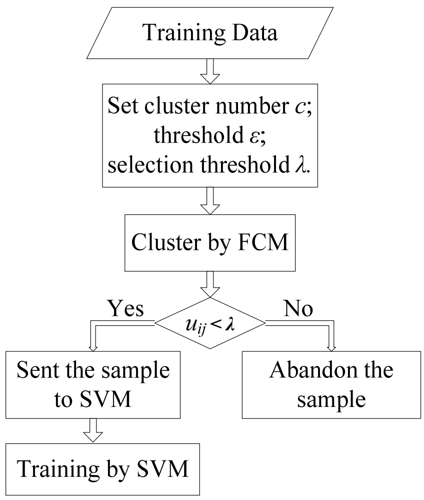


Figure 1, Block diagram of the F-SVM.

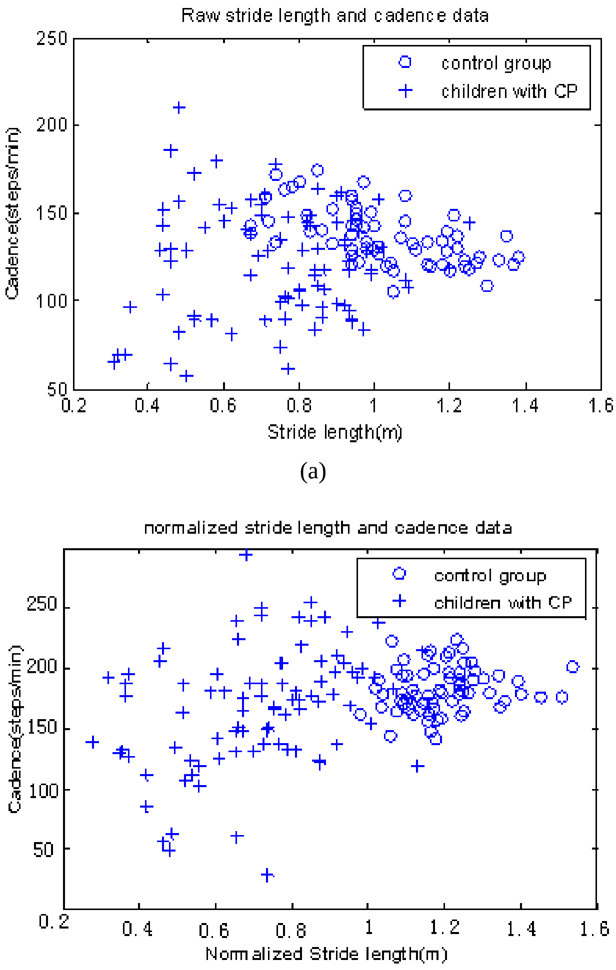
The process of F-SVM is described as follows:

1) Choose a cluster number c (2 in this study) and a threshold value ε for FCM. Choose a selection threshold λ (80% in this paper).
2) Cluster training samples by applying FCM and calculate each membership degree u_{ij} according to (9). If $u_{ij} < \lambda$, $\forall i \in \{1, \cdots, c\}$, sent the sample to SVM. Otherwise, abandon it.

3) Classify the selected training sample set by using SVM. Calculate each classification output according to (5).

III. Experiments and Analysis

In this study, the proposed algorithm F-SVM which combines SVM with FCM is applied for children gait analysis. In this section, several experiments are carried out to test the validity of F-SVM. The experimental data used in this study are obtained from a gait database in Virginia University [11]. There are two sample sets in the database including normal gait samples from 68 children and abnormal samples from 88 children with cerebral palsy (CP). The ages of these children range from 2 years old to 13 years old. Four features of gait samples are selected and they are stride length, cadence, leg length and age.



(a)

(b)

Figure 2, (a) Distribution of raw data in feature space. (b) Distribution of normalized data in feature space.

In this study, polynomial normalization in paper [12] is applied to normalize the gait samples. Figure 2 shows the contrast of samples distribution before and after normalization. As shown in Figure 2, the overlap of two sample sets is effectively reduced after normalization, which helps to improve the classification accuracy. Three kernel functions are used to build SVM classifiers in this study. By comparing the classification results of three classifiers, the most suitable kernel function may be decided for F-SVM. By applying this kernel function, the accuracy of F-SVM and standard SVM is compared at the end of this section.

Polynomial order d is an important parameter when polynomial kernel function is applied in F-SVM. The classification accuracy of gait samples by using different polynomial order is shown in Figure 3, and d is chosen from 1 to 10. As shown in Figure 3, the classification accuracy declines along with the increase of polynomial order. This is because the dimension of the feature space is high under a large polynomial order, and it leads a declining generalization capability of SVM. So, the polynomial order d is set to be 1 in the following experiments. When d is 1, the polynomial kernel function actually is a linear kernel function and the classification accuracy may reach 89.74% under it.

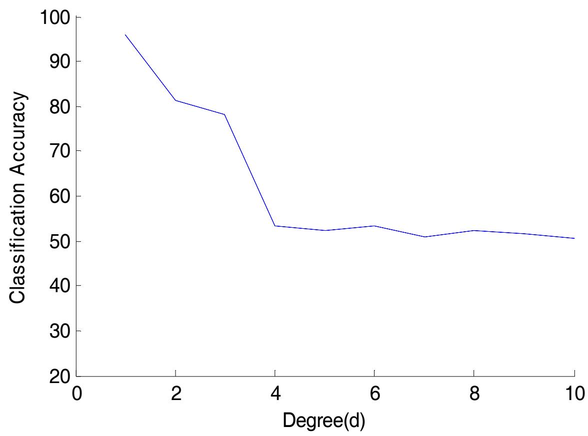


Figure 3, Relationship between classification accuracy and polynomial order

When radial basis function (RBF) is applied as the kernel function, the kernel parameter σ in RBF and regularization parameter C in equation (2) may impact the classification accuracy of F-SVM. Figure 4 shows the relationship between classification accuracy and parameters combination.

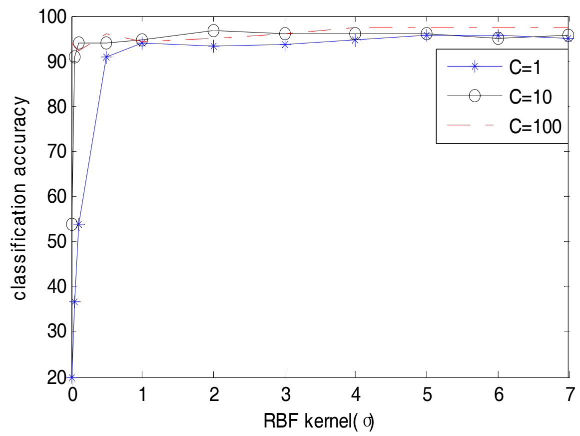


Figure 4, The relationship between classification accuracy and parameters combination

As shown in Figure 4, the generalization capability of SVM enhances along with the increase of C. This is because the regularization parameter C may adjust the ratio of confidence interval and empirical risk. The generalization capability of SVM is weak when C is small, because a small

C indicates that the punishment for empirical risk is small and the empirical risk is large. When C exceeds a certain value, the complexity of a classifier reaches the allowed maximum in the feature space. In this case, the SVM has almost no change of the empirical risk and generalization capability. Table I shows the classification accuracy of the three classifiers.

Table 1, Accuracy of three classifiers

Kernel function	Parameter	Accuracy (%)
Linear	$d = 1$	89.74
Polynomial	$d = 2$	82.69
RBF	$\sigma = 4, C = 100$	97.51

As shown in Table 1, F-SVM using RBF as the kernel function has the most accurate classification result when the kernel parameter and regularization parameter are set to be 4 and 100 respectively. Table 2 shows the comparison of F-SVM and SVM with the same kernel function and parameters.

Table 2, Comparison of SVM and F-SVM

Algorithm	Accuracy (%)
SVM	95.51
F-SVM	97.72

As shown in Table 2, classification accuracy of F-SVM is higher than the standard SVM. This is because only those important samples are sent to the classifier in F-SVM, and samples which have small possibility to be SVs. By using this method, the classification accuracy and the training time may be effectively improved.

IV. Conclusion

In this paper, a new algorithm F-SVM is proposed for children gait analysis by combining SVM with fuzzy clustering. F-SVM firstly clusters training samples into several clusters by using FCM algorithm. Only those samples that have a weak relationship with each cluster are chosen to be trained in SVM. By taking this method, the classification accuracy and the training time have been effectively improved. Simulation experiments have been conducted to show the validity of F-SVM algorithm. Experimental results show that F-SVM may obtain high classification accuracy for children gait classification even when the number of samples is small. In the experiment, the effects of three kernel functions in F-SVM have also been tested. The RBF has achieved the most accurate classification result when appropriate parameters are adopted.

References

[1] Simon S. R., *Quantification of human motion: gait analysis-benefits and limitations to its application to clinical problems.* Journal of Biomechanics, 2004(37): 1869-1880.

[2] Barton J G, Lees A, *An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams.* Gait and Posture, 1997(5): 28-33.

[3] R. K. Begg, M. Palaniswami, and B. Owen, *Support Vector Machines for automated gait classification.* IEEE Trans. Biomed. Eng., vol. 52, no. 5, pp. 828-823, May 2005.

[4] Joarder Kamruzzaman, *Support Vector Machines and Other Pattern Recognition Approaches to the Diagnosis of Cerebral Palsy Gait.* IEEE Trans. Biomed. Eng., vol. 53, no. 12, pp. 2479-2489, Dec 2006.

[5] Sheng-Wu Xiong, Hong-Bing Liu and Xiao-Xiao Niu, *Fuzzy support vector machines based on FCM clustering.* Proceedings of 2005 International Conference on Machine Learning and Cybernetics, Aug. 2005, vol. 5, pp. 2608- 2613.

[6] Dae-Jong Lee, Jong-Pil Lee, Pyeong-Shik Ji, Jae-Woon Park and Jae-Yoon Lim, *Fault Diagnosis of Power Transformer Using SVM and FCM.* 2008 IEEE International Symposium on Electrical Insulation, June 2008, pp. 112-115.

[7] Vladimir N. Vapnik, *The Nature of Statistical Learning Theory.* New York: Springer-Verlag, 1995, 187 pp.

[8] Dunn and J.C., *Some recent investigations of a new fuzzy partition algorithm and its application to pattern classification problems.* J. Cybernetics, vol. 4, pp. 1–15, 1974.

[9] Bezdek and J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum, New York, 1981.

[10] X. Wang, Y. Wang, and L. Wang, *Improving Fuzzy c-Means Clustering based on Feature-Weight Learning.* Pattern Recognition Letters, 25 (10), pp. 1123–1132, 2004.

[11] M. J. O'Malley, M. F. Abel, D. L. Damiano, and C. L. Vaughan, *Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters.* IEEE Trans. Rehab. Eng., vol. 5, no. 4, pp. 300–309, Dec. 1997.

[12] M. J. O'Malley, *Normalization of temporal-distance parameters in paediatric gait.* J. Biomech., vol. 29, no. 5, pp. 619–625, 1996.