

Southeast Europe Journal of Soft Computing

Available online: www.scjournal.com.ba



HUMAN CHROMOSOME CLASSIFICATION USING COMPETITIVE SUPPORT VECTOR MACHINE TEAMS

Ali Osman Kusakci, Sadina Gagula-Palalic

International University of Sarajevo, Faculty of Engineering and Natural Sciences, Hrasnicka Cesta 15, Ilidža 71210 Sarajevo, Bosnia and Herzegovina

Article Info

Article history: Recivied Sep 2014 Recivied in revised form Nov 2014

Keywords: Support Vector Machines, chromosome classification, competitive support vector machine teams

Abstract

Classification of chromosome is a challenging task and requires very precise autonomous classifier. This paper proposes to employ competing support vector machines (SVMs) placed in a grid. Each agent in cells of the grid is responsible to distinguish two classes. Overall output is determined by simple majority voting of SVMs. Relying same principle as the work by Palalic and Can [17], we compared the results obtained where the algorithms delivers better accuracy.

1. INTRODUCTION

Human chromosome classification represents a very important part of the human chromosome analysis process. The classification of chromosomes into classes from medical images is called as karyotyping which aims to diagnose possible disease.

Counting and karyotyping of chromosomes is very time consuming process. Since 1980s researchers are trying to develop reliable systems for automatic chromosome karyotyping and analysis. Most of these systems are semiautomated, requiring manual assistance of the laboratory technicians, especially in the images where there are touching and overlapping chromosomes.

Classification of the human chromosomes based on the classification of the chromosome features is the most studied area in the chromosome analysis. For the classification purposes, several databases have been used: Copenhagen database, Edinburg database, Philadelphia database and Denver database.

2. LITERATURE OVERVIEW

Many classification algorithms and their performances on these databases have been reported, among them: neural network based (ANN) [1], [2], distance and statistical based [3-5], neuro-fuzzy based [6], nearest-neighbor based [7], rule based [8], fuzzy-rule based [9]. The greatest success among all is a tribute to ANN with the most accurate classification rate of 93.5% on the Copenhagen data set, as reported in [10].

Recently, Ventura et al [11] have experimented classification approach of pairing chromosomes using a set of geometric and band pattern features using an algorithm based on the Bayesian framework. This method resulted in a maximum average paring rate of 92.8%.

As medical diagnosis is a challenging task which requires employing very precise classification algorithm, the mentioned correct classification rates are not truly satisfied. This necessitates more powerful algorithms to be designed.

3. METHODOLOGY

3.1 DATASET

The data used in this work is taken from Copenhagen data base with 4400 samples, 200 samples for each chromosome types. Randomly selected 80% of the data is used for training while the remaining 20% is kept for testing purposes. The dataset composed of 28 input features where the first two are the length of chromosome and the position of the centromere, called as centromere index (CI). Based on the preliminary experiments, another dimension

contributing to the success of the algorithm is taken as the ratio of the first two components, the ratio of CI to the total length. The other 25 features are the first 25 principal components obtained from the standardized gray levels of a chromosome vertical to the main axis.

One classification scheme is so called seven Denver Groups (A-G) as seen in Table1.

Table 1: Classification of chromosomes in Denver classification

Chromosome class	Denver Group
#1- #3	Group A
#4-#5	Group B
#6-#12, X	Group C
#13-#15	Group D
#16-#18	Group E
#19-#20	Group F
#21-#22, Y	Group G



Figure 1. Human Chromosome

3.2 SUPPORT VECTOR MACHINES

Support vector machines (SVMs) have gained more importance in the late decade due to its specific properties favoring them as en efficient classification tool. SVM is based on statistical learning theory pioneered by Vapnik [12] and can be used for pattern classification and nonlinear regression.

In the case of separable patterns, the main idea of SVM is to construct a hyperplane as the decision surface in such a way that the margin of separator between the two classes is maximized [13]. Hence, classification task is directly performed based on support vectors. While separating the classes, SVM performs two main mathematical operations on the data set.

- Nonlinear mapping of the input data into a high-dimensional hidden feature space which can be separated linearly.
- Construction of an optimal hyperplane for separating the features discovered in step 1 [13].



Figure 2. Transformation of input space to feature space in SVM construction (step 1) [13]

We can state the motivation for employing SVM as a classification tool as follows: the support vector classifier chooses one particular solution with the highest generalization ability: the classifier which separates the classes with maximal margin [14].



Figure 3. Linear support vector classifier [14]

Given a training set for the SVM as, $P = \{(x_i, y_i) | x_i \in \mathbb{R}^k, y_i \in \{-1,1\}\}_{i=1}^n$, where x_i stands for the n-dimensional input vector and y_i is output vector. Remembering that SVM tries to build a decision boundary with maximum possible margin between the classes, for linearly non-separable input data a transfer to a feature space with higher dimension is performed as:

 $\phi(\mathbf{x}): \mathbf{x}_i \in \mathbb{R}^{\bar{k}} \to \mathbb{R}^m, k \ll m.$

Vapnik [12] suggested selecting the transformation function from a family of universal approximation with linear separable character.

$$f(\mathbf{x}_i) = \mathbf{w}^T \phi(\mathbf{x}_i) + b \ge 1 \quad \forall i: y_i = 1$$

$$f(\mathbf{x}_i) = \mathbf{w}^T \phi(\mathbf{x}_i) + b \le -1 \quad \forall i: y_i = -1$$
(1)

The aim is to find the values of w and b for the optimal hyperplane, maximizing the margin of separation, given the training set P. The maximization of the margin is realized by minimizing the Euclidian norm of the weight vector w. For nonseparable problems, the constraints in Eq (1) can be weaken by replacing it with a soft margin involving slack variables ε_i . That is, for non-separable case, we allow that some data points may lie inside the margin area while the sum of the slack variables is used to limit the extent of violations. In this case, the optimization problem can be expressed as:

Min

$$\psi(\boldsymbol{w},\boldsymbol{\varepsilon}) = 1/2\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^N \varepsilon_i$$
 (2)

Subject to:

$$y_{i}(\boldsymbol{w}^{T}\boldsymbol{x}_{i}+b) \geq 1 - \varepsilon_{i} \text{ for } i = 1, ..., N$$

$$\varepsilon_{i} \geq 0 \text{ for all } i$$
(3)

where C is a user defined penalty parameter controlling the tradeoff between the complexity of the machine and the number of nonseparable points [13].

The inner product of two vectors induced in the feature space, $\phi(x)$ and $\phi(x_i)$, defines a kernel function [13]

$$K(\mathbf{x}, \mathbf{x}_i) = \phi(\mathbf{x})^T \phi(\mathbf{x}_i) \tag{4}$$

Radial basis, Gaussian, and polynomial kernels are mostly used kernel functions in the literature. Basically, they permit us to construct a decision surface that is nonlinear in the input space, but its image is linear in the feature space [13].

Practically, it is simpler to optimize the dual problem. With a specific kernel function K(.,.), the optimization problem can be formulated in the dual form as follows:

Given the training set $\{(x_i, y_i)\}_{i=1}^N$, find the Lagrange multipliers $\{\alpha_i\}_{i=1}^N$ that maximize the objective function

$$\sum_{i=1}^{N} \alpha_{i} - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_{i} \alpha_{j} y_{i} y_{j} K(\boldsymbol{x}_{i}, \boldsymbol{x}_{j}), \quad (5)$$

subject to the constraints:

$$\sum_{i=1}^{N} \alpha_i d_i = 0$$

$$0 \le \alpha_i \le C \text{ for } i = 1, \dots, N.$$
(6)

While designing an efficient SVM with RBFs as kernels, two parameters are of paramount importance, (C, σ^2) , the penalty parameter and the width of the RFB respectively [15]. In the formulation given in (6), *C* is also referred as box constraint parameter serving as upper bound for the constraints.

In order to find the optimal values of the couple, (C, σ^2) , for each SVM, we employed a numerical optimization

method, pattern search (PS). PS is an efficient directionbased deterministic search algorithm with no requirement of derivative information. The general scheme of all PS algorithms involves the construction of mesh of points, around the current solution. Then, the mesh is refined and the process repeated, if the current solution remained unimproved [16].

3.3 COMPETITIVE SVM TEAMS

Although SVMs are not designed for the classification problems with multiple classes as in the case of chromosome sorting, on-against-one (OAO) SVM classifiers can be trained to distinguish any pair of classes, {i, j}. However, OAO approach comes with the difficulty of combining the outputs of all SVMs and providing single final output.

Let us define a single Support Vector Machine (SVM) to be $SVM_{i,j}$ where i, j = A, B, ..., G, where $i \neq j$. The $SVM_{i,j}$ is trained to distinguish classes of type *i* and type *j*, creating the output *i*, if this new data is of type *i*, and creates the output *j*, if this new data is of type *j*. Thus, the overall structure of the classifier can be represented as a 7x7 matrix as seen in **Table 2**. Clearly, there is no SVM assigned on the diagonal axis.

Table 2: Overall structure of competitive SVM teams

Α	В	С	D	Е	F	G
Ø	SVM _{AB}					SVM _{AG}
SVM _{BA}	Ø					
		Ø				
			Ø			
				Ø		
SVM _{FA}					Ø	SVM _{FG}
SVM _{GA}					SVM _{GF}	Ø
	A Ø SVM _{BA} SVM _{FA}	A B Ø SVM _{AB} SVM _{BA} Ø SVM _{FA} SVM _{GA}	A B C Ø SVM _{AB} SVM _{BA} Ø Ø SVM _{BA} Ø J Ø SVM _{FA} Ø SVM _{GA}	A B C D Ø SVMAB SVMBA Ø Ø Ø Ø Ø Ø Ø J Ø SVMFA I I I SVMGA I I	A B C D E \emptyset SVM _{AB} SVM _{BA} \emptyset \emptyset \emptyset \emptyset \emptyset SVM _{FA} Image: SVM_GA Image: SVM_GA	A B C D E F \emptyset SVM _{AB} SVM _{BA} \emptyset \emptyset \emptyset \emptyset \emptyset SVM _{FA} \emptyset \emptyset

After the training stage, assume a new data of unknown type is to be classified. If the new data belongs to a chromosome of type k, then it is expected that all SVMs in the k-th column, $SVM_{j,k}$ j=A,B, ..., G will give an output of k. On the other hand, it is expected that all SVMs in the k-th row $SVM_{k,i}$, i=A,B, ..., G will give an output of k. Hence, the output will be a 7x7 matrix with zeros on the diagonal axis. The entries on this matrix represent the votes of each SVM assigned to the classes.

The idea of competing classifier teams was implemented by Palalic and Can [17] referring the designed classifier as Competitive Artificial Neural Network Teams (CANNT). In this work, the same structure of competing classifiers is implemented with SVMs. That is, a competition will take place between the teams of SVMs trained for a target class. Unlike [17], the overall output of the classifier will be determined by majority voting instead of Euclidean distance.

61 A.O. Kusakci & S. Gagula-Palalic / Southeast Europe Journal of Soft Computing Vol.3 No.2Sep. 2014 (58-61)

4. RESULTS

The obtained results during training and testing stages are given in Table 3 and Table 4, where the high correct classification of the proposed algorithm is clearly visible. On the training data, correct classification ratio is 99.66% while the algorithm was able to identify 99.54% of the test samples correctly. Comparing the same indicators obtained with CANNT on the same dataset, we can conclude that SVM based classifier proposed here outperforms this algorithm.

Table 3: Correct classification rates during training and testing in Denver Classification

Chromosome	Training	Training	Testing	Testing
Туре	(%)	(%) NN	(%)	(%)
	SVM		SVM	NN
А	100.00	98.00	100.00	100.00
В	100.00	100.00	100.00	98.00
С	99.73	96.00	99.64	92.00
D	100.00	100.00	100.00	100.00
E	99.58	100.00	100.00	94.00
F	99.37	92.00	97.50	100.00
G	98.44	94.00	98.75	88.00
average	99.66	97.14	99.54	96.26

Table 4: Confusion matrix on test set in Denver Classification

%	Α	В	С	D	Е	F	G
Α	100	0	0	0	0	0	0
В	0	100	0	0	0	0	0
С	0.36	0	99.64	0	0	0	0
D	0	0	0	100	0	0	0
Е	0	0	0	0	100	0	0
F	0	0	0	0	3.5	97.50	0
G	0	0	0	0	0	1.25	98.75

Confusion matrix on test set in Denver Classification

REFERENCES

- B. Lerner, H. Guterman, I. Dinstein and Y. Romem, (1995) Human chromosome classification using multilayer perceptron neural network, *Int. J. Neural Syst.*, vol. 6, pp. 359-370.
- [2] P. A. Errington and J. Graham, (1993) Application of Artificial Neural Network to Chromosome Classification", *Cytometry* 14, pp. 627-639.
- [3] J. Piper, E. Granum, D. Rutovitz, and H. Ruttledge, (1980) Automation of chromosome analysis, Signal Processing, Vol. 2, pp. 203-221.
- [4] J. Piper and E. Granum, (1989) On Fully Automatic Feature Measurement for Banded Chromosome Classification, *Cytometry* 10, pp.242-255.
- [5] Q. Wu, Z. Liu, T. Chen, Z.K.R.C. Xiong, (2005) Subspace based prototyping and classification of chromosome images, *IEEE Trans Image Process*, 14.
- [6] X.Ruan, (2000) A classifier with fuzzy Hopfield network for human chromosomes, *Proceedings of the 3rd World Congress on Intelligent Control and Automation*, China, June 28 –July 2, 2000
- [7] F. C. A. Groen, T. K. Kate, A. W. M. Smeulders, and I. T. Young, (1989) Human chromosome classification based on local band descriptors, *Pattern Recognition Letter*, vol 9, pp 211-222.

- [8] Y. Lu and Y. Ya, 1989 An expert system for banded chromosomes recognition, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
- [9] O. Sjahputera and J. M. Keller, 1999 Evolution of a fuzzy rulebased system for automatic chromosome recognition, *IEEE International Fuzzy Systems Conference Proceedings*.
- [10] J. Cho, (2000) Chromosome classification using back propagation neural networks, *IEEE Engineering Medical Biology Magazine*, 19.
- [11] R. Ventura, A. Khmelinskii, and J.M. Sanches, (2010) Classifierassisted metric for chromosome pairing, *32nd Annual International Conference of the IEEE EMBS*, Buenos Aires, Argentina, August 31 - September 4.
- [12] Vapnik, V., (2000). The Nature of Statistical Learning Theory. 2nd ed. New York: Springer-Verlag, p.314.
- [13] Haykin, S., (1999). Neural Networks: A comprehensive Foundation. Upper Saddle River, New Jersey: Prentice Hall International, Inc.S. G.
- [14] Subasi, A. and Gursoy, M.I., (2010). Expert Systems with Applications EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems With Applications*, 37(12), pp.8659–8666.
- [15] Subasi, A., (2013). Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. *Computers in Biology and Medicine*, 43(5), pp.576–586.
- [16] Kajee-bagdadi, Z., (2007). Differential Evolution Algorithms for Constrained Global Optimization. University of the Witwatersrand.
- [17] Palalic S, Can M, (2012) Denver Groups Classification of Human Chromosomes Using CANN Teams, *Southeast Journal of Soft Computing*, Vol. 1, No. 2, pp. 66-71,