International Research Journal of Mathematics, Engineering and IT
ISSN: (2349-0322)
Impact Factor- 5.489, Volume 4, Issue 12, December 2017
Website- www.aarf.asia, **Email** : editor@aarf.asia , editoraarf@gmail.com

HEART RATE SIGNAL CLASSIFICATION BY SMO ALGORİTHM

Kenan Zengin[1], Cengiz Güngör[1], Necmi Eşgi[2]
[1]Computer Engineering, Gaziosmanpaşa University, TURKEY
[2]Department of Computer Education and Educational Technology, Gaziosmanpaşa University, TURKEY

ABSTRACT

In this paper we classified heart rate signals using the WEKA data mining software developed by machine learning group at the University of Waikato. We use SMO (Sequential Minimal Optimization For Training Support Vector Machines) to classify heart rate signals. For experimental evaluation, Statlog Heart data set was selected from the University of California, Irvine (UCI) machine learning repository. The Statlog dataset contains 270 patient records, which each have 13 conditional attributes and one class attribute. We used several machine learning algorithms to classify the data and we achieved to differentiate correctly for the considered dataset using SMO algorithm.

KEYWORDS - Machine Learning, SMO, Classification

INTRODUCTION

Heart Rate Variability (HRV) analysis is based on measuring the variability of heart rate signals and more specifically, the variability in intervals between R peaks of the electrocardiogram (ECG), referred as RR intervals. Guidelines for standards of the HRV measure are summarized in [1], a summary of measures and models is presented in [2], and a review examining the physiological origins and mechanisms of heart rate can be found in [3]. In this study, we discuss the use of SMO Algorithm [4] to classify heart rate signals. Experimental results show acceptable categorization of subjects using SMO algorithm where well-known classification algorithms fail to successfully classify the input data.

METHODS

Support vectors classifiers are based on recent advances in machine learning theory. They use a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. This learning strategy is principled and very powerful method that has outperformed most other systems in a variety of applications [5].

The learning machine is given a training set of inputs belonging to two classes, with associated output values so class labels. The examples are in form of attribute vectors and the SVM finds the hyperplane separating the and being furthest from both convex hulls. If the data are not linearly separable a set of slack variables is introduced representing the amount by which the linear constrained is violated by each data point. Moreover, for many datasets, it is unlikely that a hyperplane will yield a good classifier. Instead, we want a decision boundary with more complex geometry. One way to achieve this is to map the attribute vector into some new space of higher dimensionality and look for a hyperplane in that new space, leading to kernel-based SVMs [6]. The interesting point about kernel functions is that although classification is accomplished in a space of higher dimension, any dot product product between vectors involved in the optimization process can be implicitly computed in the low dimensional space [7].

To apply support vector classifying method, we have used SMO algorithm using WEKA data mining software. For experimental evaluation, Statlog Heart data set was selected from the University of California, Irvine (UCI) machine learning repository. [8] The Statlog dataset contains 270 patient records, which each have 13 conditional attributes and one class attribute. We used several machine learning algorithms to classify the data and we achieved to differentiate correctly for the considered dataset using SMO algorithm.

This database contains 13 attributes which are ;

Attribute Information:

-- 1. age

-- 2. sex

-- 3. chest pain type (4 values)

-- 4. resting blood pressure

-- 5. serum cholestoral in mg/dl

-- 6. fasting blood sugar > 120 mg/dl

-- 7. resting electrocardiographic results (values 0,1,2)

-- 8. maximum heart rate achieved

-- 9. exercise induced angina

-- 10. oldpeak = ST depression induced by exercise relative to rest

-- 11. the slope of the peak exercise ST segment

-- 12. number of major vessels (0-3) colored by flourosopy

-- 13. thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

Variable to be predicted (class)

Absence (negative) or presence (positive) of heart disease

Figure 1 shows the graphical representation of the data. Negative values are represented in red and positive values represented as blue color. We have used several well-known classification algorithms and compared the results in Table 1. To calculate the performance of the classification algorithm we used performance measures sensitivity and specificity. This performance values is defined in Eq.(1,2)

$$\text{Sensitivity} = TP / (TP + FN) \quad (1)$$

$$\text{Specificity} = TN / (TN + FP) \quad (2)$$

Where

TP represents the True Positive count, which is calculated as the number of positive class records that the classification algorithm predicts as positive.

TN represents the True Negative count, which is calculated as the number of negative class records that the classification algorithm predicts as negative.

FP represents the false positive count, which is calculated as the number of negative class records that the classification algorithm incorrectly classifies as positive.

FN represents the false negative count, which is calculated as the number of positive class records that the classification algorithm incorrectly classifies as negative.

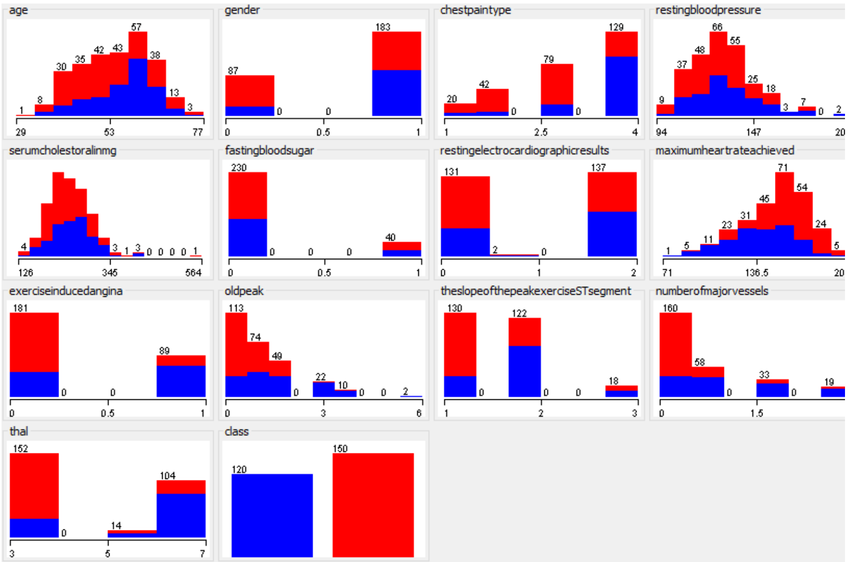

Fig. 1. Graphical representation of data.

Table 1. Comparision of classification algorithms performances on Statlog Heart data set.

Classification Algorithm	Sensitivity	Specificity	Correctly Classified Instances (%)
MLP (Multi Layer Perceptron)	0.74	0.76	75.18
K-nn	0.73	0.77	75.55
C4.5	0.75	0.83	79.26
Logistic Regression	0.79	0.86	82.96
SMO	0.8	0.87	84.07

CONCLUSIONS

In this paper we classified heart rate signals using the WEKA data mining software developed by machine learning group at the University of Waikato. We use SMO (Sequential Minimal Optimization For Training Support Vector Machines) to classify heart rate signals. For experimental evaluation, Statlog Heart data set was selected from the University of California, Irvine (UCI) machine learning repository. As a result, we achieved to differentiate correctly for the considered dataset using SMO algorithm. Compared with other classification algorithms, the SMO algorithm showed a significantly higher classification success with 84.07%. In this categorization study, it is seen that all algorithms can achieve higher specificity ratios. This means that all algorithms distinguish those who have heart disease from the healthy ones in the data set better.

REFERENCES

1. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, “Heart rate variability: Standards of measurement, physiological interpretation, and clinical use,” Eur. Heart J., vol. 17, pp. 354–381, 1996.
2. M. Teich, S. Lowen, K. Vibe-Rheymer, and C. Heneghan, “Heart rate variability: measures and models,” in Nonlinear Biomedical Signal Processing, vol. II, Dynamic Analysis and Modelling. New York: IEEE Press, 2001, pp. 159–213.
3. G. Berntson, J. Bigger, D. Eckberg, P. Grossman, P. Kaufmann, M. Malik, H. Nagaraja, S. Porges, J. Saul, P. Stone, and M. van der Molen, “Heart rate variability: Origins, methods, and interpretive caveats,” Psychophysiology, vol. 34, no. 6, pp. 623–648, Nov. 1997.
4. Platt, John. Fast Training of Support Vector Machines using Sequential Minimal Optimization, in Advances in Kernel Methods – Support Vector Learning, B. Scholkopf, C. Burges, A. Smola, eds., MIT Press (1998).
5. V. Vapnik, Statistical Learning Theory. New York: Wiley, 1998.
6. C. Cortes and V. N. Vapnik, “Support vector networks,” Mach. Learn.,vol. 20, pp. 1–25, 1995.
7. N. Christianini and J. Shawe-Taylor, Support Vector Machines and Other Kernel-based Methods. Cambridge, U.K.: Cambridge Univ. Press, 2000.
8. Uci machine learning repository is available at, .http://archive.ics.uci.edu/ml/., [Online; accessed 21.011.17] (2017).