DECISION SUPPORT SYSTEM FOR HEART DISEASE BASED ON SEQUENTIAL MINIMAL OPTIMIZATION IN SUPPORT VECTOR MACHINE

Deepti Vadicherla, Sheetal Sonawane

Department of Computer Engineering, Pune Institute of Computer and Technology,
University of Pune, Pune, India
deepti.vadicherla@gmail.com
sssonawane@pict.edu

ABSTRACT

Computer based medical decision support system (MDSS) can be useful for the physicians with its fast and accurate decision making process. Predicting the existence of heart disease accurately, results in saving life of patients followed by proper treatment. The main objective of our paper is to present a MDSS for heart disease classification based on sequential minimal optimization (SMO) technique in support vector machine (SVM). In this we illustrated the UCI machine learning repository data of Cleveland heart disease database; we trained SVM by using SMO technique. Training a SVM requires the solution of a very large QP optimization problem..SMO algorithm breaks this large optimization problem into small sub-problems. Both the training and testing phases give the accuracy on each record. The results proved that the MDSS is able to carry out heart disease diagnosis accurately in fast way and on a large dataset it shown good ability of prediction.

KEYWORDS: *Sequential Minimal Optimization, Support Vector Machine, Optimization problem, Heart disease, Medical decision support system.*

I. INTRODUCTION

At present, heart disease became a leading cause of death. It is also a major cause of disability and day by day the number of people suffering from the heart disease is rising. New data released by the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health show that especially women in older age groups are more at risk of getting heart disease. A recent study fielded by Diet Coke on behalf of *The Heart Truth*®, showed that nearly seven in 10 women mentioned heart disease as the leading cause of death among women. Heart disease can be controlled effectively if it is diagnosed at an early stage [23]. But unfortunately, accurate diagnosis of heart disease has never been an easy task. As a matter of fact, many factors can complicate the diagnosis of heart diseases, often causing the delay of a correct diagnosis decision. For instance, the clinic symptoms, the functional and the pathologic manifestations of heart diseases are associated with many human organs other than the heart and very often heart diseases may exhibit various syndromes. At the same time, different types of heart diseases may have similar symptoms. Due to this complexity, there is a need to develop medical diagnostic decision support systems which can aid medical practitioners in the diagnostic process [1],[2].
Medical decision support system is a decision-support program which is designed to assist physicians and other health professionals with decision making tasks, such as determining diagnosis of patients' data. This approach helps employees make more informed medical decisions while working with their own physician. The medical diagnosis by nature is a complex and fuzzy cognitive process, hence soft computing methods, like Support vector machine [3] have shown great potential to be applied in the

development of decision support system for heart diseases. The system uses features extracted from the ECG data of the patients.

This paper presents a medical decision support system for heart disease classification. The dataset used is the Cleveland Heart Database taken from UCI learning data set repository [22] which was donated by Detrano. In the proposed model we classify the data into two classes using SMO [4], [5] algorithm in Support Vector machine [20], [21].

The rest of the paper organized as, support vector machine described in section 2. Section 3 includes sequential minimal optimization, in which components of SMO are explained. Proposed model of MDSS and related work is mentioned and explained in section 4. Experiments and results are shown in section 5. Section 6 has conclusion followed by the future work in section 7.

II. SUPPORT VECTOR MACHINE

Support Vector Machine, is a promising method of learning machine, based on statistical learning theory developed by Vladimir Vapnik. Support vector machine (SVM) used for the classification of both linear and nonlinear data [6], [7]. It performs classification by constructing a linear optimal separating hyperplane within higher dimension, with the help of support vectors and margins, which separates the data into two categories (or classes). With an appropriate nonlinear mapping the original training data is mapped into a higher dimension. Within this the data from two classes can always be separated by a hyperplane[8].

Suppose *f* is a function for Support vector machine classification then,

$$f : I \rightarrow O$$

where, I is domain (here i.e. data set), O is a co-domain.

$$I = \{X,Y\}$$

$$X = \{ x_i \mid 1 \leq i \leq n \}$$

$$Y = \{ y_i \mid 1 \leq i \leq n \}$$

x_i is the set of 'n' training tuples with associated class labels, y_i.

Each yi can take one of the two values, either +1 or -1, which represents two classes.

$$y_i \in \{+1,-1\}$$

Set of output can be denoted as,

$$O = \{ u \mid 1 \leq i \leq n \}$$

The basic idea of SVM is to separate a set of positive examples from a set of negative examples by finding the hyperplane with maximum margin, as shown in Figure 1.

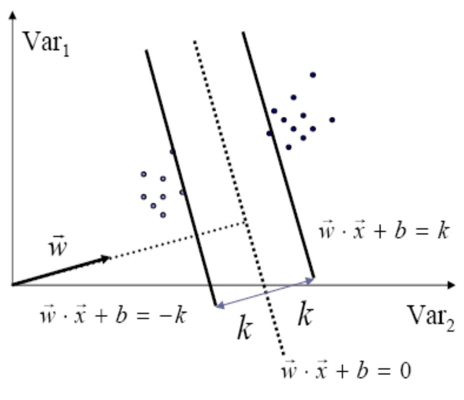

Figure 1. Optimal hyperplane with maximum margin

The support vector machine computes a linear classification of the form,

$$f(x) = wx+b$$

where *w* is a weight vector, *x* is the training example and b is bias.
The separating hyperplane can be written as ,

$$f(\mathrm{x})=0$$

Therefore we can say that, any point from one class lies above the separating hyperplane satisfies, $f(X) > 0$. In the same way any point from another class lies below the separating hyperplane satisfies, $f(X) < 0$.
Above equations were processed to make the linearly separable set D to meet the following inequality,

$$y_i\,(\, f(\mathrm{x})\,) \geq 1, \;\; \forall i$$

Here the margin *m* is , $m = \dfrac{1}{\| \mathrm{w} \|_2}$

Using above equation, maximizing margin can be written in the form of optimization problem as below:

$$\min_{w,b} \tfrac{1}{2}\|w\|^2 \quad \text{Subject to } \; y_i(w.x+b) \geq 1, \forall i$$

This optimization problem can be solved by using dual Lagrange multiplier,

$$\min_{\vec{\alpha}} \Psi(\vec{\alpha}) = \min_{\vec{\alpha}} \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j (\vec{x}_i \cdot \vec{x}_j)\alpha_i \alpha_j - \sum_{i=1}^{N}\alpha_i ,$$

For linearly separable data the support vectors are a subset of actual training tuples. Lagrangian formulation of above optimization problem contains a dot product between the support vector x_i and test tuple x_j. There is one-to-one relationship between each Lagrange multiplier [9] and each training tuple.
Not all data sets are linearly separable. There may be no hyperplane that splits the positive examples from the negative examples. SVMs can be even further generalized to non-linear classifiers. The output of a non-linear SVM is explicitly computed from the Lagrange multipliers,

$$u = \sum_{j=1}^{N} y_j \alpha_j K(\vec{x}_j, \vec{x}) - b,$$

where *K* is a kernel function. We used Radial Basis Kernel Function (RBF) [10] here, which is denoted as follow:

$$K(\mathrm{x_i}, x_j) = \exp(-\gamma \| \mathrm{x_i} - x_j \|^2),\ \gamma>0$$

The non-linearity alter the quadratic form, but the dual objective function Ψ is still quadratic in α,

$$\min_{\vec{\alpha}} \Psi(\vec{\alpha}) = \min_{\vec{\alpha}} \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j K(\vec{x}_i, \vec{x}_j)\alpha_i \alpha_j - \sum_{i=1}^{N}\alpha_i ,$$
$$0 \leq \alpha_i \leq C, \forall i,$$
$$\sum_{i=1}^{N} y_i \alpha_i = 0.$$

Sequential minimal optimization algorithm solves above quadratic programming problem.

III. SEQUENTIAL MINIMAL OPTIMIZATION

3.1. SVM training algorithms

The popularity of SVMs has led to the development of a large number of special purpose solvers for the SVM optimization problem.

Previous algorithms for training support vector machine are as below,

3.1.1. Chunking method

The chunking algorithm reduces the size of the matrix. However chunking still cannot handle large-scale training problems, since even this reduced matrix cannot fit into memory [11],[12].

3.1.2. Osuna

Osuna algorithm works by choosing a small subset from the data set and solving the related sub problem defined by the variables in the subset [13]. At each iteration, there is a strategy to replace some of the variables in the working set with other variables not in the working set and converges to the global optimal solution [14].

3.2. Need for SMO

Sequential minimal optimization (SMO) [5] is an algorithm for efficiently solving the optimization problem which arises during the training of support vector machine. At every step, SMO chooses two Lagrange multipliers to jointly optimize, finds the optimal values for these multipliers and updates the SVM to reflect the new optimal values [15].

SMO solves SVM optimization problem analytically instead of numerical QP optimization where only two parameters optimized, keeping rest fixed. Since SMO doesn't require extra matrix storage, large SVM training problems can be stored in the computer. SMO can be speed up by minimizing computation time. SMO gives good performance when many Lagrange multipliers are at bound. Due to all these features SMO became an efficient method for training SVM [16].

3.3. SMO component

We can divide the SMO in the components as below:

SMO repeatedly finds two Lagrange multipliers that can be optimized with respect to each other and analytically computes the optimal step for the two Lagrange multipliers [17].

3.3.1. Updating two Lagrange Multipliers

In the process of updating two Lagrange multipliers, SMO first computes the constraints on these multipliers and then solves for the constrained minimum [18]. SMO optimizes two Lagrange multipliers to fulfill the linear equality constraint at every step, which is not possible using one Lagrange multiplier.

$$\alpha_2^{\text{new}} = \alpha_2 + \frac{y_2(E_1 - E_2)}{\eta},$$

where, $\eta = K(\vec{x}_1, \vec{x}_1) + K(\vec{x}_2, \vec{x}_2) - 2K(\vec{x}_1, \vec{x}_2)$.

Clipping this value,

$$\alpha_2^{\text{new,clipped}} = \begin{cases} H & \text{if} \quad \alpha_2^{\text{new}} \geq H; \\ \alpha_2^{\text{new}} & \text{if} \quad L < \alpha_2^{\text{new}} < H; \\ \text{L} & \text{if} \quad \alpha_2^{\text{new}} \leq L. \end{cases}$$

where, $L = \max(0, \alpha_2 + \alpha_1 - C)$, $H = \min(C, \alpha_2 + \alpha_1)$. when y1=y2.

$L = \max(0, \alpha_2 - \alpha_1)$, $H = \min(C, C + \alpha_2 - \alpha_1)$. when y1≠y2.

The new value of α_1 is computed from the new, clipped α_2:

$$\alpha_1^{\text{new}} = \alpha_1 + s(\alpha_2 - \alpha_2^{\text{new,clipped}}).$$

where, $s = y_1 y_2$.

3.3.2. Choosing Lagrange Multipliers to optimize

In SMO, the process of choosing two Lagrange multipliers play an important role in speeding up convergence. Among two, one Lagrange multipliers is chosen first which act as the outer loop of the SMO algorithm. This loop iterates over the entire training set, selecting an example which violets the

KKT condition [19]. The example which violets KKT condition, is optimized with the first chosen Lagrange multiplier.
After one pass through the entire training set, the outer loop makes repeated passes over the non-bound examples until all of the non-bound examples obey the KKT conditions, as shown in Figure 2. The outer loop keeps alternating between single passes over the entire training set and multiple passes over the non-bound subset until the entire training set obeys the KKT conditions, whereupon the algorithm terminates.

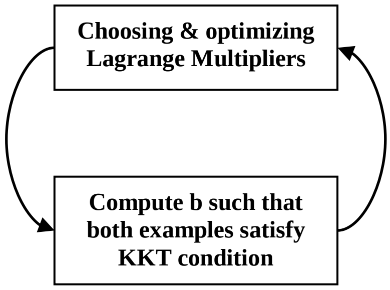

Figure 2. Components of SMO

IV. PROPOSED MODEL OF MEDICAL DECISION SUPPORT SYSTEM

4.1. Related work

Medical decision support system work has been carried on the basis of performance of different methods like SVM, Artificial neural network, Bayesian classification method, etc. [1], [2].
Neural network algorithms are inherently parallel, which result in speeding up the computation process. They have high tolerance of noisy data. The major disadvantage of neural networks is that, they have poor interpretability. Fully connected networks are difficult to articulate. Whereas various empirical studies of Bayesian classifier in comparison with decision tree and neural network classifiers have found out that, in theoretical way Bayesian classifiers have minimum error rate in comparison to all other classifiers. However, in practice this is not always the case, owing to inaccuracies in the assumptions made for its use, such as class conditional independence and the lack of available probability data.

4.2. Pre-processed data

The experiments are carried out on heart dataset using Sequential Minimal Optimization in Support Vector Machine.
Heart disease is diagnosed with the help of some complex pathological data. The heart disease dataset used in this experiment is the Cleveland Heart Disease database taken from UCI machine learning dataset repository [22]. This database contains 14 attributes as below:
1. Age of patient, 2. Sex of patient, 3. Chest pain type, 4. Resting blood pressure, 5. Serum cholesterol, 6. Fasting blood sugar, 7. Resting ECG results, 8. Maximum heart rate achieved, 9. Exercise induced angina, 10. ST depression induced by exercise relative to rest, 11. Slope of the peak exercise ST segment, 12. number of major vessels colored by flourosopy, 13. thal, 14. Diagnosis of heart disease.

4.3. Flow diagram of MDSS

The purpose of this proposed model is to diagnose the heart disease by classifying the dataset of heart disease. This classification process is shown in Figure 3.

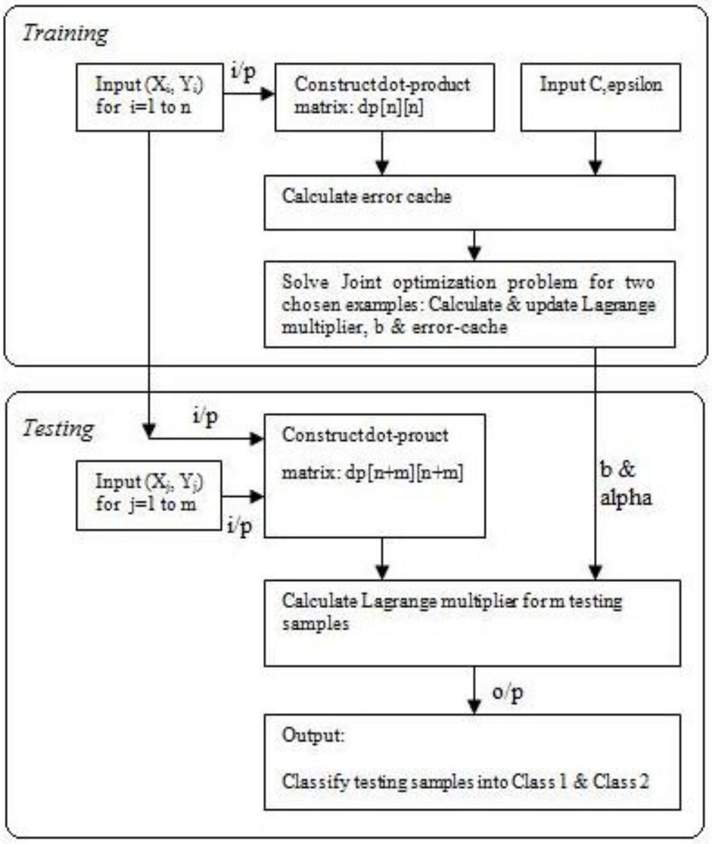

Figure 3. Flow diagram of MDSS for heart disease

V. RESULT AND DISCUSSION

In the experiment, we used dataset having 297 total number of patient records. Large part of records in the dataset is used for training and rest of them are used for testing. The main difference between the dataset given as input to training and testing is that, the input we are giving to training is the data with correct diagnosis (14th field in the dataset) and whereas the input data of testing doesn't have the correct diagnosis purposely. The Diagnosis (14th) field refers to the presence or absence of heart disease of that respective patient. It is integer valued field, having value 1(absence of disease) or -1(presence of disease). So that at the end of testing process we can check the result in the output file created after testing and verify the efficiency of the proposed model in terms of accuracy.
The experimental results are shown in the Figure 4.

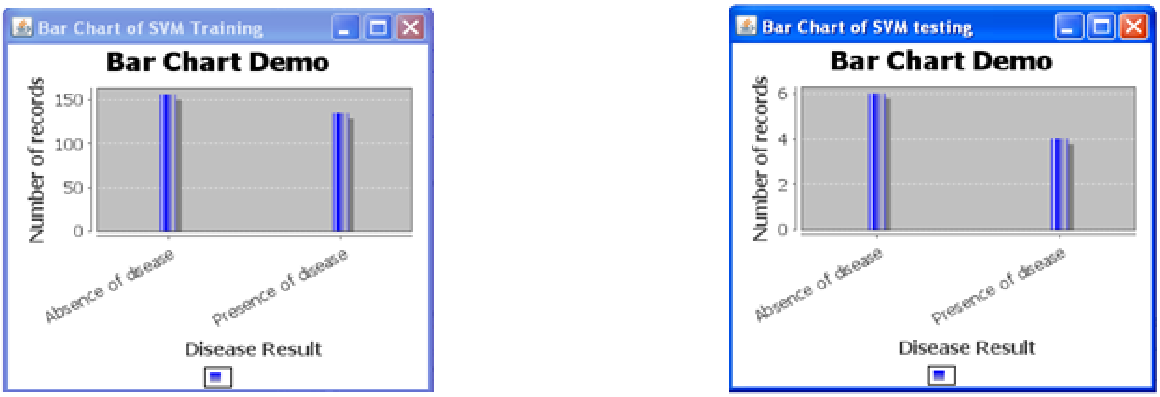

Figure 4. Results of Training and Testing

In the Sequential Minimal Optimization procedure KKT conditions are checked to be within ε of fulfillment and we set ε to 0.001 in the experiment. So it is acceptable for examples on the positive margin to have output between 0.999 to 1.001. The SMO algorithm will not converge as quickly if it is required to produce very high output.

VI. CONCLUSION

This paper proposes a model of decision support system for heart disease based on Sequential Minimal Optimization in Support Vector Machine. For training of SVM we used SMO algorithm which incorporated its features like high accuracy and high speed in the proposed model. Because of its ease of use and better scaling with the training set size, SMO is a strong candidate for becoming the standard SVM training algorithm.

VII. FUTURE WORK

In future, research issues on improving SMO using decomposition techniques can be explored on techniques like shrinking and kernel caching. Further improvements to working set selection can be done. Also the same model can be extended, for the multi level classification purpose.

REFERENCES

[1] Long Wan, Wenxing Bao, “Research and Application of Animal Disease Intelligent Diagnosis Based on Support Vector Machine” IEEE International Conference on Computational Intelligence and Security, Pages 66-70, 2009.

[2] S.N. Deepa, B. Aruna Devi, “Neural Networks and SMO based Classification for Brain Tumor”, IEEE World Congress on Information and Communication Technologies, Pages 1032-1037,2010.

[3] C. Cortes and V. Vapnik, “Support-vector network,” Machine Learning, vol.20, 1995.

[4] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy “Improvements to platt’s smo algorithm for svm classifier design,” Neural Computation, vol.13,pp:637-649, 2002.

[5] John C. Platt, “Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines”, Microsoft Research, Technical Report MSR-TR-98-14.

[6] Gong Wei Wang Shoubin, “Support Vector Machine for Assistant Clinical Diagnosis of Cardiac Disease”, IEEE Global Congress on Intelligent Systems, pp: 588-591, 2009.

[7] Ya Gao; Shiliang Sun, “An empirical evaluation of linear and nonlinear kernels for text classification using Support Vector Machines”, IEEE Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Pages: 1502-1505, 2010.

[8] Qinghua Jiang, Guohua Wang, Tianjiao Zhang, Yadong Wang, “Predicting Human microRNA-disease Associations Based on Support Vector Machine”, IEEE International Conference on Bioinformatics and Biomedicine, pp:467-472 , 2010.

[9] Browne, K.E.; Burkholder, R.J., “Nonlinear Optimization of Radar Images From a Through-Wall Sensing System via the Lagrange Multiplier Method”, IEEE Geoscience and Remote Sensing Letters, Pages: 803-807, 2012.

[10] Gao Daqi; Zhang Tao, “Support vector machine classifiers using RBF kernels with clustering-based centers and widths” IEEE Intrnational Joint Conference on Neural Networks, 2007 , Pages: 2971-2976.

[11] Olvi L.Mangasarian and Michael E.Thompson, “Massive Data Classification via Unconstrined Support Vector Machines” Journal of Optimization Theory and Applications, 131, Pages315–325, 2006.

[12] P. S. Bradley and O. L. Mangasarian. “Massive data discrimination via linear support vector machines. Optimization Methods and Software, 13:1-10, 2000.ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-05.ps

[13] Osuna, E., Freund, R., Girosi, F., “Improved Training Algorithm for Support Vector Machines”, Proc. IEEE NNSP ’97, 1997.

[14] Yiqiang Zhan, Dinggang Shen, “Design efficient support vector machine for fast classification”, The journal of pattern recognition society 38, Pages 157-161,2004.

[15] Peng Peng, Qian-Li Ma, Lei-Ming Hong, “The Research of the Parallel SMO algorithm for solving SVM”, Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, pp:1271-1274, 12-15 July 2009.

[16] Chin-Jen Lin, "Asymptotic convergence of an SMO algorithm without any assumptions", IEEE Transactions on Neural Networks, vol.13,issue 1, pp:248-250, 2002.
[17] Baxter Tyson Smith, "Lagrange Multipliers Tutorial in the Context of Support Vector Machines", Memorial University of Newfoundland St. John's, Newfoundland, Canada.
[18] Ya-Zhou Liu; Hong-Xun Yao; Wen Gao; De-bin Zhao, "Single sequential minimal optimization: an improved SVMs training algorithm", Proceedings of 2005 International Conference on Machine Learning and Cybernetics, 2005. Pages 4360- 4364.
[19] Kjeldsen, Tinne Hoff. "A contextualized historical analysis of the Kuhn-Tucker theorem in nonlinear programming: the impact of World War II". Historia Math. 27 (2000), no. 4, 331–361.
[20] Ian H. Witten, Eibe Frank, Data Mining, Elsevier Inc. 2005.
[21] Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, 2/e, Morgan Kaufmann Publishers, Elsevier Inc. 2006.
[22] UCI Machine Learning Repository: Heart Disease Data Set. http://archive.ics.uci.edu/ml/datasets/Heart+Disease
[23] Mediline Plus: Information related to Heart Diseases. http://www.nlm.gov/medlineplus/heartdiseases.html

AUTHORS

Deepti Vadicherla holds a Bachelor's degree in Computer Science Engineering from SRTMU, Nanded , Maharashtra, India. She is pursuing her Master's in Computer engineering from University of Pune. Her area of interests are Data mining and Computer algorithms.

Sheetal Sonawane holds a Master's degree in Computer Engineering from University of Pune, Maharashtra, India. She is pursuing her Ph.D. in Computer engineering. She is Associate Professor at Computer Engineering department of PICT, University of Pune. Her area of interests are data mining and biological networks.