Support Vector Machine and Neural Network Classification Improved By Bagging

Kolluru Venkata Nagendra[1], Dr.MaligelaUssenaiah[2]

[1]*Research Scholar, Vikrama Simhapuri University,Nellore.*
[2]*Assistant Professor, Department of Computer Science,VSU,Nellore.*
[1]kvnscholar@gmail.com

Abstract— Classification is supervised learning approach in Data mining used to predict group membership for data instances. A hybrid classification method encompasses the advantages of the individual classification approaches that it is built upon. In this paper we will be examining few popular algorithms used for classifying medical diagnosis data with a hybrid of support vector machines and neural networks. After that we will discuss the performance of these algorithms depending on different parameters and comparing their correct rate in different categories.

Keywords*— Classification, Data mining, Classifier, Support Vector Machines, Neural Networks, Bagging.*

*****

I. Introduction

Extracting important information from vast amounts of data is a very vital part of strategic decision making in any organization. This process is called data mining. It has three important approaches. The first one known as supervised approach helps in predicting continuous or discreet value. In this case output labels are known. Another approach called unsupervised learning groups data based on similarities. In this the output labels are not known. The third approach is called reinforcement learning which rewards right decisions and punishes wrong decisions [1].

In supervised approach there are basically two sub approaches. The first one known as regression helps in predicting continuous data and the second one known as classification help in predicting discreet data. Classification can be done using various algorithms. The prominent ones among them being K nearest neighbor, neural networks, decision tree, random forest, support vector machines etc.[2]

In this research paper we shall be discussing the implementation of a hybrid approach wherein we shall be combining the best features of neural network, support vector machines and apply the same to a bagged samples of data. The data we shall be using while building the hybrid model is from Framingham data. The objective of the use case is to predict based on some features whether or not a person is prone to a heart attack or not. We shall compare the results of this hybrid model with SVM and neural network approaches individually and try to establish if it is better than existing models or not. In the process we shall be using various statistical parameters like accuracy, sensitivity and specificity etc. to arrive at the said comparisons [3].

Our next section presents the proposed methodology. Section III describes Background Knowledge about various classifiers discussed and in section IV discussion of results is carried out. Section V summarizes the work and in section VI data source is acknowledged and in the final section references are given.

II. Proposed Methodology

The approach used in this paper has many steps as illustrated below:

A. Data Selection

The first obvious step in any data mining procedure is data selection. The data being used in this paper is from Framingham heart study. This study began in Framingham with 5209 subjects. Much of the now commonly known heart attack symptoms have come from this study [4]. It has following attributes:

- i) Gender
- ii) Age
- iii) Education
- iv) Whether smoker or not
- v) BP stats of a person
- vi) Whether a person already suffered a heart stroke
- vii) Whether a person was diabetic or not
- viii) Cholesterol levels of a person
- ix) Body Mass index of a person
- x) Heart beat rate of a person
- xi) Glucose levels of a person

ISSN: 2454-4248
125 – 130

The model tries to predict based on the above mentioned attributes whether a person is prone to suffer a heart attack or not.

B. Processing of data

The data thus collected shall have many erroneous entries like missing values, duplicate values etc. The same is first cleaned for removing data anomalies.

C. Feature Extraction and Selection

Feature selection happens to be an important step in a supervised learning process that shall be used in this paper. [5]

D. Supervised classification approach selection

Different supervised classification approaches have diverse merits and demerits [8]. The selection of an appropriate classification approach is an art rather than a science [6-7].

In this paper we shall be using hybrid model comprising SVM and neural network classification method supplemented by bagging to classify the Framingham datasets and shall be comparing its performance with individual SVM and neural network classification methods. The algorithm for the same is mentioned below:

Algorithm Hybrid SVM

Step 1: Train the SVM classifier for the data set under consideration

Step 2: Construct a new data set with the decision function of the SVM classification approach.

Step 3: Train the ANN with the new data set and validate the same using the test set

Step 4: Repeat the above steps with different samples with replacement approach.

Background Knowledge the mentioned classification methods are given in Section III.

E Training and Testing

The obtained model from the previous method is validated against the test set[8].

III. BACKGROUND KNOWLEDGE

A).OVERVIEW OF NEURAL NETWORKS

The neural network method has a layered architecture of neurons such that one layers output is fed as another following layers input [9-12].

The learning methodology of MLP classifier uses a parameter called Euclidean norm so as to diminish the error function on the dataset (xi,di) for every i= 1. . . N, as in (1)

$$E(w) = \frac{1}{2}\sum_{x=1}^{K} ||q(px, c) - Dx||^2 \qquad (1)$$

where pi=input vector with K dimensions for x=1...N

q=M-dimensional output vector

c=the vector with tailored weights

Dx=the M- dimensional preferred output vector.

The network builds predictive model using a sample data set. The MLP network demonstrated below is constructed with 3 layers: an input layer, hidden layers in between and an output layer as displayed in the below figure. Every layer comprises of neurons which through two parameters viz the bias and weight values are connected with other layers. The network learns the association amid pairs of output(responses) and inputs (factors) vectors by changing the weight and bias values.

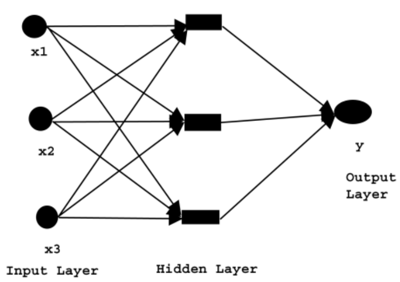


Fig 2.4.6.3.1: MLP network.

Every input layer maps to the weighted sum of the previous layer as in (2).

$$Ij=\sum Wj\text{-}1.\,Oj\text{-}1 \qquad (2)$$

Here Ij is input of current layer and Oj-1 and Wj-1 are output and weight of previous layer respectively.

The algorithm for the Multi-Layer Perceptron classification method is mentioned below:

Algorithm[3] MLP Classifier Classification Methodology

Step 1: Initialize weights

Step 2: Do

Step 3: For every tuple x part of the training dataset

Step 4: Assign O= neural-net-output and later forward the acquired outcome to the subsequent pass

Step 5: Assign T= outcome for any given tuple x

Step 6: Compute error<- (T - O) at each unit

Step 7: Compute parameter delta_wi for all weights wi initially from core hidden layers and then the output layer; Subsequently Compute delta_wi for all wi from a picky input layer to hidden layers;

Step 8: Continue

Step 9: Revise weights for all neurons in the MLP network

Step 10: Until the process is complete or exit condition gets fulfilled

Step 11: Output mlp network

MLP methodology can be executed in two modes: Sequential mode: It is also called per-pattern, stochastic or on-line mode. Herein weights revised after every pattern is used

Batch mode: It is also called per -epoch or off-line mode. In this mode the weight changes for each pattern are considered in the training set. The total change is later computed by summing the individual changes

B) Overview of the SVM

SVM is a supervised technique wherein data is mapped onto a higher order space using a kernel function: K(x,xj)=⟨ϕ(x),ϕ(xj)⟩. Decision function can be stated as[13-16] (3):

$$f(x) = \sum_{j=1}^{Sv} \alpha j y j K(x, xj) + b \qquad (3)$$

αjisLagrange multipliers, K is kernel function and b is the bias which is calculated using a support vector[17].Then, the optimum hyper-plane relates to f (x)=0. Henceforth, test data can be dented as (4):

$$x\epsilon\begin{cases} positive\ category\ if\ f(x) > 0 \\ negative\ category\ if\ f(x) < 0 \end{cases} \qquad (4)$$

The specified training set of case label pairs as delivered by equation [18] (5).

$$(x, y) = \{ (x_1, y_1), (x_2, y_2), \ldots\ldots\ldots (x_n, y_n)\} \qquad (5)$$

Where $x_n\, \epsilon R^y$ and $y_n \epsilon$ {-1, +1} and SVM desires resolution of the optimization problem$\frac{1}{2} W^T W + \sum_{i=1}^{l} \xi_i$C

Subject to as given in equation (6)

$$y_i\ (w^T \Phi\ (x_i) + b) \geq 1\text{-}\xi_i \xi_i \geq 0 \qquad (6)$$

Algorithm [2] Classification using SVM Classifier

Input: Input data matrix, label information
Output: Set of Support vectors (SV)
Initialization: Error threshold = huge value

begin
Randomly sample 2 instances belonging to different classes.
Add them to the current set of Support Vectors.
Loop to randomly sample n instances
Choose that set of n instances with which current SVs give n pt test error less than the current error threshold
Break if insufficient instances
end loop random sampling of n instances
Loop over misclassified instances
Add the point to current SVs
end Loop
Train using SVM and test over the remaining instances
end

In this paper 740 support vectors classify the data into two sets. The training error in the said classification process was 0.123109.

C) Overview of Bagging

This technique processes samples in parallel is called bagging also known as bootstrap aggregating. It plans to advance the accuracy by constructing enhanced fused classifier, by aggregating the variety of learned classifier outputs into a solitary prediction. The pseudo-code is mentioned below [19-21].

Algorithm Bagging [19]
I:- an inducer
N:-the number of iterations
S:-the training set
Output: Ci; i = 1,……,N
1: i ← 1
2: repeat
3: Si ←Sample N instances from S with replacement.
4: Construct classifier Ci using I on Si
5: ti+ +
6: until i > N

As sampling with replacement is used, a few original instances of S might come into view frequently in Si and few might not be incorporated at all. So Si are dissimilar from each other, however are surely not autonomous. To classify a novel instance, every classifier returns the class forecast for the unidentified instance. The compound bagged classifier, I1, returns the class which was forecasted frequently. The consequence is that this method generates a joint model which often fares superior to the solitary model

ISSN: 2454-4248
125 – 130

constructed from the original data. It is true particularly for unsteady inducers as bagging can get rid of their unsteadiness.

IV. Results and Discussion

The algorithms thus discussed is applied for classifying the medical diagnosis datasets into two groups that is people prone to heart attack and people not prone to heart attack while in the end performance of classification is measured by means of diverse metrics. R software is used to perform the mentioned operations [22].

A. Data Acquisition

Around4200 records of heart patient's data from Framingham are obtained.

B. Features Extraction

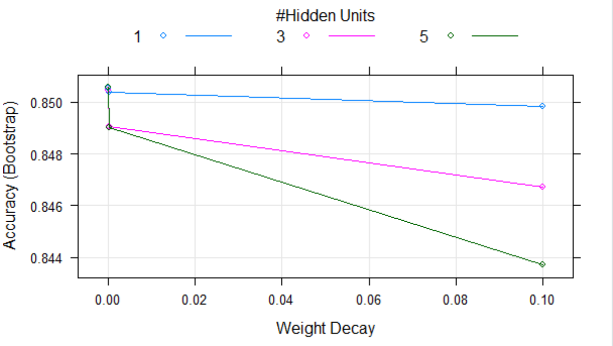
Fifteen features extracted from the data set are fed to the SVM and these are compared with other classifiers.

C .Image Grouping

Classification output can be characterized by categorizing the same with a confusion matrix as demonstrated in Table-I [23].

TABLE I

Real group	Classification consequence	
	More	Less
More	TN	FP
Less	FN	TP



(a)

Figure 1a): Plot of Neural Networks for the heart patient data

D .Evaluation metrics

Evaluation of the classification models can be carried out using the metrics defined below [24-27]. The formulae for sensitivity, specificity, prevalence, and detection rate and detection prevalence are given in (7), (8), (9), (10) and (11)

$$\text{Sensitivity} = \frac{TN}{(TN+FN)} \times 100 \qquad (7)$$

$$\text{Specificity} = \frac{TP}{(TP+FP)} \times 100 \qquad (8)$$

$$\text{Prevalence} = \frac{TN+FN}{(TP+FN+FP+TN)} \times 100 \quad (9)$$

$$\text{Detection rate}=\frac{TN}{(TP+FN+FP+TN)} \text{x}100 \quad (10)$$

$$\text{Detection Prevalence}=\frac{TN+FP}{(TP+FN+FP+TN)} \text{ X } 100 \quad (11)$$

C) Result Scrutiny

The performance of the proposed classifier hybrid SVM with neural networks is examined and likened with SVM and neural network and the results are mentioned below.

TABLE II

Performance Measures

Classifier	Accuracy	Specificity	Prevalence	Detection Rate	Detection Prevalence
SVM	86.03	100	98.7	84.7	84.7
Neural Networks	84.7	NA	100	84.7	84.7
Hybrid SVM	94	100	98.9	90.2	90.2

Hence, it is identified from the results and findings with the comparative evaluations that hybrid SVM approach is better than other classifiers.

V. CONCLUSION

The research is related to the classification of heart patient data from Framingham using hybrid SVM and neural networks. The same is compared with SVM and neural network methods. The research has concluded that the hybrid SVM approach of classification is more effective and efficient.

REFERENCES

[1] Nadine Kashmar, MirnaAtieh, Ali Haidar, Identifying the Effective Parameters for Vertical Handover in Cellular Networks Using Data Mining Techniques, Procedia Computer Science, Volume 98, 2016, Pages 91-99, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2016.09.016.

[2] Ling Chen, Xue Li, Yi Yang, Hanna Kurniawati, Quan Z. Sheng, Hsiao-Yun Hu, Nicole Huang, Personal health indexing based on medical examinations: A data mining approach, Decision Support Systems, Volume 81, January

2016, Pages 54-65, ISSN 0167-9236, https://doi.org/10.1016/j.dss.2015.10.008.

[3] Eva Armengol, DionísBoixader, Francisco Grimaldo, Special Issue on Pattern Recognition Techniques in Data Mining, Pattern Recognition Letters, Volume 93, 1 July 2017, Pages 1-2, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2017.02.014.

[4] Gangin Lee, Unil Yun, HeungmoRyang, An uncertainty-based approach: Frequent itemset mining from uncertain data with different item importance, Knowledge-Based Systems, Volume 90, December 2015, Pages 239-256, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2015.08.018.

[5] TarekHamrouni, SarraSlimani, Faouzi Ben Charrada, A data mining correlated patterns-based periodic decentralized replication strategy for data grids, Journal of Systems and Software, Volume 110, December 2015, Pages 10-27, ISSN 0164-1212, https://doi.org/10.1016/j.jss.2015.08.019.

[6] Jia Wu, Shirui Pan, Xingquan Zhu, ZhihuaCai, Peng Zhang, Chengqi Zhang, Self-adaptive attribute weighting for Naive Bayes classification, Expert Systems with Applications, Volume 42, Issue 3, 15 February 2015, Pages 1487-1502, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2014.09.019.

[7] Liangxiao Jiang, Chaoqun Li, Shasha Wang, Lungan Zhang, Deep feature weighting for naive Bayes and its application to text classification, Engineering Applications of Artificial Intelligence, Volume 52, June 2016, Pages 26-39, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2016.02.002.

[8] P. Julian Benadit, F. Sagayaraj Francis, U. Muruganantham, Improving the Performance of a Proxy Cache Using Tree Augmented Naive Bayes Classifier, Procedia Computer Science, Volume 46, 2015, Pages 184-193, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.02.010.

[9] Md. Shafiqul Islam, M.A. Hannan, Hassan Basri, AiniHussain, Maher Arebey, Solid waste bin detection and classification using Dynamic Time Warping and MLP classifier, Waste Management, Volume 34, Issue 2, February 2014, Pages 281-290, ISSN 0956-053X, http://dx.doi.org/10.1016/j.wasman.2013.10.030

[10] Hari Mohan Rai, AnuragTrivedi, ShailjaShukla, ECG signal processing for abnormalities detection using multi-resolution wavelet transform and Artificial Neural Network classifier, Measurement, Volume 46, Issue 9, November 2013, Pages 3238-3246, ISSN 0263-2241, http://dx.doi.org/10.1016/j.measurement.2013.05.021..

[11] LokanathSarangi, Mihir Narayan Mohanty, SrikantaPattanayak, Design of MLP Based Model for Analysis of Patient Suffering from Influenza, Procedia Computer Science, Volume 92, 2016, Pages 396-403, ISSN 1877-0509, http: //dx.doi.org/ 10.1016/ j.procs. 2016.07. 396.

[12] LokanathSarangi, Mihir Narayan Mohanty, SrikantaPattanayak, Design of MLP Based Model for Analysis of Patient Suffering from Influenza, Procedia Computer Science, Volume 92, 2016, Pages 396-403, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2016.07.396.

[13] JakubGajewski, David Vališ, The determination of combustion engine condition and reliability using oil analysis by MLP and RBF neural networks, Tribology International, Available online 23 June 2017, ISSN 0301-679X, https://doi.org/10.1016/j.triboint.2017.06.032.

[14] Ahmed F. Mashaly, A.A. Alazba, MLP and MLR models for instantaneous thermal efficiency prediction of solar still under hyper-arid environment, Computers and Electronics in Agriculture, Volume 122, March 2016, Pages 146-155, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2016.01.030.

[15] Yu Zhang, Shixing Wang, MLP technique based reinforcement learning control of discrete pure-feedback systems, Neurocomputing, Volume 168, 30 November 2015, Pages 401-407, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2015.05.087.

[16] TayyabWaqar, Mustafa Demetgul, Thermal analysis MLP neural network based fault diagnosis on worm gears, Measurement, Volume 86, May 2016, Pages 56-66, ISSN 0263-2241, https://doi.org/10.1016/j.measurement.2016.02.024.

[17] P.J. García Nieto, E. García-Gonzalo, J. Bové, G. Arbat, M. Duran-Ros, J. Puig-Bargués, Modeling pressure drop produced by different filtering media in microirrigation sand filters using the hybrid ABC-MARS-based approach, MLP neural network and M5 model tree, Computers and Electronics in Agriculture, Volume 139, 15 June 2017, Pages 65-74, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2017.05.008.

[18] Dewan Md. Farid, Li Zhang, ChowdhuryMofizurRahman, M.A. Hossain, Rebecca Strachan, Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks, Expert Systems with Applications, Volume 41, Issue 4, Part 2, March 2014, Pages 1937-1946, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2013.08.089.

[19] Hamid Parvin, MiresmaeilMirnabiBaboli, Hamid Alinejad-Rokny, Proposing a classifier ensemble framework based on classifier selection and decision tree, Engineering Applications of Artificial Intelligence, Volume 37, January 2015, Pages 34-42, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2014.08.005.

[20] Om PrakashMahela, Abdul GafoorShaik, Recognition of power quality disturbances using S-transform based ruled decision tree and fuzzy C-means clustering classifiers, Applied Soft Computing, Volume 59, October 2017, Pages 243-257, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2017.05.061.

[21] https://en.wikipedia.org/wiki/Ensemble_learning

[22] Shre Kumar Chatterjee, Saptarshi Das, KoushikMaharatna, Elisa Masi, Luisa Santopolo, IlariaColzi, Stefano Mancuso, Andrea Vitaletti, Comparison of decision tree based classification strategies to detect external chemical stimuli from raw and filtered plant electrical response, Sensors and Actuators B: Chemical, Volume 249, October 2017, Pages 278-295, ISSN 0925-4005, https://doi.org/10.1016/j.snb.2017.04.071.

ISSN: 2454-4248
125 – 130

[23] ArisPagoropoulos, Anders H. Møller, Tim C. McAloone, Applying Multi-Class Support Vector Machines for performance assessment of shipping operations: The case of tanker vessels, Ocean Engineering, Volume 140, 1 August 2017, Pages 1-6, ISSN 0029-8018, https://doi.org/10.1016/j.oceaneng.2017.05.001.

[24] Abdulla Amin Aburomman, Mamun Bin IbneReaz, A novel weighted support vector machines multiclass classifier based on differential evolution for intrusion detection systems, Information Sciences, Volume 414, November 2017, Pages 225-246, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2017.06.007.

[25] Michael E. Cholette, PietroBorghesani, Egidio Di Gialleonardo, Francesco Braghin, Using support vector machines for the computationally efficient identification of acceptable design parameters in computer-aided engineering applications, Expert Systems with Applications, Volume 81, 15 September 2017, Pages 39-52, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2017.03.050.

[26] Madson L. Dantas Dias, Ajalmar R. Rocha Neto, Training soft margin support vector machines by simulated annealing: A dual approach, Expert Systems with Applications, Volume 87, 30 November 2017, Pages 157-169, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2017.06.016.

[27] Sebastián Maldonado, Juan Pérez, Cristián Bravo, Cost-based feature selection for Support Vector Machines: An application in credit scoring, European Journal of Operational Research, Volume 261, Issue 2, 1 September 2017, Pages 656-665, ISSN 0377-2217, https://doi.org/10.1016/j.ejor.2017.02.037.

BIOGRAPHIES

K.Venkata Nagendra, working as Assistant Professor in the Department of Computer Science Engineering at Geethanjali Institute of Science and Technology, Nellore, Andhrapradesh, India. He has 6 years of experience in the field of teaching. He is a research scholar in Vikrama Simhapuri University. He did his M.Tech in ANU, Guntur. His areas of interests are Data warehousing and Data Mining and Cloud Computing.

Dr.Maligela Ussnaiah, working as Assistant Professor in the Department of Computer Science in Vikrama Simhapuri University, Nellore, Andhra Pradesh, India. He is having 8 years of teaching experience. He did his PhD in computer science from SriKrishna Devaraya Univeristy, Ananthapur, Andhra Pradesh. His areas of interests are Networks, Mobile Wireless Networks, Data warehousing and Data Mining and Image processing.