Least Squares Support Vector Regression for Spirometric Forced Expiratory Volume (FEV1) Values

Article · January 2013

CITATIONS
2

READS
123

1 author:

Perumal Venkatesan
Sri Ramachandra University
222 PUBLICATIONS **3,462** CITATIONS

SEE PROFILE

Volume 3 No.1, January 2013 ISSN 2224-3577

International Journal of Science and Technology

http://www.ejournalofsciences.org

Least Squares Support Vector Regression for Spirometric Forced Expiratory Volume (FEV_1) Values

G. Meena Devi[1], P.Venkatesan[2]

[1] St.Joseph's College of Engineering, Department of Mathematics, Chennai, India

[2] National Institute for Research in Tuberculosis, ICMR, Chetpet, Chennai, India

ABSTRACT

Spirometry test is an inevitable, essential screening test in the case of respiratory and lung related diseases. This work focuses on predicting FEV_1, which is the most significant and one of the deciding value in making the conclusion on respiratory related disorders by Least Squares Support Vector Machine (LS SVM) regression. This prediction of FEV_1 values will enhance the spirometric method, when the data is incomplete or poorly recorded, accuracy of diagnosis of the abnormalities can be improved using SVM based methods. In this paper, an attempt is made to predict FEV_1 values by LS SVM. The results show that the predication accuracy is very high.

Keywords: *Regression, FEV_1, LS SVM, Spirometry, FVC*

1. INTRODUCTION

Spirometry test is valuable to assess the general respiratory diseases. It is a non invasive test used as a screening test of general respiratory health in the same way as the blood pressure for general cardiovascular diseases. It is an essential tool in the diagnosis of airway obstruction. In this test the volume of air inhaled or exhaled by the patient is measured as a function of time. [1-4]. The more common lung function values measured with spirometry are FVC (Forced Vital Capacity), FEV_1(Forced expiratory volume in the first second of the forceful exhalation), PEF (Peak Expiratory Flow which), MVC (Maximum Voluntary Ventilation), $FEF_{25-75\%}$ (The average expired flow over the middle half of the FVC manoeuvre and is regarded as a more sensitive measure of small airways narrowing than FEV_1). These values forms the parameters for the diagnosis and management of disorder related to lung and other respiratory organs [5]. Broadly the respiratory disorders are classified into obstructive and restrictive. The ratio FEV_1/FVC is most significant and specific index to detect airflow obstruction. Restrictive Ventilatory defect can be identified if FVC is low and the ratio FEV_1/FVC is normal or high(typically > 80%)[2].

A low ratio of FEV_1/FVC is giving indication of obstructive Ventilatory defects. In the case, where FVC is low and low ratio of FEV_1/FVC occurs there is chance of having mixed Ventilatory defects, that is both obstructive and restrictive. By its own way these spirometric measurements play a vital role in the diagnosis and treatment process of asthma, COPD and restrictive lung diseases. Certain lung related cardiovascular diseases can be predicted by FEV_1 values. Even the intensity of these diseases can also be identified from FEV_1 values [3].

Spirometric measurements are depending on the exact and proper way of conducting the test, accuracy of equipment and performance of correct breathing manoeuvre. Therefore the test needs exact cooperation and coordination between the patient and examiner. As the test requires the effort from the patient, sometimes the test may not completed to acceptable level because of breathing difficulties, uncomfortable position, and lower confidence level of the subject which creates the incomplete data or data with some missing values [6]. To cater this requirement, the important FEV_1 value has to be predicted to make a decision on the pulmonary function disorders.
Radial Basis Function Neural Network and Support Vector Machine (SVM) have already been tested with the task of predicting FEV1 values [7-9]. However the traditional neural networks suffer with unacceptable level of generalization when compared with SVM.

2. SUPPORT VECTOR MACHINE

SVMs are one of the recently developed machine learning algorithm under the supervised learning approach, from the statistical learning theory implementing the structural risk minimization (SRM) principle[10-15]. It maps the data into high dimensional input space and constructs an optimal separating hyperplane in this space. The quality of the solution does not depend directly on the dimensionality of the input space.

It has been successful in many real world classification problems like handwritten recognition, object recognition, text categorization, image recognition, classification of gene expression and many more. Regression task in time series prediction, credit scoring has been successfully carried out by SVM[16,17].

Unlike neural networks, SVMs minimize the estimation error keeping the training error fixed [13]. There are many variants of SVM ever since its existence in the literature. The basic problem in the standard SVM formulation proposed by Vapnik is to solve the QPP, wherein the formulation of LSSVM [18] focuses on solving a set of linear equations. Thus non-linear pattern o is also done by solving a set of linear equation.

3. LEAST SQUARES SUPPORT VECTOR MACHINE

This section deals with the brief theory of LS SVM formulation for regression. Instead of slack variables for the constraints used in the ordinary SVM formulation, Suykens et al proposed to use the target values e_i. This way, the inequality constraints has been simplified as equality constraints and problem is formulated via least squares [19].

Let us consider the training set D be $\{(x_i, y_i)\}_{i=1}^{N}, x_i \in R^m$ and the output $y_i \in R$. The function to be estimated can be written as $y = y(x) = w^T \varphi(x) + b$, with w denoting the normal vector to the separating hyperplane, where $\varphi(\cdot) : R^m \to R^{n_h}$ is the mapping to the high dimensional and potentially infinite dimensional feature space. In the primal weight space this optimization problem is:

Minimize $\frac{1}{2} w^T + \frac{\gamma}{2} \sum_{i=1}^{N} e_i^{\,2}$

Subject to the equality constraints

$$y_i = w^T \varphi(x_i) + b + e_i \qquad \forall i \qquad (1)$$

If w becomes infinite dimensional, the primal problem can not be solved. By the inspiration of Lagrangian formulation we have,

$$L(w,b,e,\alpha) = \frac{1}{2} \| w \|^2 + \frac{\gamma}{2} \sum_{i=1}^{N} e_i^{\,2} - \sum_{i=1}^{N} \alpha_i [w^T \varphi(x_i) + b + e_i - y_i] \qquad (2)$$

Where α_i are the Larangian multipliers.

The conditions for optimality are:

$$\begin{aligned} \frac{\partial L}{\partial w} = 0 &\Rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 &\Rightarrow \sum_{i=1}^{N} \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 &\Rightarrow \alpha_i = \gamma e_i \qquad \forall i \\ \frac{\partial L}{\partial \alpha_i} = 0 &\Rightarrow w \varphi(x_i) + b + e_i - y_i = 0 \qquad \forall i \end{aligned} \qquad (3)$$

Equivalently with

$$Y = [y_1; y_2; \ldots ; y_n],$$
$$\alpha = [\alpha_1; \alpha_2; \ldots ; \alpha_n],$$
$$1_v = [1; 1; \ldots 1]$$

and after elimination of $w \& e$, (3) can be written as

$$\left(\begin{array}{c|c} 0 & 1_v^T \\ \hline 1_v & \Omega + I/\gamma \end{array}\right)\left(\begin{array}{c} b \\ \hline \alpha \end{array}\right) = \left(\begin{array}{c} 0 \\ \hline y \end{array}\right) \qquad (4)$$

The required solution of LS SVM for function estimation is given by

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \qquad (5)$$

with the kernel

$\Omega_{kl} = \varphi(x_k)^T \varphi(x_l) = K(x_k, x_l) \quad \forall\ k,l$ where α_i, b are the solutions of the linear systems

4. MATERIALS AND METHODS

For this study, 619 patient's spirometric measurements are considered including other covariates. The variables considered were age, height ,weight , sex and FVC values. The prediction of FEV_1 values were carried out by linear LS SVM, polynomial LS SVM and RBF LS SVM. We have used LS SVM toolbox to build the model for this classification. Matlab 7.5 has been used to construct the model[20].

The models were built with tenfold cross validation with minimal root mean square values as specification and parameters were tuned by grid search method. It is found that there is no significant difference between the linear and polynomial models in both normal and abnormal cases. From the results the predicated values in both cases are consistent and do not deviate much from the measured values.

5. RESULTS

The FEV_1 values predication for male patients is done with 258 cases for training and 111 cases for validation. Table 1 and Table 2 give the statistical analysis of the predictions. It is found that Linear LS SVM seems to be the better model even though the performances of all the models are same. Comparing with the measured values the predication of FEV_1 values by LS SVM models is a reliable and efficient method.

In case of female patients, 175 cases were used for training and 75 cases for validation. The polynomial and linear model performs in the same way whereas the RBF model over fits the data. In the abnormal female cases the RBF model slightly deviates from the measured values. The results of the predictions are presented in Table 3 and Table 4. The Fig1&2 depict comparison of measured and predicated FEV_1 values of Normal and Abnormal Female patients. From the results and

tenfold cross validation it is found that LS SVM based methods are comparatively better and efficient.

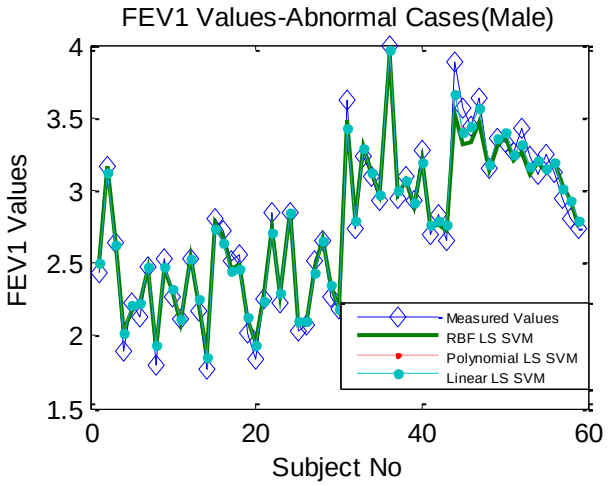

Fig1: FEV_1 values of male abnormal cases

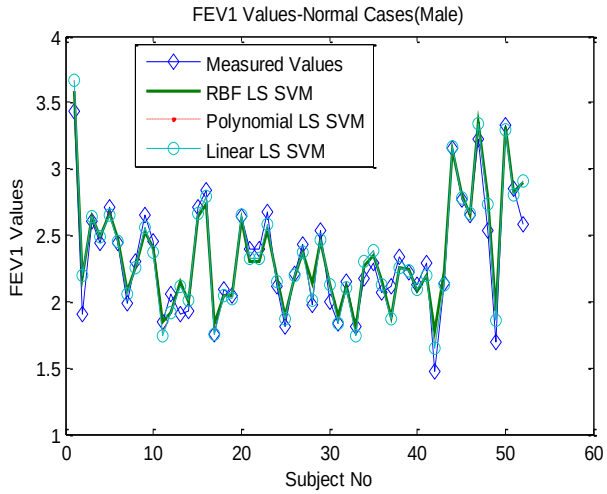

Fig 1: FEV_1 values of male normal cases

Table 1: Normal Cases (Male)

	Measured Values	RBF LS SVM	Polynomial LS SVM	Linear LS SVM
Mean	2.7585	2.7514	2.7622	2.7620
Standard deviation	0.5428	0.4872	0.4963	0.5004
t		0.9705	0.9688	0.9705
Confidence Interval on the mean of the difference		(-0.194, 0.1868)	(-0.1934, 0.1859)	(-0.1821, 0.1534)
RMSE		0.0945	0.0832	0.0817

Table 2: Abnormal Cases (Male)

	Measured Values	RBF LS SVM	Polynomial LS SVM	Linear LS SVM
Mean	2.3294	2.3567	2.3465	2.3438
Standard deviation	0.4267	0.4069	0.4327	0.4378
t		0.8657	0.8398	0.8657
Confidence Interval on the mean of the difference		(-0.1821, 0.1534)	(-0.1839, 0.1498)	(-0.1821, 0.1534)
RMSE		0.1238	0.1119	0.1118

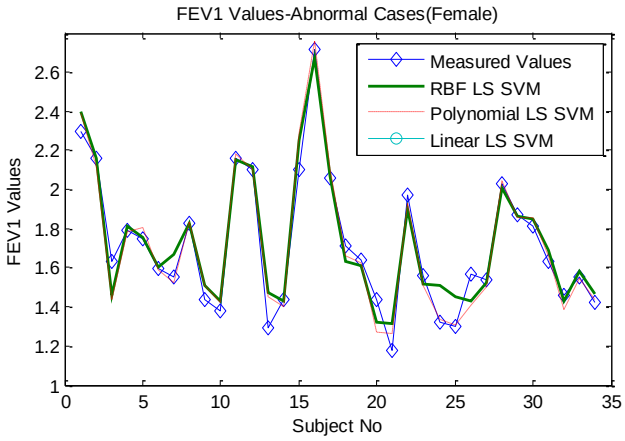

Fig 3: FEV_1 values of female abnormal cases

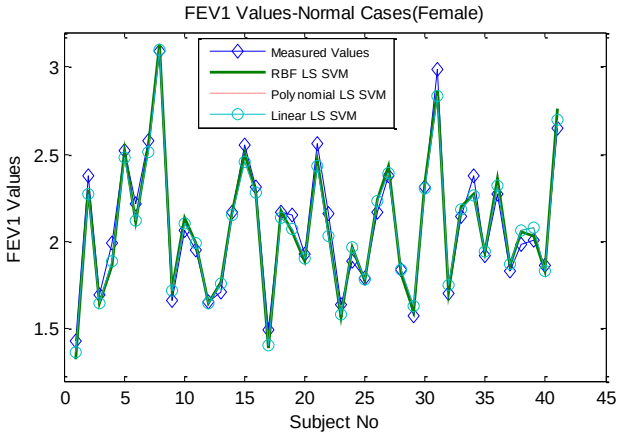

Fig 4: FEV_1 values of female normal cases

6. CONCLUSION

It is observed that the linear model better fits with respect to the root mean square value in all the cases. A similar work of FEV1 values have been carried out in [9]. But the detailed analysis of prediction of FEV1 values for males and females shows the generalization capacity of LS SVM regression. Thus this attempt of computing FEV1 values through LS SVM regression will enhance the spirometric investigations. In case of incomplete spirometric tests this method may give valuable suggestions and directions. In this work all the three models have been employed for the prediction. Parameters of the models are tuned in such a way that the mean squared errors of the fit as the cost criterion in ten folds cross validation. We observe that this methodology can be effectively implemented and gives an opening for spirometric clinical trials.

Table 3: Normal Cases (Female)

	Measured Values	RBF LS SVM	Polynomial LS SVM	Linear LS SVM
Mean	2.091	2.0808	2.0735	2.0735
Standard deviation	0.385	0.3894	0.3699	0.3699
t		0.9028	0.8322	0.8321
Confidence Interval on the mean of the difference		(-0.1597, 0.1807)	(-0.1482, 0.1837)	(-0.1482, 0.1837)
RMSE		0.0854	0.0344	0.0344

Table 4: Abnormal Cases (Female)

	Measured Values	RBF LS SVM	Polynomial LS SVM	Linear LS SVM
Mean	1.7147	1.7328	1.7152	1.7152
Standard deviation	0.3407	0.3316	0.3620	0.3620
t		0.8249	0.9952	0.9951
Confidence Interval on the mean of the difference		(-0.1809, 0.1447)	(-0.1708, 0.1697)	(-0.1708, 0.1697)
RMSE		0.0802	0.0686	0.0686

REFERENCES

[1]. Miller, M.R., Hankinson, J., Brusasco, V., Brugos, F., et al. (2005). Standardisation of spirometry. Eur. Respir. J.26 (2) , 319-338.

[2]. Pierce, R. (2005). Spirometer: an essential clinical measurement. Aust. Fam. Physician 34 (7), 535-539.

[3]. American Thoracic Society. (1991). Lung Function testing: selection of reference values and interpretative strategies. American Reviews on Respiratory Diseases, 144: 1202-18

[4]. Wagner, N. L., Beckett, W. S. and Steinberg, R. (2006). Using Spirometry result in occupational medicine and research common error and good practice in statistical analysis and reporting, Indian Journal of Medical System, 10. 5-10

[5]. David.P.J., Pierce.R (2008). Spirometry - The measurement and interpretation of ventilatory function in clinical practice. Spirometry Handbook, 3rd Edition. 1-24.

[6]. Ulmer, W.T. (2003). Lung Function-Clinical importance, problems, and new results. J. Physiol Pharmacol. 54 (1), 11-13.

[7]. Sujatha C. M., and Ramakrishnan S. (2009). Prediction of Forced Expiratory Vollume in Normal and restrictive respiratory functions using spirometry and self organzing map, Journal of Medical Engineering and Technology,33, 19-32.

[8]. Sujatha C. M., Mahesh V. and Ramakrishnan S. (2008), Comparison of two ANN methods for classification of spirometer data, Measurement Science Review,8(2),53–57.

[9]. Kavitha A., C. M. Sujatha and S. Ramakrishnan (2010), Prediction of spirometric Forced Expiratory Volume data using Support Vector Regression, Measurement Science Review, Volume 10, No. 2, 2010.

[10]. Boser, B., I. Guyon, and V. Vapnik (1992). A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory.

[11]. Vapnik V(1998), Statistical learning theory, John Wiley, New-York.

[12]. Vapnik.V (1995), The Nature of Statistical Learning Theory. Springer.

[13]. Kecman .V(2004), Learning and soft computing: Support Vector Machines, Neural Networks and Fuzzy Logic models.

[14]. Burges.C.J.C(1998), A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 2, 121-167.

[15]. Cristianini N., Shawe-Taylor J. (2000), An Introduction to Support Vector Machines, Cambridge University Press.

[16]. Kyoung - jae Kim, (2003) Financial time series forecasting using support vector machines. Neurocomputing, 55, 307 – 319.

[17]. Li T., Shiue W., and Huang H., (2006)The Evaluation of Consumer Loans Using Support Vector Machines, Computer Journal of Expert Systems with Applications, vol. 30, no. 4, pp. 772-782.

[18]. Suykens. J. A. K. , Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural Processing Letters, 9(3), 293-300

[19]. Suykens.J et al (2000), least squares Support Vector Machine classifiers: An Empirical Evaluation. K.U. Leuven ESAT-SISTA Internal Report 00-03

[20]. http://www.esat.kuleuven.be/sista/lssvmlab/