A Software Measurement Using Artificial Neural Network and Support Vector Machine

Bhavendra Kumar Sinha¹, Amit Sinhal² and Bhupendra Verma³

¹M.Tech. (IT) Student, Department of Information Technology, TIT, Bhopal
Bhavendra_sinha@yahoo.com

²Professor, Department of Computer Science & Engineering, TIT, Bhopal
amit_sinhal@rediffmail.com

³Director, TIT (Excellence), Bhopal
bk_verma3@rediffmail.com

ABSTRACT

Today, Software measurement are based on various techniques such that neural network, Genetic algorithm, Fuzzy Logic etc. This study involves the efficiency of applying support vector machine using Gaussian Radial Basis kernel function to software measurement problem to increase the performance and accuracy. Support vector machines (SVM) are innovative approach to constructing learning machines that Minimize generalization error. There is a close relationship between SVMs and the Radial Basis Function (RBF) classifiers. Both have found numerous applications such as in optical character recognition, object detection, face verification, text categorization, and so on. The result demonstrated that the accuracy and generalization performance of SVM Gaussian Radial Basis kernel function is better than RBFN. We also examine and summarize the several superior points of the SVM compared with RBFN.

KEYWORDS


1. Introduction

Software Measurement is the process which helps the organizations to improve a measurement program. It does not provide guidance for application of some measurement such as software cost measurement or to analyze software complexity. Instead of it does help on effective software measurement program and for understanding some of the key lessons that how the work. Klayman et al. (1999) reported that the measurement is accurate judgment process which depends upon confidence people but overconfidence increases with the difficulty of the task [1]. The purpose of the Software Measurement has grown up of successful measurement applications. Stone and Opel (2000) report about probability judgment accuracy by examining the effects of two different training techniques such as performance feedback or environmental feedback. They measured their improvement in calibration and discrimination as a function of feedback type which require separate training techniques for improvement [2]. It presents the specific procedures and activities which play the roles to achieve the goals and also roles of the people involved.

Measurement is the process by which numbers or symbols are assigned to attributes of entities in the real world in such a way as to characterize the attributes by clearly defined rules. Thus, measurement requires
Entities (objects of interest).
- Attributes (characteristics of entities).
- Rules (and scales) for assigning values to the attributes.

**Entity** is an object or an event in the real world. An entity is important to distinguish one entity from another by identifying characteristics.

An attribute is a feature or property of an entity. An entity defined by attributes and attribute using numbers or symbols. These numbers and symbols are abstraction that we use to reflect our perceptions of the real world. Software entities are described by attributes such as size, cost, elapsed time, effort expended, response time, transaction rates, number of defects found, and operational reliability.

In software measurement various factors have made development effort and cost difficult, complex and inaccurate. These factors influencing total development time between expert programmer and average programmer team. Software development cost measurement models can be used to determine development effort which can then transformed into the development time and cost. Although, these all are closely related with each others.

Accurate Software measurement is important for reasons such as
- It can provide the facility for classify and design different development projects for various business applications.
- The measurement model can determine how the resources can be used in well manner.
- It can be used to assess the impact of changes and support preplanning.
- The resources can be used to fulfill the needs of project application.
- Customers expect actual development costs and time to be in line with estimated costs and time.

2. A Review on Software Measurement

This software measurement based survey presents information to show the importance of measurement. It discusses significant specific strategic procedure and activity for software project measurement in terms of quality, complexity, cost, effort and schedule estimation. Software cost is related closely to software quality and productivity. Unrealistically low cost estimates frequently lead to poor product quality and low project productivity [3].

- G. R. Finnie et al. (1997) conclude that an artificial intelligence models are capable of providing adequate measurement models. Their performance is to a large degree dependent on the data on which they are trained, and the extent to which suitable project data is available will determine the extent to which adequate effort measurement models can be developed [4].
- Anita Lee et al. (1998) concluded that an integrated neural network model with cluster analysis is provide accurate measurement for software cost measurement to increase the training efficacy of the network [5].
- Hearst et al. (1998) positioned the SVM algorithm at the intersection of learning theory and practice: “it contains a large class of neural nets, radial basis function (RBF) nets, and polynomial classifiers as special cases. Yet it is simple enough to be analyzed mathematically, because it can be shown to correspond to a linear method in a high dimensional feature space nonlinearly related to input space.’’ [6].
Tony Van Gestel et. al. (2001) was described in this paper; the Bayesian evidence framework is combined with least squares support vector machines (LS-SVMs) for nonlinear regression in order to infer nonlinear models of a time series and the corresponding volatility. On the first level of inference, a statistical framework is related to the LS-SVM formulation which allows to include the time-varying volatility of the market by an appropriate choice of several hyperparameters. In the second level inferred hyperparameters, related to the volatility, are used to construct a volatility model within the evidence framework. Model comparison is performed on the third level to infer the tuning parameter of the RBF-kernel by ranking the evidences of the different models [7].

Zan Huang et. al. (2004) reported in their study a newly introduced learning method based on statistical learning theory, support vector machines, together with a frequently used high-performance method, backpropagation neural networks, to the problem of credit rating prediction. They used two data sets for Taiwan financial institutes and United States commercial banks for experiment. The results showed that support vector machines achieved accuracy comparable to that of backpropagation neural networks [8].

M. Jorgensen (2004) concludes on their review that expert estimation is the dominant strategy when estimating the effort of software development projects, and that there is no substantial evidence supporting the superiority of model estimates over expert estimates. There are situations where expert estimates are more likely to be more accurate, e.g., situations where experts have important domain knowledge not included in the models or situations when simple estimation strategies provide accurate estimates [9].

Mohammed Abdullah Al-Hajri et. al. (2005) report that in software measurement includes the standard Function Point (FP) and many different models derived from it. In this paper a new FP weights system was established using Artificial Neural Networks. This method is a modification of the complexity weights of FP measure and the results were very accurate and much suitable when they were applied on real data sets of software projects [10].

Kai-Bo Duan and S. Sathiya Keerthi (2005) stated that Multiclass SVMs are usually implemented by combining several Binary (two-class) support vector machines (SVMs) classification techniques which is a very well developed technique. Due to various complexities, a direct solution of multiclass problems using a single SVM formulation is usually avoided. The better approach is to use a combination of several binary SVM classifiers to solve a given multiclass problem. Popular methods are (WTA_SVM); (MWV_SVM). In the paper numerically study the performance of the four methods such as winner-takes-all strategy (WTA_SVM), max-wins voting (MWV_SVM), pairwise coupling with Platt’s probabilities (PWC_PSVM) and pairwise coupling with Kernel logistic regression (PWC_KLR). PWC PSVM gives the best classification results and has significantly smaller mean values of test error but in WTA SVM, MWV SVM and PWC KLR, it is hard to tell which one is better [11].

Yuming Zhou et. al. (2007) report Accurate software metrics-based maintainability prediction can not only enable developers to better identify the determinants of software quality and thus help them improve design or coding, it can also provide managers with useful information to help them plan the use of valuable resources. In this paper, The prediction accuracy of the MARS models are evaluated and compared using multivariate linear regression models, artificial neural network models, regression tree models, and support vector models. The results suggest that for one system MARS can predict maintainability more accurately than the other four typical modeling techniques [12].

YuShen Su and Chin-Yu Huang (2007) propose an artificial neural-network-based approach for software reliability measurement based on numerical analysis viewpoints of software reliability modeling and applying this approach to build a dynamic weighted combinational model (DWCM). In their experimental analysis two real software failure
data sets are analyzed by which gives you a better predictions result for proposed dynamic weighted combinational mode [13].

- Dr. Karanjeet Singh Kahlon et al. (2010) propose of measurement by selecting the most popular models among Artificial-Neural-Network Based Model (ANN) and Halstead, Walston-Felix, Bailey-Basili and Doty models to improve accuracy to utilize NASA software projects. The proposed Neuro based system is able to provide good measurement capabilities [14].
- Chintala Abhishek et al. (2010) focuses on finding a method which gives a best analysis report for effort prediction in testing phase by using large number of test cases by selecting the unique feature of learning through usage. The proposed model developed can also be evolved a minor effect on the effort measurement which affect the accuracy of model [15].

3. ANN Overview

The first neural network architecture that we have chosen is the Ward Network [16]. Neural network is a highly interconnecting network of a complex and large number of processing element called neurons. Neural network architecture classify besides processing on input and output layer and also have one or more intermediately layer called hidden layer. The input layer neurons are linked to hidden layer neurons and the hidden layer neurons are linked to output layer called Hidden output layer weights. An artificial neural network (ANN) is an efficient information processing system which resembles in characteristics with biological neural networks. ANN process large no of highly interconnected processing element called nodes or units or neurons, which usually operate in the parallel and are configured in regular architectures shown in Figure 1. Each neuron is connected with the other by a connection link. Each connection link is associated with weights which contain information about input signal. This information is used by the neurons net to solve a particular problem.

Figure 1:- Mathematical model of artificial neuron
4. Neural Network Learning Algorithm Classification

The classification of the neural network learning algorithm is illustrated as in Figure 2:

![Figure 2: Neural Network Learning Algorithm]

The neural network learning algorithms have been categorized as Supervised Learning (SL), Unsupervised Learning (USL) and Reinforcement Learning.

In Supervised Learning every input pattern is used to train the network is associated with the output patterns. During the Supervised Learning process the computation are based on actual and expected output to determine the errors and the errors can be used to improve the performance of the network model. In the Unsupervised Learning method the target output is not given to the network model. Now the system will capable to learn of its own desired pattern and observing expected structure for input pattern. In Reinforcement Learning (RL) algorithm is used both SL and USL algorithm approach for prediction of result where the various options related with correct option is given to the examiner and the examiner would capable to identify the correct expected answer by using their own approaches.

4.1 Radial Basis Function Network

RBF networks have been proven to be universal approximates and have the properties of fast convergence, easy solution to regularization, and robustness to outliers [17]. The construction of Radial Basis Function network (RBFN) based on three layer architecture namely, an input layer is made up of source nodes that connected to the network. The second layer hidden layer with non-linear Radial Basis activation function, such as Gaussian function that receives non linear transformation from the input layer to the hidden layer. The output layer implements a linear combination of Radial Basis Function that constitutes hidden layer function in a network to the activation pattern applied on it. Miyoung Shin (2000) pointed out that many desirable features and properties of RBF are flexible and potentially very useful for software measurement application [18]. The RBFN is used to applied on to focus complex pattern classification problem is basically solved by the transforming it into a high dimensional space in nonlinear manner.
Step 1: Set the weight to small random values.
Step 2: Perform Step 2-8 when the stopping condition is false.
Step 3: Perform Step 4-8 for each input.
Step 4: Each input unit receives input signals and transmits to the next hidden layer unit.
Step 5: Calculate the radial basis function.
Step 6: Select the centers for the radial basis function. The centers are selected from the Set of input vectors. It should be noted that a sufficient number of centers have to be selected to ensure adequate sampling of the input vector space.
Step 7: Calculate the output from the hidden layer unit:

\[
V_i(x_i) = \exp\left[-\sum \frac{(x_{ij} - \hat{x}_{ij})^2}{\sigma_i^2}\right]
\]

where

- \(\hat{x}_{ij}\) = Center of the RBF unit for input variables
- \(\sigma_i\) = width of ith RBF unit
- \(x_{ji}\) = jth variable of input pattern

Step 8: Calculate the output of the neural network:

\[
\gamma_{net} = \sum_{i=1}^{k} w_{im} V_i(x_i) + w_o
\]

where

- \(k\) = number of hidden layer nodes (RBF function).
- \(\gamma_{net}\) = output value of \(m\)th node in output layer for the \(n\)th incoming pattern.
- \(w_{im}\) = weight between \(i\)th RBF unit and \(m\)th output node.
- \(w_o\) = biasing term at \(n\)th output node.

Step 9: Calculate the error and test for the stopping condition. The stopping condition may be number of epochs or to a certain extent weight change.

5. Radial Basis Function Optimization using Support Vector Machine

Support vector machines (SVM) are innovative approach to constructing learning machines that Minimize generalization error. SVMs are based on very simple and intuitive concepts. They are constructed by locating a set of (hyper) planes that separate two or more classes of data. By construction of these hyper planes, the SVM discovers the boundaries between the input classes; the elements of the input data that define these boundaries called support vectors. SVMs can classify data separated by nonlinear boundaries. Through the use of kernel function, it problem is implicitly...
mapped to a higher dimensional space in which hyper planes suffice to define the boundaries.

Let us define labeled training examples \([v_i, y_i]\), an input vector \(v_i \in \mathbb{R}^n\), a class value \(y_i \in \{-1, 1\}\), \(i = 1, \ldots\). Here following SVM algorithm define hyper plane decision function for non-linearly separable case and multi class support vector classifier.

**Step 1:** Initialize Data needs to be trained under normal and faulty condition.

**Step 2:** Build the structure of SVM and Define the optimal hyper planes in terms of the support vector

\[
Y_{out} = \text{sign}(\sum_{i=1}^{t} y_i \alpha_i (v, v_i) + b)
\]  

(3)

Where \(Y_{out}\) is the outcome, \(y_i\) is the class value of the training example \(v_i\), and \(\alpha_i\) represents the inner product. The vector \(v = (v_1, v_2, ..., v_t)\) corresponds to an input and the vectors \(v_i, i = 1, \ldots, t\), are the support vectors. In Eq. (3), \(b\) and \(\alpha_i\) are parameters that determine the hyper plane.

**Step 3:** For the non-linearly separable case, a high-dimensional Version of Eq. (3) is given as follows:

\[
Y_{out} = \text{sign}(\sum_{i=1}^{t} y_i \alpha_i k(v, v_i) + b)
\]  

(4)

The function \(k(v, v_i)\) is defined as the kernel function and formula for a Gaussian radial basis function machine with kernel function is given as

\[
k(v, v_i) = (\gamma |v - v_i|^2)
\]  

(5)

Where, \(\gamma\) is radial basis function kernel.

**Step 4:** Determine the constraint \(C\) by cross validation.

**Step 5:** Construct the classification function and Compute the SVM model.

**Step 6:** Remove low rank feature and stop when features will be ended.

**6. RESULT SIMULATION**

The present work have tried to find out the Development Time (DT) by applying first the Neural Radial Bias Function Network Model and then the Support Vector Machine. Following methodology was adopted to carry out the development time prediction measurement using both the approaches. The evolution criteria of performance analysis is based on Magnitude of Relative Error (MRE), Mean Magnitude of Relative Error (MMRE) and Percentage of the Prediction (PRED) which are mentioned below as :-

\[
\text{MRE} = \frac{|ADT - EDT|}{ADT}
\]

\[
\text{MMRE} = \sum_{i=1}^{N} \frac{MRE}{N}
\]

\[
\text{PRED}(X) = \frac{P}{N}
\]
Where, ADT is actual development time and EDT is estimated development time and P is the number of instances of projects with MRE less than or equal to X which is accepted at level of 0.25 and N is the number of considered projects used. The selected model comparisons are given in Table 1 with various criteria such as Standard Deviation (SD), Mean Absolute Error (MAE) and with Min, Max and Mean error values.

Table 1: Experimental Result Evaluation of Proposed Model

<table>
<thead>
<tr>
<th>Measurement Model</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBFN</td>
<td>0.619</td>
<td>2.83</td>
<td>-5.639</td>
<td>6.59</td>
<td>2.09</td>
</tr>
<tr>
<td>SVM</td>
<td>0.299</td>
<td>2.194</td>
<td>-3.72</td>
<td>5.055</td>
<td>1.772</td>
</tr>
</tbody>
</table>

Table 2 shows the dataset used for carrying out experimentation which has Lopez-Martin dataset with results of MRE and MMRE of both model RBFN and SVM with Gaussian Radial basis kernel function.

Table 2: Detailed Lopez-Martin dataset with their standard numeric values.

<table>
<thead>
<tr>
<th>Lopez-Martin Data</th>
<th>MRE (Mean Relative Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S No  mc  dc  loc  dt</td>
<td>RBFN</td>
</tr>
<tr>
<td>1  1  0.25  4  13</td>
<td>0.476545</td>
</tr>
<tr>
<td>2  1  0.25  10  13</td>
<td>0.312803</td>
</tr>
<tr>
<td>3  1  0.333  4  9</td>
<td>0.593678</td>
</tr>
<tr>
<td>4  2  0.083  10  15</td>
<td>0.129818</td>
</tr>
<tr>
<td>5  2  0.111  23  15</td>
<td>0.049379</td>
</tr>
<tr>
<td>6  2  0.125  9  15</td>
<td>0.038338</td>
</tr>
<tr>
<td>7  2  0.125  9  16</td>
<td>0.098442</td>
</tr>
<tr>
<td>8  2  0.125  14  16</td>
<td>0.034378</td>
</tr>
<tr>
<td>9  2  0.167  7  16</td>
<td>0.100582</td>
</tr>
<tr>
<td>10 2  0.167  8  18</td>
<td>0.126528</td>
</tr>
<tr>
<td>11 2  0.167  10  15</td>
<td>0.168749</td>
</tr>
<tr>
<td>12 2  0.167  10  15</td>
<td>0.168749</td>
</tr>
<tr>
<td>13 2  0.167  10  18</td>
<td>0.026043</td>
</tr>
<tr>
<td>14 2  0.2   10  13</td>
<td>0.079027</td>
</tr>
<tr>
<td>15 2  0.2   10  14</td>
<td>0.001954</td>
</tr>
<tr>
<td>16 2  0.2   10  15</td>
<td>0.064843</td>
</tr>
<tr>
<td>17 2  0.2   15  13</td>
<td>0.189461</td>
</tr>
</tbody>
</table>
In the Table 3 MMRE and PRED (0.25) criteria is analyzed over the complete dataset. It can be seen that the SVM model outperform the RBFN model. The performance of SVM and RBFN model is compared for development time prediction problem in the present work.

Table 3: Result analysis of different prediction model with various performance criteria

<table>
<thead>
<tr>
<th>Performance Criteria</th>
<th>RBFN</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMRE</td>
<td>0.13508</td>
<td>0.108056</td>
</tr>
</tbody>
</table>

49
The results have minimized error as compared with actual value. These results emphasize the importance of software development time measurement. Figure 3 shows the behavior of measurement of the optimal software development time in available dataset used in thesis work. In the Figure 3 red fluctuated line denotes the estimated development time mean relative error (MRE) of SVM model and compares it with blue fluctuated line of RBFN model. From the Figure 3, it is found that both RBFN and SVM model fluctuated in the neighborhood of the actual development time and MRE is taken on various observation points.

Figure 3: Comparison in MRE between both Software Measurement model RBFN and SVM

Figure 4: Graph on MMRE (Mean Magnitude Relative Error) of different Prediction Model.
From the graph drown in Figure 4 and Figure 5 based on result analyzed by Table-3, it is clearly recognize that SVM approach give better performance then the existing RBN model.

5. CONCLUSION

This work illustrated a review on neural network and SVM approaches for software measurement to predict software development time. The proposed analysis gives the better result to compare their work done and result over the year. The review clearly points out the potential of artificial neural networks being used as a tool for classification and prediction problems. This work evaluates the capability of SVM using Gaussian basis kernel function which have been successfully applied for solving both classification and regression problems in many applications. The results showed that support vector machines achieved accuracy comparable to that of radial bias neural networks. The methodology can be used to helpful for proposing an efficient method using both techniques such as RBFN and SVM which can be applied over the other measurement model for software measurement with the efficient manner.

REFERENCES

Prediction Using Least Squares Support Vector Machines Within the Evidence Framework”, IEEE Transactions on Neural Networks, Vol. 12, No. 4, pp. 809 - 821


AUTHORS

Bhavendra Kumar Sinha is a M. Tech student in the Technocrats Institute of Technology, Bhopal (MP). He received Bachelor of engineering degree in Information Technology from GuruGhasiDas Vishwavidyalya, Bilaspur(CG). His main research interest include software metrics and measurement, software engineering application and application of softcomputing and data mining in software engineering.

Prof. Amit Sinhal completed his B.E. in Computer Engineering from NIT Surat in 1996, M.Tech in Computer Science & Engineering from SATI Vidisha in 2005 and is pursuing Ph.D. from Rajiv Gandhi Technical University, Bhopal. He worked in various reputed software development companies as Project Lead and University Institute of Technology, Barkatullah University Bhopal as Assistant Professor. Currently he is working at Technocrats Institute of Technology, Bhopal as Professor in the Computer Science & Engineering department. He has published/presented more than 20 research papers in International Journals/Conferences.

Dr. Bhupendra Verma completed his B.E. in Computer Engineering, M.Tech in Computer Science & Engineering from SATI Vidisha and Ph.D. from Rajiv Gandhi Technical University, Bhopal. Currently he is working in Technocrats Institute of Technology, (Excellence) Bhopal as Director and Professor in the Computer Science & Engineering department. He has published/presented more than 50 research papers in International Journals/Conferences. He has guided 20 M.Tech dissertations and is presently guiding five students for Ph.D.