EPH - International Journal of Science And Engineering
ISSN (Online): 2454 - 2016
Volume 01 Issue 01 January 2015

DOI: https://doi.org/10.53555/eijse.v1i1.15

ESTIMATION OF SOFTWARE DEVELOPMENT EFFORT WITH MACHINE LEARNING APPROACHES:A REVIEW

Sonam Bhatia[1]*, Varinder Kaur Attri[2]
*[1,2]Dept. of CSE GNDU, RC jalandhar, India, [2]Email:- varinder2002@yahoo.com.

Corresponding Author:-
Email: dbhatia297@gmail.com.

Abstract: -

For the initial steps of the software life cycle, it is essential to handle software estimation, because it assists managers bid on projects and allot resources conventionally. In software planning estimation of the effort is one of the most critical responsibilities. It is necessary to have good effort estimation in order to conduct well budget. The accuracy of the effort estimation of software projects is vital for the competitiveness of software companies. For the forecasting of software effort, it is important to select the correct software effort estimation techniques. Inaccurate effort estimation can be risky to an IT industry's economics and certainty due to poor quality or trait and stakeholder's disapproval with the software product. This paper presents the most commonly used machine learning techniques such as Multi-Layer Perceptron, linear regression, decision tree, for effort evaluation in the field of software development.

Keywords: - *Effort estimation, Decision tree, linear regression, Multi-Layer Perceptron*

I. INTRODUCTION

Software effort estimation is the forcasting about the amount of effort needed to make a software system and its duration [1] Good estimates play a very important role in the management of software projects. [2]. The effort is the most important cause that affecting the budget of a project. Estimating the effort with a high degree of accuracy is a issue which has not yet been solved and even the project manager has to deal with it since the beginning. Several parameters can affect the effort estimation. These parameters Incorporate Size, Category, Personnel Attributes, Complexity [3] Most of the effort estimation metrics takes the input as the software size, which can be measured with function point, LOC, object point. A number of models have been enlarged to provide the relation between size and effort [18]

SLOC is typically used to predict the amount of effort that will be needed to establish a program, as well as to valuation programming productivity or maintainability once the software is developed[11] Effort is measured in terms of person months and duration.[4]. More recently attention has turned to a variety of machine learning techniques to predict software development effort [7] [8]. Most of the projects are break down due to imprecise estimated effort, so the success of any software project depends on an initial and accurate effort estimation. [9] The purpose of Machine Learning is to provide increasing levels of automation in the knowledge engineering process, replacing much time consuming human activity with automatic techniques that improve reliability or efficiency by observing and manipulating regularities in training data.[5]

There are many reason for vary of effort estimation. These are Project approval, project management, defining of project task etc. The field of Machine Learning (ML) is devoted to develop computational methods that implement various forms of learning, in particular mechanisms capable of inducing knowledge from examples or data [10]. An important requirement is that the learning system should be able to deal with imperfections of the data. Many methods have been explored for software effort estimation, consisting traditional methods such as the COCOMO and, more recently, machine learning techniques such Linear regression, Multi-Layer Perceptron, Decision tree[2] Machine learning methods have been exploited to generate better software products ,to be part of software product and to make software development process more convenient and adequate .[6]

The remainder of this paper is laid out as follows. Section II describes need for effort estimation. An evaluation criterion and effort estimation using machine learning is presented in Section III. Section IV highlights the comparison of effort estimation techniques. In Section V, overviews related work. In Section VI, conclusion and future scope are presented.

II. NEED FOR EFFORT ESTIMATION

Effort estimation is necessary for many people and different departments in an organization. At various point of project lifecycle well-defined effort estimation is essential. The computation of the effort might be used as input to project plans, determining the budget and other important procedure needed for the successful release of the software. The progress or failure of projects depends on the authenticity or reliability of effort and schedule evaluations, among other things. Early effort estimation also assists the project manager to investigate whether the available resource is effective to complete the project. As software applications have grown in size and significance, the need for reliability in software cost estimating has grown, too. [19].

III.SOFTWARE PROJECT EFFORT ESTIMATION USING MACHINE LEARNING

A. Evaluation criterion

Correlation measures of the strength of a relationship between two variables, Mean Absolute error measures of how far the estimates are from actual values, Relative absolute Error (RAE) takes the total absolute error and normalizes it by dividing by the total absolute error of the simple predictor. Root Mean Square Error (RMSE): RMSE evaluates the difference between value estimated by a model and the value actually observed. [9]

B. Effort Estimation Techniques

Machine learning is a new field which is elaborating the promise of developing consistently reliable estimates. The system effectively "learns" how to evaluate from training set of finished projects.

1) Regression

Regression is a machine learning approach used to fit an equation to a dataset. Regression analysis is a statistical manner that attempts to explore and model the relationship between two or more variables. .Linear regression analysis is the most widely used of all statistical techniques. It is the study of linear, additive relationships between variables. Linear regression is an approach for modeling the relationship between a scalar dependent variable y and independent variable denoted X. The case of one explanatory variable is called simple linear regression [4][13]. For more than one independent variable, the process is called multiple linear regression [1]. Linear model considers the relationships between variables are straight-line relationships. The regression model is then used to evaluate the result of an unknown dependent variable, given the values of independent variables.

2) Multi-layer perceptronl

Neural networks are a nonlinear modeling approach influenced by the functioning of the human brain [15] and have previously been implemented in the context of software effort estimation. A multilayer perceptron (MLP) is a feed forward artificial neural network for structuring the neurons is known as multi-layer neural network. Artificial neurons are interconnected in the form of layers. The first layer called the input layer neurons that express the set of input variables. The output layer express the output variable which is the actual effort required to end the project. The connections between the neurons have weighted numerical inputs associated with them. The layers between two layers are known as hidden layers .All the neurons of one layer generate some output, which acts as input to the next layer. This `next layer' can be either the hidden layer or the output layer [14] [9]. It uses back propagation learning algorithm is used to compute the effort [9] the back-prop method works by determine the difference between output and the observed output value.

The Multi-Layer Perceptron (MLP) neural network has been applied successfully to a number of problems such as regression, time series forecasting and classification The main components that effect performance during the use of multilayer perceptron in project are: the number of hidden layers, the number of neurons of each hidden layer, the number of training epochs, and the learning rate and the momentum. [14].

3) Decision Tree

A decision tree is a logical model that is contributes in operations research, specifically in decision analysis [9]. Decision tree is a kind of tool to come out with a decision on the basis of some conditions and their possible consequences.[16] Decision tree is a procedure used for classification and regression [15].Decision tree is a flowchart like tree structure, where each internal node stand for a test on an attribute, each branch express an outcome of the test, and each leaf node holds a class label. The root node is the topmost node in a tree [17]

Decision trees are generated from training data in a top down, general to specific direction. The initial state of tree is root node that is assigned all examples from training the training set. If it is case that all the examples belong to same class then no further decision need to be made to partition the examples and the solution is complete. If example at this node belongs to two or more classes then test is made at node that will result in split. The process is recursively repeated for each intermediate node until completely discriminating tree is obtained.M5P is powerful because it implements as much decision trees as linear regression for predicting a continuous variable. This algorithm is a multivariate tree algorithm which is appropriate for noise removal and also applies for large database The M5P Introduced by Quinlan, the model tree technique (M5) can be recognized as an extension to CART. A model tree will fit a linear regression to the observations at each leaf rather of allowing a single value like CART. The M5P algorithm has three stages: building a tree, pruning the tree and smoothing.

[4] COMPARISON OF TECHNIQUE - ADVANTAGES AND DISADVANTAGES

With increases in size, software effort also increases because effort depend on size Each technique has own advantages and disadvantages. Decision tree needs little data preparation and use white boxmodel. Nonlinearity is very important propertyin case of feed forward neural network MLF neural networks are very robust, i.e. their outcomes deteriorated in the presence of increasing amounts of noise .Regression is applicable is on the normalized data.

Table i: advantages and disadvantages of techniques

Techniques	Advantages	Disadvantages
Regressions	The advantages of a linear model are its simplicity and ease of use. Analysts can use linear regression together with manner such as variable recoding, transformation, or segmentation	Linear regression is restricted to estimate numeric output. The major drawback of linear regression is its high model bias: if the underlying function is not well approximated by a linear function, then linear regression generates poor outcome.Density of Branching
Multi Layer Perceptron	They have small memory requirements and efficiently classify new data and exhibit good generalized capability. Multi layer perceptron Can easily be parallelized	Computationally expensive learning process because Large number of iterations needed for learning, not suitable for real-time learning There is a scaling problem ,it is hard to scale
Decision trees	It is simple to understand and interpret and achieves well with huge data in short in a short time Decisions. It is able to handle both numerical or categorical data	It is Not good for predicting the value of a continuous class attribute. Most of the algorithms (like ID3 and C4.5) require that the target attribute will have only discrete values.

IV. Related work .

Jyoti Shivhare [9] in 2014 presented a paper and described an technique for estimation based upon various feature selection and machine learning techniques for non-quantitative data and is investigated in two phases. In the first phase of method three feature selection techniques, such as Rough Reduct, RSA-Rank and Info Gain, are applied to the dataset to find the optimal feature set. The second phase include effort estimation for reduced dataset using machine learning techniques like FFNN, RBFN, FLANN, LMNN, NBC, CART and SVC .

Sumeet Kaur Sehra [7] in 2011 described that the Radial basis neural network gives more reliable results as compared to intermediate COCOMO Model and fuzzifying size and cost drivers by using Gaussian MF. The accuracy of effort estimation can be improved and the estimated effort is very close to the actual effort. Also explained genetic programming based effort model provides results which are more robust and accurate.

Abdulbasit S. Banga[14]in 2011 explained algorithmic cost model and machine learning techniques and also described advantages , disadvantages of each techniques and models and the underlying aspects in preparing cost estimates. Comparison of various estimation techniques and models also described by author.

Neha Saini [16] in 2014 evaluated various machine learning techniques for software effort estimation like bagging, decision trees, decision tables, multilayer perceptron and RBF networks. Two different datasets i.e. heiatheiat dataset and miyazaki94 dataset have been used in research. Decision trees are good for evaluating the software effort. Also author described that Decision trees perform best among a other models in term of MMRE value.

Karel Dejaeger [15] in 2012 presented a paper and explained that ordinary least squares regression in combination with a logarithmic transformation performs best. By selecting a subset of highly predictive attributes, typically a significant increase in estimation accuracy can be achieved. These results also demonstrate that data mining approches can make a valuable comission to the set of software effort estimation techniques, but should not change expert judgment.

Evandro N. Regoli [10] in 2003 explored two ML techniques, GP and NN. Author described that both techniques perform well in the regression problem. GP is able to investigate the correct functional equation that fits the data and its appropriate numerical coefficients. NN gives a net that express a complete mathematical formula, without a direct interpretation.

Ruchika Malhotra [4] in 2011 presented a paper and estiamte, compares the potential of Linear Regression, Artificial Neural Network, Decision Tree, Support Vector Machine and Bagging on software project dataset. The dataset is obtained from 499 projects. The results show that Mean Magnitude Relative error of decision tree method is only 17.06%. Thus, the performance of decision tree method is better than all the other compared methods.

Sweta Kumari [18] in 2013 provided a comparative study on support vector regression (SVR), Intermediate COCOMO and Multiple Objective Particle Swarm Optimization (MOPSO) model for effort estimation and SVR gives better results.

VI. CONCLUSION AND FUTURE SCOPE

The main contribution of this review is to deliberate different machine learning techniques engaged in effort estimation the paper also gives a relative comparison of approaches based on their advantages and limitations. Researchers have developed different models for estimation but there is no estimation approach which can compute the best estimates in all various situations and each technique can be suitable in the special project. Software development in this stage is at ambitious phase, and evaluation of effort in this area always remains an open issue and considered to be a complex assignment.

Many applications will be designed which have expected to have less effort so that software complexity can be reduced. This survey indicates directions for further research. Trying to enhance the performance of existing methods and familiarize the new methods for estimation based on today's software project requirements can be future works in this area so the research is on the way to combine different approaches for calculating the best estimate.

ACKNOWLEDGEMENT

We would like to thank acknowledge almighty for his constant blessings. Then we like to thank our family and friends for helping and supporting us throughout the making of this paper.

REFERENCES

[1]. Olga , "Software Effort Estimation with Multiple Linear Regression: review and practical application Journal Of Information Science And Engineerin (2011)

[2]. Petrônio L. Braga and Adriano L. I. Oliveira, ",Software Effort Estimation using Machine Learning Techniques with Robust Confidence Intervals",19th IEEE International Conference on Tools with Artificial Intelligence

[3]. Ali Bou Nassif, " Software Size and Effort Estimation from Use Case Diagrams Using Regression and Soft Computing Models" ,MAY 2012

[4]. Ruchika Malhotra, "Software Effort Prediction using Statistical and Machine Learning Methods", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.1, January 2011,.

[5]. Yogesh Singh, Pradeep Kumar Bhatia & Omprakash Sangwan, "A Review Of Studies On Machine Learning Techniques" International Journal of Computer Science and Security, Volume (1) : Issue (1)

[6]. Zhang, "Advances in Machine Learning Applications in Software Engineering

[7]. Sumeet Kaur Sehra1, Yadwinder Singh Brar2, and Navdeep Kaur3 , "Soft Computing Techniques For Software Project Effort Estimation",International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624.Vol 2, Issue 3, 2011, pp 160167
[8]. Prasad Reddy P.V.G.D, Sudha K.R, Rama Sree P and Ramesh "Software Effort Estimation using Radial Basis and Generalized Regression Neural Networks", Journal of Computing, Volume 2, Issue 5, pp 87-92,. 2010
[9]. Jyoti Shivhare, "Effectiveness of Feature Selection and Machine Learning Techniques for Software Effort Estimation" June 2014
[10]. Evandro N. Regolin Gustavo A. de Souza,"Exploring Machine Learning Techniques for Software Size Estimation" ,International Conference of the Chilean Computer Science Society (SCCC'03)1522 4902/03 $ 17.00 © 2003 IEEE
[11]. Kaushal Bhatt, Vinit Tardy, Pushpraj Patel , "Analysis Of Source Lines Of Code(SLOC) Metric", International Journal of Emerging Technology and Advanced Engineering(ISSN 2250-2459, Volume 2, Issue 5, May 2012)
[12]. Geetika Batra, Kuntal Barua , "A Review on Cost and Effort Estimation Approach for Software Development" International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 4, October2013
[13]. Mohd. Sadiq, Aleem Ali, Syed Uvaid Ullah, Shadab Khan, and Qamar Alam, " Prediction of Software Project Effort Using Linear Regression Model"International Journal of Information and Electronics Engineering, Vol. 3, No. 3, May 2013
[14]. Abdulbasit S. Banga "Software Estimation Techniques" National Conference; INDIACom-2011Computing for Nation Development, March,2011
[15]. Karel Dejaeger, Wouter Verbeke, David Martens, and Bart Baesens, "Data Mining Techniques for Software Effort Estimation: A Comparative Study", IEEE Transaction ON Software Engineering, VOL. 38, NO. 2, MARCH/APRIL 2012
[16]. Neha Saini ,Bushra Khalid, "Empirical Evaluation of machine learning techniques for software effort estimation" IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. IX (Feb. 2014), PP 34-3
[17]. Jiawei Han Data Mining: Concepts and Techniques Second Edition,2011 [18] Sweta Kumari "Comparison and Analysis of Different Software Cost Estimation Methods"(IJACSA) International Journal of Advanced Evandro N. Regolin Computer Science and Applications, Vol. 4, No.1, 2013 Gustavo A. de Souza, "Exploring Machine Learning Techniques for Software Size [19] Estimation",International Conference of t_{2003}http://www.qualitymanagementconference.com/effort_estimation.php IEEE he Chilean Computer Science Society (SCCC'03)1522-4902/03 $ 17.00 ©