

A Comparative Study of Genetic Algorithm, Support Vector Regression and Artificial Neural Network for software cost estimation

Devbrat Sahu
M.Tech Research Scholar, CSE
Chatrapati Shivaji Institute of Technology
Durg, India
devbratsahu89@gmail.com

Mrs. Deepty Dubey
Associate Professor, CSE
Chatrapati Shivaji Institute of Technology
Durg, India
deepty.shukla@gmail.com

Abstract— Software Cost Estimation (SCE) is one of the important topics in producing software in recent decades. Real estimation requires cost and effort factors in producing software by using cost estimation algorithms or Artificial Intelligent (AI) techniques. It uses a hypothesis space of linear functions in a high dimensional space, trained with a learning algorithm from optimization theory. ANN are learning frameworks enlivened by natural neural frameworks, which are known to be exceptionally effective at figuring out how to take care of issues. The SVR is executed as the Support Vector Machine (SVM) for relapse. Genetic Algorithms (GA) was produced as option strategies for handling general advancement issues with expansive spaces. In this paper, different data mining techniques to estimate software costs are presented and then the results of each technique like SVR, ANN, GA are evaluated and compared. SVM generalization execution (estimation precision) relies on a decent setting of meta-parameters C, and the part parameters. Execution of Artificial Neural Networks is Spatial Data Analysis. Genetic Algorithm's approach can be used for modeling the Performance of Memory-bound Computations. The objective of this paper is to compare and evaluate Artificial Neural Network, Support Vector Regression and Genetic Algorithm based on software cost estimation.

Keywords- Artificial Neural Network, Support Vector Regression, Genetic Algorithm.

I. INTRODUCTION

Software cost estimation can be defined as the approximate judgement of the costs for a project. Cost estimation will never be an exact science because there are too many variables involved in the calculation for a cost estimate, such as human, technical, environmental, and political. Furthermore, any process that involves a significant human factor can never be exact because humans are far too complex to be entirely predictable. Furthermore, software development for any fair-sized project will inevitably include a number of tasks that have complexities that are difficult to judge because of the complexity of software systems.

Cost estimation is usually measured in terms of effort. The most common metric used is person months or years (or man months or years). The effort is the amount of time for one person to work for a certain period of time.

An artificial neuron network (ANN) is a computational model based on the structure and functions of biological neural networks. Support Vector Regression is a technique use to build support vector machines for regression with the possibility to add or remove samples without training the machine from the beginning. Genetic Algorithm (GA) is a search heuristic that mimics the process of natural selection. This heuristic (also sometimes called a meta heuristics) is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover.

Input-

Cost drivers
Constraints

Other inputs

Output-

Effort
Loading
Duration

II. PROPOSED METHODOLOGY

A. ARTIFICIAL NEURAL NETWORK

ANN model is one of building design that is utilized all the more as a part of SCE. ANNs have three layers that are interconnected. The first layer consists of input neurons. Those neurons send data on to the second layer, which in turn sends the output neurons to the third layer. ANN model proposed to investigate the relationship between software size (expressed in LOC or FP) and effort, by conducting a series of experiments. We are concerned with inspecting the predictive ability of the ANN with respect to the attribute counting the size of the software developed for each project in each dataset.

B. SUPPORT VECTOR REGRESSION

SVR is based on the Support Vector Machine (SVM) for regression. It uses a hypothesis space of linear functions in a high dimensional space, trained with a learning algorithm from optimization theory, and it implements a learning bias derived SVR is based on the Support Vector Machine (SVM) for regression. It uses a hypothesis space of linear functions in a high dimensional space, trained with a learning algorithm from optimization theory, and it implements a learning bias derived from statistical learning theory[2]. For predicting software costs, SVR applies a linear model to implement non-linear class borders. It maps nonlinear info vectors (comprising of EM and Size of the activities) into a high dimensional qualities space by method for pieces. In this theme, part is made out of poly piece. At that point, the help vectors are connected to create an ideal straight dividing hyper plane (in an instance of

example distinguishment) or a direct relapse capacity (on account of relapse) in this feature space.

C. GENETIC ALGORITHM

Genetic Algorithms (GA) were developed as an alternative techniques for tackling general optimization problems with large search spaces. Genetic Algorithm (GA) is constructed to calibrate the model's architecture. Genetic Algorithms (GA) are also widely used for accurate effort estimation.

The basic process of the genetic algorithm is as follows, although a number of variations are possible:-

1. Generate at random a population of solution ,i.e. a family of chromosomes.
2. Create a new population from the previous one by applying genetic operation to the fittest chromosomes, or pairs of fittest chromosomes of the previous population.
3. Repeat step(2), until either the fitness of the best solution has converged or a specified number of generations have been produced

III. EXPECTED OUTCOME

After the implementation of proposed research work, the expected outcome includes table form generated on the basis of the following:-

- A. Correlation Coefficient
- B. Mean absolute error
- C. Root mean squared error
- D. Relative absolute error
- E. Root relative squared error

IV. CONCLUSION

Artificial Neural Network, Genetic Algorithm and Support Vector Regression all the three algorithms are proposed for the software cost estimation. Fundamentally all the three algorithms have been studied. It is to be proved that Genetic Algorithm is Better than SVR and ANN.

REFERENCES

- [1] J. S. Pahariya, V. Ravi, M. Carr, M. Vasu, Computational Intelligence Hybrids Applied to Software Cost Estimation, International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), Vol. 2, pp. 104-112, 2010
- [2] Zeynab Abbasi khalifelu, Farhad Soleimanian Gharehchopogh, Comparison and evaluation of data mining techniques with algorithmic models in software cost estimation, 2012
- [3] <http://www.computing.dcu.ie/~reanaat/ca421/report.html>
- [4] <http://www.techopedia.com/definition/5967/artificial-neural-network-Ann>.
- [5] www.sciencedirect.com