International Journal of Pure and Applied Mathematics
Volume 118 No. 20 2018, 3851-3862
ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

A CRITIQUE ON SOFTWARE COST ESTIMATION

K.Rajeswari, Assistant Professor,
Dept of Computer Science,
Kongunadu Arts and Science college,
Email Id: rajeswari13@yahoo.com

Dr.R.Beena, Associate Professor& Head,
Dept of Computer Science,
Kongunadu Arts and Science College,
Email Id: beenamridula@yahoo.co.in

Abstract

Software Engineering is playing predominant role in various developing fields of computer science and engineering. It is a platform which fuses software with hardware based on required specification. Software is a simple executable code which should have minimum complexity and low implementation cost. Though lot of methods has been incorporated for cost estimation of software, but it still remains an alarming issue in software industry. Overall cost estimation of the software includes Software Plan, Resources, Design, Development, Coding and testing. The two parameters which plays vital role in cost estimation is workload and work progress during the course of development and implementation. So the appropriate recurring and non- recurring cost of software is indispensable in the field software engineering. This paper provides an overview of software cost estimation in various fields of science and technology.

Keywords: Software Cost Estimation (SCE), Algorithmic methods, Non-algorithmic methods, Machine-learning methods.

Introduction

Many modern organizations rely on a software development for their growth and sustainability in current developing IT industry. Cost estimation is a challenging issue for many technological firms. The cost of a software should be estimated in such a way that it must not compromise quality, timelines and efficiency.

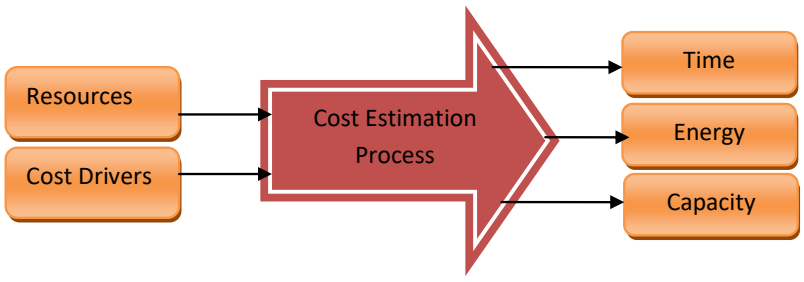

Figure 1: Process of Cost Estimation

SCE can be implemented by using three methods algorithmic, non-algorithmic and machine-learning methods. The former technique is about the mathematical modeling based on prior project works which depends upon source line of code (SLOC), Function points and other cost drivers. The primary focuses of the algorithmic model is design, analysis, implementation and optimization. The secondary focus is on the experimental evaluation and practical applications of the software. The most frequently used models are COCOMO, function point, Putnam's models.

The Constructive cost model (COCOMO) is commonly used method for estimating the time and effort of the SCE. In current trends, the COCOMO-II model plays an essential role in the software projects.

In Non-algorithmic technique, the former data is used for estimating the cost of the existing projects with proposed software projects. It is easy to learn and it should have complete information of the past projects. The recent methods of estimation are based on analogy technique, Judgment, Price to win etc...

Accurate software effort is predicted by machine learning techniques because of its nature learning. It is merged with data mining techniques. Its main focuses are on data analysis and known as unsupervised learning.

2. Literature Review

Xiaotie Qin et al [1] emphasized management is an important factor. To estimate a project cost the basic requirement are resource, planning, analysis, design, coding and testing. Several methods are integrated in SCE. These methods are compared for perfect cost and the modification is easy because each step is documented. It is dependent on the precedent data and the accuracy can be estimated in the final stage of lifecycle.

SCE plays an important vital role in software industry to produce the accurate budget of a project. Snehal A. Deshmukh et al [2] say an unexpected budget increase and lack of project completion can leads to lose the opportunities of business in IT. To overcome this problem classification of data mining techniques is used to compare with the SCE model (COCOMO-II) [1]. It identifies the cost drivers of the historical data and it is a multiplicative factors. K-means algorithm is used to find the total variance. Cost drivers are very useful to predict the cost of a new project.

V.R. Arulmozhi et al applied various methods to estimate a cost of software. It predicts size, cost, effort and project artifacts. A new data mining technique approach Least square-support vector machine technique is used to improve the accuracy of SCE [3].It is a regression function to minimize the squared error. There are two datasets. Testing dataset, test a minor part (20%) of the project and training dataset is a major (80%) are compared with the past datasets of COCOMO-II. The correct result is implemented in Artificial Intelligence techniques. Expert driven or analytical methods are effective for the estimation. To improve the overall cost accuracy there are many methods.

Zeynab Abbasi Khalifelu et al [4] proposed an empirical role based on data mining techniques Linear Regression (LR), Artificial Neural Network (ANN), Squared Vector Regression (SVR)

and K-Nearest Neighbors (KNN) are compared with COCOMO model and is implemented using AI techniques to estimate the accurate cost of the software project. Testing and training datasets are used to compare the actual energy and estimated energy. The result is evaluated by Magnitude of relative error (MRE), coefficient correlation, mean absolute error; root mean squared error using NASA dataset and is implemented by WEAK data mining tools. The metrics to estimate a actual cost are SLOC, function point and effort multipliers. LR and KNN have less efficiency when compared with other methods. So ANN and SVR method is best to achieve high performance and more efficient method.

Shivangi Shekhar et al [5] say software cost estimation use two major software engineering techniques are algorithmic and non-algorithmic. Some of the algorithmic methods are COCOMO, Putnam model, function point. It is calculated by the mathematical formulas. The non-algorithmic methods are analogy method, Expert Judgment method, Prince to win estimation etc…It represents the historical data and should have the previous information of the project. A project can be selected in two ways. When the project is large and knowledge of the project is minimum it is better to choose algorithmic methods. A detailed knowledge of the project can choose non-algorithmic methods. For better estimation result we can compare the techniques to produce accurate software cost.

Alifia Puspaningrum et al [6] insisted the optimal value of software using hybrid Cuckoo search (CS) and Harmony search (HS) of machine learning method to estimate the effort and time. CS is an optimization algorithm, it can get most favorable solution but searching is done by infinite variance. So to overcome this problem HS is combined with CS by using NASA 93 dataset and it is measured by magnitude of relative error (MRE) and mean magnitude of relative error (MMRE). CS provides the optimal result by local and global search but HS uses the pitch adjustment to change the value of the variables in each direction to achieve the local solution and randomization is more efficient in global search. The different elements of HS and CS are compared with COCOMO-II coefficients.

Poonam Rijwani et al [7] developed multi layered feed forward neural network of Artificial neural network (ANN) technology of machine learning methods which are used to measure the skill of the project. MSE and MMER are the evaluation methods and to reduce the count of the network, a testing and training datasets, COCOMO datasets and mathematical tools are used to improve the accuracy of economic growth in data mining and machine learning algorithms. The scale factors, line of code and cost drivers are the parameters of COCOMO-II and inputs are compared with the hidden layer and effort is produced in the output layer.

Zdenka Prokopova et al anticipated the clustering methods to improve the software size. K-means, hierarchical and density-based clustering are used to compare the size with algorithmic optimization method (AOM) which is based on Use Case Point (UCP).UCP is a part of unified modeling language based on object oriented. AMO is on UCP and Multiple Least Square Regression [8]. The evaluation methods are gap, silhouette (graphical illustration) and calinski-harabasz (optimal cluster). The effect of clustering is compared with AOM algorithm with the measures to find the software size.

Monika et al [9] proposed a comparison of machine learning techniques like Artificial neural network (ANN), Fuzzy logic, Analogy estimation, Hybrid of ANN, Fuzzy, Genetic algorithm, Support vector machine (SVM) and Bayesian Network are used to estimate the effort. The validation methods are cross validation, iterative method, back propagation, square of correlation coefficient by using MSE, MMRE, MRE, MdMRE, Pred., RMS /RMSSE as a performance measures and the effect is compared on their functions, methods and procedures to reduce the size of the project.

Densumite et al developed an Activity Point (AP) method to estimate the size of a project using activity diagram which presents a chart demonstration of series actions and flow of control in a system [10]. The Technical Complexity Factor (TCF), Environment Complexity Factor (ECF) and People Risk Factor (PRF) are the adjustment factors used to predict the size of software. AP uses UCP in transactions, scenarios and relations to discover the effort of software.

3. Cost Estimation Technique

In software industry, successful estimation software is very critical activity and a challenging task. Cost estimation is a systematic approach that impress upon the cost model parameters on effort and schedule of the project.

A project management includes seven steps

1. Creating database of complete projects
2. Size estimation
3. Productivity estimation
4. Phase distribution
5. Risk and sensitivity analysis
6. Validation
7. Calibration

Many of the tools applying SCE models are lying in Calibration, Sensitivity and Risk analysis area. The attributes of the software is based on the productivity estimation and the efficiency required for the software development. The main activity for cost estimation is cost drivers.

In software industries the accurate Cost estimation is really important because to identify the matched resources with the requirements, to improve the business plan of software with better resources, to decide which resources are required and to produce, verify and validate the software products with the resources [11].

Cost estimation is to estimate the cost of software based on time, schedule and effort for low, medium and high projects. SCE is divided into 3 categories.

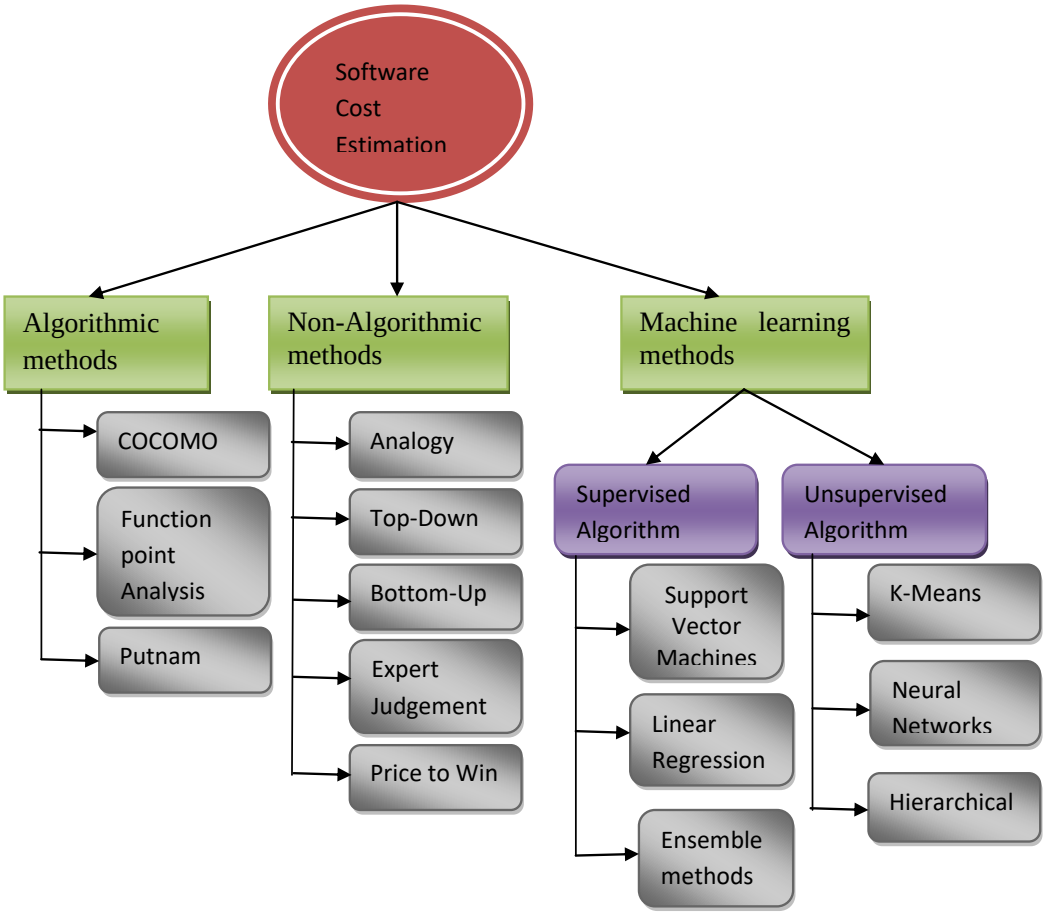

Figure 2: The techniques for SCE

3.1 Algorithmic methods

Algorithmic method uses the chronological cost information that reveals some software metric (lines of code) to project cost. Effort is predicted by the estimation of metrics. It uses the statistical equations to measure the cost.

The main advantages are creating repeatable estimations, modification is easy for input data, easy to process and modify formulas, and experience is just sequential. Incomparable conditions are unable to deal, some experience and factors cannot be quantified and algorithms can be recover are the cons of this model. Some Algorithmic methods are COCOMO method, Function point method, Putnam's method.

3.1.1 COCOMO

COCOMO – II is one of the contenders for the number of tools to utilize in SCE based on available literature. The original model was developed by Boehm in 1981. Basic COCOMO, Middle COCOMO, Detailed COCOMO is the levels of COCOMO. Organic, embedded and Semi-detached are the developed methods for COCOMO.

It expresses the size in thousands of SLOC KLOC). There are many cost drivers for COCOMO. A cost driver is a factor which changes in the cost of any activity. Some of the drivers are size of a database, software complexity, memory constrains, resources, tools, quality, constraints of a project etc…

As reveal [1] the input to the project is selection, follow-up and assessment for estimation. The regression-based estimation approaches include most conventional parametric estimation models like COCOMO based on 63 selected projects. Effort is a measure to indicate whole working time for software development. Man power is also an important factor in software cost estimation. It is easy to understand and more efficient because COCOMO-II is based on the man-hour or man-month to estimate the effort of the software.

3.1.2 Putnam's Model

The Macro estimation model estimates the time and effort established by Rayleigh or Norden function of large projects in software lifecycle management (SLIM) .The parameters are COCOMO81 and COCOMO II. The software research projects and human power are the basis for the Putnam's model.

3.1.3 Function Point analysis Method

The functionality of a project and size is estimated by the function point method. The Complexity is reduced and SLOC is used as a metric by Albrecht in 1983. Single unit cost is considered from the early phases [12]. The weight and complexity of software is measured by inputs-outputs of external, external inquiries, external interface and logical files. Function count (FC) is used to measure the size of the software by counting the functions of user's and adjusting the process complexity. The models for function point analysis are SPQR/20 and ESTIMACS.

3.2 Non-algorithmic method

Non-algorithmic method is used to estimate the project foregoing experiences and earlier project data. The integrated instructions of a previous software project can analysis by Expert judgement method, Analogy method, Price to win estimation, Top-down estimation and Bottom-up estimation.

3.2.1 Analogy Cost Estimation

It compares the similarity of previous and current project. The limited information of the current project is used to calculate the cost of the project. It is easy to implement because it is used in the

system or component level. The success rate is approximately 60%. The metrics are size, complexity and weight. The requirements are past data, approximate project cost, design and operations. The main pros of this method are to improve the accuracy of software by the experience of historical data. The steps for Analogy method are

a) Analogy Selection
b) The similarities and differences among the projects
c) Inspecting the quality of the analogy
d) Determining the estimation

3.2.2 Expert Judgement method

Human experts use their experience in the development of software and application domain to predict the cost and schedule of software. They use the knowledge of similar effort to understand the projects. They are educated guess about the energy to develop the software. Work breakdown structure (WBS) is and activity for the project works. The most common techniques for judgement method is WBS, Wideband Delphi technique, Three-point, PERT.

3.2.3 Top-Down Method

In top-down method the total cost of a project is determined from overall properties and it is splitted into various factors of the system. The detailed information is not available in the past stage because system-level function costs are not missed and it requires minimal project details. It is most constructive method and Putnam model is the foremost method.

3.2.4 Bottom-Up Method

For each individual component the cost is derived and combined to produce the global values of the project. This method uses small project s to estimate the cost. The estimated errors are balanced in various components of a project. COCOMO detailed model is the important model.

3.3 Machine learning methods

The machine learning is a data processing element to make predictions of cost. Training dataset is used to estimate the effort. It is broadly grouped as Supervised learning and Unsupervised learning.

The Supervised learning method is a fastest and accurate predictive method which has both input and output data of anticipating model. It is sub divided as Classification (Response value) and Regression (Statistical value). Some of the algorithms are Linear Regression, Support Vector Machine (SVM), Decision Trees, Nearest Neighbor, Naïve Bayes and Neural Networks.

The Unsupervised learning is an unlabeled data and it is pattern detection and descriptive modeling. It should discover the result. It is classified as Clustering (hidden pattern) and Association rule (discover the rule). The algorithms are K-means clustering, hierarchical and association rules.

3.3.1 Artificial Neural Network

ANN is a supervised learning models based on mathematical formulas, input, process and output for the cost estimation process. It is easy to modify the load of the software project. Neural Network and Hopfield network are the better method to estimate the cost. It has a reasoning power but it requires generous amount of training data and there is no result for designing [7]. A trial and error approach is used to find the best network architecture. A random selection is made to predict the effort with a minimum number of layers and nodes to build up the adaptability of the network.

3.3.2 Linear Regression

Linear Regression is used in predictive analysis and gives a statistical relationship between the dependent and independent variables. Some of the estimation techniques are Least-squares estimation, Bayesian linear regression, Quantile regression, mixed models etc…

3.4 Software Measures

Software measure gives a difference between the actual effort and the predicted effort. Some of the cost estimation measures are

- MRE (Magnitude Relative Error)
- MMRE (Mean Magnitude Relative Error)
- MdMRE (Medina Magnitude Relative Error)
- MMER (Mean Magnitude of Error Relative)
- MBER (Mean Balanced Error)
- Pred (Predicition performance)
- MAE (Mean -Absolute -Error)
- RMSE (Root- Mean -Square error)
- RAE (Relative-Absolute Error)

3.5 Software Metrics

Software metrics is a standard measure which is used to assess the growth of the software. Good metrics should contribute to the development models that are capable of predicting software product and processes. The metrics and models can be used to measure the productivity and product quality by estimating the product cost and schedule. The frequently used size and overall functionality are used to develop the software based on metrics. It is divided into three types.

1. Process Metrics (used for software development and maintenance)
2. Project Metrics (Used to estimate the cost and effort)
3. Product Metrics (Quality metrics, Architecture ,length metrics and testing)

Some of the SCE metrics are

- Source line of code
- Cyclomatic Complexity
- Code Coverge
- Function Point
- Effort Multipliers
- Scale Factor

3.6 Study of Software Cost Estimation

Technology	Method	Benefits	Limitations
Algorithmic Method	COCOMO	Estimating the cost is simple and gets the clear result	Details of past project are required
	Function point analysis	Tools, methods and language are independent to achieve the fast result	Time, Quality and manual work are not considered
	Putnam's Model	Time, Size and Effort are easily collected for past projects	It does not consider the other phases of SDLC
Non-Algorithmic Method	Analogy based Estimation	Experience and knowledge are used for actual projects	Attributes are required
	Expert Judgement	New technology, domain and architecture are the basis to estimate the cost	Experience of similar projects
	Top-Down	Requires minimum amount of details of the projects	Low level problems are difficult to identify
	Bottom-Up	Errors are estimated and very stable	Time and system-level activities are not considered

Machine learning methods	Linear Regression	Statistical models	Relationships between dependent and independent variables
	Support Vector Machine	Flexibility, Robustness, Unique solutions	Computation is expensive, binary classifier
	Neural Network	Powerful method, mathematical formula, ease to use	Large complexity of network structure
	K-Means	Fast Result, Easy to implement	Difficult to predict K-value, Global clusters
	Hierarchical cluster	Easy to decide the clusters	Time complexity, Not possible to undo the previous step

Conclusion

This paper presents the all essential synopsis required for software cost estimation, which defines and decides the sustainability of the software in the market. Conventional techniques available for SCE are algorithmic, non- algorithmic and machine learning approaches.

From various research works it is found that COCOMO dataset is used to compare the accuracy of the software cost. The attributes of the software depends on the correlation of proper knowledge and techniques, which ensures the reliability of software. Proper integration of techniques is the basic pre-requisite for the better quality and cost in data mining techniques. The relative study is based upon the effectiveness of the SCE methods.

References

[1] Qin, Xiaotie, and Miao Fang. "Summarization of Software Cost Estimation." Procedia Engineering 15 (2011): 3027-3031.

[2] Snehal A. Deshmukh, Prof.S.W. Ahmad. "Implementation of using classification data mining techniques for software cost estimation", (2016): 2321-8169.

[3] V.R.Arulmozhi, B.Vijaya Nirmala, N.Deepa . "A software cost estimation approach using Least Square-Support Vector Machine Technique", (2015): 2395-695X.

[4] Khalifelu, Zeynab Abbasi, and Farhad Soleimanian Gharehchopogh. "Comparison and evaluation of data mining techniques with algorithmic models in software cost estimation." Procedia Technology 1 (2012): 65-71.

[5] Shekhar, Shivangi, and Umesh Kumar. "Review of Various Software Cost Estimation Techniques." International Journal of Computer Applications 141.11 (2016): 00975-8887.

[6] Puspaningrum, Alifia, and Riyanarto Sarno. "A Hybrid Cuckoo Optimization and Harmony Search Algorithm for Software Cost Estimation." Procedia Computer Science 124 (2017): 461-469.

[7] Rijwani, Poonam, and Sonal Jain. "Enhanced software effort estimation using multi layered feed forward artificial neural network technique." Procedia Computer Science 89 (2016): 307-312.

[8] Prokopová, Zdenka, Radek Silhavy, and Petr Silhavy. "The effects of clustering to software size estimation for the use case points methods." Computer Science On-line Conference. Springer, Cham, 2017

[9] Sangwan, Om Prakash. "Software effort estimation using machine learning techniques." Cloud Computing, Data Science & Engineering-Confluence, 2017 7th International Conference on. IEEE, 2017.

[10] Densumite, S., and P. Muenchaisri. "Software Size Estimation UsingActivity Point." IOP Conference Series: Materials Science and Engineering. Vol. 185. No. 1. IOP Publishing, 2017.

[11] Iqbal, Syed Zaffar, et al. "Comparative Analysis of Common Software Cost Estimation Modelling Techniques." 2017.

[12] Nerkar, L. R., and P. M. Yawalkar. "Software Cost Estimation using Algorithmic Model and Non-Algorithmic Model a Review." IJCA Proceedings on Innovations and Trends in Computer and Communication Engineering 2 (2014): 4-7.

[13] Sharma, Narendra, Aman Bajpai, and Mr Ratnesh Litoriya. "A Comparison of software cost estimation methods: A Survey." The International Journal of Computer Science and Applications (TIJCSA) 1.3 (2012).

[14] Kobayashi, Toru. "Data Mining Technique Effectiveness for Improving Software Productivity and Quality." Computer Software and Applications Conference (COMPSAC), 2012 IEEE 36th Annual. IEEE, 2012.

[15] Snehal A. Deshmukh, Prof.S.W. Ahmad. "Using classification data mining techniques for software cost estimation", (2016): 2229 – 5518.