A NEW MODEL FOR SOFTWARE COSTESTIMATION USING HARMONY SEARCH

Sahar Sadouni

Department of Computer Engineering, Urmia Branch, Islamic Azad University, Urmia, Iran

ABSTRACT

Accurate and realistic estimation is always considered to be a great challenge in software industry. Software Cost Estimation (SCE) is the standard application used to manage software projects. Determining the amount of estimation in the initial stages of the project depends on planning other activities of the project. In fact, the estimation is confronted with a number of uncertainties and barriers', yet assessing the previous projects is essential to solve this problem. Several models have been developed for the analysis of software projects. But the classical reference method is the COCOMO model, there are other methods which are also applied such as Function Point (FP), Line of Code(LOC); meanwhile, the expert`s opinions matter in this regard. In recent years, the growth and the combination of meta-heuristic algorithms with high accuracy have brought about a great achievement in software engineering. Meta-heuristic algorithms which can analyze data from multiple dimensions and identify the optimum solution between them are analytical tools for the analysis of data. In this paper, we have used the Harmony Search (HS)algorithm for SCE. The proposed model which is a collection of 60 standard projects from Dataset NASA60 has been assessed.The experimental results show that HS algorithm is a good way for determining the weight similarity measures factors of software effort, and reducing the error of MRE.

KEYWORDS

Software Cost Estimation, COCOMO, Harmony Search.

1. INTRODUCTION

The delivery of software products with required budget and with the acceptable level of quality on time has always been one of the main concerns of Software Companies. Accurate and realistic estimation of the cost along with the attempt for software development and effective human resources, plus the allocation of resources and scheduling the projects have a very crucial role in the early stages [1]. Inaccurate estimation of attempt and cost can affect the software quality and the commercial reliability of software companies in a negative way. On the other hand, high estimation may result in the loss of customer and competitive market place when software projects trustee bids. Hence, improving the accuracy of estimation of the cost and attempt, especially in the early stages of projects is always believed to be an important goal in the software industry. Software companies use various techniques such as COCOMO [2,3] and FP [4] to achieve this goal. In many companies, COCOMO model is used for developing and manufacturing, evaluating and completing the software besides the adaptability of SCE with the actual amount. For SCE, many intelligent models have also been established [5,6].Expert

opinion, reasoning and similarity are the most popular ones among the other estimation procedures. Estimation based on similarity means comparing software projects intended for estimation with the previous ones. In other words, this procedure is systemic mode of the expert opinion, since the method of expert opinion requires that the expert match the previous similar projects with each other and finally does the estimation of attempts of the projects. Attempt estimation model based on similarity and reasoning is known as the identification of the previous projects which are similar to the current developing ones.

SCE deals with the preparation of budget, manpower supply, software support and maintenance, ordering the process of supporting, plus preserving software. Preliminary tasks are important for efficiency of the software development, support and implementation with low and reasonable prices.Most of the commercial projects which have been used by traditional cost estimation method are faced with the following problems [7, 8].

- Collecting software data for measuring LOC can be complicated.
- Measuring instructions and tools may not be stable.
- Ensuring the quality and timeliness of measurement data can be.
- Lack of communication and inconsistencies between project staff can reduce the efficiency.

One of the most important and most complicated tasks in the software development process is SCE. Over the past three decades, a growing trend in the use of various models of SCE has been observed in various software development processes. Along with this tremendous growth, we have noticed the need for different SCE models and preparing the schedule more rapidly. Plenty of the research time and funds has been dedicated to improve the accuracy and the exactness of estimation models. Due to the inherent uncertainty in software projects, such as the inherent complexity of software, the pressure on the standardization and the lack of software information, extremely accurate estimation of software development attempt is thought to be unrealistic. Since the accuracy and reliability of estimation attempt is very important for the competitiveness of software, companies and researchers try their best to develop accurate modals in order to estimate near-surface attempts. The specifications in software development and production projects include [9, 10, and 11]:

- Parameters: size, personnel, complexity, database, platform and application.
- Prediction: effort, time, manpower and schedule.

The SCE of a process or an approximation of the potential costs of a product or project is available based on the information. Exact cost estimation for any type of project is very important; if we do not estimate the cost of software projects properly the cost of projects will be about 25% or 75% more than the original price, subsequently [11].In that case it will be necessary to estimate the project correctly. The main cost of many models is measured using LOC and FP.Software cost estimation is an essential component of software development process. Accurate assessment of development cost is important for guaranteeing the success of the project. However, there are many inadequacies in current assessment methods. Therefore, in this paper, we will analyze the SCE using the COCOMO model and HS [12] algorithm.

The overall structure of this paper is organized as follows: In Section 2, we will discuss the work done in the field of SCE; in Section 3, describes HS algorithm; in Section 4, we will explicate the

proposed model; Section 5, we will evaluate and the results of the proposed model explain, and finally in Section 6, we will illustrate conclusion and future work.

2. RELATED WORKS

Reviewing most of the SCE models, we will conclude that Artificial Intelligence (AI) models have found a great place in software testing and evaluation, and are used extensively at various stages of application such as the estimation of attempt and cost. In this section, we will analyze the evaluated SCE models.

Hybrid PSO-FCM and PSO-LA models have been proposed for the SCE [13]. Evaluation was performed on Dataset NASA60.In the hybrid model of PSO-FCM, minimum of inter-cluster distance and the total of intra-cluster distance and the number of clusters have been used as fitness and algorithm-efficiency parameters. Using FCM (Fuzzy C-Means) makes the particles gather in the best cluster and causes the fitness function to have many local optimum points. In order to improve the performance of the PSO algorithm, LA (Learning Automata) is used to adjust the behavior of particles. In the hybrid PSO-LA model, all particles simultaneously search for a place in search space. In the hybrid PSO-LA model, LA models strategy according to reward criterion for algorithm PSO provides the opportunity for particles to achieve local optimums. Their experimental results show that the hybrid PSO-FCM model has less MRE error rate than the PSO-LA hybrid models. MMRE error in the PSO-FCM model is 25.36, 24.56, 24.22 and 23.86, and in the PSO-LA model is 26.32.The accuracy of PRED (25) in COCOMO Model is 40, and is 61.6 ،58.3 و68.3 in PSO-FCM model. In the PSO-LA model, it is 63.3.

Using PSO algorithm, a new estimation attempt model for software projects has been proposed [14].Parameters which are effective for attempt estimation have been analyzed using PSO algorithm. Assessing KEMERER dataset has been conducted using 15projects. Test results have shown that the MMRE error rate in proposed model is 56.57, and in COCOMO model, it is 245.39.

RBF-GA hybrid model is proposed for the SCE [15]. Evaluation was performed on Dataset COCOMO81.Training and testing data of the network is respectively considered to be 80% and 20%. GA algorithm has been used for the efficiency of training and testing data of RBF network. The results show that MdMRE, MMRE and MSE errors in COCOMO model is respectively 113.2, 2.92 and 0.0287. In RBF model, MdMRE is respectively 0.0077 and 0.0137 in training and test steps. MMRE error is respectively 0.4220 and 0.9665. . In the RBF-GA hybrid model, MdMRE is respectively 0.0125 and 0.0044 over the training and test steps. MMRE error is respectively 0.1636 and 0.5485. In the hybrid model the criteria of MMRE and MdMRE are more accurate in training and test steps. Comparisons between COCOMO and RBF-GA models have shown that the error is less in hybrid model.

ACO and GA hybrid model have been proposed for SCE based on training and testing the software projects[16]. Evaluating the hybrid model has been done on the Dataset NASA60. Training the data has been conducted using ACO, and testing the data using GA. The results with 10 projects show that in the hybrid model, the MRE error is about 0.9 of the projects less in comparison to COCOMO model. The MMRE error for 60 projects in the hybrid model is 27.53 and 29.64 in the COCOMO model. Hybrid model has reduced about 1.07 of the amount of MMRE error.

Case-Based Reasoning (CBR), Artificial Neural Networks (ANN), Decision Trees (DT), Bayesian Networks (BN), Support Vector Regression (SVR), Genetic Algorithms (GA), Genetic Programming (GP), Association Rules (AR) models, which are the techniques of machine learning have been proposed for the SCE [17]. Evaluation has been performed on Desharnais, COCOMO, ISBSG, Albrecht, Kemerer, NASA, Tukutuku datasets. The results on datasets show that MMRE error in SVR is lower compared to other models. The amount ofDT error is less than the BN, and ANN is less than CBR and DT models. The error value of BN model is less than GP mode. The accuracy of PRED (25) and DT SVR model is better than the other models. SVR model is more accurate than GP model. CBR model has better accuracy than the ANN model. The BN model is more accurate than the DT model. MdMRE in the ANN and CBR models is less compared to other models. The results show that CBR and BN models perform better on small projects.

The techniques of Liner Regression (LR), ANN, SVR, and K Nearest Neighbors (KNN) are used for SCE [18]. LR model can be used to determine the dependence of the effective traits in SCE.LR model finds the relationship between the independent and dependent factors in the data. ANN tries to reduce the amount of MRE using training and testing the data. SVR model is used for the optimization of the factors influential in SCE. KNN is a technique for data-analyzing which is useful for classifying data in a collection of previously specified applications whose properties have been identified. Using KNN, the weight of effective characteristics in SCE is determined. Prediction accuracy on the training data in LR, ANN, SVR and KNN models is respectively 74%, 87%, 95% and 68%. Also, prediction accuracy on the tested data in LR, ANN, SVR and KNN models is respectively 60%, 95%, 80% and 60%. MAE in the LR, ANN, SVR and KNN models is respectively 39.17%, 12.07%, 36.32% and 1.12%. The amount of error in the RRSE LR, ANN, SVR and KNN models is respectively 11.6%, 2.37%, 8.20% and 0.77%. Their results show that KNN model has less amount of MAE than the other models.

ANN-MLP model is of the conventional methods in the SCE [19]. In order to demonstrate the performance of ANN, from 63 projects in NASA software dataset that have been tested and trained using ANN, 11 projects compared with COCOMO model they have been shown that the MRE value in COCOMO model is more than ANN model. 80% of projects have been used for training and 20% for testing. The results show in 90% of the projects, ANN model is presented the best estimation than COCOMO model.

3. META-HEURISTIC ALGORITHMS

Among the models based on AI, the ability of meta-heuristic algorithms has been demonstrated very efficient in solving optimization problems [20]. Meantime, there are no complicated relationships between input and output parameters of meta-heuristic algorithms. Meta-heuristic algorithm is based on the initial population, whose function is dependent on the stages of algorithm and its parameter settings.

3.1. Harmony Search

Harmony search algorithm which is one of the heuristic algorithms is musically inspired and is based on possible search [12]. Harmony search algorithm is based on the process of music with generation of solution vectors in terms of memory space begin to search problem-solving space and moves to optimal space on the basis of probabilistic approach.

In the HS algorithm, the Harmony Memory (HM) can be used to preserve the best of former solutions. HM is implemented according to the Eq. (1) in the form of a matrix each row of whose is a solution. Therefore, the number of the columns of this matrix shows the dimensions of the solution. The last column of the matrix is used to store the amount of the fitness function in each row. The number of response vectors in HM is indicated through Harmony Memory Size (HMS).The value of fitness function of any vector is expressed through f(x).

$$HM = \left[\begin{array}{cccc|c} x_1^1 & x_2^1 & \dots & x_n^1 & f(x^1) \\ x_1^2 & x_2^2 & \dots & x_n^2 & f(x^2) \\ \dots & & & & \dots \\ x_1^{HMS} & x_2^{HMS} & \dots & x_n^{HMS} & f(x^{HMS}) \end{array}\right] \tag{1}$$

The way by which HS algorithm works is that first the fitness function and the appropriate values are set for Harmony Memory Considering Rate (HMCR), HMS and Pitch Adjustment Rate (PAR) parameters. Then all the rows of HM are randomly generated, after that the fitness function value for each row is calculated and stored in the last column of the HM matrix. Then, considering the number of iterations needed or until the occurrence of a certain condition, all HM is scanned for each row and according to the parameters of the HS algorithm, the suitable value for each element is set. Afterwards, if the value of the fitness function created is better than the value of the fitness function which is the worst solution in HM, it will be replaced. Finally, the solution whose fitness function value is better will be known as the best solution available. By the way, the possibility of the use of the variables available in HM will be clarified by the help of HMCR. The possibility of changing the variables of HMCR is decided through PAR. The numerical of the parameters of HMCR and PAR are between 0, 1. Number Iteration (NI) in HS is expressed by the NIparameter.In the HS algorithm, one variable can randomly take values. Randomization is used to increase the diversity of solutions. Although, PAR plays a similar role this task makes algorithm search locally, and causes it to be mired down in local efficiencies. Using randomization leads algorithm to the global search and to find a global efficient answer.

4. PROPOSED MODEL

While SCE, the factors of software projects affecting the cost of software must be identified. The values of these factors are different among the projects. In the past, SCE was used widely with the algorithm models. The disadvantages of algorithmic models include weak level of accuracy in estimating the cost and inefficiency in measuring factors and changes in properties. One of the major disadvantages of COCOMO model is its need for certain mathematical relationships for the cost function. So, in this paper, we have used the HA algorithm which is one of meta-heuristic algorithms for optimality and finding the relationships of factors of NASA60 Dataset that are effective in determining the cost and attempt [21]. A flowchart of the proposed model is shown in Figure (1).

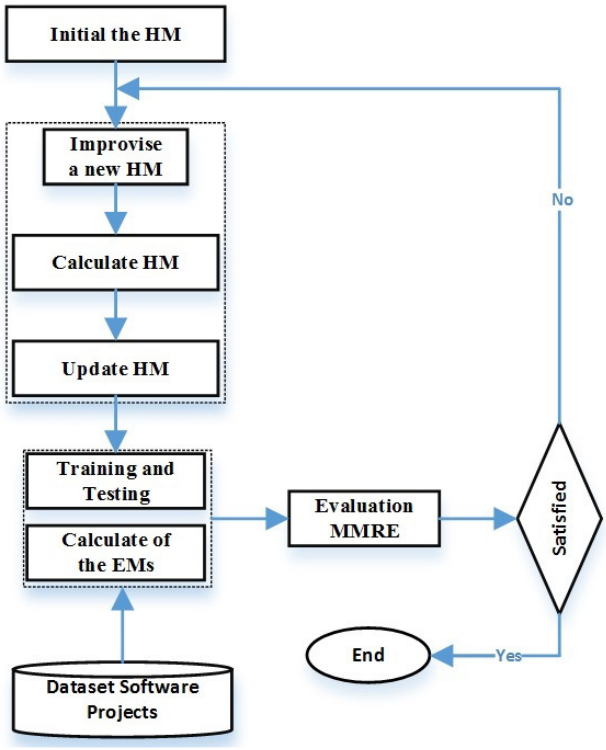

Figure 1: Flowchart of the Proposed Model

In the proposed model, the initial results are stored in a virtual memory pick, then using the approach of neighborhood the search for the optimal solution which has the condition of acceptability will be placed near the primary responses. This process continues until the algorithm reaches the final condition. Searching approach is used based on PAR in a way that new values are randomly generated and are replaced with the worst members in the index of the array. Searching approach designed from intelligent search method which is based on avoiding a return to the answers found in the former repetitions works using a virtual directory of memory called HM.The HM memory used depends on the amount of search and prevents the repetition of search over the course of repetition. To achieve the desired responses from the points of quality and optimality, we have used the elitist approach of search, in which only the rows of the matrix which are more suitable for testing and training EM factors are selected. In the proposed model, MMRE has been considered as the fitness function. The objective of the Fitness Function is to minimize the amount of MMRE. In the proposed model, the value of MMRE is repeated until reaching the desired amount. Fitness Function for the proposed model is defined in the form of the Eq. (3) [22].

$$MRE_i = \frac{|act_i - est_i|}{act_i} \times 100 \tag{2}$$

$$MMRE = \frac{1}{n}\sum_{i=1}^{n} MRE_i, i = 1,2,...,n \tag{3}$$

Using Eq. (3), we can compare the total error of the estimation made. Evaluating the estimation criterion, the model with less MRE is better than the one with higher MRE;the model with less MMRE is better than the one with higher MMRE.

5. EVALUATION AND RESULTS

Assessment and results of the models have been carried out on 60 projects of NASA60 dataset projects. Simulation of the proposed model has been performed using VC#.NET 2013 programming. In the proposed model, factors such as the initial values for parameters and fitness function are taken into consideration because the impact of these factors is the key to achieve the convergence and optimal solutions. In the proposed model, the values of the HMS, HMCR and PAR parameters are considered to be respectively 50, 0.6 and 0.4 and the NI is thought to be 20. In Figure (2), the MRE diagram of the proposed model is compared with COCOMO model.

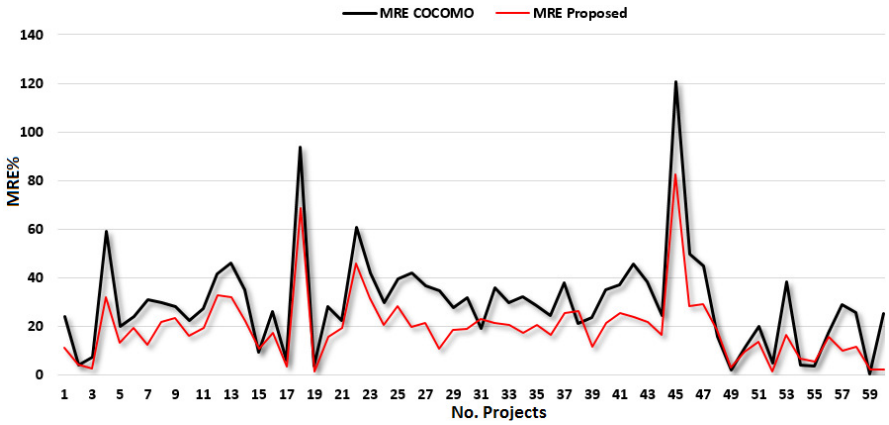

Figure 2: Diagram of Comparing MRE of the Proposed Model with COCOMO Model

Results of Figure (2) shows that the proposed model has better performance compared with the COCOMO model and that in the majority of projects, ithas lower amount of MRE.The MMRE error in the hybrid model is19.38; it is 29.64 in the COCOMO model. The parameter of MMRE in the proposed model is less than in the COCOMO models, by the way; it has managed to reduce the amount of MMRE to about 1.52. In Figure (3), the MMRE diagrams of the models have been compared.

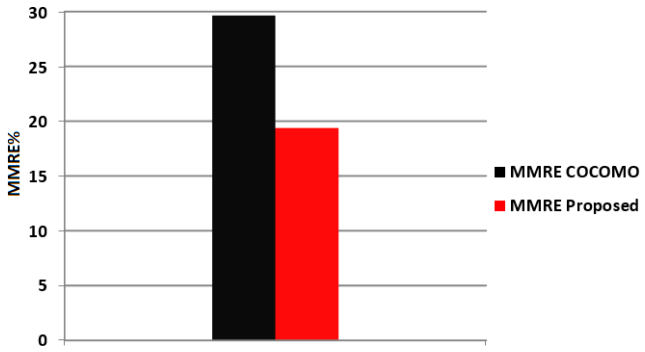

Figure 3: Comparing MRE of the Proposed Model with COCOMO Model

Figure (4) indicates the MRE diagram of the proposed model compared with COCOMO model on the basis of the number of iterations. As observed, MRE value reduces as the number of repetitions increases; this is the result of training and testing most of the data.

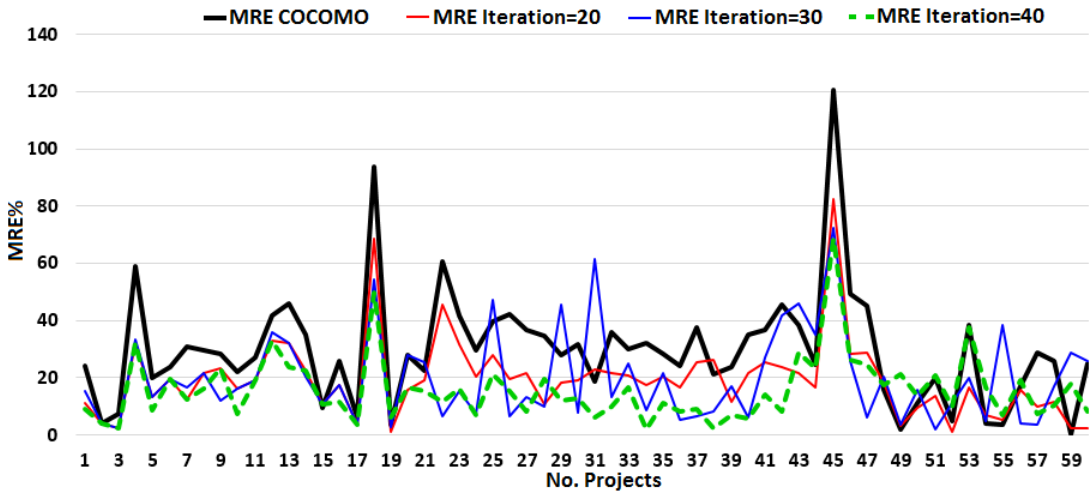

Figure 4: Comparing MRE of the Proposed Model with COCOMO Model Based on Iteration

6. CONCLUSION AND FUTURE WORKS

With regard to financial resources for the strategies of the managers and participating in tenders, SCE in the initial stages of project is a key factor in determining the software cost and attempt. Accurate estimation is one of the important factors in the ultimate success of the project while being impressive also in its various stages.SCE is widely done with Algorithmic models such as COCOMO with low accuracy. In this paper, we have used the HA algorithm to estimate in a more accurate way and to reduce the errors of MRE and MMRE criteria. The results show that the proposed model has lower amount of MMRE in comparison to COCOMO model, and has reduced 1.52 times of that. Through this study, we hope that we will be able to offer more accurate estimations using a combination of AI techniques and COCOMO models.

REFERENCES

[1] I. Maleki, L. Ebrahimi, F.S. Gharehchopogh, "A Hybrid Approach of Firefly and Genetic Algorithms in Software Cost Estimation", MAGNT Research Report, 2(6): 372-388, 2014.

[2] B.W. Boehm, "Software Engineering Economics", Prentice-Hall, Englewood Cliffs, New Jersy, 1981.

[3] B.W. Boehm, "Software Cost Estimation with COCOMO II", Prentice Hall PTR, Englewood Cliffs, New Jersy, 2000.

[4] A.J. Albrecht, J. Gaffney, "Software Function, Source Lines of Code, and Development Effort Prediction: a software science validation", IEEE Transactions on Software Engineering SE, 9(6), pp. 639-648, 1983.

[5] F.S. Gharehchopogh, I. Maleki, A. Kamalinia, H. Mohammad Zadeh, Artificial bee colony based constructive cost model for software cost estimation, Journal of Scientific Research and Development 1 (2): 44-51, 2014

[6] Z.A. KHALIFELU, F.S. GHAREHCHOPOGH, "A Survey of Data Mining Techniques in Software Cost Estimation", AWERProcedia Information Technology & Computer Science Journal, Vol. 1, pp. 331-342, 2012.

[7] P.C. Pendharkar, "Probabilistic Estimation of Software Size and Effort", Expert Systems with Applications, Vol. 37, Issue 6, pp. 4435-4440, 2010.
[8] Gharehchopogh F.S., Maleki I., Sadouni S., Artificial Neural Networks Based Analysis of Software Cost Estimation Models, MAGNT Research Report, 2 (6): 597-505, 2014.
[9] Y. Lotfi, S. Govahi, S. Sadouni, I. Maleki, Hybrid Genetic Algorithm with Linear Regression and Nonlinear Regression for Software Cost Estimation, MAGNT Research Report, 2 (6): 784-794, 2014.
[10] I. Maleki, L. Ebrahimi, S. Jodati, I. Ramesh, Analysis of Software Cost Estimation Using Fuzzy Logic, International Journal in Foundations of Computer Science & Technology (IJFCST), 4(3), 27-41, 2014.
[11] F.S.Gharehchopogh, A. Talebi, I. Maleki, Analysis of Use Case Points Models for Software Cost Estimation, International Journal of Academic Research, Part A, 6(3), 118-124, 2014.
[12] Z.W. Geem, J.H. Kim, G.V. Loganathan, A New Heuristic Optimization Algorithm: Harmony Search, Simulation, 76(2), 60-68, 2001.
[13] F.S. Gharehchopogh, L. Ebrahimi, I. Maleki, S.J. Gourabi, A Novel PSO based Approach with Hybrid of Fuzzy CMeans and Learning Automata in Software Cost Estimation, Indian Journal of Science and Technology, 7(6), 795-803, 2014
[14] F.S. Gharehchopogh, I. Maleki, S.R. Khaze, A Novel Particle Swarm Optimization Approach for Software Effort Estimation, International Journal of Academic Research, Part A, 6(2), 69-76, 2014.
[15] M. Molni, A. Ghaffari, A. Jafarian, A New Approach to Software Project Cost Estimation using a Hybrid Model of Radial Basis Function Neural Network and Genetic Algorithm, Indian Journal of Science and Technology, 7(6), 833-838, 2014.
[16] I. Maleki , A. Ghaffari, M. Masdari, A New Approach for Software Cost Estimation with Hybrid Genetic Algorithm and Ant Colony Optimization, International Journal of Innovation and Applied Studies, 5(1), 72-81, 2014.
[17] J. Wena, Sh. Li, Z. Lin, Y. Hu, C. Huang, Systematic Literature Review of Machine Learning based Software Development Effort Estimation Models, Information and Software Technology, Vol. 54, 41-59, 2012.
[18] Z.A. Khalifelu, F.S. Gharehchopogh, Comparison and Evaluation Data Mining Techniques with Algorithmic Models in Software Cost Estimation, Elsevier, Procedia-Technology Journal, ISSN: 2212-0173, Vol. 1, pp. 65-71, 2012.
[19] F.S. Gharehchopogh, Neural Networks Application in Software Cost Estimation: A Case Study, 2011 International Symposium on Innovations in Intelligent Systems and Applications (INISTA 2011), pp. 69-73, IEEE, Istanbul, Turkey, 15-18 June 2011.
[20] F.S. Gharehchopogh, I. Maleki, M. Farahmandian, New Approach for Solving Dynamic Traveling Salesman Problem with Hybrid Genetic Algorithms and Ant Colony Optimization, International Journal of Computer Applications (IJCA), Vol. 53, No.1, pp. 39-44, September 2012.
[21] T. Menzies, D. Port, Zh. Chen, J. Hihn, Validation Methods for Calibrating Software Effort Models, ICSE ACM, USA, 2005.
[22] S.G. MacDonell, A.R. Gray, A Comparison of Modeling Techniques for Software Development Effort Prediction, in Proceedings of International Conference on Neural Information Processing and Intelligent Information Systems, pp. 869-872, 1997.