International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:2, No:10, 2008

Software Effort Estimation Using Soft Computing Techniques

Parvinder S. Sandhu, Porush Bassi, and Amanpreet Singh Brar

***Abstract*—**Various models have been derived by studying large number of completed software projects from various organizations and applications to explore how project sizes mapped into project effort. But, still there is a need to prediction accuracy of the models. As Neuro-fuzzy based system is able to approximate the non-linear function with more precision. So, Neuro-Fuzzy system is used as a soft computing approach to generate model by formulating the relationship based on its training. In this paper, Neuro-Fuzzy technique is used for software estimation modeling of on NASA software project data and performance of the developed models are compared with the Halstead, Walston-Felix, Bailey-Basili and Doty Models mentioned in the literature.

***Keywords*—**Effort Estimation, Neural-Fuzzy Model, Halstead Model, Walston-Felix Model, Bailey-Basili Model, Doty Model.

I. Introduction

ACCURATE software cost estimates are critical to both developers and customers. Underestimating the costs may result in management approving proposed systems which can exceed their budgets, with underdeveloped functions and poor quality, and failure to complete on time. Overestimating may result in too many resources committed to the project, or, during contract bidding, result in not winning the contract, which can lead to loss of jobs. So accurate cost estimation is important and Software cost estimation involves the determination of effort (usually in person-months), project duration (in calendar time) and cost (in dollars). Most cost estimation models attempt to generate an effort estimate, which can then be converted into the project duration and cost. In the last three decades, many quantitative software cost estimation models have been developed. Most cost models are based on the size measure, such as Line of Code (LOC) and Function Point (FP), obtained from size estimation. The accuracy of size estimation directly impacts the accuracy of cost estimation.

A review of the literature revealed that there are two major types of cost estimation methods Algorithmic and Non-algorithmic models as discussed in [3, 4, 5, 6, 7, 9, 11, 12, 13].

The remainder of this paper can be described as follows: The next section contains a description of the methodology used and Sections III discusses the implementation results of proposed and existing models. At last, conclusions are drawn.

II. Methodology Used

The following steps of the methodology are proposed for modeling of effort estimation:

A. Data Collection

First, Survey of the existing Models of Effort Estimation is to be performed and Secondly, Historical Data being used by various existing models for the cost estimation is collected.

B. Neuro Fuzzy Modeling

Neuro Fuzzy computing is a popular framework for solving complex problems [2]. If one has knowledge expressed in linguistic rules, one can build a Fuzzy Inference System (FIS), and if one has data, or can learn from a simulation (training) then one can use Artificial Neural Networks (ANNs). An analysis reveals that the drawbacks pertaining to these approaches seem complementary and therefore it is natural to consider building an integrated system combining the concepts. While the learning capability is an advantage from the viewpoint of FIS, the formation of linguistic rule base will be advantage from the viewpoint of ANN.

In the simplest way, a cooperative model can be considered as a preprocessor wherein ANN learning mechanism determines the FIS membership functions or fuzzy rules from the training data. Once the FIS parameters are determined, ANN goes to the background. The rule based is usually determined by a clustering approach or fuzzy clustering algorithms. Membership Functions are usually approximated by neural network from the training data [10].

In a concurrent model, ANN assists the FIS continuously to determine the required parameters especially if the input variables of the controller cannot be measured directly. In some cases the FIS outputs might not be directly applicable to the process. In that case ANN can act as a postprocessor of FIS outputs [2]. On wide categorization there are two basic types of Neuro-Fuzzy systems: Mamdani Neuro-Fuzzy System and Tagaki-Sugeno Neuro-Fuzzy System.

The Sugeno based NF Model has the following advantages over Mamdani Model:

a. Expressing a clearer concept of inference process.

Parvinder S. Sandhu, PhD, is working as Professor with CSE & IT Department, Rayat & Bahra Institute of Engg. & Bio-Technology, Sahauran, India (parvinder.sandhu@gmail.com).

Amanpreet Singh Brar is Asstt. Professor & Head (Computer Science & Engineering Department), Guru Nanak Dev Engg. College, Ludhiana (Punjab), India.

Porush Bassi is doing his Masters from Computer Science & Engineering Department, Guru Nanak Dev Engg. College, Ludhiana (Punjab)-India.

b. Less number of fuzzy rules.
c. Faster learning.
d. Less memory space.

So, we have only tried a Takagi-Sugeno Neuro-Fuzzy systems. In this Neuro-fuzzy system the first Sugeno Based Fuzzy Inference System is designed that needs the initialization of the Membership Function of the different attributes and linear Membership Function for the output and deducing the fuzzy rules from the data. Then the Sugeno Based FIS is trained with the neural Network using the hybrid training algorithm. In the forward pass, Backpropagation learning algorithm and in the backward pass Least Mean Square Error (LMS) learning algorithm is used to update the non-linear and linear parameters of the Neuro-fuzzy system respectively.

The detailed functioning of each layer of Takagi Sugeno Neuro-Fuzzy system [1] is as follows:

Layer-1 (Input Layer): No computation is done in this layer. Each node in this layer, which corresponds to one input variable, only transmits input values to the next layer directly. The link weight in layer 1 is unity.

Layer-2 (Fuzzification Layer): Each node in this layer corresponds to one linguistic label (excellent, good, etc.) to one of the input variables in layer 1. In other words, the output link represents the membership value, which specifies the degree to which an input value belongs to a fuzzy set, is calculated in layer 2. A clustering algorithm will decide the initial number and type of membership functions to be allocated to each of the input variable. The final shapes of the MFs will be fine tuned during network learning.

Layer-3 (Rule Antecedent Layer): A node in this layer represents the antecedent part of a rule. Usually a T-norm operator is used in this node. The output of a layer 3 node represents strength of the corresponding fuzzy rule.

Layer-4 (Rule strength normalization): Every node in this layer calculates the ratio of the ith rule's firing strength to the sum of all rules firing strength.

$$\bar{\omega}_i = \frac{\omega_i}{\omega_1 + \omega_2}, i = 1, 2... \tag{1}$$

Layer-5 (Rule consequent layer): Every node i in this layer is with a node function:

$$\bar{\omega}_i f_i = \bar{\omega}_i (p_i x_i + q_i x_2 + r_i) \tag{2}$$

Where, w_i is the output of layer 4, and $\{f_i; q_i; r_i\}$ is the parameter set. A well established way is to determine the consequent parameters using the least means squares algorithm.

Layer-6 (Rule inference layer): The single node in this layer computes the overall output as the summation of all incoming signals Overall output =X_i

$$Overall\ output = \sum_i \bar{\omega}_i f_i = \frac{\sum_i \bar{\omega}_i f_i}{\sum_i w_i} \tag{3}$$

C. Comparison with Existing Models

The following modeling approaches are used for comparison:

- Mamdani Fuzzy Inference System with 4 clusters
- Mamdani Fuzzy Inference System with 15 clusters
- Sugeno Based Neuro-Fuzzy Model
- Halstead Model
- Walston-Felix Model
- Bailey-Basili Model
- DotyModel
- Genetic Algorithm Based Model

The comparison of the results is made on the basis of the following factors:

RMSE is frequently used measure of differences between values predicted by a model or estimator and the values actually observed from the thing being modeled or estimated. It is just the square root of the mean square error as shown in equation given below:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{3}$$

Where y_i represents the i^{th} value of the effort and $\hat{y}_i$ is the estimated effort.

MMRE is another measure and is the percentage of the absolute values of the relative errors, averaged over the *N* items in the "Test" set and can be written as:

$$MMRE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \Big/ y_i \tag{4}$$

PRED(N) is the third criteria used for the comparison and this reports the average percentage of estimates that were within N% of the actual values [3].

D. Conclusion

In the last step on the basis of the errors in calculated effort conclusions are made.

III. Results & Discussion

The dataset of NASA [8] is used for the comparison of different models. In this dataset, there is empirical data in terms of *DKLOC*, *Methodology* and *Effort* values of 18 projects as shown in Table I.

The Sugeno based Fuzzy Inference system is developed and in order to train the Sugeno FIS, Adaptive Neuro-Fuzzy system (ANFIS) is designed that makes use of the Sugeno FIS Structure as shown in Fig. 1. The following the structure parameters of the Neuro-fuzzy system:

- Number of nodes: 21
- Number of linear parameters: 12
- Number of nonlinear parameters: 12
- Total number of parameters: 24
- Number of training data pairs: 18
- Number of checking data pairs: 0
- Number of fuzzy rules: 4

TABLE I
NASA DATA [8] OF EFFORT ESTIMATION

Project No.	DKLOC	Methodology	Actual Effort
1	90.2	30	115.8
2	46.2	20	96
3	46.5	19	79
4	54.5	20	90.8
5	31.1	35	39.6
6	67.5	29	98.4
7	12.8	26	18.9
8	10.5	34	10.3
9	21.5	31	28.5
10	3.1	26	7
11	4.2	19	9
12	7.8	31	7.3
13	2.1	28	5
14	5	29	8.4
15	78.6	35	98.7
16	9.7	27	15.6
17	12.5	27	23.9
18	100.8	34	138.3

The NF system is trained for 500 epoch and tested. The plot of data index v/s expected output and actual output is shown in Fig. 2.

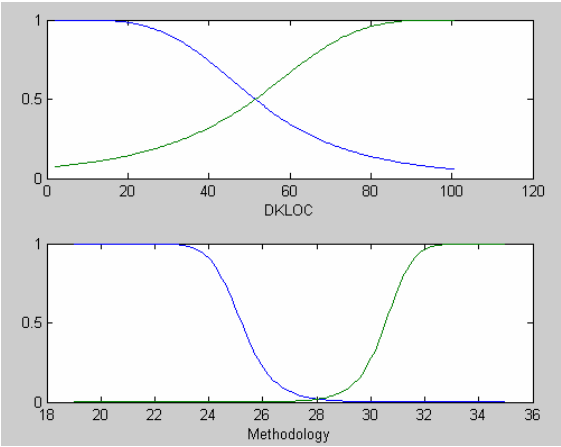

Fig. 2 Changed Parameters of Membership Functions of the input attributes

In the following Tables II and III, the actual measured effort over the given 13 projects is shown, in the training case, and the 5 projects in the testing (i.e. validating) case for all models. It is tried to evaluate the performance of various models on basis of *RMSSE* and *MMRE*. This helps in comparing the results for each developed model. In Table IV *RMSSE* and *MMRE* criteria is computed over the complete data set. It can be seen that the Neuro-Fuzzy model outperform the Halstead, Walston-Felix, Bailey-Basili and Doty models.

IV. CONCLUSION

The performance of the Neuro-fuzzy based effort estimation Model and the other existing Halstead Model, Walston-Felix Model, Bailey-Basili Model and Doty Model

TABLE II
COMPUTED EFFORT FOR NASA SOFTWARE PROJECTS-TRAINING CASE

Project No.	Actual Effort	Neuro-Fuzzy Model	Halstead Model	Walston-Felix Model	Bailey-Basili Model	Doty Model
1	115.8	115.7	599.66	312.78	140.82	589.38
2	96	96.069	219.82	170.15	67.774	292.53
3	79	78.903	221.96	171.15	68.243	294.52
4	90.8	90.813	281.64	197.75	80.93	347.78
5	39.6	38.875	121.41	118.69	44.848	193.29
6	98.4	98.447	388.2	240.25	102.18	435.08
7	18.9	20.789	32.056	52.913	19.55	76.303
8	10.3	9.9815	23.817	44.186	16.666	62.012
9	28.5	29.343	69.784	84.825	31.142	131.33
10	7	7.3786	3.8207	14.559	8.2121	17.288
11	9	8.6934	6.0252	19.194	9.3574	23.759
12	7.3	8.9732	15.249	33.714	13.41	45.427
13	5	4.6705	2.1302	10.215	7.2262	11.499

TABLE III
COMPUTED EFFORT FOR NASA SOFTWARE PROJECTS-TESTING CASE

Project No.	Actual Effort	Neuro-Fuzzy Model	Halstead Model	Walston-Felix Model	Bailey-Basili Model	Doty Model
14	8.4	10.129	7.8262	22.494	10.222	28.518
15	98.7	113.81	487.79	275.95	120.85	510.27
16	15.6	18.126	21.147	41.112	15.685	57.074
17	23.9	22.647	30.936	51.783	19.169	74.431
18	138.3	141.6	708.42	346.06	159.43	662.09

TABLE IV
COMPUTED RMSSE AND MMRE CRITERION FOR ALL MODELS

Performance Criteria	Model Used				
	Neuro-Fuzzy Model	Halstead Model	Walston-Felix Model	Bailey-Basili Model	Doty Model
RMSSE	7.0731	308.71	123.46	13.877	299.47
MMRE	0.11943	1.7566	1.5556	0.1595	3.025

models is compared for effort dataset available in literature [8]. The results show that the Neuro-fuzzy system has the lowest *MMRE* and *RMSSE* values i.e. 0.11943 and 7.0731 respectively. The second best performance is shown by Bailey-Basili software estimation system with 0.1595 and 13.877 as *MMRE* and *RMSSE* values. Hence, the proposed Neuro-fuzzy based system can be used for the software effort estimation of all types of the projects.

REFERENCES

[1] A. Abraham, Adaptation of Fuzzy Inference System Using Neural Learning, Springer Berlin, ISSN: 1434-9922 (Print) 1860-0808 (Online), vol. 181, 2005.
[2] A. Abraham and M.R. Khan, Neuro-Fuzzy Paradigms for Intelligent Energy Management, Innovations in Intelligent Systems: Design,

Management and Applications, Studies in Fuzziness and Soft Computing, Springer Verlag Germany, Chapter 12, pp. 285-314, 2003.
[3] B. W. Boehm, Software engineering economics, Englewood Cliffs, NJ: Prentice-Hall, 1981.
[4] C. E. Walston, C. P. Felix, A method of programming measurement and estimation, IBM Systems Journal, vol. 16, no. 1, pp. 54-73, 1977.
[5] G.N. Parkinson, Parkinson's Law and Other Studies in Administration, Houghton-Miffin, Boston, 1957.
[6] L. H. Putnam, A general empirical solution to the macro software sizing and estimating problem, IEEE Trans. Soft. Eng., pp. 345-361, July 1978.
[7] J. R. Herd, J.N. Postak, W.E. Russell, K.R. Steward, Software cost estimation study:. Study results, Final Technical Report, RADC-TR77-220, vol. I, Doty Associates, Inc., Rockville, MD, pp. 1-10, 1977.
[8] J. W. Bailey and V. R. Basili, "A meta model for software development resource expenditure," in Proceedings of the International Conference on Software Engineering, pp. 107–115, 1981.
[9] R. E. Park, PRICE S: The calculation within and why, Proceedings of ISPA Tenth Annual Conference, Brighton, England, pp. 231-240, July 1988.
[10] R. Jang, Neuro-Fuzzy Modeling: Architectures, Analyses and Applications, Ph.D. Thesis, University of California, Berkeley, 1992.
[11] R.K.D. Black, R. P. Curnow, R. Katz, M. D. Gray, BCS Software Production Data, Final Technical Report, RADC-TR-77-116, Boeing Computer Services, Inc., March, pp. 5-8, 1977.
[12] R. Tausworthe, Deep Space Network Software Cost Estimation Model, Jet Propulsion Laboratory Publication 81-7, pp. 67-78, 1981.
[13] W. S. Donelson, Project Planning and Control, Datamation, pp. 73-80, June 1976.

Parvinder S. Sandhu is working as Professor and Chair of the Department of Computer Science and Engineering and Information Technology, Rayat & Bahra Institute of Engineering & Bio-Technology, India and previously he was with Guru Nanak Dev Engineering College, Ludhiana, India. He is doctorate from Guru Nanak Dev University (India) and member of IASTED Technical Committee of Software Engineering. He has also chaired a number of sessions of International Conferences. He is Master of Engineering in Software Engineering (Thapar University, Patiala), M.B.A. and Bachelor in Computer Engineering from National Institute of Technology (NIT), Kurukshetra. He has published 18 research papers in referred International journals and more than 20 papers in renowned international conferences. His current research interests are Software Reusability, Software Cost Estimation, Bio-informatics, Software Maintenance and Machine Learning.