

Software Effort Prediction -A Fuzzy Logic Approach

Sanjay Kumar¹, Jaya Pal² and Vandana Bhattacharjee³

¹ Birla Institute of Technology, Ranchi

Received: 13 December 2012 Accepted: 2 January 2013 Published: 15 January 2013

Abstract

Accuracy in the estimation of software Effort/Cost is one of the desirable criteria for any software cost estimation model. The estimation of effort or cost before the actual development of any software is the most crucial task of the present day software development project managers. Software project attributes are often measured in terms of linguistic values such as very low, low, Average, high and very high. The imprecise nature of such attributes constitutes uncertainty and vagueness in their subsequent interpretation. In this paper we propose a Fuzzy logic based model for software effort prediction. We feel that fuzzy Software cost estimation Model should be able to deal with imprecision and uncertainty associated with various parameter values. Fuzzy analogy model has been developed and validated upon student data.

Index terms— software cost estimation, effort prediction, fuzzy logic, linear regression.

1 Introduction

Accurate and timely prediction of the development effort and schedule required to develop a software system is one of the most critical activities in managing software projects. In addition software estimation has been identified as one of the three great challenges for half-century-old computer science. [19] In the last 30 years many different studies have been done in the area of Software Cost Estimation to improve the estimation accuracy and so many models are introduced. The rest of the paper contains the following sections as follows: section II represents Research Method, section III represents Experimental Results, and section IV represents Conclusion and Future Scope. a) Fuzzy Logic Intelligent Systems provide alternative paradigms aimed at facilitating the representation and manipulation of uncertain, incomplete, imprecise or noisy data. Fuzzy logic is a form of many-valued logic or probabilistic logic; it deals with reasoning that is approximate rather than fixed and exact.

The traditional approach to building system controllers requires a prior model of the system. The quality of the model, that is, loss of precision from linearization and/or uncertainties in the system's parameters negatively influences the quality of the Authors : Department of CSE, Birla Institute of Technology Ranchi, India. E-mails : k_sanjay71@yahoo.co.in, jayapal@bitmesra.ac.in, vbhattacharya@bitmesra.ac.in resulting control. It is well known that the fuzzy theory not only provides natural tool for describing quantitative data but also generally produces good performance in many applications. In addition, fuzzy rules allow us to effectively classify data having non-axis-parallel decision boundaries, which is difficult for the conventional attribute-based methods. However, one of the difficulties with fuzzy decision trees is determining an appropriate set of membership functions representing fuzzy linguistic terms. Usually membership functions are given manually, however, it is difficult for even an expert to determine an appropriate set of membership functions when the volume and dimensionality of data are large.

At the same time, methods of soft computing such as fuzzy logic possess non-linear mapping capabilities, do not require an analytical model and can deal with uncertainties in the system's parameters. Although fuzzy logic deals with imprecise information, the information is processed in sound mathematical theory [40]. Based on the nature of fuzzy human thinking, Lofti Zadeh originated the "fuzzy logic" or "fuzzy set theory", in 1965. Fuzzy logic deals with the problems that have fuzziness or vagueness. In fuzzy set theory based on fuzzy logic a

particular object has a degree of membership in a given set that may be anywhere in the range of 0 (completely not in the set) to 1 (completely in the set) [41].

For this reason fuzzy logic is often defined as multi-valued logic (0 to 1), compared to bi-valued Boolean logic [42]. Specifically, Fuzzy Logic offers a particularly convenient way to generate a keen mapping between input and output spaces thanks to fuzzy rules' natural expression [2]. Fuzzy logic has been used in [36] [37][38] [39]. Fuzzy set theory and fuzzy logic are a highly suitable and applicable basis for developing knowledge-based systems This paper presents a fuzzy rule based system having two fuzzy inputs, namely Line of code (LOC) and Adjusted difficulty level (Adj.diff.level) and one output Development time (Devtime) as shown in Figure 1. ii. Knowledge Base That consists of two parts: database that defines linguistic variables conditional statements (Fuzzy sets, and rule base that represents the mapping of fuzzy input set into a fuzzy output set. Rules are fuzzy implications). Fuzzy sets, and rule base that represents the mapping of fuzzy input set into a fuzzy output set. Rules are fuzzy conditional statements (implications).

2 iii. Decision Logic

That simulates human decision making based on fuzzy concepts. Conclusion of certain condition is derived by decision making logic.

3 iv. Defuzzification

That converts rule base fuzzy outputs into nonfuzzy (.numerical) values.

Central mechanism of knowledge base and decision making logic considers the fuzzy extension of conventional rule inferencing concept to fuzzy rules inferencing. Premises and conclusions of rules now contain fuzzy values. These facts by definition describe practically continual input set of characteristics. In this manner, one rule can replace more conventional rules. Fuzzy inferencing rules generally connect m conditional variables X_1, \dots, X_m to n consequent variables Y_1, \dots, Y_n in form of: IF (X_1 is A_1 and $...X_m$ is A_m) THEN (Y_1 is B_1 and $... Y_n$ is B_n).

Where A_1, \dots, A_m and B_1, \dots, B_n are linguistic terms of linguistic variables X_1, \dots, X_m and Y_1, \dots, Y_n , respectively.

The IF part is called the "antecedent" and the THEN part is called the "consequent". To make a decision based on a set of rules, a rules-based system follows these steps: 1. All the rules that apply are invoked, using the membership functions and truth values obtained from the inputs (by a process called fuzzification), to determine the result of the antecedent. 2. This result in turn will be mapped into a membership function and truth value controlling the output variable. This process is known as implication. Two of the more common implication functions are: clipping (the fuzzy set is clipped to a value given by the level of activation of the input variables) and scaling (the fuzzy set is multiplied by a value given by the level of activation of the input variables). 3. These results are combined by a process called aggregation. One common approach for the aggregation involves using the "maximum" of the implicated sets. 4. Finally, a process known as defuzzification is used to compute a single value that is representative of the aggregated fuzzy set.

4 c) Multiple Regressions

A linear equation with three independent variables (multiple regressions) may be expressed as: $y = b_0 + b_1 x_1 + b_2 x_2$ (1)

Where b_0 , b_1 , and b_2 are constants; x_1 , and x_2 are the independent variables, and y is the dependent variable. The values of b_0 , b_1 , and b_2 of the multiple regression. Equation may be obtained solving the following system of linear equations $y = nb_0 + b_1(x_1) + b_2(x_2)$ (2) $x_1y = b_0(x_1) + b_1(x_1^2) + b_2(x_1x_2)$ (3) $x_2y = b_0(x_2) + b_1(x_1x_2) + b_2(x_2^2)$ (4)

5 d) Evaluation Criteria

A common criterion for the evaluation of software effort models is the Magnitude of Relative Error (MRE) which is defined as follows:

The MRE value is calculated for each observation whose devtime is predicted. The A complementary criterion is the prediction at level l, $Pred(l) = k/N$, where k is the number of observations where MRE is less than or equal to l, and N is the total number of observations. Thus, $Pred(0.25)$ gives the percentage of development time of software which were predicted with a MRE less or equal than 0.25.

6 II.

7 Research Method a) Metrics Used

The following metrics have been used Line of Code (LOC), and Adjusted Difficulty Level (ADJ.DIFF.LEVEL) which is served as input to the Fuzzy Logic System. And one output Development Time (DEVTIME).

Description of Metrics 1. Line of Code (LOC): Loc is the total number of lines of code used to develop the software excluding the comment lines. This metric was measured on the scale of 0-60.

2. Adjusted Difficulty Level (ADJ.DIFF.LEVEL): This is the difficulty level of the programmers to develop the software which is further adjusted with the help of expert judgments. This metric was measured on the scale of 0-6.

8 b) Data Gathered

The proposed model was validated by a data set collected from the BIT, students of MCA. This data set consists of 10 project data. The data set is applied to the proposed fuzzy model is shown in the Table 1.

9 c) Fuzzy Rules

The term fuzzy identification usually refers to the techniques and algorithms for constructing fuzzy models from data. The expert knowledge in a verbal form is translated into a set of if-then rules. A certain model structure can be created, and parameters of this structure, such as membership functions and weights of rules, can be tuned using input and output data. This paper is based on five fuzzy rules as follows: 1. If ADJ.DIFF.LEVEL is (Very Low) then (DEVTIME (Very Low)). (1). 2. If (loc is Low) and (ADJ.DIFF.LEVEL is (Average)) then (DEVTIME is (Low)) (1). 3. If (loc is Average) and (ADJ.DIFF.LEVEL is (Average)) then (DEVTIME is (Average)) (1). 4. If (loc is High) and (ADJ.DIFF.LEVEL is (Average)) then (DEVTIME is (High)) (1). 5. If (loc is Very High) and (ADJ.DIFF.LEVEL is (High)) then (DEVTIME is (Very High)) (1). The weight of all the rule is 1.

Input and output Membership Functions (MF) are depicted in Table 2. All are triangular and their scalar parameters (a, b, c) are defined as follows: Where MF is membership function.

The membership function plots corresponding to Table 2 The same dataset has been used in Multiple Regression model to estimate the development time, which is to be used to develop a software. The estimated value of development time is compared with the actual value of development time, and with the help of this, using the evaluation criterion the MRE, MMRE and the Pred (25) value is also calculated.

10 e) Evaluation criteria

For this model the same evaluation criterion is used. The criterion which is used to evaluate the fuzzy model.

11 III.

12 Experimental Results

The results show more accuracy in case of effort estimation by the proposed fuzzy model. The result is shown in the Table 3, Table 4 and Table 5. We have compared the actual development time with the predicted development time given by the model for each data set and found that difference between the actual devtime and predicted devtime. Then we calculated the MRE of each project and $MMRE = 0.1762$ and $pred(0.25\%)$ which is 0.6. The same dataset has been tested using multiple regression model and the calculated MMR was used to further find out the $MMRE = 0.5358$ and $Pred(0.25\%) = 0.3$. after going through the results we conclude that the proposed fuzzy model gives the better accuracy.

13 IV.

14 Conclusion and Future Scope

The main benefit of this model is its good interpretability by using fuzzy rules and another great advantage of this research is that it can put together expert knowledge (Fuzzy rules) and project data into one general framework that may have a wide range of applicability in software estimation.

Further the comparison with multiple regression model to fuzzy logic model, and the results support the fuzzy logic model.

In our future work we will test this model upon different real-time datasets. Datasets have been collected from the software engineering data repository [34]. These datasets have been used by various researchers in their work [4] [34] [35].

15 Multiple Regression

Fuzzy

1 2

¹Software Effort Prediction -A Fuzzy Logic Approach

²CSoftware Effort Prediction -A Fuzzy Logic Approach



1

Figure 1: Figure 1 :

1

	LOC	ADJ.DFF.LEVEL	DEVTIME(ACT)	
1	16		1.5	3
2	32		1.8	4
3	23		2.3	10
4	18		2.8	15
5	16		1.5	3
6	32		1.3	4
7	16		1.3	4
8	18		1.4	4
9	19		1.8	7
10	22		1.9	5

Figure 2: Table 1 PROG_ID

2

Variable Name	Range	MF	Parameters		
			a	b	c
LOC	0-60	VL	0	10	20
		L	10	20	30
		AV	20	30	40
		H	30	40	50
		VH	40	50	60
		VL	0	1	2
ADJ.DIFF.LEVEL	0-6	L	1	2	3
		AV	2	3	4
		H	3	4	5
		VH	4	5	6
Output					
Variable Name	Range	MF	Parameters		
			a	b	c
DEVTIME	0-24	VL	0	4	8
		L	4	8	12
		AV	8	12	16
		H	12	16	20
		VH	16	20	24

Figure 3: Table 2 :

3

	DEVTIME (ACT)	DEVTIME USING FUZZY CALCULATED LOGIG	(PRED)	MRE
1	3	4		0.3333
2	4	4		0.0000
3	10	10		0.0000
4	15	8		0.4667
5	3	4		0.3333
6	4	4		0.0000
7	4	4		0.0000
8	4	4		0.0000
9	7	4		0.4285
10	4	4		0.0000

Figure 4: Table 3 PROG_ID

4

PROG_ID	DEVTIME (ACT)	DEVTIME (PRED) CALCULATED USING LINEAR REGRESSION	MRE
1	3	4.2048	0.4016
2	4	5.7181	0.4295
3	10	9.8191	0.1808
4	15	13.7431	0.0838
5	3	4.2048	0.4016
6	4	2.0155	0.4961
7	4	2.7237	0.3191
8	4	3.3757	1.5606
9	7	6.2935	1.0092
10	4	6.9013	0.4753

Figure 5: Table 4

5

Figure 6: Table 5 :

Min(MRE)	0.0838	0.0000
Max(MRE)	1.5606	0.4667
MMRE	0.5358	0.1762
Pred(25)	0.3	0.6

[Note: © 2013 Global Journals Inc. (US)Global Journal of Computer Science and TechnologyVolume XIII Issue X Version I]

Figure 7:

.1 Acknowledgment

- [Bhattacharjee et al. ()] 'A Case Study on Estimation of Software Development Effort'. V Bhattacharjee , S Kumar , E Rashid . *Proceedings of the 3 Technology, Hyderabad.(A National Conference ICACT-2008, (the & Technology, Hyderabad.(A National Conference ICACT-2008India) 2008. Gokaraju Rangaraju Institute of Engineering,P.)*
- [Idri and Abran ()] 'A Fuzzy Logic Based Measures for Software Project Similarity: Validation and Possible Improvements'. A Idri , A Abran . *Proceedings of the 7th International Symposium on Software Metrics, (the 7th International Symposium on Software MetricsEngland, UK) 2001. IEEE Computer Society. p. .*
- [Cuahtemoc Lopez Martin et al. ()] 'A Fuzzy Logic Model Based Upon Reused and New & Cahnged Code for Software Development Effort Estimation at Personal Level'. Cornelio Cuahtemoc Lopez Martin , Agustin Yanez-Marquez , Gutierrez-Tornes . *Proceedings of the 15th International conference on Computing (CIC'06), (the 15th International conference on Computing (CIC'06)) 2006. IEEE Computer Society.*
- [Pal and Bhattacharjee ()] *A Fuzzy Logic System for Software Quality Estimation*, J Pal , V Bhattacharjee . 2009. p. . (in proceedings ICIT 2009)
- [Bhattacharjee and Kumar ()] 'An Expert-Case Based Frame work for Software Cost Estimation'. V Bhattacharjee , S Kumar . *Proceedings of the National Conference on Soft Computing Techniques for Engineering Application, (the National Conference on Soft Computing Techniques for Engineering ApplicationNIT Rourkela) 2006. SCT-2006*
- [Kumar and Bhattacharjee (2007)] 'Analogy and Expert Judgment: A Hybrid Approach to Software Cost Estimation'. S Kumar , V Bhattacharjee . *Proceedings of the National Conference on information Technology: Present practice and Challenge, (the National Conference on information Technology: Present practice and ChallengeNew-Delhi, India) 2007. Sep'07.*
- [Desharnais ()] *Analyse statistique de la productivite des projets informatique a partie de la technique des point des foncti\on*, J M Desharnais . 1989. University of Montreal (Masters Thesis)
- [Bhattacharjee and Pal (2010)] 'Applying Fuzzy Clustering To Predict Software Usability'. V Bhattacharjee , J Pal . *International Journal of Applied Research on Information Technology and Computing (IJARITAC)* Sept-Dec, 2010. (1) p. .
- [Briand] Lionel C Briand . *On the many ways Software Engineering can benefit from Knowledge Engineering*,
- [Bhattacharjee et al. ()] *Case Based Estimation Model using Project Feature Weights*, V Bhattacharjee , S Kumar , Ekbal Rashid , ; Ism , Dhanbad Jharkhand . 2009.
- [Aamodt and Plaza ()] *Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches*, A Aamodt , E Plaza . 1994. AI Communications, IOS Press. 7 p. .
- [Idri et al. ()] 'COCOMO Cost Model Using Fuzzy Logic'. Ali Idri , Alain Abran , Laila Kjiri . *Proc. 7th Intl. Conf. On Fuzzy theory & Technology, (7th Intl. Conf. On Fuzzy theory & Technology) 2000. 2000.*
- [Bhattacharjee et al. (2009)] 'Complexity Metric For Analogy Based Effort Estimation'. V Bhattacharjee , P K Mahanti , S Kumar . *Journal of Theoretical and Applied Information Technology* 2009. 31st july 2009. 6 (1) .
- [Shepperd and Schofield (1997)] 'Estimating Software Project Effort Using Analogies'. M Shepperd , C Schofield . *IEEE Transactions on Software Engineering* 1997. November 1997. 23 (12) p. .
- [Bhattacharjee et al. (2008)] 'Estimation of Software Development Effort in University Setting: A Case Study'. V Bhattacharjee , S Kumar , Ekbal Rashid . *Presented at National Conference on Architecturing Future IT Systems (NCAFIS '08), (Indore) 2008. October 2008.*
- [Kadoda et al. ()] 'Experiences Using Case-Based Reasoning to Predict Software Project Effort'. G Kadoda , M Cartwright , L Chen , M Shepperd . *Proceedings of EASE, (EASEKeele, UK) 2000. p. .*
- [Rush and Roy] *Expert Judgement in cost estimating: Modelling reasoning process*, Christopher Rush , Rajkumar Roy .
- [Bose (1994)] 'Expert systems, fuzzy logic, and neural network application in power electronics and motion control'. B K Bose . *Proceeding of the IEEE, (eeding of the IEEE) Aug. 1994. 82.*
- [Idri et al. ()] *Fuzzy Case-Based Reasoning Models for Software Cost Estimation*, Ali Idri , Alain Abran , T M Khoshgoftaar . 2002. 2002. Theory and Applications Published by Springer-Verlag. (to appear in the book Soft Computing in Software Engineering)
- [Kumar and Bhattacharjee (2005)] 'Fuzzy Logic Based Model for Software Cost Estimation'. S Kumar , V Bhattacharjee . *Proceedings of the international Conference on information Technology, (the international Conference on information TechnologyLudhiana, India) 2005. Nov'05.*
- [Bhattacharjee and Pal ()] 'Fuzzy Rule Based System for Fine Prediction ction in Accidents'. V Bhattacharjee , J Pal . *proceedings of Conference on information Science, Technology and Management, (Conference on information Science, Technology and ManagementGurgaon, India) July 13-15, 2009. (CISTM 2009, MDI)*

- [German Workshop on Case-Based Reasoning] *German Workshop on Case-Based Reasoning*,
 [Idri et al.] Ali Idri , Alain Abran , Taghi M Khoshgoftaar . *Computational Intelligence in Empirical Software Engineering*,
 [Auer and Biffl ()] 'Increasing the Accuracy and Reliability of Analogy -Based Cost Estimation with Extensive Project feature Dimention Weighting'. Martin Auer , Stefan Biffl . *International Symposium on Empirical Software Engineering (ISESE'04)*, 2004. IEEE Computer Society.
 [Jane Delany et al. ()] Sarah Jane Delany , Padraig Cunningham , Wolfgang Wilke . *The Limits of CBR in Software Projects Estimation*, 1998. 1998. (Presented at)
 [Srinivasan and Fisher (1995)] 'Machine Learning approaches to Estimating Software Development Effort'. Krishnamoorthy Srinivasan , Douglas Fisher . *IEEE transactions on Software Engineering*, 1995. February 1995. 21.
 [Math Works, Fuzzy Logic Toolbox User's Guide (1998)] *Math Works, Fuzzy Logic Toolbox User's Guide*, Jan., 1998.
 [Rajasekaran et al. ()] *Newral Networks Fuzzy logic, and Genetic Algorithms Synthesis and Applications*, S Rajasekaran , G A Vijayalakshmi , Pai . 2003. (PHI. 34. Software Engineering Data Repository)
 [Auer et al. (2006)] 'Optimal Project Feature weights in Analogy -Based Cost Estimation: Improvement and Limitations'. Martin Auer , Adam Trendowicz , Bernhard Graser , Ernst Haunschmid , Stefan Biffl . *IEEE transactions on Software Engineering*, 2006. February 2006. 32.
 [Jargensen ()] 'Practical Guidelines for Expert-Judgment-Based Software Effort Estimation'. Magne Jargensen . *Proceedings of the 7th ACIS International Conference on Software Engineering Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD'06)*, (the 7th ACIS International Conference on Software Engineering Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD'06)) 2005. IEEE Computer Society. p. .
 [Attarzadeh et al. (2010)] 'Soft Computing Approach for Software Cost Estimation'. Imran Attarzadeh , Siew Hock , Ow . *In international Journal of Software Engineering, IJSE* 2010. January 2010. 3 (1) .
 [Leung and Fan] *Software Cost Estimation*, Hareton Leung , Zhang Fan .
 [Bhattacharjee and Kumar ()] 'Software Cost Estimation and its Relevance in The Indian Software Industry'. V Bhattacharjee , S Kumar . *Proceedings of the International Conference on Emerging Technologies IT Industry*, (the International Conference on Emerging Technologies IT IndustryLudhiana, India) 2004.
 [Nisar and Elahi ()] 'Software Development Effort Estimation Using Fuzzy Logic -A Survey'. M Wasif Nisar , Yong-Ji , Wang , Manzoor Elahi . *5th International Conference on Fuzzy Systems and Knowledge Discovery*, 2008. IEEE Computer Society.
 [Cuauhtemoc Lopez Martin et al.] 'Software Development Effort Estimation using Fuzzy Logic: A Case Study'. Jerome Cuauhtemoc Lopez Martin , Cornelio Leboeuf Pasqueir , M Yanez , Agustin Gutierrez , T . *Proceedings of the 6th International conference on Computer Science (ENC'05)*, (the 6th International conference on Computer Science (ENC'05) IEEE Computer Society.
 [Braz and Vergilio ()] 'Software Effort Estimation Based on Use Cases'. Marcio Rodrigo Braz , Silvia Regina Vergilio . *Proceedings of the 30th Annual International Computer Software and Applications Conferences (COMPSAC'06)*, (the 30th Annual International Computer Software and Applications Conferences (COMPSAC'06)) 2006. IEEE Computer Society.
 [Fairley ; T M H ()] *Software Engineering Concepts*, R Fairley ; T M H . 1997.
 [Boehm ()] *Software Engineering Economic*, B W Boehm . 1981. Prentice -Hall.
 [Tsoukalas and Using ()] L H Tsoukalas , R E Using . *Fuzzy and Neural Approaches in Engineering*, (NY) 1997. John Wiley.
 [Braz and Vergilio ()] 'Using Fuzzy Theory for Effort Estimation of Object -Oriented Software'. Marcio Rodrigo Braz , Silvia Regina Vergilio . *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence*, (the 16th IEEE International Conference on Tools with Artificial Intelligence) 2004. IEEE Computer Society. (ICTAI 2004)