

Contents lists available at Sjournals



Journal homepage: www.Sjournals.com



Original article

The comparison of software cost estimation methods using fuzzy sets theory

H. Nosrati Nahook^a

^aInstructor, Department of Computer Engineering, Payame Noor University, Iran.

*Corresponding author; Email: hsn.nosrati@gmail.com. Tel: +989155481075.

ARTICLE INFO

ABSTRACT

Article history,

Received 27 August 2015

Accepted 20 September 2015

Available online 27 September 2015

Keywords:

Triangular fuzzy number

Membership function

Fuzziness

Effort estimation

Kloc

Estimation by analogy is one of the expedient techniques in software effort estimation field. However, the methodology utilized for the estimation of software effort by analogy is not able to handle the categorical data in an explicit and precise manner. Early software estimation models are based on regression analysis or mathematical derivations. Software effort estimation is the process of predicting most realistic use of effort required to develop or maintain software based on incomplete and uncertain input. There are various methods suggested by researchers for calculating effort. The best result are achieved by using soft computing technique. In this paper we have represented size in KLOC as a triangular fuzzy number. Fuzzy-based methods compare with common methods. MATLAB is used for tuning the parameters of famous various cost estimation methods. On published software projects data, the performance of the method is evaluated. Comparison of results from SCEFL (Software Cost Estimation using Fuzzy Logic) methods with existing ubiquitous methods is done.

© 2015 Sjournals. All rights reserved.

1. Introduction

The ability to estimate accurately the size, effort and duration of a software development project is crucial to project success. Inaccurate software estimate causes trouble in business process related to software development .As the demand for software application increases the software companies need accurate estimation of project under development. Cost Estimation is achieved in terms of person-months (PM), which can be translated into actual dollar cost. The concept of software cost estimation has been growing rapidly due to practicality and

demand for it. Today the peoples are expecting high quality software with a low cost, the main objective of software engineering. So many popular cost estimation methods like COCOMO,FP, Delphi, Halsted Equation, Bailey - Basili, Doty, barry boehm and Anish Mittal Method had came into existence. These methods are created as a result of regression analysis and power regression analysis methods applied to historical data. Today most of the software companies follow COCOMOII for estimating the cost of products; we found some variations in this method (Robert, 1987), (Kim Johnson, 1998; G. Witting, 1997; Baiely, 1981; B. Boehm, 1995).

Fuzzy logic with its features of a powerful linguistic representation can signify imprecision in inputs and outputs, while providing a more expert knowledge based approach to method building. The fuzzy logic method uses the fuzzy logic concepts introduced by Lofti A. Zadeh (Zadeh, 1965; Jose Galindo, 2008; Lotfi Zadeh, 1994).

Cost estimation is a process in which certain parameters are taken as input and merging them with the knowledge of past to develop the estimated cost. Software cost estimation (SCE) involves the determination of one or more of the following estimates and on the basis of it cost is estimated.

- Effort(in persons per month)
- Project duration
- Cost (in currency i.e. Dollars, Euro, Rupee etc.)

There are two different approaches to cost estimation. The older approach is called LOC estimation and the newer one is Counting Function point (David A,2003).

Our cost estimation method consider cost estimation method generate cost in terms of effort estimate (in person per month) using fuzzified KLOC. This can be converted into cost by calculating average salary per unit time of the staff involved and multiplying this by estimated effort required.

Various methods of software cost estimation are provided by various researchers to provide better cost estimation In (H. Zeng, 2004) give the use of fuzzy set in COCOMO. In (S. Kumar, 1994) many software methods are evaluated and presented. In (A. C. Hodgkinson, 1999) many methods were explored to provide better effort estimation. Fuzzy sets and neural network is used in Software project management. In (Musilek, 2000) , f-COCOMO method using fuzzy sets is purposed.

2. Related Terms

2.1. software cost estimation

Several Algorithmic manual methods (Linda M, 2006) are purposed for effort estimation. They are in the form of $Effort = a \times (SIZE)^b$, Where 'a' and 'b' are empirically determined constants, Size is length of thee code in KLOC. There are various effort estimation equations Based on KLOC Such As:

Halsted Equation:	$Effort = 5.2(KLOC)^{1.50}$	(1)
Bailey - Basili:	$Effort = 5.5 + 0.73(KLOC)^{1.16}$	(2)
Doty:	$Effort = 5.288(KLOC)^{1.047}$	(3)

One Traditional Parametric method COCOMO (Constructive Cost Method) was given by Boehm . It is the most famous and popular algorithmic effort estimation method. In it the software projects are grouped into three classes of the projects 1. Organic 2. Semidetached 3. Embedded. The COCOMO Method General equation comes in the form of:

$$Effort = a(KLOC)^b \quad (4)$$

The values of 'a' and 'b' depends upon the class of the project it belongs. Values of 'a' and 'b' in various classes is given in Table 1. The Effort is computed in Persons per Month.

Table 1

Cocomo Values of ‘a’ and ‘b’ in various classes.

Project No.	a	b
Organic	2.4	1.05
Semidetached	3.0	1.12
Embedded	3.6	1.2

2.2. Fuzzy Logic

Fuzzy logic is a methodology, to solve problems which are too complex to be understood quantitatively, based on fuzzy set theory (Jose Galindo,2008). Use of fuzzy sets in logical expression is known as fuzzy logic. A fuzzy set is characterized by a membership function, which associates with each point in the fuzzy set a real number in the interval [0,1], called degree or grade of membership. The membership function may be triangular, trapezoidal, parabolic etc. A triangular fuzzy number (TFN) is described by a triplet (α, m, β), where m is the method value , α and β are the right and left boundary respectively. Size of the project, especially in the beginning of the project, cannot be taken precisely. It can be taken as a fuzzy number. The effort is estimated in terms of distribution described by membership function of the effort.

Fuzzy numbers are one way to describe data vagueness and imprecision. A fuzzy number is an extension of a regular number in the sense that it does not refer to one single value but rather to a connected set of possible values, where each possible value has its own weight between ‘0’ and ‘1’. This weight is called the membership function. The membership function is increasing towards the mean and decreasing away from it. The Fuzzy number can be of three types 1. Triangular fuzzy Number 2. Trapezoidal fuzzy number 3. Bell shaped fuzzy number Figure 1 shows The three curves (Harish Mittal, 2009).

Fuzziness in a fuzzy set is characterized by its membership functions. A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The input space is sometimes referred to as the universe of discourse, a fancy name for a simple concept. It classifies the element in the set, whether it is discrete or continuous. The graphical representations may include different shapes. There are certain restrictions regarding the shapes used. The “shape” of the membership function is an important criterion that has to be considered. There are different methods to form membership functions. Zadeh proposed a series of membership functions that could be classified into two groups: those made up of straight lines, or “linear,” and Gaussian forms, or “curved.”. Based on this criteria the membership function can be of the following types (Jose Galindo, 2008).

1. Triangular(Figure 1 and 2)

Defined by its lower limit a, its upper limit b, and the modal value m, so that a < m < b. We call the value b – m margin when it is equal to the value m – a.

$$T(x) = \begin{cases} 0 & x \leq a \quad \text{or} \quad x \geq b \\ \frac{(x-a)}{(m-a)} & x \in (a,m) \\ \frac{(b-x)}{(b-m)} & x \in (m,b) \end{cases} \quad (5)$$

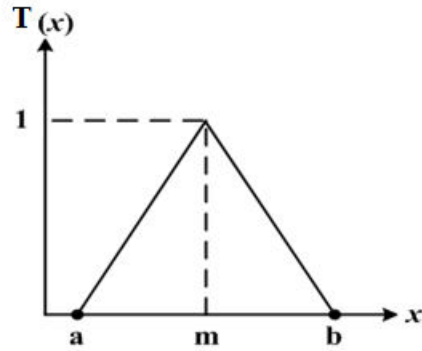


Fig. 1. Triangular Fuzzy Asymmetrical.

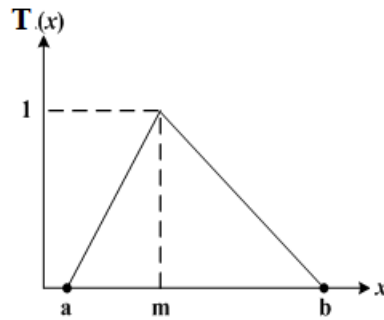


Fig. 2. Triangular Fuzzy symmetrical.

2. Trapezoidal (Figure 3)

Defined by its lower limit a , its upper limit d , and the lower and upper limits of its nucleus or Kernel b and c respectively:

$$TP(x) = \begin{cases} 0 & \text{if } x \leq a \text{ or } x \geq d \\ \frac{(x-a)}{(b-a)} & \text{if } x \in (a, b) \\ 1 & \text{if } x \in (b, c) \\ \frac{(d-x)}{(d-c)} & \text{if } x \in (c, d) \end{cases} \quad (6)$$

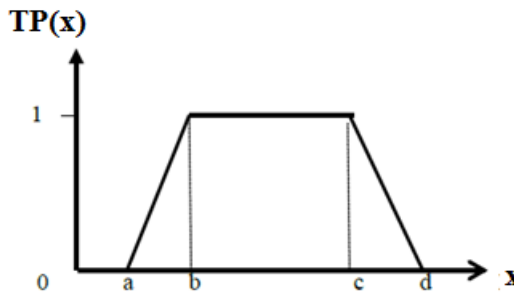


Fig. 3. Trapezoidal Fuzzy.

Fuzziness results from imprecise boundary of fuzzy set. fuzziness of any set is measured by metric distance between its membership grade function and its nearest crisp set. A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The graphical representations may include different shape. The “shape” of the membership function is an important criterion that has to be considered. There are different methods to form membership functions. Zadeh proposed a series of membership functions that could be classified into two groups: those made up of straight lines, or “linear,” and Gaussian forms, or “curved” (Jyh – shing Roger Jang, 1997).

$$T(F) = G(F) = \frac{\beta - \alpha}{2m}, 0 < F < 1 \quad (7)$$

2.3. Criterion for Measurement of Software Cost Estimation Method Performance

The mean magnitude of relative error (MMRE) as the main performance measure. Using MMRE we evaluate the impact of estimation accuracy using evaluation criteria, for each method. Method with minimum MMRE value gives better estimation result. The mean magnitude of relative error (MMRE), defined as:

$$MMRE = \frac{\sum_{i=1}^N \frac{|E_{actual} - E_{estimated}|}{E_{actual}}}{N} \quad (8)$$

3. Proposed Methods Using Fuzzy Numbers

3.1. Fuzzification

using Triangle fuzzy number:

We Use Triangular Fuzzy number T(S) which is defined as Follows:

$$T(s) = \begin{cases} 0 & s \leq \alpha \quad \text{or} \quad s \geq \beta \\ \frac{(s - \alpha)}{(m - \alpha)} & s \in (\alpha, m) \\ \frac{(\beta - s)}{(\beta - m)} & s \in (m, \beta) \end{cases} \quad (9)$$

Where S is size as input, E is effort as output, α , m and β are the parameters of membership function T(S), m is the method value, α and β are the right and left boundaries respectively. Let (m, 0) divides, internally, the base of the triangle in ratio k : 1, where k is real positive number. So that:

$$m = \frac{\alpha + k\beta}{k+1} \quad (10)$$

As by definition of fuzziness:

$$F = \frac{\beta - \alpha}{2m} \quad \text{so}$$

$$\alpha = \left(1 - \frac{2kF}{k+1}\right) m \quad (11)$$

$$\beta = \left(1 + \frac{2F}{k+1}\right) m \quad (12)$$

Similarly, the $\mu(E)$ is defined as, (Figure 4)

$$\mu(E) = \begin{cases} 0 & E \leq a\alpha^b \text{ or } E \geq a\beta^b \\ \frac{\left(\left(\frac{E}{a}\right)^{\frac{1}{b}} - \alpha\right)}{(m - \alpha)} & E \in (a\alpha^b, am^b) \\ \frac{\left(\beta - \left(\frac{E}{a}\right)^{\frac{1}{b}}\right)}{(\beta - m)} & E \in (am^b, a\beta^b) \end{cases} \quad (13)$$

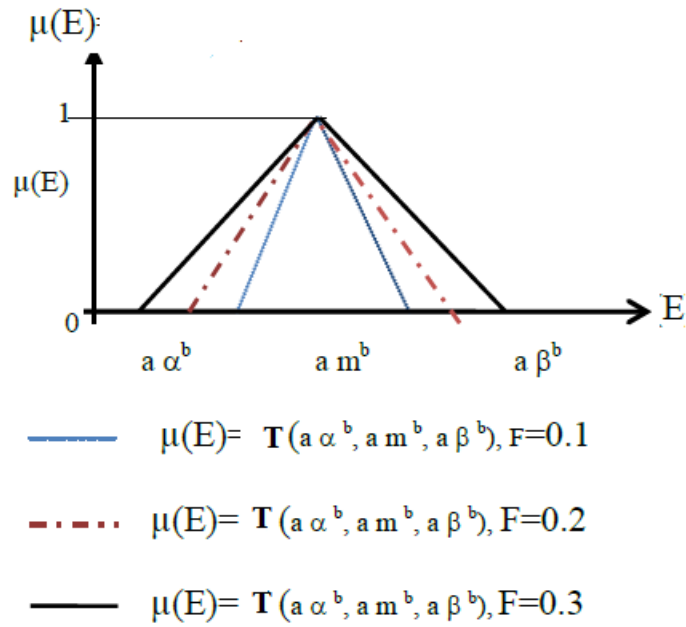


Fig. 4 . Representation of $\mu(E)$.

Table 2, gives values of α and β for $F=0.1, 0.2$ and 0.3 for various values of k using equations (11) and (12), where m , size estimate in KLOC.

Table 2
Values of α and β for various values of F and k .

F	k	1
0.1		$\alpha = 0.9m$
		$\beta = 1.1m$
0.2		$\alpha = 0.8m$
		$\beta = 1.2m$
0.3		$\alpha = 0.7m$
		$\beta = 1.3m$

3.2. Defuzzification

Fuzzy effort estimate (E) is given as:

$$E = \frac{W_1(a\alpha^b) + W_2(am^b) + W_3(a\beta^b)}{W_1 + W_2 + W_3} \quad (14)$$

Where, W_1, W_2 and W_3 are weights of the optimistic, most likely and pessimistic estimate respectively. Maximum weight should be given to the most expected estimate (Roger S. Pressman, 2005), (Anish Mittal, 2010). where $a=6, b=0.6$ obtained by using MATLAB, m represents the size in KLOC. Here w_1, w_2 and w_3 are arbitrary constants. The effort is estimated in man months (MM).

4. Experimental

The Data is taken from (Robert W, 1987) and given in Table 3. Let $F = .03, k = 1$, then from Tables 2 and 3, $\alpha = 0.7m, \beta = 1.3m$. We have taken $W1 = 1, W2 = 4$ and $W3=1$ for our Method. Table 5 below gives the experimental results.

Table 3
Data set.

s. no	Project no	Size (kloc)	Actual effort
1	2	40.5	82.5
2	6	50	84
3	11	128.6	230.7
4	9	39	72
5	12	161.4	157
6	13	164.8	246.9
7	7	200	130.3
8	1	253.6	287
9	4	214.4	86.9
10	10	254.2	258.7

Table 4
Effort estimated of various methods

Method	actual effort	cocomo semidetached	cocomo embedded	Halsted method	Doty Estimate	Barry boehm	SCE method using triangle fuzzy number(SCE - TFN)
Project 2	82.5	189.43	305.66	1340.25	254.8571697	155.94	231.48
6	84	239.86	393.61	1838.47	317.7701076	194.56	62.65
11	230.7	690.98	1222.90	7583.50	1743.854488	524.62	72.08
9	72	181.59	292.13	1266.48	1994.580128	149.88	212.55
12	157	891.19	1606.15	10662.40	854.4108222	665.96	135.10
13	246.9	912.24	1646.83	11001.10	1083.843164	680.69	157.13
7	130.3	1133.11	2077.48	14707.20	1243.572167	834.12	121.45
1	287	1478.31	2762.36	21000.38	1739.54517	1070.30	212.22
4	86.9	1224.87	2258.25	16324.50	1459.095749	897.29	189.79
10	258.7	1482.23	2770.20	21074.90	1266.485594	1072.90	159.33

Comparison of various methods on the basis of various performance criterions for software cost estimation is given in below Table 5.

Table 5
comparison among various methods and fuzzy proposed methods

Method Name	MMRE
Cocomo semidetached	0.414
Cocomo embedded	0.837
Halsted method	6.4
Doty Estimate	8.1
Barry boehm	0.281
SCE method using triangle fuzzy number(SCE - TFN)	0.146

The following figure 5 shows the comparison of effort calculated among various methods .

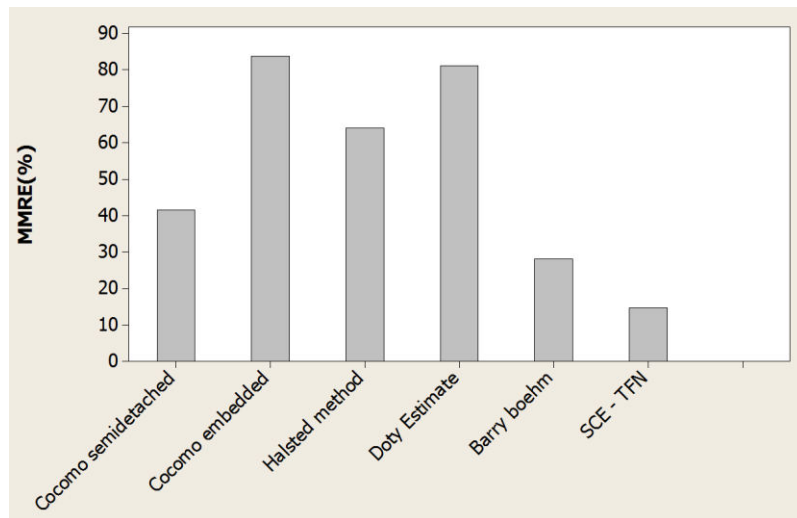


Fig. 5. MMRE(%) comparison of various methods.

5. Conclusion

In this paper we compare a method that performs better than other methods in achieving the accuracy of effort estimation. We described fuzzy software cost estimation method that handles ambiguousness, obscurity and then compared with other popular software cost estimation methods. From the experiments evaluation we concluded that: SCE method using fuzzy number showed better software effort estimate in view of the MMRE evaluation criteria as compared to the traditional estimation methods. We observed that the fuzzy method based on fuzzy numbers provide better results. The methodology of fuzzy sets used for developing the method is sufficiently general and can be applied to other methods of software cost estimation methods based on function point methods and to other areas of quantitative software engineering. The future work may also be done by using hybrid model neuro fuzzy model that combine fuzzy logic with neural network for more accurate estimation result.

References

- A. C. Hodgkinson and P. W. Garratt, (1999). "A neuro-fuzzy cost estimator," in Proceedings of the Third Conference on Software Engineering and Applications: 401–406.
- Anish Mittal, Kamal Parkash, Harish Mittal, (2010), "Software Cost Estimation Using Fuzzy Logic", ACM SIGSOFT Software Engineering, 35(1) : 1-7.
- B. Boehm., (1995). Cost Models for Future Life Cycle Process: COCOMO2. Annals of Software Engineering.
- B. Boehm, (1981). Software Engineering Economics Englewood Cliffs, NJ, Prentice Hall.
- Baiely, j.w Basili, (1981). "A Metamedel for Software Development Resource Expenditure." Proc. Intl. Conference Software Egg : 107-115.
- David A. Gustafson, (2003). Theory and problems of software engineering, TMH.
- G. Witting and G. Finnie, (1997). "Estimating software development effort with connectionist models," in Proceedings of the Information and Software Technology Conference: 469–476.
- H. Zeng and D. Rine, (2004). "A neural network approach for software defects fix effort estimation," in Proceedings of the Eighth IASTED International Conference Software Engineering and Applications: 513– 517.
- Harish Mittal, (2009). Pardeep Bhatia," Software Maintainability Assessment based on fuzzy logic Technique "ACM SIGSOFT 34 (3).
- Jose Galindo, (2008). ".Handbook of Research in Fuzzy Information Processing in Databases", Information science Reference.
- Jyh – shing Roger Jang, Chuen – Tasi Sun, Eiji Mizutani, (1997). "Neuro – Fuzzy and Soft Computing" A Computational Approach to Learning and Machine Intelligence, Prentice Hall Upper Saddle River, 07458 : 47–50.
- Kim Johnson, (1998). Dept of Computer Science, University of Calgary, Alberta, Canada, "Software Cost Estimation: Metrics and Models" : 1 - 17.
- Linda M. Laird, M. Carol Brennan, (2006). " Software Measurement & Estimation: A Practical Approach, Wiley Interscience.
- Lotfi Zadeh, A., (1994). Fuzzy Logic, Neural Networks and Soft Computing, Communication of ACM., 37(3): 77-84.
- Musilek ,p.Pedrucz,W.succi,g. & reformat,m., (2000)"Software Cost Estimation with Fuzzy Models." ACM SIGAPP Applied Computing Review, 8(2) : 4-29.
- Pankaj jalote, "An Integrated Approach for Software Engineering.", Third Edition. ISBN: 978-81-7319-702-4.
- Robert W. Zmud, Chris F. Kemerer, (1987). "An Empirical Validation of Software Cost Estimation Models" Communication of the ACM 30 (5).
- Robert W. Zmud, Chris F. Kemerer, (1987). "An Empirical Validation of Software Cost Estimation Models" Communication.
- Roger S. Pressman, (2005). Software Engineering; A Practitioner Approach, Mc Graw-Hill International Edition, Sixth Edition.
- S. Kumar, B. A. Krishna, and P. Satsangi, (1994). "Fuzzy systems and neural networks in software engg. project management," Journal of Applied Intelligence, 4: 31–52.
- Zadeh, L.A., (1965). Fuzzy sets, Info and Control, 8: 338-353.