

A Proposed Novel Framework for Early Effort Estimation using Fuzzy Logic Techniques

Roheet Bhatnagar¹, Vandana Bhattacharjee², Mrinal Kanti Ghose³

GJCST Classification
1.5.1

Abstract-Estimating software development effort is an important task in the management of software projects. The task of effort estimation is challenging and is an important area of research in the field of Software Project Management. A number of estimation models exist for effort prediction. However, many newer models are still being proposed and researched upon to obtain more accurate estimations. The development of software has always been characterized by parameters that possess certain level of fuzziness. This requires that some degree of uncertainty be introduced in the models, in order to make the models realistic. Fuzzy logic techniques come to play in such a situation to tackle the uncertainty issues. Besides, fuzzy logic had been combined with algorithmic, non-algorithmic effort estimation models as well as a combination of them to deal with the inherent uncertainty issues. The primary purpose of this paper is to review the studies carried over the years for estimating the software development effort using Fuzzy Logic Techniques in order to improve prediction accuracies. In this paper we have also tried to propose a novel fuzzy logic based framework to estimate the effort at the early stages of software project development. The paper discusses the theoretical foundations and concepts for the framework and in subsequent publications we will be discussing the results obtained from our ongoing experimental work.

Keywords- software development effort, effort estimation, fuzzy logic techniques, estimation models, effort estimation framework.

I. INTRODUCTION

Estimating software development effort remains a complex problem that attracts considerable research attention. Improving the estimation techniques that are currently available to project managers would facilitate increased control of time and money in software development. Furthermore, any improvement in the accuracy of predicting the development effort can significantly reduce the costs from errors, such as estimating inaccurately, misleading tendering bids, and disabling the monitoring progress [1]. Software development effort estimates are the basis for project bidding and planning (both are critical practices in the software industry). The consequences of poor budgets and plans can be dramatic: if they are too pessimistic, business opportunities can be lost, while over optimism may be followed by

significant losses [2]. Software effort estimation has even been identified as one of the three great challenges in Computer Science [3]. During the development process, the cost and time estimates are useful for the initial rough validation and monitoring of the project's after completion, these estimates may be useful for project productivity assessment for example. Software effort estimation models are divided into two main categories: algorithmic and non-algorithmic. The most popular algorithmic estimation models include Boehm's COCOMO [4], Putnam's SLIM [5] and Albrecht's Function Point [6]. These models require as inputs, accurate estimate of certain attributes such as line of code (LOC), complexity and so on which are difficult to obtain during the early stage of a software development project. The models also have difficulty in modeling the inherent complex relationships between the contributing factors, are unable to handle categorical data as well as lack of reasoning capabilities [7]. The limitations of algorithmic models led to the exploration of the non-algorithmic techniques which are soft computing based. New paradigms offer alternatives to estimate the software development effort, in particular the Computational Intelligence (CI) that exploits mechanisms of interaction between humans and processes domain knowledge with the intention of building intelligent systems (IS) [8]. Amongst IS, fuzzy logic may be a convenient tool for software development effort estimation. Fuzzy logic-based cost estimation models are more appropriate when vague and imprecise information is to be accounted for. Fuzzy systems try to behave just like the processes of the brain with a rule base. The basic concept is inspired by the human processes where the decisional criteria are not clear cut, but blurred and it is difficult to find objective to make the decisions more precise and clear. Use of fuzzy sets in logical expression is known as fuzzy logic. Fuzzy decision systems are based on fuzzy logic that tries to reproduce the fuzzy human reasoning. Fuzzy logic is extended to take care of the partially truth values. Fuzzy logic as the name describes, the mode of reasoning to be approximate rather than exact. In fuzzy approach there are four stages: Fuzzification, Extending complexity matrix, determining the productivity rate and the effort, Defuzzification.

Why Fuzzy Logic?

Fuzzy logic effort prediction brings numerous benefits. Undeniably the development of software is characterized by parameters that possess certain level of fuzziness. This requires that some degree of uncertainty be introduced in the models, in order to make the models realistic.

About¹- Associate Professor in the Department of Computer Science & Engineering at Sikkim Manipal Institute of Technology, Sikkim, India.

(telephone: +91 9475078498 email: roheetbhatnagar@yahoo.com) About²- Associate Professor in the Department of Computer Science & Engineering at BITEC, Lalpur, Birla Institute of Technology, Mesra, Ranchi, India. (email: vbhattacharya@bitmesra.ac.in)

About³ - Professor & Head, Department of Computer Science & Engineering at Sikkim Manipal Institute of Technology, Sikkim, India. (email: mkghose2000@yahoo.com)

Moore and Edwards discovered that the accepted level of estimation accuracy was typically $\pm 20\%$ [9]. This shows that software project managers are aware of the irrefutable uncertainty in estimation and therefore expect some inaccuracy in the prediction output. Besides, the use of formal models did not improve the accuracy of estimation [10]. Fuzzy logic enables linguistic representation of the input and output of a model to tolerate imprecision. It is particularly suitable for software estimation as many software attributes are measured on nominal or ordinal scale type which is a particular case of linguistic values [11]. The use of fuzzy set satisfies the first criterion of soft computing which is the tolerance of imprecision, as defined by Zadeh [12]. More importantly, the model should support the fact that human reasons in fuzziness. The use of fuzzy set supports continuous belongingness (membership) of elements to a given concept (such as small software project) thus alleviating a dichotomy problem (yes/no) [13] that caused similar projects having different estimated efforts. Apart from that, fuzzy logic approach is less dependent on historical data [14] [15] [16]. Fuzzy logic models can be constructed without any data or with little data [14][15]. This makes fuzzy logic superior over data-driven model building approaches such as neural network, regression and case-based reasoning. In addition, fuzzy logic models can adapt to new environment when data become available [7]. Another advantage of fuzzy logic model is that it has the

ability to represent different levels of uncertainty for the inputs and outputs whilst inferring based on the same model (rules and membership functions). Consequently it is able to cater for the needs of different level of precision for the different stages of development life cycle [14][15].

II. FUZZY LOGIC IN SOFTWARE EFFORT ESTIMATION

A fuzzy model is a modeling construct featuring two main properties [17]: (1) It operates at a level of linguistic terms (fuzzy sets), and (2) it represents and processes uncertainty. Fuzzy logic offers a particularly convenient way to generate a keen mapping between input and output spaces thanks to the natural expression of fuzzy rules. In software development effort estimation, two considerations justify the decision of implementing a fuzzy model: first, it is impossible to develop a precise mathematical model of the domain [18]; second, metrics only produce estimations of the real complexity. Thus, according to the previous assertions, formulating a tiny set of natural rules describing underlying interactions between the software metrics and the effort estimation could effortlessly reveal their intrinsic and wider correlations. In Table 1 below we present a tabular view of a group of papers regarding aspects related to previous research on software development effort estimation based on Fuzzy Logic techniques and concepts.

Table I
Research On Software Development Effort Estimation Based On Fuzzy Logic Techniques And Concepts

Authors	Year	Related Work Done	Results Reported
Fei and Lui [19]	1992	Introduced the f-COCOMO model which applied Fuzzy Logic to the COCOMO model for software effort estimation.	Since there was no comparison of results between the f-COCOMO and other effort estimation models in their study the estimation capability of their model is unknown.
Kumar et al [20]	1994	Had applied fuzzy logic in Putnam's manpower buildup index (MBI) estimation model. MBI selection process was based upon 64 different fuzzy associative memory (FAM) rules.	The work showed how fuzzy FAM's can be effectively applied to the domain of software project management and control for the estimation of the MBI.
Gray and MacDonell [21] [15]	1997	Compared Function Point Analysis, Regression techniques, feedforward neural network and fuzzy logic in software effort estimation.	Their results showed that fuzzy logic model achieved good performance, being outperformed in terms of accuracy only by neural network model with considerably more input variables.
	1999	Developed FULSOME (Fuzzy Logic for Software Metrics) which is a set of tools that helps in creating fuzzy model.	The automatically generated fuzzy model performs acceptably when compared to regression-based models.
Ryder [22]	1998	Researched on the application of fuzzy logic to COCOMO and Function Points models.	Result showed Fuzzy Logic is good at making effort estimations.
Musflek et al [13]	2000	Worked on fuzzifying basic COCOMO model without considering the adjustment factor. In their simple f-COCOMO model, the size input into the COCOMO model is represented by a fuzzy set, while a and b coefficients are crisp values. Besides the size, augmented f-COCOMO also fuzzified both the coefficients	They concluded that (a) fuzzy sets help articulate the estimates and their essence (by exploiting fuzzy numbers described by asymmetric membership functions) and (b) they generate a feedback as to the given uncertainty (granularity) of the results.

		related to the development mode. Triangular membership functions are used in this study.	
Idri et al [23] [11]	2000	Proposed fuzzy intermediate COCOMO'81. The FLM is based upon trapezoidal membership functions. The dataset is randomly generated and compared with actual data of COCOMO 81. The effort multiplier for each cost driver is obtained from fuzzy set, enabling its gradual transition from one interval to a contiguous interval (such as from high to very high).	Validation results showed that the fuzzy intermediate COCOMO'81 can tolerate imprecision in its input (cost drivers) and generate more gradual outputs. Thus fuzzy intermediate COCOMO'81 is less sensitive to the changes in the inputs as compared to intermediate COCOMO'81.
	2002	Proposed an approach based on fuzzy logic named Fuzzy Analogy. Its dataset is that of COCOMO 81.	Taking into account their results, they suggested the following ranking of the four techniques in terms of accuracy and adequacy to deal with linguistic values: 1. Fuzzy Logic, 2. Fuzzy intermediate COCOMO'81, 3. Classical intermediate COCOMO'81, and 4. Classical Analogy.
Huang et al. [24]	2003	Proposed a model combining fuzzy logic and neural networks. The dataset was obtained from the original COCOMO (1981).	The results of the fuzzy logic model were better than those of the COCOMO equations. The FLM was based upon triangular membership functions. The main benefit of this model is its good interpretability by using the fuzzy rules.
Saliu et al. [7]	2004	They fuzzyfied the two different portions of the intermediate COCOMO model i.e. nominal effort estimation and the adjustment factor. They proposed a fuzzy logic framework for effort prediction by integrating the fuzzified nominal effort and the fuzzified effort multipliers of the intermediate COCOMO model.	This approach is able to deal with uncertainty, provides transparency on prediction rationale through rules, incorporate experts knowledge in the definition of membership functions and rules, as well as adaptable to new data by changing the parameters of membership functions.
Ahmed et al. [25]	2004	Presented a FLM based upon triangular membership functions. The dataset for validating the FLM was (a) generated randomly and (b) that of COCOMO 81 was used.	Results showed that the FLM was slightly better than COCOMO equations. In addition, they reported promising experimental summary results in spite of the little background knowledge of the rule base and training data.
Crespo et al. [26]	2004	Explored fuzzy regression techniques based upon fuzzification of input values. Project database of COCOMO-81 are used.	fuzzy regression is able to obtain estimation models with similar predictive properties than existing basic estimation models.
Braz et al. [27]	2004	Applied Fuzzy Logic for effort estimation of object-oriented software. FUSP (Fuzzy use case size points) metric allows gradual classifications of use case size points in the effort estimation by using fuzzy numbers.	Results showed that FUSP fares better than USP.
Xu and Khoshgoftaar [28]	2004	Presented a fuzzy identification cost estimation modeling technique to deal with linguistic data, and automatically generate fuzzy membership functions and rules. A case of study based on the COCOMO'81 database compared the proposed model with all three COCOMO'81 models (basic, intermediate and detailed).	It was observed that the fuzzy identification model provided significantly better cost estimations than the three COCOMO'81 models.
Cuauhtemoc et	2006	Carried out a study to compare personal Fuzzy	Results show that a FLS can be used as

al. [29]		Logic Systems (FLS) with linear regression using evaluation criteria which is based upon ANOVA of MRE and MER, as well as MMRE, MMER and pred(25)	an alternative for estimating the development effort at personal level.
Moon Ting Su et al. [30]	2007	Proposed an enhanced fuzzy logic model for the estimation of software development effort. The model <i>Fuzzy Logic Model for Software Development Effort and Cost Estimation</i> (FLECE) possesses similar capabilities as the previous fuzzy logic model. In addition to that, the enhancements done in FLECE improved the empirical accuracy of the previous model in terms of MMRE (Mean Magnitude of Relative Error) and threshold-oriented prediction measure or prediction quality (pred).	The analysis of the results shows that FLECE is able to obtain more accurate results in the estimation of software development effort when compared to the previous fuzzy logic model. Hence, the enhancements to FLECE are truly useful and had given better performance to the model.
Iman Attarzadeh and Siew Hock Ow [31]	2009	Proposed an enhanced Fuzzy Logic approach for the estimation of software development effort.	Results showed that the value of MMRE applying their Fuzzy Logic model was substantially lower than MMRE values as calculated by applying other Fuzzy Logic models.

III. PROPOSED FRAMEWORK FOR EARLY STAGE EFFORT ESTIMATION USING FUZZY LOGIC

Our proposed Early Stage Software Development Effort Estimation (ESSDEE) Fuzzy Logic Model (as shown in Figure 1 below) imbibes all the benefits of Fuzzy Logic applications in developing estimation models as discussed in the previous section. This model is a novel proposal and the work to validate the model based on project data of (B.Tech. Final Year) students is under progress and it is planned to establish the model with input data from industry sources as well at a later stage. In this model we have considered Entity Relationship Diagrams, Use-Case Diagrams and Class Diagrams as the conceptual modeling tool/technique used by students to capture the project requirements.

Students were monitored during their major projects and timesheet was maintained which gave us their actual development time (Effort in months/hours). As the prediction accuracy of effort estimation at the early stage of development is difficult due to the uncertainties involved we will consider the linguistic values for rule based inference component of our model. The model consists of three important components namely, Fuzzification block, Rule based inference and Defuzzification block. The inputs to the fuzzification block will be from various conceptual modeling parameters taken from the student's major project data. Linguistic values are supplied to the Rule Based Inference component and we can obtain a Fuzzy Effort as another linguistic value or a Crisp Effort after Defuzzification.

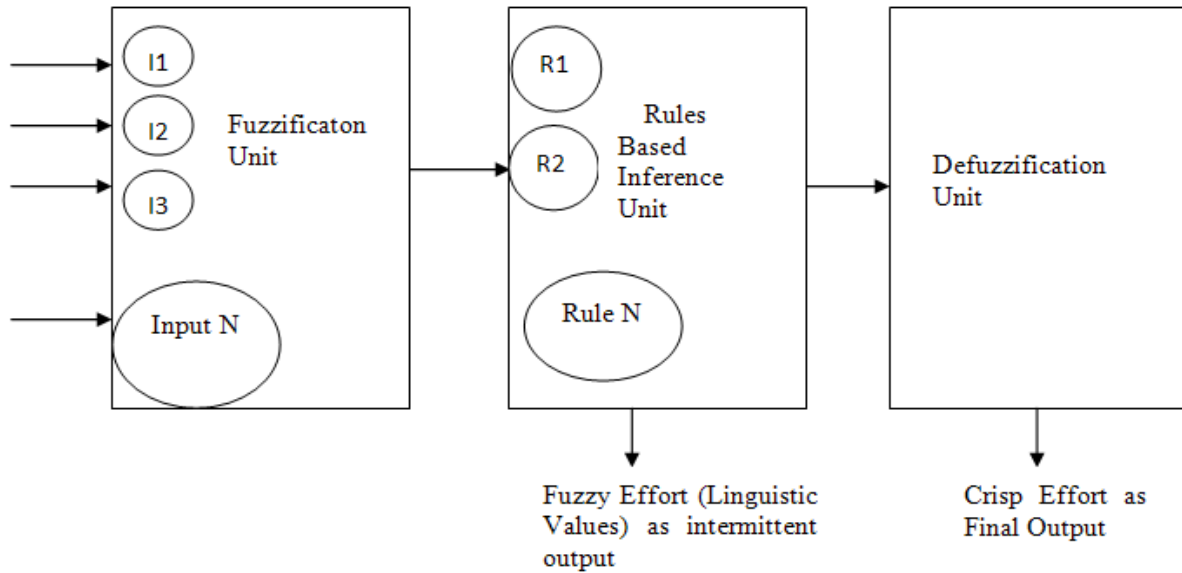


Figure 1: Proposed Fuzzy Logic based Framework to estimate the Effort of Software Project development

Following Table 2 describes the Data Capture Format from student's data.

Table 2
Student Data Collection Format

Conceptual Modeling Technique		
ER Diagrams	Use Case Diagrams	Class Diagrams
Total Number of Entity Sets	Total Number of Use-Cases	Total number of Classes
Total Number of Attributes	Total Number of Actors	Total count of attributes of a class
Relationship Types and their total count		
1:1	Type and Number of Communications <<extend>> <<include>> relations	Class Relationships
1:M		
M:1		
M:M		
Number of Weak Entity sets	Project Duration in Hours	Project Duration in Hours
	Project Category	Project Category
	CGPA of student	CGPA of student
Project Category		
CGPA of student		
Project Duration in Hours		

Based on the above data collection format student's data is collected and linguistic values (*High, Medium, Low*) will be established for each of the parameters for different conceptual modeling techniques. The proposed model is

diagrammatically shown in Figure 1 above which will make use of the above data from Table 2 for making out the effort

estimation in project development at the early stages of development.

IV. CONCLUSION

In this paper we have presented a review on the Fuzzy Logic applications in Software development effort estimation models development. We also discussed the various advantages of Fuzzy Logic for developing prediction models. Finally we propose a novel method for carrying out

the effort estimation at the early stages of software development namely the Requirements Gathering/Analysis and Design phases of Software Development Life Cycle. We plan to validate our model based on the data collected from student's project data over a span of past 3 years and will fine tune our model with large input sets. Finally we intend to establish the model based on the actual data from the software development industry.

V. REFERENCES

- 1) Luiz Fernando Capretz and Venus Marza, Improving Effort Estimation by Voting Software Estimation Models Hindawi Publishing Corporation Advances in Software Engineering Volume 2009
- 2) Grimstad, S., Jorgensen, M., Molokken-Ostfold, K., Software Effort Estimation Terminology: The Tower of Babel. Information and Software Technology. Elsevier, 2005.
- 3) Brooks Jr., F.P., Three great challenges for half-century-old computer science. Journal of the ACM, 2003.
- 4) B. W. Boehm, Software Engineering Economics, Englewood Cliffs, NJ, Prentice Hall, 1981.
- 5) L. H. Putnam, A general empirical solution to the macrossoftware sizing and estimating problem". IEEE Transactions on Software Engineering, SE-4(4), 1978, pp 345-361.
- 6) Attar Software. Fuzzy Logic in Knowledge Builder", White Paper, <http://www.intellicrafters.com/fuzzy.htm>, 2002.
- 7) M.O. Saliu, M. Ahmed and J. AlGhamdi. Towards adaptive soft computing based software effort prediction" Fuzzy Information, 2004. Processing NAFIPS '04. IEEE Annual Meeting of the North American Fuzzy Information Processing Society, 27-30 June 2004, Vol.1, pp. 16-21.
- 8) Witold Pedrycz. Computational Intelligence as an Emerging paradigm of Software Engineering. SEKE 2002, Ischia, Italy
- 9) T.T. Moores, and J.S. Edwards, Could Large UK Corporations and Computing Companies Use Software Cost Estimating Tools? - A Survey". European Journal of Information Systems, Vol.1, No.5, 1992, pp.311-319.
- 10) F.J. Heemstra, R.J. Kusters. "Controlling Software Development Costs: A Field Study". International Conference on Organisation and Information Systems, Bled, Yugoslavia, 1989.
- 11) A. Idri, A. Abran,, T.M. Khoshgoftaar. Estimating software project effort by analogy based on linguistic values" in.Proceedings of the Eighth IEEE Symposium on Software Metrics, 4-7 June 2002, pp. 21 – 30.
- 12) L.A. Zadeh, Fuzzy Logic, Neural Networks, and Soft Computing", Communication of ACM, Vol. 37, No.3, March 1994, pp.77-84.
- 13) P. Musflek, W. Pedrycz, G. Succi, M. Reformat, "Software Cost Estimation with Fuzzy Models", Applied Computing Review, Vol. 8, No.2, 2000, pp. 24-29.
- 14) A. R. Gray, S. G. MacDonell, Fuzzy Logic for Software Metric Models throughout the Development Life-Cycle" in Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society – NAFIPS, 10 – 12 June 1999, New York, USA, pp. 258 – 262.
- 15) S. G. MacDonell, A. R. Gray, J. M. Calvert. FULSOME: Fuzzy Logic for Software Metric Practitioners and Researchers" in Proceedings of the 6th International Conference on Neural Information Processing ICONIP'99, ANZIIS'99, ANNES'99 and ACNN'99, Perth, Western Australia, IEEE Computer Society Press, 1999, pp.308-313.
- 16) J. Ryder, Fuzzy Modeling of Software Effort Prediction" in Proceeding. of IEEE Information Technology Conference, Syracuse, NY, 1-3 Sept 1998, pp: 53-56.
- 17) Pedrycz, W., Gomide, F., 1998. An Introduction to Fuzzy Sets. The MIT Press.
- 18) Lewis, J.P., 2001. Large limits to software estimation. ACM Software Engineering Notes 26 (4), 54–59.
- 19) [Fei Z, Liu X, f-COCOMO Fuzzy Constructive Cost Model in Software Engineering in IEEE international conference on Fuzzy systems, 1992, pp. 331-337.
- 20) S. Kumar, B.A. Krishna, and P.S. Satsangi, "Fuzzy systems and neural networks in software engineering project management". Journal of Applied Intelligence, Vol. 4, 1994, pp. 31-52.
- 21) A. R. Gray, S. G. MacDonell, Applications of Fuzzy Logic to Software Metric Models for Development Effort Estimation". Fuzzy Information Processing Society 1997 NAFIPS' 97, Annual Meeting of the North American, 21 – 24 September 1997, pp. 394 – 399.
- 22) J. Ryder, Fuzzy Modeling of Software Effort Prediction" in Proceeding. of IEEE Information Technology Conference, Syracuse, NY, 1-3 Sept 1998, pp: 53-56.
- 23) A. Idri, A. Abran, L. Kjiri. „COCOMO cost model using Fuzzy Logic", 7th International Conference on Fuzzy Theory & Techniques, 27 Feb – 3 March 2000, Atlantic City, New Jersey.
- 24) Huang, X., Capretz, L.F., Ren, J., Ho, D.A., Neuro-Fuzzy Model for Software Cost Estimation. In: Proceedings of the Third International Conference on Quality Software, 2003.
- 25) Ahmed, M.A., Saliu, M.O., AlGhamdi, J., Adaptive Fuzzy Logic-Based Framework for Software

- Development Effort Prediction, Information and Software Technology. Elsevier, 2004.
- 26) Crespo, F.J., Sicicila, M.A., Cuadrado, J.J., On the use of fuzzy regression in parametric software estimation models: integrating imprecision in COCOMO cost drivers. WSEAS Transactions on Systems, 2004.
 - 27) M.R. Braz, S.R. Vergilio, Using fuzzy theory for effort estimation of object-oriented software". 16th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2004, 15-17 Nov 2004, pp. 196 – 201.
 - 28) Xu, Z., Khoshgoftaar, T.M., 2004. Identification of fuzzy models of software cost estimation. Elsevier Fuzzy Sets and Systems 145 (1), 141–163.
 - 29) L.M. Cuauhtemoc, Y.M. Cornelio, G.T. Agustin. Fuzzy Logic Systems for Software Development Effort Estimation Based Upon Clustering of Programs Segmented by Personal Practices", Electronics, Robotics and Automotive Mechanics Conference (CERMA'06), 2006, pp. 367-372.
 - 30) Moon Ting Su, Teck Chaw Ling, Keat Keong Phang, Chee Sun Liew, Peck Yen Man, ENHANCED SOFTWARE DEVELOPMENT EFFORT AND COST ESTIMATION
 - 31) USING FUZZY LOGIC MODEL, Malaysian Journal of Computer Science, Vol. 20(2), 2007
 - 32) Iman Attarzadeh and Siew Hock Ow, Software Development Effort Estimation based on a New Fuzzy Logic Model, International Journal of Computer Theory and Engineering, Vol 1. No. 4 October 2009.