Software Effort Estimation Framework in Iterative and Incremental Development; Issues and Proposed Solution

Abou Bakar Nauman, Rommana Aziz

Sarhad University of Science and Information Technology,
Peshawar, Pakistan
Dept of Computer Sciences COMSATS Institute of Information Technology Islamabad, Pakistan,
Email: anauman@comsats.edu.pk, abubakar.csit@suit.edu.pk

ABSTRACT

Software effort estimation is now dynamic issue, with the values of factors are updated frequently specially in iterative projects. The cone of uncertainty in project estimates is a reality, however the project does not follow the cone it self. The visibility of cone of uncertainty is more in iterative projects than in sequential, which require use of data from current project. There exist some models which provide the capability to learn and update these factors. However these models have some limitations, which should be addressed to develop an intelligent effort estimation model. This paper highlights some important issues in IID effort estimation and also provides a review of existing models. This paper also proposes a framework/ model to implement idioms in dynamic Bayesian networks for development of estimation model for iterative projects.

Keywords-Iterative development; software estimation, uncertainty.

I. INTRODUCTION

The IID project management divides the artifacts in layers with prime objective of risk management. The risks involved with effort and schedule require an adaptive planning to mitigate. Rather than a single layer the management can have more than one layer, with each layer the set of artifacts are repeated [1]. Top three layers, Program Management, Project management and the day to day development management, are concerned with the whole project. However the iteration level management is focused on the ensuring the control and monitor the evolution of product deliveries. In IID project this layer is the most active layer. All the work is being done in the iteration layer, for one iteration the data being produced and is used for next iteration. The management at this layer is also carried out in the same manner as in project management, i.e. going through management cycle of Agree, Execute and Assess. However the cycle operates in relatively shorter time periods. The goals of each cycle are also affected by the risk reduction at each cycle [1,2].There exist many variations of agile or Iterative development methods, which have their own set of management cycles; however in each method the management cycle works in a loop [2,3]. Agree and execute phases are product driven, however the assessment is a task which can be ignored. This assessment phase is as crucial as the other phases of management cycles. Thus the project assessment is a very crucial part of iterative development. This depends on the "feedback" mechanism which is one of the most significant natural phenomenon. The assessment enables the team to learn from previous data and create the opportunities to keep the project on track. Without proper assessment procedures, the whole project can collapse. As we have discussed, estimation is foundation of project management, and this is more significant in iteration management layer and management cycles.

The evolving understanding of the requirements, the architecture, and the team's ability enables additional estimating models and methods to be applied. Collection of metrics from the completed iterations enables the estimating models to be calibrated and future trends to be predicted. The figure below illustrates the trends we expect to see for revised estimates produced as the evolution progresses through the iterations and phases [4].

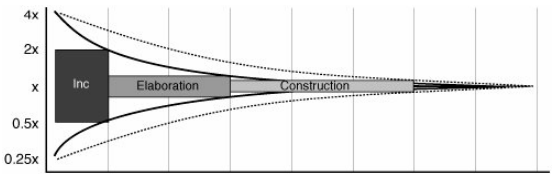

Figure 1. Cone of Uncertainty

The researchers have presented modified models of traditional estimation models to adjust in the Iterative developments e.g. [5]. Some researchers have also worked to develop the estimation model with iterations of data calibration and variable inclusion [6,7,8]. The tool SEER-PPP have the capability to adjust the further estimations based on latest data provided.

II. MAJOR ISSUES OF EFFORT ESTIMATION IN IID PROJECTS

Some burning issues in estimation research are development of new models, metrics conversion, uncertainty, missing data, intelligent decision support and models for new life cycles [4,5,6, 8-11]. In this section we try to identify and discuss various issues which arises in effort estimation in general and IID effort estimation in particular. Effort estimation is based on two major factors; size and productivity, thus issues regarding effort estimation are normally concerned with size or with productivity. It has been established that the estimation is an important part of project planning in case of linear project, and if the project is iterative the estimation becomes more important as after each iterations the project plan is to be revised according to new assessments. We can say that the issues which obstacles in the linear project estimation become more in volume as we step into Iterative development paradigm. Some issues also become more controlled in the case of Iterative development as the size of project becomes smaller in iterations [10].

Effort estimation is active area of research, where each year there are number of novel dimensions and ideas are presented for solution of burning issues. Some research groups are active in identifying the degree of need in different areas of research in estimation process.

a. Accuracy:

The issue of highest priority in any estimation model is accuracy. There are number of methods and research contributions in development of assessment models for measuring and benchmarking the accuracy of effort estimation models. Magnitude relative error, Basic relative error and pred(n) are most commonly used accuracy assessment techniques. Then there are some researches which have tried to improve the MRE and BRE models to improve the assessment methods. Biasness is also a related issue to accuracy. Biasness represents the degree of underestimation or over estimation. This measure is useful in calibration the estimation for providing better estimated effort. Calibration is a process of tuning or optimizing the estimation model on the basis of accuracy measures.

b. Data sets:

Another issue is availability of data set. It has been matter of great concern for many years that not sufficient data was present for software development projects. It is interesting to note that some major estimation models are developed using relatively smaller data sets. Recently some major data collection and benchmarking services have increased their effort in this area. These organizations have presented multidimensional data for software development projects. Meanwhile with the variation in development methods, tools and lifecycles there exists a strong need for collecting recent data of software development projects. In case of IID projects, the issue is to collect most recent data of current project and update the estimations.

c. Learning:

One of the reasons for failure of projects is failure to execute current estimation. Many of the factors involved in estimation process, get changed in value and importance, by the time, and some remain consistent [10]. Estimation must also be seen as a range of values, rather than a single value. Learning, or the ability to update the model with latest data, is considered as key attribute of an estimation model. The term learning is also used to represent the intelligence in a model to select the appropriate classification of available data to process the estimation. Some models have used Case-base reasoning, Neural networks and Fuzzy logics to implement the learning process [6,7]. The results of accuracy measures are used to calibrate the estimation models in these techniques. The implementation of this technique has two major drawbacks, first the learning process is not an ongoing process, and secondly its implementation in IID projects is not analyzed.

d. Uncertainty:

Uncertainty is another factor which needs to be represented in estimation models [3-11]. An estimate must present the degree on belief with the most likely value. Prediction interval is one of the ways to represent the uncertain values. With this interval the accuracy measures are more practical. One of the major issues is processing the uncertain values in mathematical or classification models. To manage uncertainties fuzzy functions and Bayesian networks are successful. These are capable of receiving, processing and producing results with uncertain data. In IID the uncertainty reduces with

each of the iteration completed, however it doesn't completely vanishes until the project is completed. The estimation models hence require representing the change in uncertainty. The cone of uncertainty, which also exist in the iterative development, becomes narrower with the iterations developed. As the scope of earlier iterations is also smaller, the volume of uncertainty is also less in these iterations in comparison with linear projects. In the case of IID estimation, the issues which have been discussed above have their own significance. The data set of multiple iterative projects is hard to find or data sets available are too small to represent broadly. Learning also has to be done by the latest data of current project. The process of learning also has to be quick and automated. Rather than an offline learning process, the model should be able to learn from most recent data. The degree of certainty also has to be increased with each actual data item close to estimated values.

e. Size:

Size is basic factor in effort estimation. However the actual size is not known until the project is completed, hence requiring effort estimate to base on another estimated quantity. In IID the whole project is divided into smaller projects hence enabling the estimators to learn more certain values of size. The size in Iterative projects is however influenced by some breakage or integration costs, which increases the size-effort ratio. The usable size of project decreases, however the size required to be built remain same. Another issue is reusability. Some of the modules developed earlier in the project can be reused in the later stages of project, which make a positive effect on the productivity. At initial stages of project e.g. Inception, the architectural/functional complexity of project is higher than the next phases, which also affect the functional size f the project. Thus contrary to a single value of size, the size estimate also possesses uncertainty which can be represented by probability statement. An estimation model should be capable of receiving estimated size with some uncertainty value.

III. RELATED WORK

A number of researches have been conducted to improve learning and intelligence in estimation models [6,7,12,13]. However these learning models are not for Iterative and Incremental development. These models don't provide the evidence of their capability of working in IID projects. One of the major draw back in these models is the lack of adaptive-ness, i.e. "Adaptability to cope with continuous change in development technologies and environments" [12]. In case of IID project, there can be frequent changes and adaptive-ness can help the model to learn the values of significant variables. However lack of this characteristic is evident in several models. The model presented by [13] is also not developed and tested for IID estimation. There are some general reviews on software estimation by [7-9]. We conduct this review on the basis of proposed models for iterative estimation model and their capability to provide learning mechanism. Interestingly there are quite a few researches for estimation iterative development. The table below summarizes the researches available on digital libraries for the terms "Iterative" AND "Software" AND "Estimation". It is evident that need for IID estimation models with learning capability is highly desirable. The framework proposed by [15] and the implementation in regression is significant in the area.

Table: 1. REVIEW FINDINGS

Title	Findings
"COCOMO-Based Effort Estimation for Iterative and Incremental Software Development"[5]	o Highlighted the need for effort estimation in IID. o Proposed an enhancement of COCOMO model. o Introduction of new variables for IID projects o No automatic learning.
"Project Estimation: A Simple Use-Case-Based Model"[14]	o Highlighted the need for estimation in iterative development o Proposed the framework to refine the estimation model on the basis of latest results. o Proposed Use case point for effort estimation.
"Effort Prediction in Iterative Software Development Processes – Incremental Versus Global Prediction Models"[15]	o Highlighted the need IID estimation model. o Proposed mechanism of calibration of estimation model o Used Regression models to calibrate o Raised the need for application of Bayesian networks to develop such models.
"Adaptive fuzzy logic-based framework for software development effort prediction"[12]	o Highlighted the need for adaptive learning in Estimation models o Proposed a fuzzy logic based estimation model.

	o The model is trained by latest data o Model is not tested or validated for IID projects.
"TUPUX: An Estimation Tool for Incremental Software Development Projects"[16]	o Not learning based o Based on Function point o Addresses the issue of increments in size
"The size and effort estimates in iterative development"[17]	o UML based o Size estimate is improved after first iteration o Feedback based o Mathematical computation required by the user to improve the estimates
"Development of a Hybrid Cost Estimation Model in an Iterative Manner"[18]	o Highlighted the need for calibration of estimation model. o Proposed the framework for calibration in iterative manner. o Calibrated the causal model. o Factors of calibration are increased with each iteration (The term iteration is used for calibration cycle)

Some significant limitations in these researches are:

"The learning mechanism is passive"

OR

"Evidence of learning models in IID projects is not available"

OR

"Uncertainty is not considered".

Thus there is not a single research which deals with all of issues in a single model. It can be easily understood that Iterative and Incremental development require frequent assessment of project, and feedback to next iteration plan. Primary techniques required for estimation in IID projects not only include Initial estimates but Calibration by historic data and use of data from current project is also required [7]. The estimation tools claiming to support the iterative development, their support is either up-to introduction of new factors or changed values for these factors. There is no intelligent enough estimation tool which promises to learn automatically from latest data. The learning mechanism provided by some tools is based on personal learning of the project, or training of model on non-related data set. For example in COCOMO, the project manager can revise the selected values of parameters, however he can't be sure how much his selection is correct. Similarly the Wide band Delphi method requires the estimators to revise the estimated values of effort required on the basis of new data or information. The models provided by [14] and [19] support the IID effort estimation; however these models don't provide the feature of learning form latest data. Thus we conclude that there exists a strong need to develop an effort estimation model which is capable of learning from completed iterations in IID project but also provide initial estimate and calibration on historic data.

IV. PROPOSED FRAMEWORK

Need of implementation of estimation models with causal models and Bayesian Networks has already been highlighted by Fenton in his various researches. Pekka Abrahamsson et. al. [15] has also called for implementation of IID estimation model with Bayesian networks. As discussed earlier the need of uncertainty management and that of learning is significant for an effort estimation model. It is interesting to note that Bayesian Statistics and Bayesian Networks are being used in managing uncertainty, calibration, learning and prediction with respect to effort estimation. However evidence of learning the productivity of a team using Bayesian networks is not available. The application of Bayesian Networks, Dynamic Bayesian Networks and Hybrid networks has also been encouraged in decision support systems and multi-criteria decision making. This research is significant in this context that it provides the evidence of implementing the problem of effort estimation in IID projects with Bayesian Networks.

Prime objective of an effort estimation model is to provide accurate estimates. In iterative development the estimates are

not only required at initial stage of project but also after each of the iteration completed, estimates for further iterations can be revised [7]. Estimation tools are built on the basis of industrial data which can be further calibrated to provide more accurate estimates for specific type of projects [7]. Factor used in software estimation range up to 100 technical or organizational factors, however a count of 15-20 factors is found in some traditionally used models [20,26]. For implementation of Bayesian Nets the issue of sensitivity (how much effect a node can make with small change of its value) is significant. As the number of nodes (a node representing one factor or combination) increase there are two type of effects on the inference; firstly the time required for propagation is increased, secondly the sensitivity of nodes is decreased. Hence in a large model, there can be nodes which actually reflect no significant change even if these are part of the causality.

We advocate the need of development of some useful components, which can be joined together to develop a dynamic Bayesian network to be implemented for estimation in IID projects. The figure below (2) provides a conceptual diagram for selecting the components and using them for estimation at different stages of project planning.

V. CHARACTERISTICS

Modeling a dynamic Bayesian network require understanding of available of objects and links inside and beyond the time slices. While developing the Bayesian network for effort estimation in Iterative and Incremental development projects one need to join these objects and links such that solution to our research question can be found. We thus propose that the implementation of Bayesian Networks should be made such a way that smaller components are identified and structured according to project management need. The estimation should be available for project level as well as Iteration level. There should be mechanism for evidence recording for completed iterations. The results of completed iterations (I 0---n-1) to be forwarded to estimation process of next iterations (I n). however the learning should also be controlled intelligently.

The development of such model also includes implementation of project scenarios, validation of results and development of GUI for easy result interpretation. The figure 3 is proof of implementation of this framework using AgenaRisk software. The model is developed using an idiom of Measure-Size-Effort nodes. The nodes are connected with a series of productivity nodes for iterations.

Table 4.1. PROPOSED FRAMEWORK/MODEL **features.**

Sr #	Title	Objectives
1	Composition	Identification of units of proposed model to be used in IID effort estimation model development. A basic model can be used multiple times.
2	Structuring	The different units of the model can be joined together using dynamic Bayesian networks
3	Project Level Estimate	Project level estimates can be collected before start of project.
4	Evidence Recording	The evidence of latest actual effort can be submitted in model
5	Results Interpretation	The results from the model can be interpreted using probability distributions.
6	Results forwarding	The results from one time slice can be forwarded to next slices using dynamic Bayesian Networks
7	Unrolling	The procedure to use multiple instances of basic model.
8	Learning	Options to control the degree of learning from latest evidences will be provided with help of weight parameters and variance.
9	Project Scenarios	The model should be implemented in different project scenarios to see its applicability.
10	Validation	To analyze capability of model, this model should be implemented in industrial project.

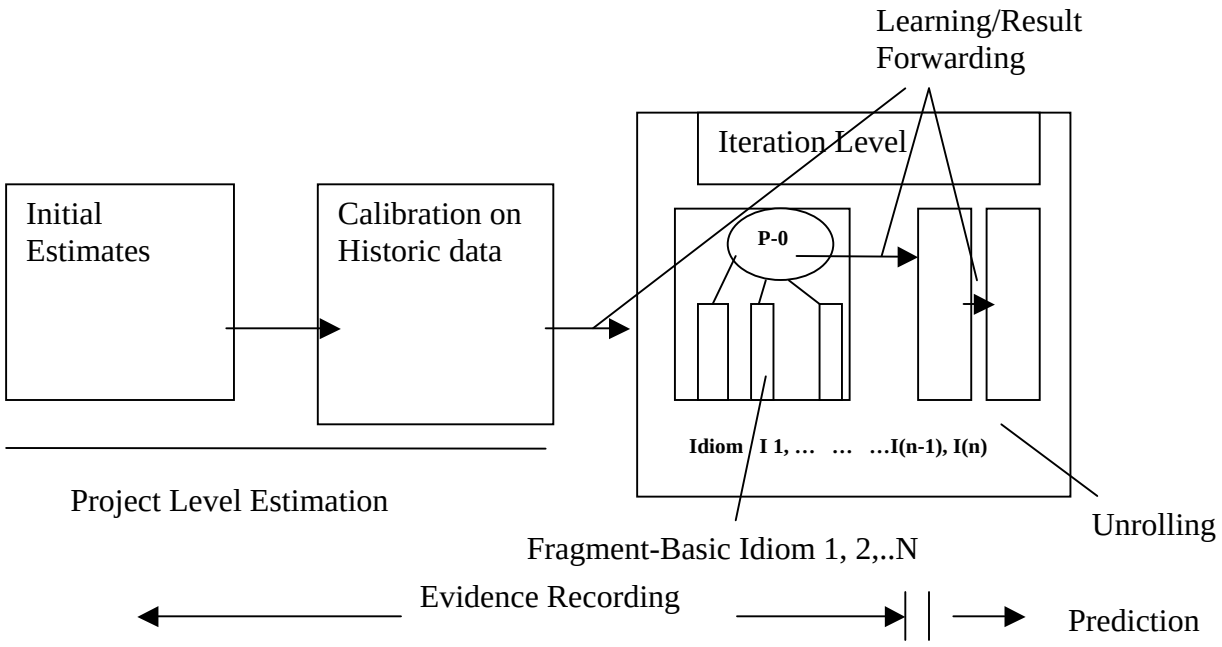

Figure 2. Framework for IID DBN development

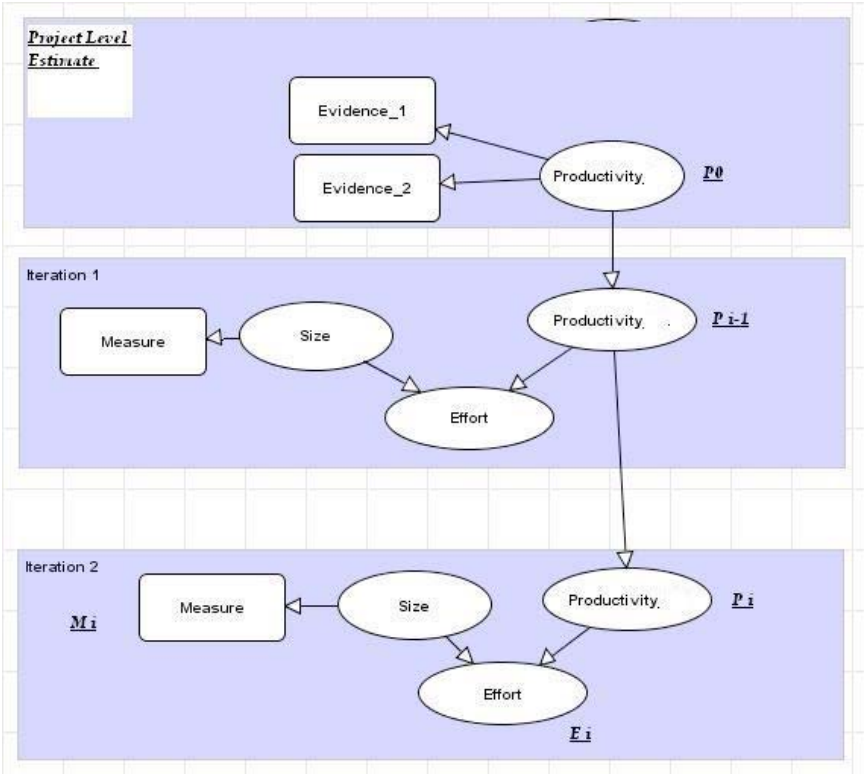

Figure 3. Implementation in AgenaRisk

VI. CONCLUSION:

By the review of literature two significant issues are highlighted; First the issue of Iterative project estimation is not similar to traditional software project estimation. We need intelligent estimation models which have to have learning characteristics with ability to manage uncertainty. Secondly there is hard to find an estimation model which posses an active learning mechanism while managing uncertainty and providing evidence of learning in IID projects. There exist a gap in knowledge and the area requires further research. The proposed framework is significant in development of the dynamic Bayesian network in IID estimation process. This framework attempts to bridge the knowledge gap by providing a roadmap for model development.

REFERENCES

[1]. Steve McConnell, "Software Estimation: Demystifying the Black Art", Microsoft Press,2006.

[2]. Walker Royce, " Software Project Management, A Unified Frame work" Pearson Education, 2000

[3]. Richard Bechtold, 1999, "Essentials of software project management", Management Concepts inc.

[4]. Little T, "Schedule Estimation and Uncertainty surrounding the cone of uncertainty", IEEE Software may/June 2006

[5]. Oddur Benediktsson, Darren Dalcher, Karl Reed, Mark Woodman "COCOMO Based Effort Estimation for Iterative and Incremental Software Development, Software Quality Journal, 11, 265–281, 2003

[6]. Mohammad Azzeh et al. "Software Effort Estimation Based on Weighted Fuzzy Grey Relational Analysis", ACM 2009

[7]. Andrew R. Gray, Stephen G. MacDonell, A comparison of techniques for developing predictive models of software metrics, Information and Software Technology 39 (1997) 425-437

[8]. Boehm B., C. Abts and S. Chulani 2000, "Software development cost estimation approaches—A survey", Annals of Software Engineering 10, pp. 177–205.

[9]. Jorgensen, M.; Shepperd, M. 2007, "A Systematic Review of Software Development Cost Estimation Studies", IEEE Transactions on Software Engineering.

[10]. Kurt Bittner, Ian Spence,(2006) "Managing Iterative Software Development Projects", Addison Wesley Professional.

[11]. Liu, Qin; Mintram, Robert C. (2005), "Preliminary Data Analysis Methods in Software Estimation", Software Quality Journal.

[12]. Moataz A. Ahmeda, Moshood Omolade Saliub, Jarallah AlGhamdia, "Adaptive fuzzy logic-based framework for software development effort prediction", Information and Software Technology 47 (2005) 31–48

[13]. G. Witting and G. Finnie, "Estimating software development effort with connectionist models," in Proceedings of the Information and Software Technology Conference, pp. 469--476, 1997.

[14]. Ray Ashman, "Project Estimation: A Simple Use-Case-Based Model", IT Professional, v.6 n.4, p.40-44, July 2004

[15]. Abrahamsson, P.; Moser, R.; Pedrycz, W.; Sillitti, A.; Succi, G.; VTT Electron., Oulu, "Effort Prediction in Iterative Software Development Processes -- Incremental Versus Global Prediction Models", Empirical Software Engineering and Measurement, 2007. ESEM 2007.

[16]. Balbin, D.; Ocrospoma, M.; Soto, E.; Pow-Sang, J.A.; "TUPUX: An Estimation Tool for Incremental Software Development Projects", International e-Conference Advanced Science and Technology, 2009. AST '09.

[17]. Marjan Heričko, Ales ivkovič , "The size and effort estimates in iterative development", Information and Software Technology Journal, Volume 50 Issue 7-8, June, 2008

[18]. Adam Trendowicz, Jens Heidrich, Jürgen Münch, Yasushi Ishigai, Kenji Yokoyama, Nahomi Kikuchi, "Development of a hybrid cost estimation model in an iterative manner" ICSE'06, May 20-28, 2006, Shanghai, China.

[19]. Parastoo Mohagheghi, Bente Anda, Reidar Conradi. "Effort Estimation of Use Cases for Incremental Large-Scale Software Development", 2005, Proc of the 27th Int'l Conf. on Software Engineering. St Louis