

EXTENDED ANGEL: KNOWLEDGE-BASED APPROACH FOR LOC AND EFFORT ESTIMATION FOR MULTIMEDIA PROJECTS IN MEDICAL DOMAIN

Sridhar S

Associate Professor, Department of Information Science and Technology,
Anna University, Chennai, India.
ssridhar2004@gmail.com

ABSTRACT

Developing good multimedia applications is expensive, mostly in terms of time and degree of difficulty. Estimating the cost and effort for multimedia application development is a great challenge. One of the useful techniques for predicting effort and lines of code (LOC) required for developing multimedia projects is by using the technique of “estimation by analogies”. This technique involves characterizing the project for which an estimate is required. This characterization forms the basis for finding similar or analogous projects that have been completed before and for which effort and LOC values are known. The tool ANGEL uses analogy for cost and effort estimation. However the accuracy of estimates generated through this tool is poor under certain conditions. It is proposed to overcome the limitations of ANGEL and to improve the accuracy of estimation by using a knowledge based technique. In knowledge-based estimation, estimates are generated using techniques of both ANGEL and knowledge rules.

To analyze the performance of the proposed knowledge based estimation tools, the source code of 20 multimedia projects in medical domain have been taken as test cases. For these projects various metric attribute values useful for cost and effort estimation are computed and the estimation is done using the tool ANGEL. Then the estimation is improved using the knowledge rules. The results of the knowledge based tool and ANGEL tool are compared to assess the performance improvement in the estimation process of the proposed technique.

KEYWORDS

Knowledge based estimation, ANGEL, Estimation by Analogy, LOC and Effort Estimation

1. INTRODUCTION

Effective Project management requires the correct estimation of cost and effort. Most of the software projects fail due to the wrong estimates of cost and effort. So much emphasis is placed on the estimation process related to cost and effort. The objective of the work is to develop knowledge-based LOC and effort prediction model for multimedia applications. The proposed model improves the accuracy of estimates generated by ANGEL tool using the knowledge rules. Authoring of multimedia applications encompasses the management activities for the actual

International Journal of Software Engineering & Applications (IJSEA), Vol.2, No.4, October 2011
 content and application development. Thus multimedia application development covers the methods used for generating the structure and functionality of the multimedia applications. The metric attributes that could influence the effort and LOC value of the applications are classified into five groups [1] and are listed below.

LENGTH SIZE METRICS	EFFORT METRICS
<ul style="list-style-type: none"> • Page count • Media count • Program count • Total page allocation • Total media allocation • Total code length 	<ul style="list-style-type: none"> • Structuring effort • Inter linking effort • Interface planning • Interface building • Link-testing effort • Media-testing effort • Total effort
REUSABILITY METRICS	COMPLEXITY AND SIZE METRICS
<ul style="list-style-type: none"> • Reused media count • Reused program count • Total reused media allocation 	<ul style="list-style-type: none"> • Connectivity • Connectivity density • Total page complexity • Cyclomatic complexity
CONFOUNDING FACTORS	
ENTITY TYPE	METRIC
<ul style="list-style-type: none"> • Developer 	<ul style="list-style-type: none"> • Experience
<ul style="list-style-type: none"> • Tool 	<ul style="list-style-type: none"> • Type

The aim is to extract these attribute values from the real multimedia applications in medical domain. The tool ANGEL is then used to predict the cost and effort estimation. Then the proposed knowledge based system is used to tune the estimated values by using knowledge rules. The knowledge rules are obtained by comparing the estimated values and actual values. It is expected that the usage of the knowledge rules help to increase the performance of the cost and effort estimates. The level of improvement is investigated in this paper.

2. ESTIMATION BY ANALOGIES

After identifying the metric attribute values, the effort and LOC estimate can be predicted using analogy. The technique of "Estimation by analogies " is discussed in the literature [2]. LOC is considered fundamental of all estimation process. The problems associated with LOC are discussed in the Literature [14]. ANGEL is an automated environment is used to predict effort for the new project using analogies and estimates that are generated from completed projects. The details of ANGEL tool is provided in the Literature [3]. ANGEL uses the historical data of effort and LOC of the completed past projects to predict the estimate of the new project.

The available software cost estimation techniques are described in the Literature [11]. The survey of the cost estimation techniques is provided in the literature [6]. Wrong estimates often lead to the failure of the projects. The empirical study linking the prediction and project success is discussed in the Literature [8]. Cost estimation can influence the other phases of the software development. The relation between project cost and effort estimate and project maintenance is discussed in Literature [10, 12]. Some of the latest techniques used include Support Vector machines [7]. Lots of studies have been carried out to compare the various estimation tools [13, 12] to find optimal tools for effect cost estimation. But none of the tools are satisfactory. ANGEL has proved to be effective tool [3]. But still there are some situations where the tool performs badly. It is thus proposed to tune the ANGEL tool by using knowledge rules. Knowledge rules are obtained using regression analysis.

2.1. ANGEL Estimation Process

The steps used by the tool ANGEL for cost and effort estimation are as follows

- Identification of multimedia components and collection of metrics
- LOC and effort estimation for the new project using analogy

Typically ANGEL uses historical data of all completed projects. Then the similarity between new project for which cost and effort estimate is required is determined by using "nearest neighbour algorithm".

Each project is plotted as a single point in an N-dimensional space where "N" is the number of attributes extracted from the project. ANGEL minimizes the Euclidean distance in N-dimensional space [5]. Thus by using the distance measure, the new estimates are predicted.

The following listed issues are must for successful estimation process using the tool ANGEL. The issues are

1. Best description of projects.
2. Confidence mechanism for analogies.
3. Mechanism to derive an effort estimate of new project from known effort values.

2.2. Limitations of Angel

However ANGEL is not a perfect tool. The following are some of the disadvantages of ANGEL.

- The effort estimate generated for the new project is the average effort value of the analogous projects
- Attribute values are standardized based on the assumption that each attribute contributes equally to the target attribute value
- No specific estimation mechanism to handle out of range values.
- No specific estimation mechanism to handle "multiple-exact-match" condition
- No estimation mechanism to handle "no-match" condition

As the result of the above mentioned limitations ANGEL tool does not provide valid estimate for out of range values, so the level of accuracy of estimate generated is poor. To overcome this limitation it is proposed to develop knowledge based estimation mechanism that makes use of a knowledge rules for estimation.

2.3. Knowledge Based Estimation

The basic idea behind knowledge-based estimation is generating knowledge rules.

- Analysis of performance of ANGEL for various test cases.
- Identify the cases for which ANGEL gives good estimate and the cases for which the estimate generated is poor
- Identify the reasons for the deviation of results from the actual values
- Formulate rules to improve accuracy of the estimate
- Develop knowledge based effort and LOC estimation
- Compare the estimate obtained from knowledge based estimation mechanism with the results of ANGEL

2.3.1 Rules Extraction Process

The rules are generated using the following method of Identifying the relationships of each predictor attributes with the target attribute.

In this process, the relation between the predictor and target attributes are identified using a tool named as DataFitX 1.0.X [5]. DataFitX is a COM component that can be used to perform nonlinear regression analysis. Since it is developed using COM technology, it can be used from within any development environment that supports COM, including Microsoft Visual Basic, Microsoft Office Products, Visual C++ and Delphi [5].

International Journal of Software Engineering & Applications (IJSEA), Vol.2, No.4, October 2011
The method of using regression analysis is listed below

- Identify the interdependencies between the predictor attributes
- Identify the most influencing factors and less influencing factors for each target attribute
- Fix the intervals for each predictor attributes based on the data base entries
- Generate estimate for out of range values by using regression analysis.
- Generate estimate for new pattern by using relationships obtained from regression analysis

This knowledge is converted into rules which are then used to estimate effort and LOC for new projects. The thorough knowledge of multimedia applications, Regression analysis are expected to overcome the weakness of ANGEL project leading to effectiveness of these rules.

The general block diagram of the proposed work is shown in Figure 1.

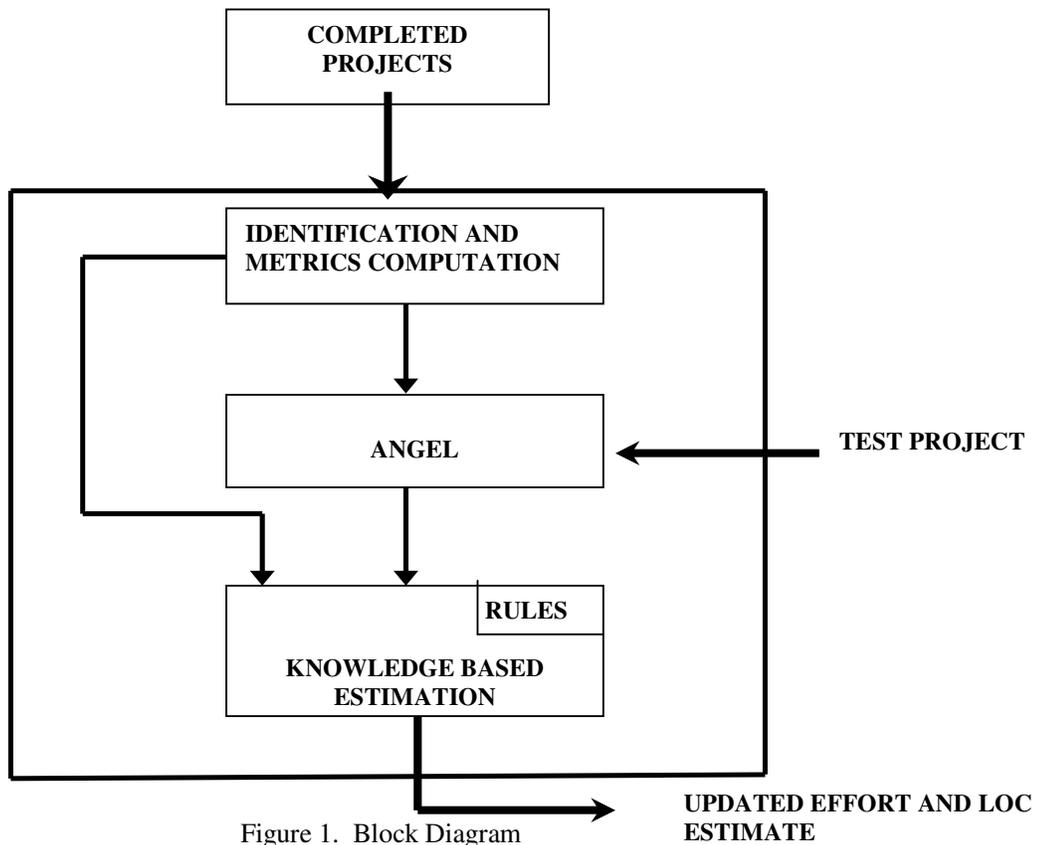


Figure 1. Block Diagram

3. EXPERIMENT

To analyze the performance of knowledge based estimation the source code of 20 multimedia projects has been taken related to medical domain. For these projects various metric attribute values are computed and the database is populated with these computed values. For various test cases the estimates generated through ANGEL and through knowledge bases estimation are recorded. These estimates are then compared with the actual effort and LOC values.

The results obtained for 6 sample test cases through ANGEL and through knowledge based estimation are shown in the following Table 1 and Table 2.

TABLE 1 RESULTS OBTAINED USING ANGEL TOOL

S.NO	PROJECT NAME	ESTIMATED LOC	ESTIMATED EFFORT	ACTUAL LOC (Calculated After project is over)	ACTUAL EFFORT (Calculated After project is over)
1	Renal	4796.67	1.54	850	1.24
2	Liver	72276.33	46.54	66840	38.24
3	Renal calculi	26220.33	19.40	29330	19.29
4	Gall stone	26220.33	19.40	38532	27.36
5	Ureteric stone	4796.67	2.52	1320	4.09
6	Heart plaque	4796.67	2.52	1350	3.57

TABLE 2. RESULTS OBTAINED BY KNOWLEDGE BASED ESTIMATION

S.NO	PROJECT NAME	ESTIMATED LOC	ESTIMATED EFFORT	ACTUAL LOC (Calculated After project is over)	ACTUAL EFFORT (Calculated After project is over)
1	Renal	797.13	1.438	850	1.24
2	Liver	89201.47	29.602	66840	38.24
3	Renal calculi	34277.53	20.762	29330	19.29
4	Gall stone	36662.89	22.046	38532	27.36
5	Ureter Stone	1861.82	7.080	1320	4.09
6	Heart plaque	2285.03	5.310	1350	3.57

International Journal of Software Engineering & Applications (IJSEA), Vol.2, No.4, October 2011
 Using the results mentioned in the above table graphs are plotted, shown in Figure 2 and 3. Each point in the graph corresponds to percentage of accuracy of estimate being generated for specific test case.

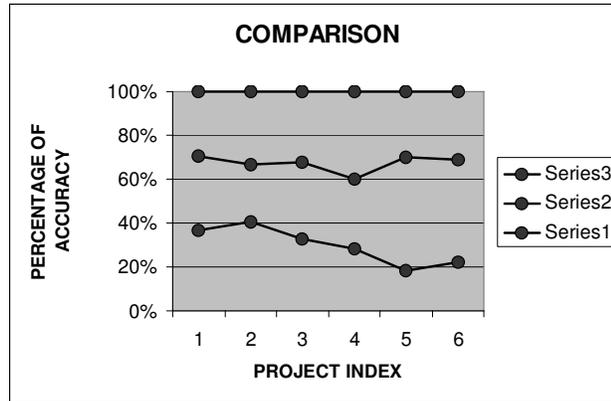


Figure 2. Comparison for LOC Estimates

Series 1: % of accuracy of estimate generated through ANGEL

Series 2: % of accuracy of estimate generated through knowledge-based estimation

Series 3: Ideal Estimate

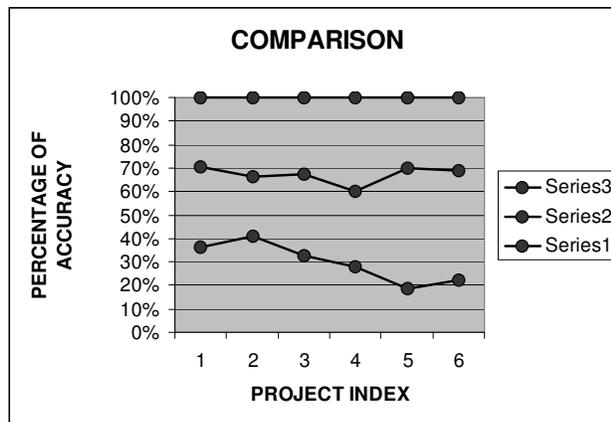


Figure 3. Comparison for Effort Estimates

Series 1: % of accuracy of estimate generated through ANGEL

Series 2: % of accuracy of estimate generated through knowledge-based estimation

Series 3: Ideal Estimate

4. CONCLUSION

Project Management is a complicated task. Predicting LOC is a tough task. Wrong estimates of LOC may lead to the wrong estimation of cost and effort. Often the project failure is attributed to the wrong prediction of cost estimate.

While there are many tools available, none of the tools are satisfactory. In this aspect, the tool ANGEL is satisfactory. In spite of its usefulness, there are many circumstances where ANGEL performs badly in the boundary values. So the proposed work started by improving the existing tool. From the above graphs, one can conclude that the accuracy of estimation can be improved to 20-40% in case of LOC estimation and 30-40% in case of effort estimation using knowledge rules. Thus the objective of the improving performance is carried out by combining the knowledge rules obtained using regression process along with the estimation by analogy process. The result can be observed by the improvement of the level of accuracy of cost and effort estimates.

One of the major problems is obtaining the historical data of the completed projects. For this work, all the project data used are pertaining to multimedia applications. This work can be extended by using a lot of project data of projects related to other domain also. One can expect better estimates by using a lot of historical data. This project has been tested and the performance of the tool compares well with the real time estimates.

This tool is useful for academic purposes and can be used for research purposes. The work is suitable for project managers also. This work can be extended by using data mining techniques.

REFERENCES

- [1] Emila Mendes, Nile Mosley, Steve Counsell (2001), "Web Metrics- Estimating Design and Authoring Effort", IEEE Multimedia, Vol. 8, No. 1, pp50-57.
- [2] Martin Sheppard and Chris Schofield (1997) " Estimating Software Project Effort Using Analogies ", IEEE Transactions on software Engineering, Vol. 23, No. 11, pp 736-743.
- [3] Shepperd M.J., Scholfield C and Kitchenham, (1996), "Effort Estimation by analogy", Proc Intl. Conference Software Engineering, IEEE Computer Soc. Press, Los Alamitos, California, US.
- [4] <http://dec.bmth.ac.uk/ESERG/ANGEL/>
- [5] <http://www.curvefitting.com/>
- [6] Ali Arifoglu (1993), "A methodology for software cost estimation", *SIGSOFT Softw. Eng. Notes*, 18(2):96-105.
- [7] Oliverira A (2006), "Estimation of Software Project Effort with Support Vector Machine", *Neurocomputing*, Vol. 69, No. 13-15., pp. 1749-1753.
- [8] Raymond L and Bergeron F (2008), "Project Management Information Systems: An Empirical Study of their impact on Project Managers and Project Success", *International Journal of Project Management*, Vol. 26, No. 2, pp 213-220.
- [9] Chris F. Kemerer (2005), "An Empirical Validation of Software Estimation Models", *Commun. ACM*, Vol. 30, No. 5, pp 416-429.

International Journal of Software Engineering & Applications (IJSEA), Vol.2, No.4, October 2011

- [10] Sneed H.M. (2004), “A cost model for software maintenance & evolution”, *Software Maintenance, In Proceedings. 20th IEEE International Conference on* (2004), pp. 264-273.
- [11] Norman E. Fenton and Shari L. Pfleeger (1998), “*Software Metrics: A Rigorous and Practical Approach*”, PWS Publishing Co., Boston, MA, USA,
- [12] Fioravanti F and P. Nesi (2001), “Estimation and prediction metrics for adaptive maintenance effort of object-oriented systems”, *IEEE Transactions on Software Engineering*, 27(12):1062–1084, 2001.
- [13] Rüdiger L et al. (2008), “Comparing software metrics tools”, In *Proceedings of the 2008 international symposium on Software testing and analysis, ISSTA '08*, pages 131–142, New York, USA, 2008.
- [14] Rosenberg J (1997), “Some misconceptions about lines of code”, *IEEE Computer*, Vol. 0, pp 137–142, Los Alamitos, CA, USA.

Author

Dr. S. Sridhar is an Associate Professor in the Department of Information Science and Technology at College of Engineering, Guindy. His area of interest includes Image Processing, Pattern Recognition, Software Engineering and Algorithms.

