

A NOTE ON COST ESTIMATION BASED ON PRIME NUMBERS

Surya Prakash Tripathi¹, Kavita Agarwal² & Shyam Kishore Bajpai³

It is a difficult task to calculate cost (effort) of the software at the beginning of the software development life cycle. Function points can be calculated at the requirement phase and are independent of the tools and techniques used for software development. In this paper we map Function points to prime numbers and call it Prime function points then use these prime function points to calculate cost (effort) of the software. Conversely, given cost we can find the function points.

Keywords: Cost Estimation, Function Point, Prime Numbers, Software Effort Estimation.

1. INTRODUCTION

While cost is being developed or posed to be developed, cost measurement activity is seen to be difficult task a priori having no knowledge of size and other major features [10], [11]. Many software cost estimation techniques have been proposed based on the experience of the developer. Some of the known techniques are based on Algorithmic model [6], Expert judgement, Analogy, Parkinson's view, Price to win, or on the basis of top down or bottom up considerations [7],[9]. All the above methods are better from one another depending upon what ground that has been used on. Our aim in writing this paper is to establish a mathematical criterion for estimating cost of any project no matter what environment has been chosen except for traditional views used in Function points.

Cost estimation in software engineering is very important due to the fact software being intangible i.e. a priori not known, nor visible, nor predictable. So far in literature cost estimation is being visualized on experience basis of similar programs. This leads to errors in estimation from 500% onwards except for FP considerations where error has been drastically reduced to 105% onwards [4]. We found that FP alone is not sufficient to reduce errors but some more factors are essential in reducing errors. One of our point of view is entirely new and not used before to best of our knowledge, is based on the fact that if we limit our considerations of cost evaluation based on several

natural findings in computer science from algorithm analysis and design to DBMS and other development, then we find that prime numbers play a specific role however nobody has ever thought that prime numbers have any role in cost estimation. It is not only prime numbers, we made observation based on several well known iterative and recursive series. Our observation leads that if we intermingle prime numbers with FP then error is drastically reduced in case of prime numbers from 42% to 103%.

Few considerations we are going to give in the following pages which will reflect our view. The case of 42% error has been considered in another investigation (submitted for consideration under the title of "Prime number cost estimation criterion"). Our investigation recently has reflected more interesting nature. For example intermingling FP and Prime numbers error is 103% where as in case of square series ($1^2, 2^2, \dots$) error is 66% approximately. We focus on prime numbers because recently prime numbers have been studied in detail for many purposes [13].

We further observe that for small number of prime numbers the case is near linear whereas for large numbers prime is non linear.

Before we propose the following preposition we explain the difference between FP and Prime FP.

FP (function point) is the function point calculated at requirement phase.

Prime FP (prime function point) is the FP^{th} prime number i.e. if FP is 5 then Prime FP is 7 (we used a program in java to calculate this).

We propose the following preposition:

Given FP we calculate Prime FP and denoted it by X, then cost estimate denoted by Y will be determined by the following equation:

$$-14350.06 + 5.8846X = Y \quad (1)$$

¹Asst. Professor & Head, Department of Computer Science & Engineering, Institute of Engineering & Technology, Lucknow, Uttar Pradesh, INDIA

²Asst. Professor, Department of Computer Science & Engineering, BBDNITM, Lucknow, Uttar Pradesh, INDIA

³Ex-Professor & Head, Department of Computer Science & Engineering, Institute of Engineering & Technology, Lucknow, Uttar Pradesh, INDIA

Email: tripathee_sp@yahoo.co.in¹, kavitalucknow@gmail.com², skbajpaiiet@gmail.com³

Conversely, given cost denoted by X we calculate Prime FP (denoted by Y) using the following equation:

$$2669.353196 + 0.192256588X = Y \quad (2)$$

Then calculate FP by counting number of prime numbers less than equal to Prime FP.

Our method apriori need the knowledge of how function points criterion has been established for evaluating cost estimation. We give brief account of this method and then used function points on the basis of the equations illustrated in the above preposition.

We have used most of the available data cost estimation and found that above preposition develops the best way of finding cost estimation at the requirement phase itself. Our method will be appreciated if judgement is drawn on some worked out project and their cost estimation. We considered the data available and cost available and compare the evaluation based on our method, remembering that our cost will be unique as prime numbers are unique. Before reporting our results to the software community, we contacted several industries to provide actual data they have used for cost estimation. This proposal has not worked out well as industrial people involved have not provided their data and cost estimation but whatsoever data was available we have given our cost estimation. We have given cost of software project apriori without the knowledge of the software, by simply converting prime number cost to match it to function points.

2. FUNCTION POINTS

We will illustrate function points as our method depends totally on this concept [1],[3]. Since theory of Function point is well known to the reader, we briefly account the necessary parts in the following.

The function point measure is done in three steps:

- a) classify and count the five user function points
 - i. external input types;
 - ii. external output types;
 - iii. logical internal files;
 - iv. external interface file types;
 - v. external inquiry types.

Each user function point is counted, for example external input types. Suppose there are 10 external input types. Then these 10 function points are classified as simple, average and complex. Suppose 2 are simple, 3 are average and 5 are complex, then for simple we multiply 2 by 3, for average 3 is multiplied by 4 and for complex 5 is multiplied by 6, as indicated by the fig 1. Then total = $2*3+3*4+5*6$.

Similarly for other function points is calculated as indicated by the fig 1.

Description	Simple	Average	Complex	Total
External Input	*3=	* 4 =	* 6 =	–
External Output	*4 =	* 5 =	* 7 =	–
Logical internal file	* 7 =	* 10 =	* 15 =	–
External interface file	* 5 =	* 7 =	* 10 =	–
External Inquiry	* 3 =	* 4 =	* 6 =	–

Fig 1

In the above figure concept of TOTAL is sum of all the values in the three columns in the row.

Total Unadjusted Function Point (UFP) = Sum of all the Totals

- b) adjust for processing complexity
 - i. The degree of influence of each of 14 general characteristics namely: Fourteen characteristics are :-Data communication, Distributed functions, Performance, Heavily used configuration, Transaction rate, Online data entry, End user efficiency, Online data update, Complex processing, Reusability, Installation ease, Operational ease, Multiple sites and Facilitate change, is taken.
 - ii. The degree of influence of each of above 14 general characteristics given in (i) is estimated on a scale of 0 to 5 where 0 is no influence and 5 is maximum influence. For example where online data entry is not required there online data entry will be scale as 0 and thus no influence. The influence on scale 1-4 shall be dependent on the experience of the person who is calculating the cost but this experience factor by the concerned person is dependent on the knowledge of theory of function point and its use. On this point to make it more mathematical in nature we propose a separate investigation which is not a part of this paper.
 - iii. All influences are summed (PC, processing complexity) and an adjustment factor is developed. Where Processing complexity adjustment (PCA)= $0.65+(0.01 * PC)$.
 - iv. Finally 14 degrees of influence is summed and adjustment factor ranging from 0.65 to 1.35 is calculated where 0.65 is minimum and 1.35 is maximum[1].
- c) Make the function point calculation.

- i. Multiply function points in (a) to processing complexity adjustment in (b). Thus Function Point (FP) = UFP*PCA.

3. THE IDEA

A. We compute equation (1) using following algorithm 1 and then calculate the cost of the software using algorithm 2 on certain specific data given in fig 2 [4].

Using data [4]

Project number	Actual MM	Function point
1	287	1217
2	82.5	507
3	1107.31	2306
4	86.9	788
5	336.3	1337
6	84	421
7	23.2	100
8	130.3	993
9	116	1592
10	72	240
11	258.7	1611
12	230.7	789
13	157	690
14	246.9	1347
15	69.9	1044

Fig 2

MM = number of man months
 (=152 working hours) [4]

Using data given in fig 2 for deriving equation 1 we use the following Algorithm 1:

Step 1: Take function point (FP) calculated in requirement phase.

Step 2: Find the FPth prime number and call it prime function point(denoted by X).

Step 3: using linear regression [12] we relate prime function point(denoted by X) and actual MM*152(denoted by Y) and we get equation (2).

In linear regression[12] Straight line equation is taken to be: $a_0 + a_1 * X = Y$;

Where $a_0 = (\sum Y_i \sum X_i^2 - \sum X_i \sum (X_i Y_i)) / (n \sum X_i^2 - (\sum X_i)^2)$

And $a_1 = (n \sum X_i Y_i - \sum X_i \sum Y_i) / (n \sum X_i^2 - (\sum X_i)^2)$

For our data $a_0 = -4350.06$ and

$a_1 = 5.8846$

final equation is $-14350.06 + 5.8846X = Y$

Now we obtain cost estimation using Algorithm 2 as follows:

Step 1: Find function point in the requirement phase.

Step 2: Map each FP to a prime number in sequence (find FPth prime number). This prime number is the prime function point.

Step 3: Calculate ESTIMATED MM (denoted by Y) using equation (1) using prime function point as X.

Step 4: Error percentage is calculated as

$((\text{estimated MM} - \text{actual MM} * 152) / \text{actual MM} * 152) * 100$

Step 5: Average error is calculated taking absolute values.

Result:

In fig 3

FP corresponds to the Function point in fig 2 prime FP(denoted by X) – Prime function point corresponds to the prime number where FP is the number of prime numbers less than Prime function point i.e. FPth prime number.

MM*152 is calculated by multiplying actual MM (fig 2) by 152 to get effort in hours.

Estimated MM(denoted by Y) is calculate using equation (2) and using prime function point as X.

Error % is calculate as $((\text{estimated MM} - \text{MM} * 152) / \text{MM} * 152) * 100$.

B . We compute equation (2) using following algorithm 3 and then calculate the FP of the software using algorithm 4 on certain specific data given in fig 2 [4].

Using data given in fig 2 for deriving equation 2 we use the following algorithm 3:

Step1: Given function points we calculate prime function points (for this we used a program in Java).

Step 2: Prime function points are denoted by Y and cost(effort) is taken as X.

Step 3: We use linear regression to relate X and Y to get equation 2.

In linear regression [12] Straight line equation is taken to be: $b_0 + b_1 * X = Y$;

Where $b_0 = (\sum Y_i \sum X_i^2 - \sum X_i \sum (X_i Y_i)) / (n \sum X_i^2 - (\sum X_i)^2)$

And $b_1 = (n \sum X_i Y_i - \sum X_i \sum Y_i) / (n \sum X_i^2 - (\sum X_i)^2)$

For our data $b_0 = 2669.353196$ and $b_1 = 0.192256588$

final equation is $2669.353196 + 0.192256588 X = Y$

FP	Prime FP		EST	
	MM*152		MM(Y USING EQUATION)	ERROR %
1217	9859	43624	43666.92596	0.0984
507	3623	12540	6970.107703	44.417
2306	20407	168311.12	105738.4524	37.17679
788	6043	13208.8	21211.01536	60.58246
1337	11027	51117.6	50540.22354	1.129506
421	2909	12768	2768.451476	78.31727
100	541	3526.4	-1166.45321	416.6531
993	7867	19805.6	31944.65816	61.29104
1592	13441	17632	64745.82316	267.2063
240	1511	10944	-458.320801	149.875
1611	13627	39322.4	65840.37226	67.43732
789	6047	35066.4	21234.55405	39.44473
690	5189	23864	16185.50497	32.17606
1347	11117	37528.8	51069.84407	36.08174
1044	8317	10624.8	34592.76083	225.5851
SUMS	121525	499883.92	700784.7751	40.1895
AVG ERROR(ABS VAL)			103.8441	

Fig. 3

We now obtain function points using algorithm 4 as follows:

Step 1: Given cost (effort) of a software denoted by X, calculate Prime function points (denoted as Y) using equation 2.

Step 2: Count number of prime numbers less than equal to Prime function points (we used a program in Java for this). This count becomes the function points.

Step 3 : Error percentage is calculated as ((calculated FP-FP)/FP)*100

Step 4: Average error is calculated taking absolute values.

Results:

In Fig 4:

FP is the function points in fig 2.

Prime FP is the corresponding prime function point (FPth prime number).

MM*152 is the effort in man hours in fig 2.

FP	Prime FP (Y)	MM*152 (X)
1217	9859	43624
507	3623	12540
788	6043	13208.8
1337	11027	51117.6

421	2909	12768
100	541	3526.4
993	7867	19805.6
1592	13441	17632
240	1511	10944
1611	13627	39322.4
789	6047	35066.4
690	5189	23864
1347	11117	37528.8
1044	8317	10624.8

Fig 4

In fig 5.

Cal FP is the calculate FP obtained after using algorithm 4.

Error is the absolute error %.

cal FP	Error
1340	10.10682
680	34.12229
694	11.92893
1493	11.66791
686	62.94537
473	373
841	15.30715
791	50.31407
642	167.5
1255	22.09808
1164	47.52852
929	34.63768
1220	9.428359
636	39.08046
average error	63.54755

Fig 5

4. OBSERVATION

The results obtained are quite encouraging. For the same data average error using function points it is 102.74%, for COCOMO basic it is 610.09%, for COCOMO intermediate it is 583.82% and for COCOMO detailed it is 607.85% [4]. In our idea we mapped function points to prime numbers and the average error % was 103.84.

Our point of view is straight forward and will be advantageous in most of the cases. However Function points used may vary from one person to another performing the analysis, thus it contributes to the variations in the results [2], [5]. Which result will be optimal is still a point of consideration but we feel that if larger set of data is used the above result will be better.

5. CONCLUSION

Finally we arrive to a conclusion that based on our idea the cost estimation at requirement phase will be given by the following theorem:

Theorem: Given FP we calculate Prime FP (FPth prime number) and denoted it by X, then cost estimate denoted by Y will be determined by the following equation:

$$a_0 + a_1X = Y$$

Where actual calculations of a_0 and a_1 can be done using our algorithm 1.

Conversely, given cost denoted by X we calculate Prime FP (denoted by Y) using the following equation:

$$b_0 + b_1X = Y$$

Then calculate FP by counting number of prime numbers less than equal to Prime FP.

Where actual calculations of b_0 and b_1 can be done using our algorithm 3.

REFERENCES

- [1] Allan J. Albrecht and J.E. Gaffney, "Software Function, Source Lines of Code, and Development Effort Prediction: a Software Science Validation", IEEE Transactions on Software Engineering, SE-9, No. 6, 1983, pp.639-648.
- [2] B.A. Kitchenham, E. Mendes, G.H. Travassos, "Cross Versus within-company Cost Estimation Studies: A Systematic Review", IEEE Transactions on Software Engineering , 33, No. 5, 2007, pp. 316-329.
- [3] Chris F Kemerer, "Software Project Management, Readings and Cases" McGraw Hill Company, Irwin Book Team-1997.
- [4] Chris F Kemerer, "An Empirical Validation of Software Cost Estimation Models", Communications of ACM, 30, No. 5, 1987, pp. 416-429.
- [5] Charles R Symons, "Function Point Analysis: Difficulties and Improvement", IEEE Transactions of Software Engineering, 14, No. 1, Jan 1988, pp. 2-11.
- [6] Ian Sommerville, "Software Engineering", 5th Edition, Addison-Wesley.
- [7] Bob Hughes and M. Cotterell, "Software Project Management", 3rd Edition, Tata Mcgraw-Hill.
- [8] K. Chandrasekharan, "Introduction to Analytic Number Theory", Springer-Verlag, New York Inc. 1968.
- [9] Richard Fairley, "Software Engineering Concepts", Tata McGraw Hill.
- [10] S. Grimstad, M. Jorgensen, "A Framework for Analysis of Software Cost Estimation Accuracy", ISESE'06, Brazil, Copyright 2006 ACM, pp. 58-65.
- [11] Walker Royce, "Software Project Management A Unified Framework", Addison-Wesley.
- [12] V. Rajaraman, "Computer Oriented Numerical Methods", 2nd Edition, PHI.
- [13] M Agrawal, N Kayal and N Saxena, "Primes is in P", http://www.cse.iitk.ac.in/news/primalty_v3.ps, Feb, 2003.