The Evaluation of Useful Method of Effort Estimation in Software Projects

Amin Moradbeiky, Vahid Khatibi Bardsiri

Kerman Branch, Islamic Azad University, Kerman, Iran
moradbeigi@csri.ac.i

Kerman Branch, Islamic Azad University, Kerman, Iran
kvahid2@live.utm.my

Abstract

Software projects atmosphere experiences much complexity and ambiguities putting software project managers in several challenges during making decisions. Different methods have been proposed helping managers’ decisions toward projects’ important parameters. The present study presents some of the most applicable methods of software development effort estimation. In addition, the proposed methods are compared and tested on three real databases. We made use of the most applicable and valid databases and available evaluation standard methods for carrying experiments. The results well familiarize the researcher with the capabilities of the available methods in different databases.

Keywords: development effort estimation, software project, analogy-based method, regression

I. Introduction

Methods of software development effort estimation are divided into two groups of algorithm and non-algorithm. Algorithm approaches are created based on predefined parameters. They are defined as similar as a math formula based on some cost parameters (Boehm, 1981). As indicated by the results, algorithm methods enjoy a math-like and stable structure (Boehm, 1981). Non-algorithm approaches are another group of effort estimation methods. These approaches estimate the effort based on the condition of the project resulting in more dynamicity of non-algorithm approaches and more adaptability with software project dynamicity (Huang and Chiu, 2006) (Somerville, 2001). This study first elaborates on useful algorithm and non-algorithm methods. Some of the synthetic capabilities of these methods are also explained. Then, all proposed methods are tested on three datasets and the obtained results are analyzed.

II. Methods

The proposed methods are among the most applicable approaches in development effort estimation explained as follows
A. Step-wise Regression (SWR)
Step-wise regression is a statistical technique creating a predictive model and presenting the relationship between independent and dependent variables. Step-wise regression is used to investigate the influence of some independent variables on dependent variables. In other words, SWR’s goal is to see which independent variable can best predict the dependent variable, how the share of each variable is in this regard, and generally, how all together can make the prediction. This technique aims at finding some independent variables explaining the dependent variable variance. Step-wise regression is applied both as forward and backward aiming at finding the optimal number as the best reaction toward variable. We start with an empty set of variables for forward mode and with a full set of variable for backward mode so that variables are orderly added or removed if the prediction progresses. This function is systematically used for adding variables. Forward stepwise fit (FSWF) is a technique generating models through adding independent variable in each step and with the highest adaptability with dependent variable while considering all available variables. Forward approach starts modeling with no variables. It makes variables involved if they are statistically proved important. This selection is carried for estimating the required effort for the study project among similar projects. It is called backward stepwise fit (BSWF) (Nagpal et al., 2014).

B. Multiple Linear Regression (MLR)
Multiple linear regression is among algorithm techniques. According to Yung and Fun’s studies, this model is experimental requiring the data of former projects for evaluating current projects. Bouhem, et.al. and Sing, et.al. consider MLR as one of the effort estimation techniques showing the relationship between dependent (Y) and independent (Xi) variables. MLR model is defined in equ. 1

\[ Y = \beta_0 + \beta_1 X_1 + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon \]  

(1)

In which, X1,X2,…,Xn are independent variables, withdrawal parameter, and \( \beta_1, \beta_2, \ldots, \beta_n \) are regression coefficients responsible for estimating dependent variable, and \( \varepsilon \) is error component.

In order to evaluate the accuracy level of this model, the coefficient of determination (R2) is used. This coefficient evaluates the general validity of independent variable in average level gotten from multiple regression model. As mentioned before, each technique owns specific and unique attributes making it suitable for solving specific problem. MLR is mostly used when

1- number of cases is much more than the number of parameters of the study case
2- data are consistent
3- there are few number of removed data
4- number of independent variables was enough for linear output variables so that their outputs get interpretable.
Regression is applied when simple analyzing tools and models of effort estimation are required, and MLR application requires different hypothesis the most important of which include

1- linearity i.e. the relationship between $X_i$ and $Y$ should be linear so that the model can explain the data efficiently.
2- the error component should be independent and the distribution is supposed to be normal without constant variance and zero mean.

Available problems in this dataset and inaccurate application of the model can result in appearing some mistakes in hypotheses (Fedotova et al., 2013).

C. Classification and Regression Tree (CART)

CART method was introduced by Briman, et al. as a non-algorithm method for effort estimation. Tree methods are naturally divided into two methods of algorithm and non-algorithm suitable for classification. Trees with numerical attributes are mostly called regression trees while those with categorization attributes are called classification trees. Researcher made use of this model for analyzing former software projects to make a regression tree whose leaves represent the projects required effort. In this method, all data are divided into subsets that are usually binary. The division principles are simply based on if-then rules. The procedure continues until the stop criterion is recorded. In this tree, the procedure moves from roots to leaves based on the study project’s attributes. CART model is suitable for solving irrelevant problems. In fact, CART is an algorithm selecting suitable decision tree model for classification adapting it to consistent objectives. Simple structure of regression tree is presented in figure 1 (Breiman et al., 1984).

Fig. 1: regression tree for effort estimation
D. Stages of regression and classification tree
1- The available data from the former carried projects are used for generating a model
2- A regression and classification tree model is generated according to the obtained data from former stage
3- The previously generated model receives the estimated value and the cost related to the new project
4- The model estimates the effort. If the used tree is a regression tree, the effort estimated as a numerical value. However, if the tree is a classification tree, the effort is estimated as low or high effort.

The stages of this model are similar to the stages of the algorithm techniques. Algorithm techniques make use of data for generating equations applied in effort estimation while regression and classification tree models use data for creating a binary tree finally applied in effort estimation. The comprehensiveness of the regression trees can be considered as strong point of this technique. In order to identify the effort required for a new project, we should only select the appropriate leaves based on the attributes of the new project (Breiman et al., 1984).

E. Analogy-Based Estimation (ABE) method
As mentioned before, estimation methods are divided into two categories of algorithm and non-algorithm. Algorithm methods are not suitable for dynamic environment of software projects. Therefore, non-algorithm methods are used in these environments. ABE is one of the most applicable methods in this regard. ABE method is used for estimating an unclear important attribute (such as effort or cost) of a project in relation to some other projects. This method comprises of some stages explained in the following section (Shepperd and Schofield, 1997).

F. Similarity Function
Similarity function identifies projects similarity through studying their attributes with certain value. For this, two methods of Euclidean similarity method (Eq. 2) and Manhattan method (Equ. 3) are used. Project attributes include two groups of numerical and non-numerical groups. In Euclidean and Manhattan methods and for numerical attributes, the numerical attribute distance is estimated for identifying the difference between two projects. Also in non-numerical attributes, difference is estimated according to the value 0 or 1. Euclidean and Manhattan methods are different in the manner of estimation of the numerical attribute values. In the following equations 2 and 3, p and p' are study projects. f_i and f'_i are respectively the ith attributes of p and p'. The obtained results are indicative of the similarities between two projects.

\[
sim(p, p') = \frac{1}{\sqrt{\sum_{i=1}^{n} w_i \cdot Dis(f_i, f'_i) + \delta}} \\
\delta = 0.0001
\]

(2)
Dis \left( f_i, f'_i \right) = \begin{cases} \left( f_i - f'_i \right)^2 & \text{if } f_i \text{ and } f'_i \text{ are numerical or ordinal} \\ 0 & \text{if } f_i \text{ and } f'_i \text{ are nominal and } f_i = f'_i \\ 1 & \text{if } f_i \text{ and } f'_i \text{ are nominal and } f_i \neq f'_i \end{cases}

sim \left( p, p' \right) = \frac{1}{\delta} \left[ \sum_{i=1}^{n} w_i \cdot Dis \left( f_i, f'_i \right) + \delta \right] 

\text{with } \delta = 0.0001

\left(3\right)

G. Solution Function

Solution function is used for estimating the effort attribute of one project according to the size of that project in k more similar projects. In solution function, a value coefficient is given to the attribute size to be estimated according to the similarity of goal project with one of the K projects. Different values have so far been proposed for k value in function. Solution function is presented in equation 4.

\begin{equation}
C_p = \sum_{k=1}^{K} \frac{Sim \left( p, p_k \right)}{\sum_{k=1}^{K} Sim \left( p, p_k \right)} C_{p_k}
\end{equation}

In this equation, p is a project that we aim to estimate its effort. Pj is the jth from the more similar k project. Cpj is the attribute from the jth more similar projects that we aim to estimate it.

H. the best value for K

The k value used in effort estimation greatly influences the effort estimation accuracy. The suitable value of k considerably depends on the study projects. The high value of k in a group of study projects that are mostly different from each other reduces the accuracy. In this
situation, the effective projects are too different in the final stage of estimation. On the other hand, when study projects are similar to each other, the low value of k leads in the negligence of similar project investigation while these projects could demonstrate a positive influence on the accuracy of results in the final stage. This positive impact is caused through reducing the noise in calculation.

III. Efficiency criteria

In order to evaluate the estimation results, we need suitable criteria. In addition, to compare the results of our study with that of others, the selected criterion is supposed to prove efficient in similar domains. Anyway, an efficiency criterion should suitably be calculable and comparable. Its required parameters should be easily accessible. The efficiency criterion applied in this study involves four criteria including Relative Error (RE), Mean Relative Error (MRE), Mean Magnitude Relative Error (MMRE), and Prediction Percent (PRED). They are shown as follows (Shepperd and Schofield, 1997)

\[
RE = \frac{Estimate - Actual}{Actual}
\]  
(5)

\[
MRE = \frac{|Estimate - Actual|}{Actual}
\]  
(6)

\[
MMRE = \frac{1}{N} \sum_{i=1}^{N} MRE
\]  
(7)

\[
PRED(X) = \frac{A}{N}
\]  
(8)

\[
MdMRE = Median(MRE)
\]  
(9)

In equation 8, A is the number of projects with MRE size less than or equal to X, and N is the estimated projects. An acceptable value of X in software effort estimation is 0.25 according to which criteria are compared. On the other hand, MMRE and MdMRE should be reduced to their minimum extent, and PRED is supposed to increase to the maximum level.

IV. Evaluation of the Results

In most of the effort estimation, the arrangement of projects influences accuracy. Result evaluation methods are applied by effort estimation experts to prove the reliability of the results of their studies so that they can show that their presented model enjoys acceptable accuracy in projects with different arrangement (Hayes, 1994). Some of these methods include 3-fold and 10-fold approaches. In 3-fold approach, projects are divided into three equal parts. The project is gone under three tests each of which is carried on one of the three parts as test set. After testing,
the mean is calculated from the obtained results through three steps applied as the results of testing.

In 10-fold approach, the steps similar to the previous approach are run. In this method, projects are divided into 10 equal parts. In each step of testing, one group is introduced as the test group, and the other 9 groups are applied in training stage. This procedure is repeated nine more times in each of which one group is taken as the testing set. Finally, the mean obtained from 10 turns are considered as the general results.

As indicated by the results, this approach lacks accuracy and reliability since the result accuracy changes if the order of arrangement of groups changes. Therefore, the results do not fulfill the required reliability (Kocaguneli and Menzies, 2013).

Another approach used for estimation evaluation is Leave-One-Out method. In this method and in every step, one of the projects is selected as the test set and all other remaining projects are applied in training stage. The number of testing stages is equal to the number of total projects. The result is obtained from the result mean of each step. Kocaguneli and Menzies proved the high accuracy of this method in their study (Kocaguneli and Menzies, 2013).

Applying Leave-One-Out approach with different orders always demonstrates similar results to the previous result since in this method; the results are not dependent on the arrangement of a project with high or low estimation ability. Another advantage of this method is that the maximum number of projects is used for training resulting in more accurate results. Comparing 10-fold method with Leave-One-Out approach, we conclude that this method takes higher running time because of more training and testing turns (equal to all projects).

V. Testing the explained methods

In order to investigate the accuracy of the proposed method, three datasets are tested. The first dataset is Cocomo consisting of 63 projects with 17 attributes. The second one is Desharniz consisting of 77 projects with 10 attributes, and the third is Maxwell dataset consisting of 62 projects with 26 attributes.

A. Testing Cocomo dataset

The first test was carried on all Cocomo dataset projects. Figure 2 shows the results related to the first 16 projects of this testing. According to the results, in most of the projects, the strongest method was ABE (k=5) and the weakest SWR. As indicated by the results, the accuracy of ABE method is similar for different Ks, and in most of the projects, CART method demonstrates a close accuracy to ABE method. Moreover, the accuracy fluctuation was high in some of these methods indicating inconsistency of these methods in datasets.
Table 1 shows the values of MdMRE, MMRE, and Pred (0.25) parameters for Cocomo dataset. According to MdMRE and MMRE, the best methods was ABE (k=5). Given the equality of the best method for both criteria of MdMRE and MMRE, we understand that the accuracy fluctuation of ABE method (k=5) was suitable for its accuracy mean. However, the best method for Pred was MLR so that MLR demonstrates more accurate estimation.

<table>
<thead>
<tr>
<th>approach</th>
<th>MdMRE</th>
<th>MMRE</th>
<th>Pred</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABE K=2</td>
<td>0.8056</td>
<td>1.6063</td>
<td>0.1270</td>
</tr>
<tr>
<td>ABE K=3</td>
<td>0.8013</td>
<td>1.4658</td>
<td>0.1111</td>
</tr>
<tr>
<td>ABE K=4</td>
<td>0.7959</td>
<td>1.4763</td>
<td>0.0952</td>
</tr>
<tr>
<td>ABE K=5</td>
<td>0.7679</td>
<td>1.3070</td>
<td>0.1429</td>
</tr>
<tr>
<td>CART</td>
<td>0.8597</td>
<td>2.8002</td>
<td>0.1587</td>
</tr>
<tr>
<td>MLR</td>
<td>1.0064</td>
<td>6.3143</td>
<td>0.1746</td>
</tr>
<tr>
<td>SWR</td>
<td>10.6590</td>
<td>29.4099</td>
<td>0.0476</td>
</tr>
</tbody>
</table>

B. Testing Desharniz dataset
The next testing is carried on all Desharniz dataset projects. In order to present a better estimation accuracy, the results of this testing is presented in figure 3 for the first 20 projects of this dataset. As indicated by the results, CART method was shown as the best method and SWR as the weakest. Additionally, the obtained accuracy has high fluctuation.
Table 2 shows the values of MdMRE, MMRE, and Pred (0.25) parameters for Desharniz dataset. According to MdMRE and MMRE, the best methods were ABE (k=4) and CART, respectively. Given the difference of the best method for criteria of MdMRE and MMRE, we understand that the accuracy fluctuation of ABE method (k=4) was more than the CART method. However, the best method for Pred was ABE (k=4) so that ABE (k=4) demonstrates more accurate estimations relative to other methods.

Table 2: comparing the efficiency of different evaluation methods in Desharniz dataset

<table>
<thead>
<tr>
<th>approach</th>
<th>MdMRE</th>
<th>MMRE</th>
<th>Pred</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABE K=2</td>
<td>0.3311</td>
<td>0.8444</td>
<td>0.3896</td>
</tr>
<tr>
<td>ABE K=3</td>
<td>0.3139</td>
<td>0.7509</td>
<td>0.4286</td>
</tr>
<tr>
<td>ABE K=4</td>
<td>0.3027</td>
<td>0.8043</td>
<td>0.4675</td>
</tr>
<tr>
<td>ABE K=5</td>
<td>0.3131</td>
<td>0.7591</td>
<td>0.4286</td>
</tr>
<tr>
<td>CART</td>
<td>0.4280</td>
<td>0.7222</td>
<td>0.2857</td>
</tr>
<tr>
<td>MLR</td>
<td>0.4140</td>
<td>0.7417</td>
<td>0.2727</td>
</tr>
<tr>
<td>SWR</td>
<td>0.6557</td>
<td>1.1694</td>
<td>0.1169</td>
</tr>
</tbody>
</table>

C. Testing Maxwell dataset
The next testing is carried on all Maxwell dataset projects. In order to present a better estimation accuracy, the results of this testing is presented in figure 4 for the first 16 projects of this dataset. As indicated by the results, ABE (k=3) was shown as the best method and MLR as the weakest. Additionally, the accuracy fluctuations are clearly identified.
Table 3 shows the values of MdMRE, MMRE, and Pred (0.25) parameters for Maxwell dataset. According to MdMRE and MMRE, the best methods were ABE (k=3) and ABE (k=2), respectively. Given the difference of the best method for criteria of MdMRE and MMRE, we understand that the accuracy fluctuation of ABE (k=3) was more than ABE (k=2). However, the best method for Pred was CART so that CART demonstrates more accurate estimations relative to other methods.

Table 3: comparing the efficiency of different evaluation methods in Maxwell dataset

<table>
<thead>
<tr>
<th>approach</th>
<th>MdMRE</th>
<th>MMRE</th>
<th>Pred</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABE K=2</td>
<td>0.5659</td>
<td>0.7226</td>
<td>0.2258</td>
</tr>
<tr>
<td>ABE K=3</td>
<td>0.4777</td>
<td>0.7823</td>
<td>0.2097</td>
</tr>
<tr>
<td>ABE K=4</td>
<td>0.5069</td>
<td>0.8022</td>
<td>0.1774</td>
</tr>
<tr>
<td>ABE K=5</td>
<td>0.5536</td>
<td>0.8917</td>
<td>0.2097</td>
</tr>
<tr>
<td>CART</td>
<td>0.5652</td>
<td>0.9683</td>
<td>0.2581</td>
</tr>
<tr>
<td>MLR</td>
<td>1.7900</td>
<td>4.1311</td>
<td>0.0484</td>
</tr>
<tr>
<td>SWR</td>
<td>1.3495</td>
<td>1.8201</td>
<td>0.1129</td>
</tr>
</tbody>
</table>
VI. Conclusion

This study elaborated on some estimation methods of development efforts. In order to test the proposed methods, three datasets were introduced and the explained methods were tested on them. As indicated by the results, in most cases, the ABE method enjoyed the best or at least the nearest to the best answer. However, we cannot expect the best answer from ABE method for all datasets. On the other hand, as indicated by figures 2 and 4, ABE method demonstrated a high consistency and low fluctuation relative to other methods in Cocomo and Maxwell datasets indicating ABE’s higher accuracy in these datasets relative to other methods.

References


