International Journal of Engineering Research and General Science Volume 5, Issue 6, November-December, 2017
ISSN 2091-2730

Analysing the Calibration Curve in Different Stage from Project management of Software Development Life Cycle Using Isotonic Regression Classifier

Dr. S.Veni[1], Aparna Srinivasan[2]
[1]Professor and Head: Department of Computer Science
[2]Research Scholar: Department of Computer Science
Karpagam Academy of Higher Education
Karpagam University, Coimbatore, Tamil Nadu, India
research.igate@gmail.com

Abstract— The aim of this research is to identify the deviations from metrics and rectify the deviations using data mining and six sigma techniques. In this research work investigates the effectiveness of using computer-based machine learning algorithms of isotonic regression classifier has to be predicting performance by given data for analyzing the project task of actual and estimated value for planning stage, Designing stage, Building stage, User Acceptance test (UAT) stage, System Integrating Test (SIT) stage, Integration testing and implementation stage in Software Development Life Cycle (SDLC).

Keywords— Isotonic Regression classifier, Calibration curve, Six sigma metrics, project development stages, Pair-adjacent violators algorithm

INTRODUCTION

Using Six Sigma methodology of fig 1 represents the metric to improve the methods by identified and controlled so that there is minimal damage to the project management in SDLC [1] to [3].

Fig 1 Six Sigma metric is to reduce the deviation factor by tracking it from beginning stage of the project

- To predict the success of software projects based on information related to Estimation of planning task and an actual task is deemed to be one of the vital activities in software engineering research.
- The variance is an important metrics parameter which needs to be more focused and optimization of effort and cost variance which gives the significant factor influences internal organization was driven and customer-driven goals based on estimated value and actual value.

SCOPE OF THE RESEARCH

An Isotonic Regression is more powerful when there is sufficient data to prevent from fitting based on Pair-adjacent violators (PAV) algorithm is used to fit the training set according to this mean square error criteria which can analyse by an estimated and actual value of planning, designing and building stage, SIT stage, UAT stage, Integrate testing stage and implementation stage were calibrated at fitted point in the plot.

Table 1 Collections of Qualitative Data and Data Preparation

Stages of development	Sub Attributes
Planning Stage	Estimated Day and Actual day/ Estimated Effort and Actual Effort/ Estimated Cost and Actual Cost
Designing stages	Estimated Day and Actual day/ Estimated Effort and Actual Effort/ Estimated Cost and Actual Cost
Building stages	Estimated Day and Actual day/ Estimated Effort and Actual Effort/ Estimated Cost and Actual Cost
User Acceptance test (UAT) stage	Estimated Day and Actual day/ Estimated Effort and Actual Effort/ Estimated Cost and Actual Cost
System Integrating Test (SIT) stage	Estimated Day and Actual day/ Estimated Effort and Actual Effort/ Estimated Cost and Actual Cost
Integration testing	Estimated Day and Actual day/ Estimated Effort and Actual Effort/ Estimated Cost and Actual Cost
Implementation stage	Estimated Day and Actual day/ Estimated Effort and Actual Effort/ Estimated Cost and Actual Cost

Data collected from an estimated and actual task of each stage of project development as shown in table 1 such as planning , Designing, building, User Acceptance test, System Integrating Test, Integrating test and implementation stages.

An estimated value and actual value parameters of training data set can be analysed by isotonic regression analysis the scheduled for a 50 project task of performance and variance of estimated value for planning stage, Designing stage, Building stage, User Acceptance test (UAT) stage, System Integrating Test (SIT) stage, Integration testing and implementation stage.

RESEARCH METHODOLOGY- ISOTONIC REGRESSION

Isotonic regression depends on the regression metric [4] to [9] and the partially ordered set, an approach using pair adjacent violators (PAV), can be used to determine a non parametric method which leads to a stepwise constant mapping function. This method is more general in that the only restriction is that the mapping function was isotonic and it calibrated predictions from decision trees.

The predictions f_i from a model and the true targets y_i, the basic assumption of Isotonic Regression is that:

$$y_i = h(f_i) + \varepsilon_i \qquad (1)$$

Where h is the isotonic function and ε is an individual error term. A non-decreasing mapping function h can be found given a training set with learned membership values f_i and binary class labels y_i so that h holds the equation of 4.2

$$h = \arg\min_{k} \sum_{i=1}^{n} (y_i - k(f_i))^2 \qquad (2)$$

Pair-adjacent violators (PAV) algorithm is used to fit the training set according to this mean square error criterion. It has been shown that isotonic regression based calibration using PAV algorithm. A learning curve analysis shows that isotonic regression are prone to over fitting when data is scarce.

EXPERIMENT AND ANALYSIS

From fig 2, represents the schedule, effort and cost actual and predicted fit in the plot. shows the isotonic fit gives a new calibration curve in planning stage of SDLC where the actual value is equal to the estimated value. It may note that the confidence level higher of 99.03 %, it's come out with better estimation, effort and cost task process in maintenance.

From fig 3, represents the schedule, effort and cost actual and predicted fit in the plot. shows the isotonic fit gives a new calibration curve in Designing stage of SDLC where the actual value is predicted from estimated value. It may note that the confidence level higher of 99.01%, it's come out with better estimation, effort and cost task process in maintenance.

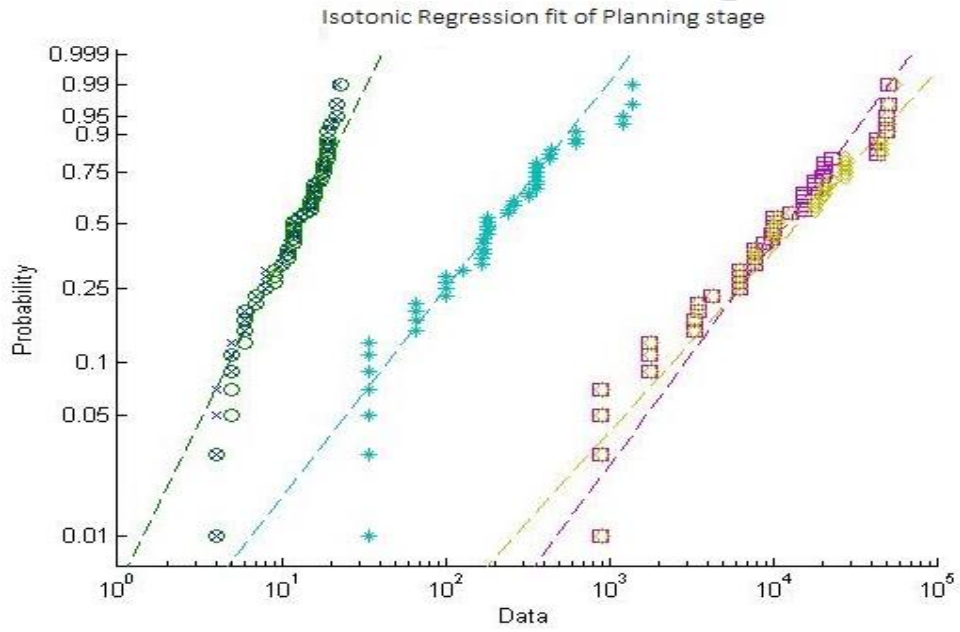

Fig. 2 Calibration curve of planning stage

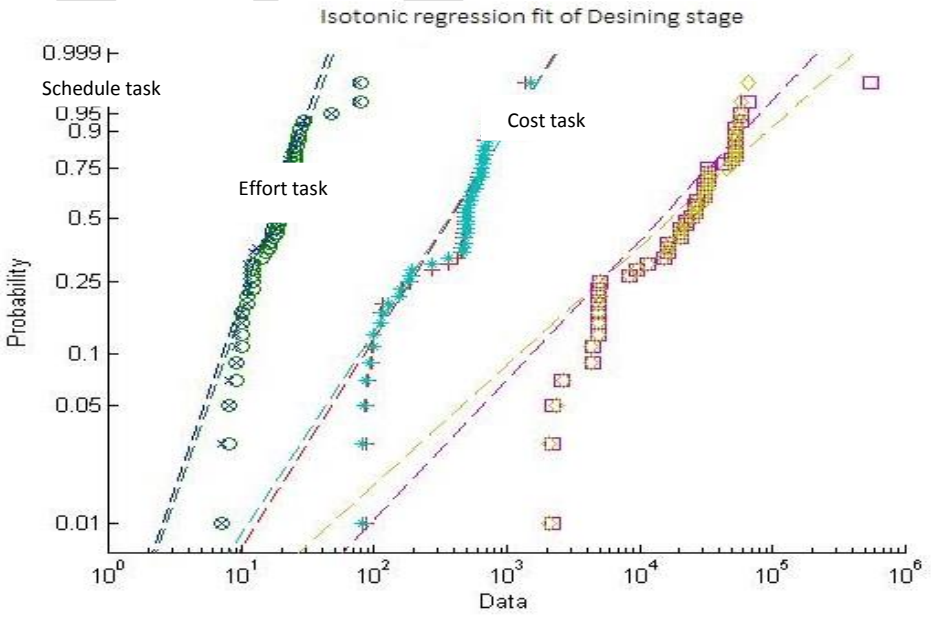

Fig. 3 Calibration curve of Designing stage

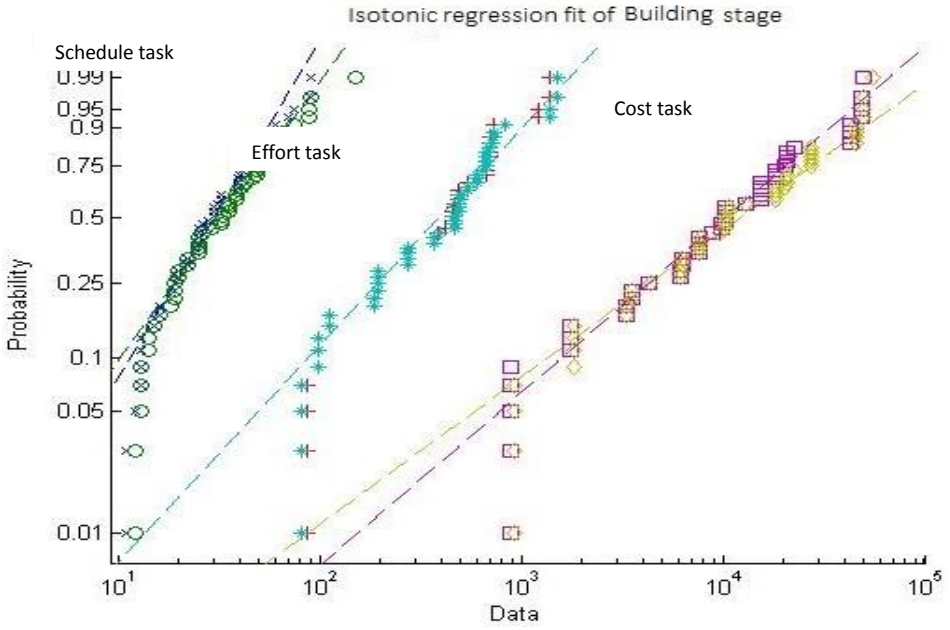

Fig. 4 Calibration curve of Building stage

From fig. 4, represents the schedule, effort and cost actual and predicted fit in the plot. shows the isotonic fit gives a new calibration curve in building stage of SDLC where the actual value is equal to the estimated value. It express goodness of fit well calibrated curves can also indicate the estimation quality. The confidence level is 99.65 %, it's come out with better estimation, effort and cost task process in maintenance.

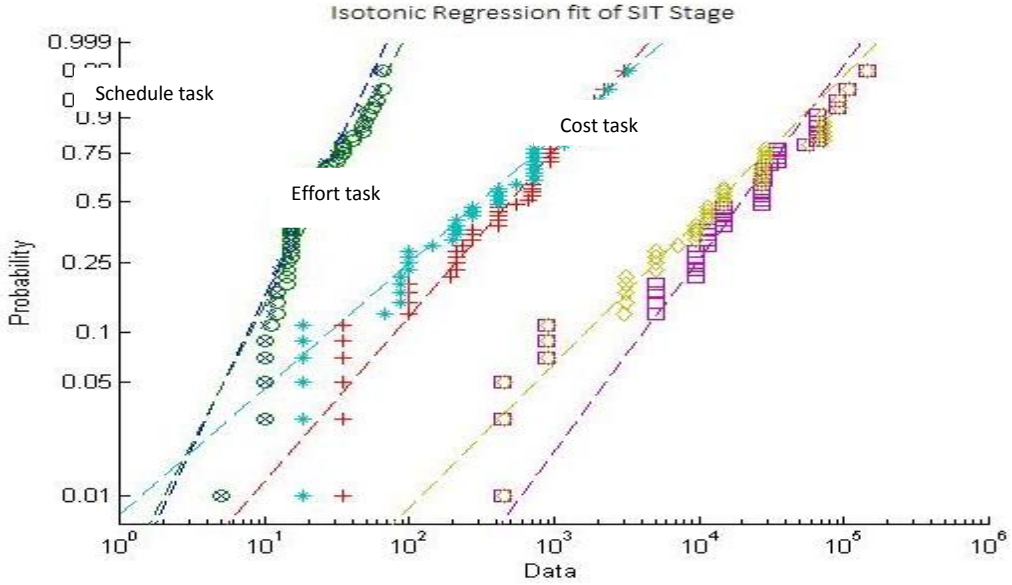

Fig.5 Calibration curve of SIT stage

From fig. 5, represents the schedule, effort and cost actual and predicted fit in the plot. shows the isotonic fit gives a new calibration curve in SIT stage of SDLC where the actual value is equal to the estimated value. It express goodness of fit well calibrated curves can also indicate the estimation quality. The confidence level is 99.79 %, it's come out with better estimation, effort and cost task process in maintenance.

From fig. 6, represents the schedule, effort and cost actual and predicted fit in the plot. shows the isotonic fit gives a new calibration curve in UAT stage of SDLC where the actual value is equal to the estimated value. It express goodness of fit well calibrated curves can also indicate the estimation quality. The confidence level is 99.72 %, it's come out with better estimation, effort and cost task process in maintenance.

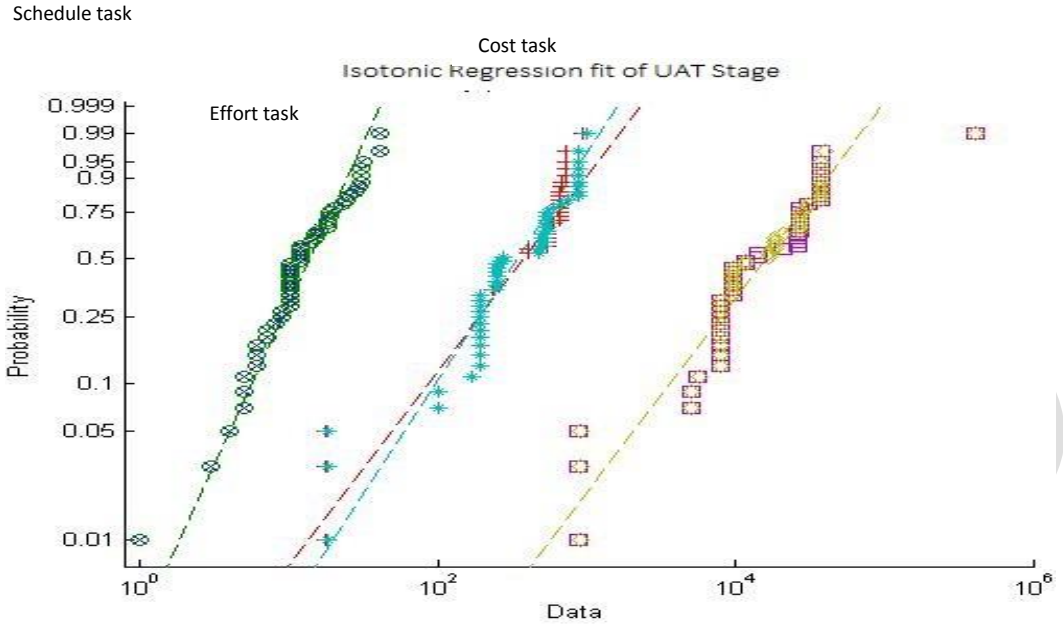

Fig. 6 Calibration curve of UAT stage

From fig. 7, represents the schedule, effort and cost actual and predicted fit in the plot. shows the isotonic fit gives a new calibration curve in integrating testing stage of SDLC where the actual value is equal to the estimated value. It express goodness of fit well calibrated curves can also indicate the estimation quality. The confidence level is 99.70 %, it's come out with better estimation, effort and cost task process in maintenance.

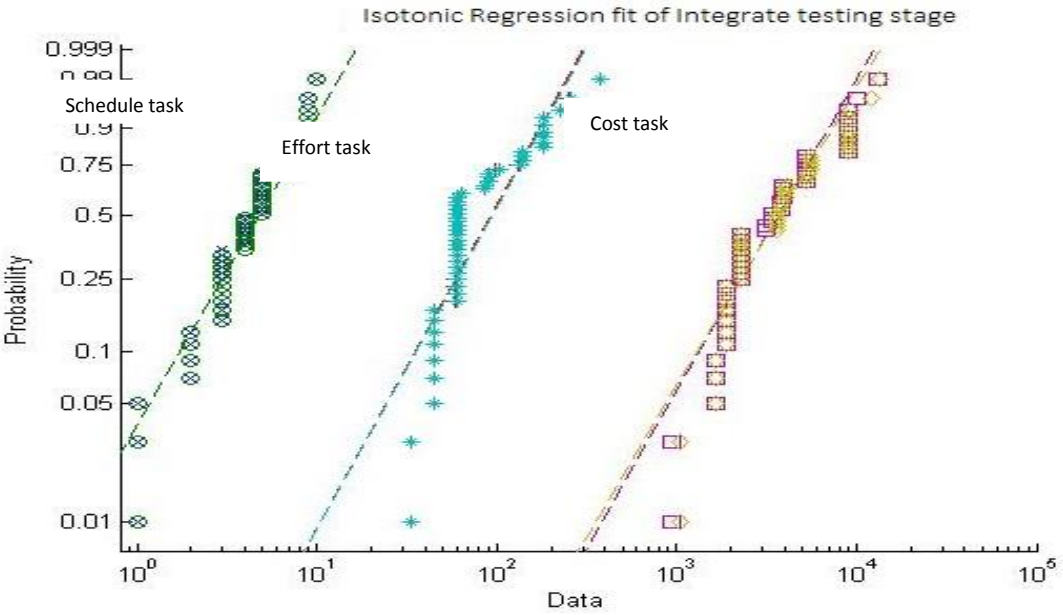

Fig.7 Calibration curve of Integrating Testing stage

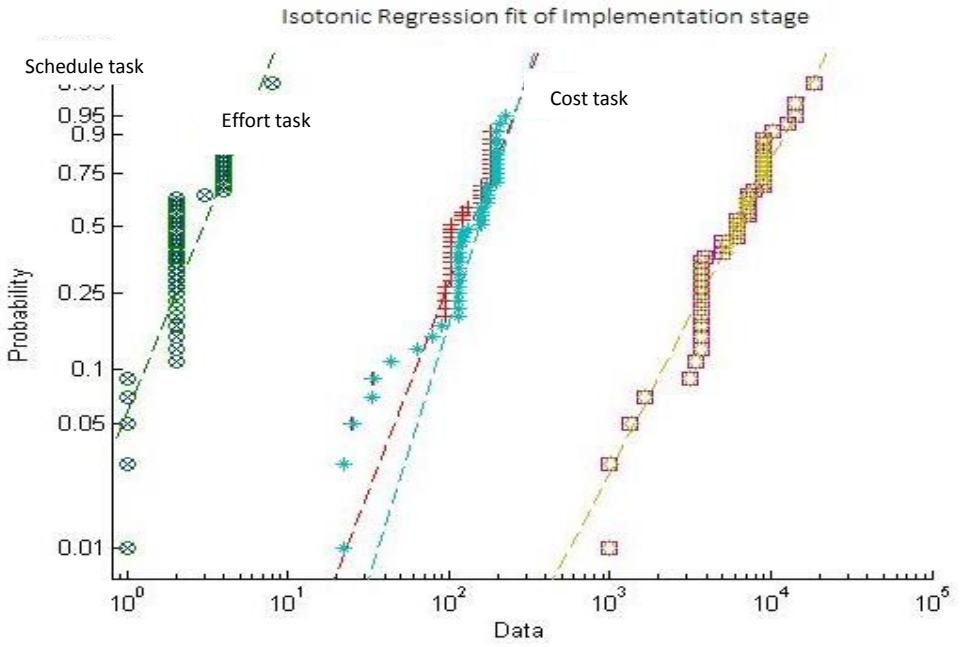

Fig. 8 Calibration curve of Implementation stage

From fig. 8, represents the schedule, effort and cost actual and predicted fit in the plot, shows the isotonic fit gives a new calibration curve in implementation stage of SDLC where the actual value is equal to the estimated value. It express goodness of fit well calibrated curves can also indicate the estimation quality. The confidence level is 99.78 %, it's come out with better estimation, effort and cost task process in maintenance.

CONCLUSION

- In this research work, it can be concluded that an estimated and actual value of planning, designing and building stage are plotted and shown the calibration curve of data points which has fitted in the plot.
- In this research all stage of scheduling day, actual effort and actual cost of software development life cycle represents positively fitted which is a data point above the graph express goodness of fit in the calibration curves.
- Mostly the values are fitted from the plot analyzed as in positive where the data point is above the graph of each stage of the project work. Isotonic Regression is more powerful when there is sufficient data to prevent over fitting based on Pair-adjacent violators (PAV) algorithm is used to fit the training set according to this mean square error criterion.
- From the estimated and actual values are plotted as scatter diagram and isotonic regression analysis is made. It is found that isotonic regression fits for schedule, effort and cost task as well as shown in Fig. 2 to Fig. 8 representing the planning stage, designing stage, building stage, SIT stage, UAT stage, Integrate testing stage and implementation stage.
- An Isotonic regression line shows a new calibrated curve where the actual value is equal to the estimated value which express goodness of fit of the calibration curves can also indicate the estimation quality. The data points fitted positively which can data points above the graph.

REFERENCES:

[1] Stephen H.Kan, "Metrics and Models in Software Quality Engineering", 2nd edition, , Addison-Wesley, 2003, ISBN 0-201-72915-6.

[2] M. Weske, "Business Process Management: Concepts, Languages, Architectures". Springer.

[3] William A. Florac and Anita D. Carlton , "Measuring the Software Process: Statistical Process Control for Software Process Improvement", Addison-Wesley, 1999, ISBN 0-201-60444-2.

[4] R. L. Dykstra and T. Robertson. An algorithm for Isotonic Regression for two or more independent variables. The Annals of Statistics, 10(3):pp. 708– 716, 1982.

[5] S. M. Mwagha, M. Muthoni, and P. Ochieg, "Comparison of nearest neighbor (ibk), Regression by discretization and isotonic regression classification algorithms for precipitation classes prediction," International Journal of Computer Applications, vol. 96, pp. 44-48, 2014.

[6] Han, J. &Kamber, M. (2012). Data Mining: Concepts and Techniques. 3rd.ed. Boston: Morgan Kaufmann Publishers.

[7] Kaner, Cem and Walter P. Bond, Software Engineering Metrics: What Do They Measure and How Do We Know?

[8] Kalai, AT and Sastry, R (2009), "The Isotron algorithm: high-dimensional isotonic regression", Proc. Comp. Learning Theory (COLT) 2009.

[9] Salanti, G and Ulm, K (2001), "Multidimensional isotonic regression and estimation of the threshold value", Discussion paper 234, Institute fu¨r Statistik, Ludwig-Maximilians Universita¨t, Munchen.