RESEARCHERID
THOMSON REUTERS

ISSN: 2277-9655
[Veni* *et al.*, 6(12): December, 2017] Impact Factor: 4.116
IC™ Value: 3.00 CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

COMPARISON OF LINEAR REGRESSION AND ISOTONIC REGRESSION ANALYSIS IMPLEMENTED FOR PROJECT MANAGEMENT IN SOFTWARE DEVELOPMENT LIFE CYCLE

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

DOI: Will get Assigned by IJESRT Team

ABSTRACT

In this research work, a decision making database of supervised learning is proposed, that is a set of possible estimator are working together to estimate the project completion task in it. Further, predicting the success of software projects based upon information related to estimated task and actual task is deemed to be one of the vital activities in software engineering research. In this paper to focuses an algorithm of isotonic regression classifier give better result when compare with Linear regression.

KEYWORDS: Isotonic Regression; Linear regression; Effort and cost estimation; project management

I. INTRODUCTION

Software effort and cost estimation plays an important role in software development [1] [2] [6], often determining the success or failure of contract negotiation and project execution based on the estimation plan. Predicting the success of software projects based upon information related to estimation task and actual task is deemed to be one of the vital activities in software engineering research. In this research work, isotonic regression classifier can be implemented for analyzing the estimation quality based on estimation value and actual value to achieve better result from the model based upon information related to Estimation task and actual task in IT project management.

II. LINEAR REGRESSION ANALYSIS

Linear regression analysis is the most widely used of all statistical techniques: it is the study of linear, additive relationships between variables. Let Y denote the "dependent" variable whose values you wish to predict, and let $X_1, \ldots, X_k$ denote the "independent" variables from which you wish to predict it [5], with the value of variable X_i in period t (or in row t of the data set) denoted by X_{it}. Then the equation for computing the predicted value of Y_t is:

$$\hat{Y}_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + \ldots + b_k X_{kt}$$

This formula has the property that the prediction for Y is a straight-line function of each of the X variables, holding the others fixed, and the contributions of different X variables to the predictions are additive. The slopes of their individual straight-line relationships with Y are the constants $\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_k}$, the so-called coefficients of the variables. That is, $\mathbf{b_i}$ is the change in the predicted value of Y per unit of change in X_i, other things being equal. The additional constant $\mathbf{b_0}$, the so-called intercept, is the prediction that the model would make if all the X's were zero (if that is possible). The coefficients and intercept are estimated by least squares, i.e., setting them equal to the unique values that minimize the sum of squared errors within the sample of data to which the model is fitted. And the model's prediction errors are typically assumed to be independently and identically normally distributed. Essentially, regression is the "best guess" at using a set of data to make some kind of prediction. It's fitting a set of points to a graph.

III. RESEARCH METHODOLOGY - ISOTONIC REGRESSION CLASSIFIER

Isotonic regression belongs to the family of regression algorithms an equation represents in equation (1). Formally isotonic regression [3] [4] is a problem where given a finite set of real numbers $Y=y_1,y_2,...,y_n$ representing observed responses and $X=x_1,x_2,...,x_n$ the unknown response values to be fitted finding a function that minimizes with respect to complete order subject with
respect to complete order subject to $x_1 \leq x_2 \leq ... \leq x_n$ and $y_1 \leq y_2 \leq ... \leq y_n$ where w_i are positive weights.

$$f(x) = \sum_{i=1}^{n} w_i \left(y_i - x_i \right)^2 \quad (1)$$

The resulting function is called isotonic regression and it is unique. It can be viewed as least squares problem under order restriction. Essentially isotonic regression is a monotonic function best fitting the original data points.

Instead of binning, a more reasonable solution seems to be a piecewise linear approximation. Just fit linear functions through each bin that remains isotonic. The prediction model is devised on the newly formed training set with a classification method of user's choice. Based on the built classifier's predictions for examples in the calibration set, a procedure for isotonic regression is called, returning calibrated probabilities.

The complexity of determining an isotonic regression depends on the regression metric and the partially ordered set. For example, for a a linear order it is well-known that a simple left-right scanning approach using pair adjacent violators (PAV), can be used to determine the L2 isotonic regression in Θ(n) time, L1 in Θ(n log n) time, and L∞ on unweighted data in Θ(n) time. An algorithm taking Θ (n log n) time is given for L∞ isotonic regression on weighted data.

The sigmoid transformation works well for some learning methods, but it is not appropriate for others successfully used a more general method based on Isotonic Regression to calibrate predictions from decision trees [7] [8] [9]. The probability of y_i is an isotonic (monotonically increasing) function of the values of the learners decision function. This is 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 be isotonic (monotonically increasing). That is, given the predictions f_i from a model and the true targets y_i , the basic assumption in Isotonic Regression is that:

$$y_i = h(f_i) + \varepsilon_i \quad (2)$$

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 2 and 3.

$$h = \arg\min_{k} \sum_{i=1}^{n} (y_i - k(f_i))^2 \quad (3)$$

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 is prone to over fitting when data is scarce.

Pseudo code for Pair-Adjacent Violator Algorithm

1. **Input:** Training data set of project management of (f_i, y_i) sorted according to f_i
2. Initialize $\hat{h}_{i,i} = y_i, w_{i,i} = 1$
3. While $\exists$ i s.t. $\hat{h}_{k,i-1} \geq \hat{h}_{k,l}$

   Set $w_{k,l} = w_{k,i-1} + w_{i,l}$

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

$$\hat{h}_{k,l} = \frac{\left(w_{k,i-1} \hat{h}_{k,i-1} + w_{i,l} \hat{h}_{i,l} \right)}{w_{k,l}}$$

Set

Replace $\hat{h}_{k,i-1}$ and $\hat{h}_{i,l}$ with $\hat{h}_{k,l}$

4. Output the step wise constant function

$$\hat{h}(f) = \hat{h}_{i,j} \quad \text{For } f_i < f < f_j$$

From the pseudo code, first assigns a probability of one to each positive instance and a probability of zero to each negative instance, and puts each instance in its own group. At each iteration, for adjacent violators: adjacent groups whose probabilities locally increase rather than decrease. When it finds such groups, it pairs them and replaces their probability estimates with the average of the group's values. It continues this process of averaging and replacement until the entire sequence is monotonically decreasing. The result is a sequence of instances, each of which has a score and an associated probability estimate, which can then be used to map scores into probability estimates.

The estimated and actual values are calibrated using isotonic regression which provides estimates of values of the dependent variables from values of the independent variable. The device used to accomplish this estimation procedure is the regression line. This method is more general in that the only restriction is that the mapping function be isotonic.

IV. IMPLEMENTATION AND EXPERIMENTAL ANALYSIS OF ISOTONIC REGRESSION VIA PAIR ADJACENT VIOLATOR ALGORITHM.

In this research work, the evaluation can be carried out in all stage of Software Life Cycle Development such as planning, designing, building for project development. The collection of fifty projects of data can be imported into the data repository file for analyzing the predicted true values in calibration curve. Here the sample training data values as shown in table 1. Similarly, the data of other two stage of project work can be collected and it can be analyzed based on isotonic regression classifier.

Table 1. Sample Training Data set of Business Project Development Task Values in planning stage.

Planning					
Planning Estimated day	Planning Actual day	Planning Effort Planned	Planning Actual Effort	Planning _Planned Cost $	Planning _Actual Cost $
22	22	437.7142857	437.7142857	20627.73714	20627.73714
22	22	328.2857143	328.2857143	15470.80286	27406.20286
15	16	1382.964	1382.964	49095.222	50372.157
15	15	1222.1975	1222.1975	43388.01125	45536.22275
8	9	365	365	12967.52	13114.6
18	18	101.4	101.4	4263.53	4361.53
18	18	169	169	7700.32	7777.32
11	11	259.995	259.995	10220.4	10220.4
19	19	363.7	363.7	18171.378	18171.378
12	12	363.7	363.7	18171.378	18171.378
16	16	636	636	22896	22896
15	15	173.33	173.33	6228	6228
4	5	244.196	244.196	10207.18	10207.18

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

5	6	66.1	66.1	3335.898	3335.898

The estimated and actual values are calibrated using isotonic regression which provides estimates of values of the dependent variables from values of the independent variable. The device used to accomplish this estimation procedure is the regression line. This method is more general in that the only restriction is that the mapping function is isotonic.

When compared with the existing method of the linear regression equation, isotonic regression classifier gives 99.79% of the confident level for analyzing the performance by applying in software metric.

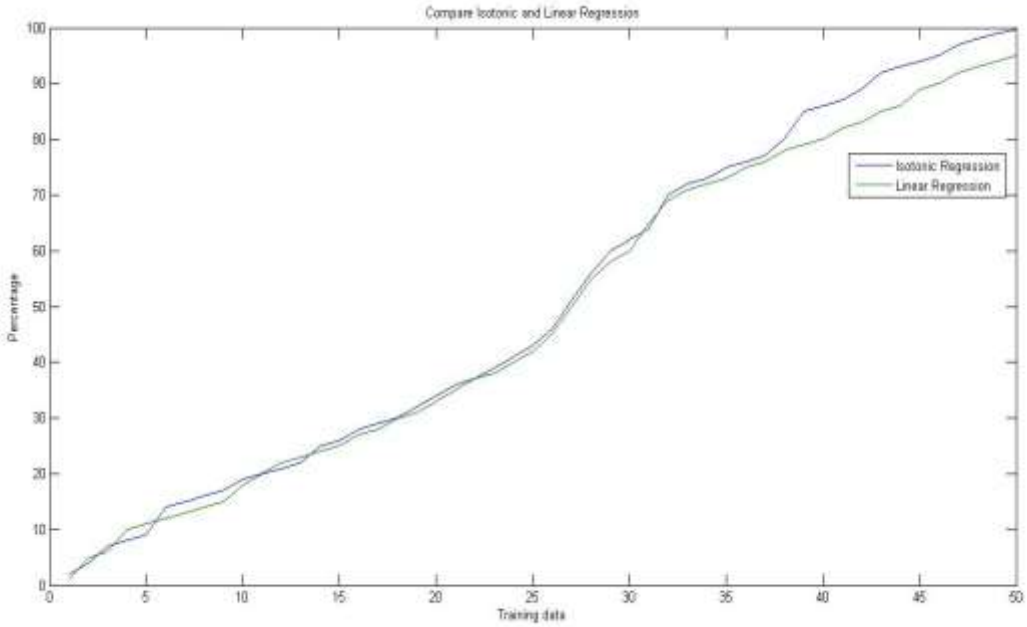

Fig. 1 Comparison of Isotonic regression and linear regression Algorithm

V. CONCLUSION

An experiment demonstrates that feature selection and isotonic regression methods improve the accuracy of predictions for match performance of based on each stage data of project development stage, compared to regression methods alone.

This work investigates the effectiveness of using computer-based machine learning isotonic regression methods to predict performance data for project development based on parameters collected on each stage of planning, building and designing task of estimated and actual values.

In this research, it can be concluded that to apply calibration of an intuitive model for analyzing the estimation quality based on estimation value and actual value to achieve better result from the model based on isotonic regression classifier.

VI. REFERENCES

[1] M. Weske, “Business Process Management: Concepts, Languages, Architectures”. Springer.
[2] 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.
[3] 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.
[4] 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.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

[5] Han, J. &Kamber, M. (2012). Data Mining: Concepts and Techniques. 3rd.ed. Boston: Morgan Kaufmann Publishers.
[6] Kaner, Cem and Walter P. Bond, Software Engineering Metrics: What Do They Measure and How Do We Know?
[7] Kalai, AT and Sastry, R (2009), "The Isotron algorithm: high-dimensional isotonic regression", Proc. Comp. Learning Theory (COLT) 2009.
[8] 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.
[9] S. M Kakade, V. Kanade, O. Shamir, and A. Kalai. Efficient learning of generalized linear and single index models with Isotonic Regression. In NIPS. 2011.

CITE AN ARTICLE:

Veni, S., Dr, & Srinivasan, A. (n.d.). COMPARISON OF LINEAR REGRESSION AND ISOTONIC REGRESSION ANALYSIS IMPLEMENTED FOR PROJECT MANAGEMENT IN SOFTWARE DEVELOPMENT LIFE CYCLE. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY*, *6*(12), 159-163.