EXTREME PROGRAMMING PROJECT PERFORMANCE MANAGEMENT BY STATISTICAL EARNED VALUE ANALYSIS

Wei Lu, Duke University
Li Lu, University of Electronic Science and Technology of China

ABSTRACT

As an important project type of Agile Software Development, the performance evaluation and prediction for eXtreme Programming project has significant meanings. Targeting on the short release life cycle and concurrent multitask features, a statistical earned value analysis model is proposed. Based on the traditional concept of earned value analysis, the statistical earned value analysis model introduced Elastic Net regression function and Laplacian hierarchical model to construct a Bayesian Elastic Net model fitted for project performance evaluation and prediction. The model is demonstrated with the JAX Laboratory software development project data. With simulated coefficients estimation, we realized an empirical data support for project performance assessment.

JEL: C35, C63, M15

KEYWORDS: Project Management, Performance, Prediction, Earned Value

INTRODUCTION

There has been growing tendency for the usage of the Agile Software Development paradigm these years, due to its claimed lower costs, better performance, productivity, business satisfaction features by Mishra (2011). This paradigm can be applied to supply chain management (SCM), a complex software development project. Considering its wide scope and complex requirements, predictable models for software development process are not fitted. Targeting such problems involving variability and uncertainty, agile methodologies are adaptive rather than predictive. Thus, for project management, how to evaluate and manage the process performance aiming this development mode is an important issue.

Project Management Body of Knowledge (PMBOK), presented by U.S.A Project Management Institution (PMI), regards earned value management as a key method for project performance management. For eXtreme Programming (XP) software project, the difficulty for spreading the application lies in the earned value determination during the process. Especially due to the variability of targets on a later stage of the project, the determination of earned value seems more difficult. Kim and Reinschmidt (2011) proposed a probabilistic cost forecasting method based on self-adaptive inside view (the bottom-up estimate), combining with outside (top-down) view of project cost estimates using Bayesian inferences and model averaging technique. Its precision is high under the condition of linear cumulative cost curve, but models with highly nonlinear features will affect its suitability.

In this paper, we introduced a concept of statistical earned value analysis (SEVA). Targeted on the concurrent features of short release cycle, programming, test, and maintaining units for software extreme programming, an analytical model is proposed. This model is able to combine the data of project implementation process with Bayesian theory, and conduct posterior analysis according to the nonlinear characteristics, thus provide timely performance evaluation data analysis to project manager.

The remainder of the paper is organized as follows. In the next section, we review relative literatures; the data and methodology section, a short introduction to EVA is presented, and then an explanation of the Bayesian Elastic Net model followed by a description of the statistical inference methods employed. In the results and discussion part, the project case used is illustrated; subsequently, the analytical results are presented.

LITERATURE REVIEW

With increasing levels of complexity and uncertainty, project management is a vital task. Due to its contemporary and unique features, it is imperative to gauge the performance. Much work has been made in this area. For example, Hillman (1994) uses the EFQM Model to suggest a practical framework for progress self-assessment. Based on it, Bryde (2003) proposed a project management performance assessment (PMPA) model, using six criteria for assessing PM performance: "project management leadership; project management staff; project management policy and strategy; project management partnerships and resources; project life cycle management processes; and project management key performance indicators". Qureshi, Warraich and Hijazi (2009) further examined the impact level of these criteria and their association scope.

As is put forward by Wysocki (2009), two variables can be used to define the project landscape: "goal" and "solution". With constant variability of goals and corresponding solutions, Extreme Project Management is an important category of PM. There is a wide range of work addressing the issues related to Extreme Programming (XP) and its process management, e.g., the work of Beck (1999) and Darwish (2011). According to PMBOK, earned value management is an efficient approach for project performance management. By conducting progress and cost analysis through Schedule Performance Index (SPI) and Cost Performance Index (CPI), this method is effective conditioned on two premises: first, earned value is convenient for acquiring and estimating, such as engineering project instance, completed progress's proportion of the total work amount would be easier to estimate; second, the ultimate goal of the project is relatively fixed, which would make the earned value estimation benchmark more constant.

However, as is discussed by Yap (2006) and Zhai etc. (2011), the difficulty of earned value determination during the process for XP software projects lays in its requirement frequent variability, deficiency and flood characteristics. The probabilistic cost forecasting method proposed by Kim and Reinschmidt (2011) is based on self-adaptive inside view combining with outside view of project cost estimates using Bayesian inferences and model averaging technique. During the execution phase, through incorporating actual predictive performance and pre-project cost estimation modification, this method can sufficiently use available prior information. It is precise for linear cumulative cost curve, but not for models with highly nonlinear features.

Targeting those unique features of XP performance management, an adaptive model should be implemented. To analyze earned value data and give predictions for future performance, we consider a form of regression. The problem of interest is the estimation of the regression coefficient parameters as well as the prediction accuracy. Tibshirani (1996) proposed the Lasso method to address the feature selection problem and is widely applied in a variety of statistical analysis. However, many project process features has high correlation, while Lasso tends to arbitrarily select just a few of those relevant features and ignore the rest. This property of Lasso will undermine model's interpretability and robustness. Thus, a naïve Elastic Net estimator is proposed by Zou and Hastie (2005), which yield sparse and grouped variable selection. Sequencing work by Park (2008) and Chen etc. (2011) provided the Bayesian and hierarchical from of Elastic Net method to further enhance the flexibility and accuracy of the model.

DATA AND METHODOLOGY

The nature of the XP problem is shown in Figure 1; from which we can see that the prophase and interim performance prediction analysis, with the prophase performance benchmark obtained at early stage, will gradually be adjusted during the development process when performance information and clearer project requirements are gathered from clients and developers.

The data is collected during a three-month software-developing project at JAX Laboratory in 2012. Through work breakdown structures (WBS), the total project quantities are determined. The weekly meetings provide assessments of performance for finished tasks and newly added workload. So as to realize plan updating, actual completion degree estimation and sample data collection.

Figure 1: Performance Evaluation and Prediction before and after Requirements Change
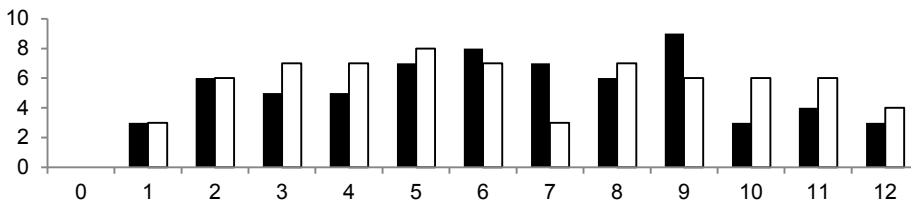
This figure shows the idea behind XP project where darker bars indicate expected performance while lighter color bars are performance evaluations after requirements change

Considering the differences between base and empirical performance for XP projects, the life cycle data is regarded as having truncated features. That is to suppose that only partial life cycle of subroutine individual (explicit requirement) is less than certain value, while the residual potential demand development time exceeds certain value. Since distinct truncation will lead to differences, the SEVA methods are based on Bayesian survival analytical theories. Using Elastic Net method and MCMC steady-state simulation method or variational Bayesian (VB) inference algorithms as derived by Park and Casella (2008), we constructed project Bayesian survival analysis model, in order to solve the difficulties of high dimensional numerical integration. By constructing XP project life period regression model, the influence of environmental and target conditions variability on the project performance is effectively reflected. The key steps for this analytical model, which are derived in detail by Chen, Carlson and Zaas (2011), are as follows:

Step 1: Construct Elastic Net regression target function to solve multi-linear problem as a substitution of Principal factor analysis;

$$\hat{\beta}(Naive\ ENet) = arg\ min\ \|y - X\beta\|_2^2 + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2 \tag{1}$$

This is the naïve elastic net criterion, which can be viewed as a penalized least squares method and a convex combination of the lasso and ridge penalty.

Step 2: Construct Laplacian Hierarchical Model with the following Laplace prior form;

$$p(\beta|\tau,\gamma) = \prod_{j=1}^p \frac{\sqrt{\gamma_j\tau}}{2} exp\left(-\sqrt{\gamma_j\tau}|\beta_j|\right)$$
$$= \prod_{j=1}^p \int N\left(\beta_j; 0, \tau^{-1}\alpha_j^{-1}\right) InvGa\left(\alpha_j; 1, \frac{\gamma_j}{2}\right) d\alpha_j \tag{2}$$

Step 3: Construct the complete Bayesian Elastic Net Model. A Gamma prior is imposed on each individual γ_j to avoid tuning;

$$y \sim N(y; \mathbf{X}\beta, \tau^{-1}\mathbf{I})$$
$$\beta_j \sim N\left(\beta_j; 0, \tau^{-1}(\alpha_j + \lambda_2)^{-1}\right)$$
$$\tau \sim Ga(\tau; c_0, d_0)$$
$$\alpha_j \sim \eta(\alpha_j/(\alpha_j + \lambda_2))^{\frac{1}{2}} InvGa\left(\alpha_j; 1, \frac{\gamma}{2}\right) Ga(\gamma_j; a_0, b_0) \tag{3}$$

Where η is a normalizing constant, and λ_2 is a parameter tuned by cross validation.

Step 4: Put forward the specific hyper-parameters of fully extended model.

$$p(\mathbf{y}, \beta, \tau, \alpha, \gamma) \propto N(y; \mathbf{X}\beta, \tau^{-1}I) Ga(\tau; c_0, d_0)$$
$$\times \prod_{j=1}^{p} N\left(\beta_j; 0, \tau^{-1}(\alpha_j + \lambda_2)^{-1}\right)(\alpha_j/(\alpha_j + \lambda_2))^{\frac{1}{2}}$$
$$\times InvGa\left(\alpha_j; 1, \frac{\gamma_j}{2}\right) Ga(\gamma_j; a_0, b_0) \tag{4}$$

RESULTS AND DISCUSSION

The results for the JAX Laboratory case study are presented in this section. This is a comprehensive system development project which requires data preprocessing, singular value decomposition (SVD), single-locus regression, Extreme Value Decomposition (EVD) test, pair-wise linear regression analysis, error propagation calculation, permutation test and networking construction in a 3-month period. Using XP mode, this project gradually perfects the identification and range definition of final target. Thus, an EVA method combining with dynamic prediction is required. Due to the uncertainty in initial phases of development, there are several modifications of bench plan during the process requiring prediction analysis to determine the best action strategy.

Here, we use character A-J to represent each step of JAX project. Each of them contains about 50 activities. Performance analysis is conducted on a daily basis. By adopting statistical earned value analysis (SEVA) methods, we applied Bayesian posterior model to performance prediction. Based on Bayesian Enet model presented in the previous section, SEVA results during the project implementation process are shown in Table 1. After simulating with Schedule Performance Index (SPI) and Cost Performance Index (CPI) analysis, we can obtain the project performance evaluation for each stage (Table 2).

We applied the Bayesian Elastic Net model introduced in the previous section to 12 groups of simulated data, where each group includes 10 earned values of sequential steps A-J. The simulation results are shown in Figure 2. The circles represent our prior expectations of the project performance, while the crosses represent posterior performance estimations. After using five samples for regression coefficient training, we modified our performance beliefs for the left seven samples. From the figure, we can see that the residuals are within an acceptable range, which proves the effectiveness of SEVA method.

Table 1: Statistical Earned Value Analysis during the Project Implementation Phases

| A	B	C	D	E	F	G	H	I	J
0.077	0.096	0.161	0.064	0.161	0.096	0.080	0.064	0.125	0.074
0.073	0.091	0.152	0.061	0.152	0.091	0.076	0.061	0.119	0.070
0.073	0.118	0.210	0.085	0.152	0.091	0.076	0.061	0.119	0.070
0.069	0.112	0.199	0.080	0.144	0.087	0.072	0.058	0.113	0.066

This table shows the earned value of each project stage, where A-J represents the steps in sequence.

Table 2: SPI and CPI Analysis during the Implementation Process

SPI	1.7	3.6	1.5	2.2	0.5	1.4	1.6	2.2	2.9	1.4
CPI	0.5	1.0	1.4	1.9	0.3	1.4	1.8	1.1	3.5	3.2

This table shows the SPI and CPI values of each project stage.

Figure 2: Performance Prediction Using Bayesian Enet Model
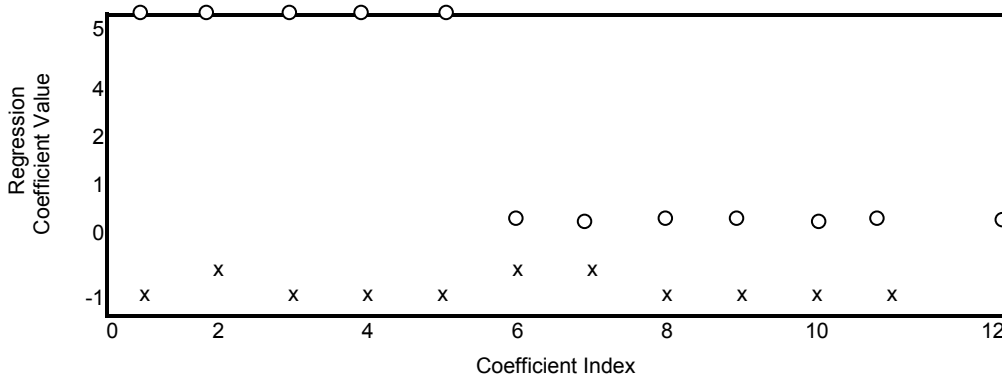
This figure shows the Bayesian Enet estimation for performance prediction. Circles represent prior beliefs, while crosses indicate the posterior performance values for each simulated sample. The first five samples are used for regression coefficient training, and the last seven ones are used for prediction testing.

CONCLUDING COMMENTS

A unique feature of XP project is its changing target with increasing clarity. This requires a correspondent reliable performance management mode. To solve the prediction difficulty occurring in traditional earned value analysis due to constant change of targets, an SEVA method is proposed in this paper and a case study based on performance data collected at JAX Laboratory is presented to illustrate the efficiency of the method.

After simulating with Schedule Performance Index and Cost Performance Index analysis, we obtain the project performance evaluation for each stage. By implementing the hierarchical Bayesian Elastic Net model and using the collected sample earned value at each project stage as our regression data matrix, the performance beliefs are updated. With regression coefficient residuals within an acceptable range, the SEVA method is proved an efficient and adaptive approach for dynamic project prediction and strategy management.

The SEVA method proposed in here is analyzed based on the work breakdown structure of a typical XP project. Thus, its limitations are inevitable due to the specific feature and duration of the project. When applying to multi-projects in large software companies or other industries, more attention should be paid on the data gathering part. Through more targeted definition of performance indexes such as time, cost and quality, we can improve the effectiveness of Enet model by enlarging project data sampling, to enhance the reliability of this method.

REFERENCES

Mishra, A. Mishra (2011) "Complex Software Project Development: Agile Methods Adoption," *The Journal of Software Maintenance and Evolution Research and Practice*, vol. 23, April, p. 549-564

Project Management Institute (PMI) (2008) "A Guide to the Project Management Body of Knowledge," 4th ed., Newtown Square, PA

Kim & Reinschmidt (2011) "Combination of Project Cost Forecasts in Earned Value Management," *The Journal of Constr. Eng. Management*, vol. 137, June, p. 958-966

Hillman, G.P. (1994) "Making Self-assessment Successful," *The TQM Magazine*, vol. 6 No. 3, pp. 29-31

Bryde JD (2003) "Modeling Project Management Performance," *The International Journal of Quality Reliability Manage*, vol. 20(2), p. 225–229

Qureshi, Tahir Masood; Warraich, Aamir Shahzad; Hijazi, Syed Tahir (2009) "Significance of Project Management Performance Assessment (PMPA) Model," *The International Journal of Project Management*, vol. 27, p. 378–388

Wysocki, Robert K. (2009) "Effective Project Management : Traditional, Agile, Extreme," 5th ed., Wiley Pub.

Beck, Kent (1999) "Extreme Programming Explained: Embrace Change," *Addison Wesley*

Darwish, Nagy Ramadan (2011) "Improving the Quality of Applying eXtreme Programming (XP) Approach," *The International Journal of Computer Science and Information Security*, vol. 9 (11), p. 16.

Yap, Monica (2006) "Value Based Extreme Programming," *Agile Conference,* July, p. 184-191

Zhai, Li-li; Hong, Lian-feng & Sun, Qin-ying (2011) "Research on Requirement for High-quality Model of Extreme Programming," *International Conference on Information Management, Innovation Management and Industrial Engineering*, vol.1, p. 518 – 522

Tibshirani, R. (1996) "Regression Shrinkage and Selection via the Lasso," *The Journal of Royal Statistical Society Series B*, vol. 58, p. 267–288

Zou, H.; Hastie T. (2005) "Regularization and Variable Selection via the Elastic Net," *The Journal of Royal Statistical Society Series B*, vol. 67, p. 301–320

Chen, Minhua; Carlson, David & Zaas (2011) "Detection of Viruses via Statistical Gene-Expression Analysis," *IEEE Transaction of Biomedical Engineering*, vol. 58(3), March, p. 468-479

Park, Casella (2008) "The Bayesian Lasso," *The Journal of the American Statistical Association*, Vol. 103, p. 681–686

BIOGRAPHY

Wei Lu is Electrical and Computer Engineering graduate focus on statistical data mining at Duke University, who can be reached at wei.lu@duke.edu.

Li Lu is Associate Professor of Management and Economics School at University of Electronic Science and Technology of China, who can be reached at luli@uestc.edu.cn.