



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Approaches to Process Performance Modelling

B R Sharma

Department of Computer Science
HMR Institute of Technology and Management
Indraprastha University New Delhi, Hamidpur
New Delhi- 110 038, INDIA

Abstract: Project Management improves by adopting better processes models. It is vital to improved project management that efficient process models need to be developed. These processes be evaluated on continuous basis and improved with changes in operating environment and new learning. The quality metrics can be pre-identified for achieving the goals that may include delivered defect density, rejection rate, time estimation etc. In this approach to build PPM, the emphasis is on process enhancement so that efficiency can be increased rather relying fully on various components. The various process models discussed are Regression, Bayesian, Queuing, Neural networks etc. With details of theories and application areas these may be applied to.

I. Introduction

The CMMI defines Process Performance Model (PPM) as “relationships among attributes of a process and those of its work products. These work products are developed from historical process-performance data, and further calibrated, by using process and product output data results obtained from the project. PPMs are used for result prediction to be achieved by following a process.”

In any business case, a model on an organizational level is needed that is capable of predicting one or more business target results. This is called a process performance model and can plan and track organizational quality performance. The adoption of better PPMs helps understand or even predict the uncertainties better in CMMI Level 4 and 5 processes namely Quantitative Project Management and Qualitative Process Management. Thus it helps in reducing risk by controlling relevant process/sub-processes. Thus, PPMs are useful tools for managing projects/ processes. PPMs tools enable prediction of process performances, by Project Managers with a level of confidence as they better identify and understand risks with these tools. “What-if” analyses on PPMs enables to see impact various courses of action on their projects by varying a factor under control.

PPM models are by definition mathematical and statistical in nature. PPMs are empirical in nature and are never perfect in prediction but quite accurate to use available data and make a fair prediction about a process. The process of building PPMs involves following steps:

- Identify or Reconfirm Business Goals.
- Identify the sub-processes/process
- Identify Outcomes to Predict (y's)
- Identify Controllable factors (x's) to predict outcomes
- Include Uncontrollable x factors
- Collect Data
- Assess Data Quality and Integrity
- All data types of outcomes y and x factors needs to be identified.
- Create Process Performance Baselines (PPBs)
- Select the right analytical techniques and also type of regression equation.
- Based on analysis and make predictions with appropriate prediction intervals.
- Statistically manage sub processes with PPMs
- Take Action Based on PPM Predictions
- Maintain PPMs including calibration and reconfirming relationships

There are challenges of building PPMs due to limited quantitative data and low variability of available data. The PPM helps in predicting Quality and Process performance/ Business objectives. Once an organization decides its goals, it starts looking for PPMs

II. Types of Models

PPMs Models can be classified as follows:

- Physical Models represent physical reality, for understanding, by scaling it down by building a prototype, for example a prototype of building.
- Mathematical Modelling involves representation of reality by a mathematical model in terms of a mathematical equation or other similar models.

- Process Modelling will have the Entire flow of Process with factors and conditions modelled to understand the bottlenecks in the process/system for correction.

III. Process Modelling –Steps, Validation, Tools

PPMs can help in analysis whether the selected process for a project will be able to meet the project's performance objectives. A good PPM should have modelling characteristics as follows:

- PPMs can be statistical, probabilistic and simulation types that link past performance with predictable outcomes. PPM should have one or more measurable attributes. This attributes represent a controllable input to a process/sub-process and helps in monitoring performance of a process/ sub-process with “What if analysis” for planning, dynamic re-planning, problem resolution. Thus prediction accuracy of model is important.
- PPMs should give insight to predicted range of variation expected from predicted results with variation of input factors. PPMs can identify potential sources of the problems or defect. Thus calibration abilities of model have bearing on performance of model.
- Robustness and Data type of the variable and factors involved in a model have bearing on performance of the model. There is need of availability of relevant tools for building the model. Model to have ability to include all critical factors in primary data type.
- PPM should enable simulation of models and help enable finding optimization of input factors for planning process/sub-process and thus there is need for flexibility in varying factors in model.

IV. How are PPMs Used

- Projects can use PPM for predicting the process performance after estimation and analysis of the defined processes. It can assist in selecting defects for analysis and also performance analysis for selecting processes or sub processes.
- PPMs can be used to understand the impact of improvements, related interaction of these improvement and resulting costs effect, and other benefits. They can be used to assess the ROI for process improvement work.
- PPMs help in estimation progress towards achieving planned quality for project and process performance objectives with the use of right tools. These estimates can be done in advance and not wait until a future phase of life cycle of project.
- When a sub-process is initially executed, based on prior instances of suitable data of similar sub-processes, process performance baselines or PPMs help to understand the impact/benefits of new innovations. PPMs provides insight into effect on process capability and performance due to effect of process changes. Also, PPMs can provide a means for analyzing possible effects due to effected changes to process elements

V. Modelling Techniques in Real World

A) Regression based Models

Regression based process models estimate relationship in a mathematical equation among dependent and independent variables.. The model $Y = f(x) + \text{Error (unknown parameters)}$ assumes that a sample data represents whole population, variables are random and have random errors, no correlation amongst independent variables and unknown factors represent a sub process/process, else these are a factors that affect data set, project or sample. A regression model can be developed with relevant indicators and that has characteristics of data, continuous or categorical.

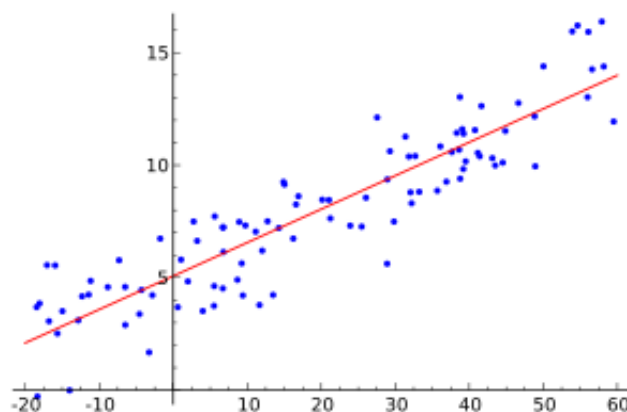


Figure 1: Illustration of linear regression on a data set.

Scatter plots amongst X versus Y helps to see if there is relationship (correlation). The coefficient of determination, R^2 is a number that indicates how well data fit a statistical model, a line or curve. R^2 values greater than 0.7, provides a better fit to the outcome. This explains how much the dependence Y has on X's. The fitment of model can also be checked by P values of independent variables X to be less than 0.05. This shows a significant relationship exists with Y.

R Systems, an Indian company, developed the PPMs using this procedure for a Regression model. R Systems used ANOVA to check the homogeneity and divided the data collected in various domains and technologies. This was to understand whether there is any variance in data. R System did the following:

- Performed normality test to find the distribution of data
- Use correlation analysis between each X factor and Y. Then verify the “Coefficient of Regression”- R Square, the result was more than 80% between X and Y.

R-Systems work resulted in a multiple regression equation, with values **R-square 83.5%, R Square (Adj) 81.3%**. The p values less than **.05 at 95% confidence level**. This gives a significant model for DDD (Delivered Defect Density).

B) Bayesian Belief Networks (BBN)

BBN is a probabilistic graphical model. BBN represents a set of variables and their probabilistic independencies (for example the probabilistic relationships between student and score). A BBN model can give quick insight into potential areas of improvements based on relevant quality factors and the current performance level of the organization. Bayesian Belief Networks (BBNs) offer another approach to modelling the quality of software products. These are probabilistic models with sets of nodes connected with arrows. These nodes represent certain relationships like: i) cause and effect ii) Without a confirmed cause and effect (probable) iii) Statistically observed correlation between two nodes.

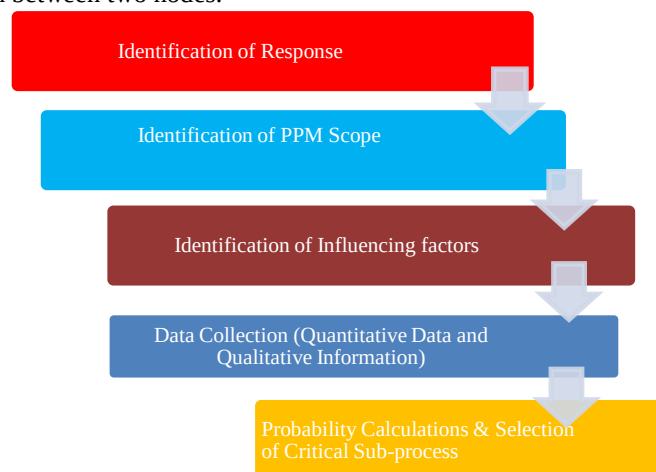


Figure 2: PPM Development through BBN

In practice, BBNs can be constructed to cover a time dimension such as a life cycle. An iterative process is conducted by identifying what factors occurring just prior to the terminal child nodes, outcome might explain the child nodes. The iteration continues in that each parent node can be explained in terms of its own parent nodes occurring earlier in the life cycle.

A BBN can be an invaluable management tool by enabling a “what-if” capability during project planning, and subsequently during all processes and sub-processes in the life cycle. The discussion on Quantitative Project Management (QPM) and Organizational Process Performance (OPP) process areas for probabilistic nodal relationships leads to conclusion that BBN is ideal approach to process modelling related to these requirements. BBN is frequently used in the Causal Analysis and Resolution and also in Organizational Innovation and Deployment process areas.

In Software Process Modelling with Bayesian Belief Networks is a model usable for effort estimation/ defect prediction. With some adaptations this model can be used to estimate quality within a project, BBN model's ability to predict quality improvement at an organizational level remains unanswered question.

A Bayesian model to develop business solutions has number of merits, it takes into account prior knowledge of problem, applicable in situations with scarce failure data. This model can include decision nodes. This model can also predict situations that are new and can formally incorporate objective/subjective information.

Strengths of Bayesian model is its ability to use prior information for initial estimation and to use information as it becomes available to improve and adjust estimates. This model can predict design and test work and can be used to investigate and decide on process improvements. BBNs look promising for modelling quality on an organizational level. Some attempts have been made to build probability models for testing using BBNs.

C) Neural Networks

Input/ Output nodes of a network are connected by links and hidden nodes. A human brain designs its reaction based on training simulation received by neurons under various instances/situations. A network similarly learns the input. The network produces its output, based on algorithm used and machine learning. There are number of network architecture exists, single layer, multi layer feed forward and recurrent architecture.

The network is supposed to learn and develop the patterns. Once a network learns right patterns, it reduces the overall network error. A sample data validates the network and checks the accuracy of its learning. A stable network has small learning and validation total mean squared error.

Normally continuous variable are used, however with new tools discrete data is also supported. Artificial Neural Networks is based on black box technique. In this approach inputs are used to determine the outputs, and normally no mathematical relationships can be defined between Input and Output. This approach gives better results than linear models.

If in a Modelling of manufacturing processes it is possible to define set of relationships which relate input and outputs of the process, then these descriptions are very useful in terms of optimization and controlling of the process. This is true for manufacturing problems related to process control, process optimization, system design and system planning. Artificial neural networks are thus a computational tool used in artificial intelligence for modelling manufacturing processes.

D) Reliability Modelling

The ability to perform at designed level of a software product, without any failure, indicates its reliability under execution. The scale 0 to 1 is used to measure reliability of any product or activity. In software product development processes, technique known as Reliability Modelling is used. This can be used for defect prediction based on arrival of defects, in each phase of development. The technique can be also be used for testing pattern of arrival of defects. In warranty defect analysis and also forecasting reliability.

Reliability can be time-dependent and also is non- time dependent. In first case time is an important measure as the occurring defects wear out with time. The defect occurs by executing faulty codes or programs in certain time frame and recorded as and when they occur. This concept is much in use in IT industry for MTTR to know Incident Arrival Rate, etc.

Reliability models normally have distribution curve of defects the shape of which reduces from peak to a lower flatter trajectory. The curve most commonly used is Weibull, logistic, small extreme value to fit in probability distributions. It is possible that each phase of software development may have different probability distributions. The defect data is measured in terms of number of defects in a period (ex: 15 /30/45 per day) or time difference of defect arrival (ex: 20, 35, 50 minutes. The PDF, CDF (Probability and Cumulative Distribution Function) are tools to understand the pattern of defects. Pre-Queuing system entity enables the creation of demand on arrival that is to be served by limited resources of the system. The discrete events in the system are managed by its resources at any given point in time. The model captures time stamps of different discrete events and models their variation with the queue system. This model can used to understand the resource utilization of servers, idle time, bottlenecks in the system events, etc.

Reliability Modelling is used in IT Industry for application maintenance, incident and problem handling. This is also applicable in cases involving Standard change Requests or contexts where the arrival rate of defects and team size is important factor for delivery on time.

E) Queuing Process Modelling

Queuing theory approach is applied to different types of problems, such as scheduling, resource allocation, and traffic flow. Queuing system model results in entity arrival that creates demand. This demand is to be serviced by limited resources assigned to the system. The Queuing system manages its resources to service various events at any given point in time. Queuing system is dependent on arrival of elements, utilization of servers, wait time/time spent in the system flows. It helps to understand the resource utilization of servers, bottlenecks in the system events, and idle time, etc.

In software Industry Queuing system can be simulated for use in application maintenance, incident/problem handling. This is also applicable in dedicated service teams /functions (ex: technical review team, Procurement). This model is applicable in a situation requiring standard change request, in which the arrival rate and team size has significance, in delivering on time. Queuing system modelling involves understanding of the tool "Process model".

Setting up flow: The actual flow of activities and resources in a system needs to be mapped and make a graphical flow and verify it. The distribution of time, arrival pattern of entities, input and output queue for each entity, resource capacity and assignment should be known. These can be solved by Time motion study for the first time. Once System knows pattern of entity arrival then by just adding storage the entities will be retained till resolution. Resources can be given in different shifts, by use of necessary functions and describing various scenarios. Their conditions of service can be changed to meet the real conditions.

Simulation: Simulation process is carried out to predict the outcomes of proposed queuing process. Replication of the simulation process can be carried out over a time period. The simulation results are displayed as output. Simulation reports can be customized by addition of new metrics and formulas. A Simulation situation can be run with or without animation. The output summary refers to idle time of entities or waiting time in queue. This is important area to work on process change and improve the condition. If there is no Hot Spot or idle time of entities, then there is need to study the activities with High Mean or Standard deviation or both of individual activities. These are critical sub processes to control.

Optimization: It is important to check that the system replicates the real life condition by comparing actual value with predicted values. This difference should be less than 5 %. It is desired to find the best combination of resource assignment along with different activities. The earlier defined scenarios are going to be the control factors and a certain range is provided in the tool (LSL and USL). The optimization is performed, with default value of convergence, simulation length left undisturbed. The tool looks at different combination of values with existing system. It selects the one which meets best our target, these values can be made use of in composition of processes (resource skill, activity and time taken etc).

F) Fuzzy Logic

Fuzzy Logic is a way of representation of a of problem expressed in a language with different variables that are fuzzy/vague in terms of their value, to take decisions. The variables do not have sharp boundaries in Fuzzy set theory and it permits overlapping. In Fuzzy systems membership values of fuzzy set are in 0 to 1 range with 0 for false value and 1 for absolute truth.

Fuzzy set theory is different from normal set theory for it allows each element of the given set belongs to that set to some degree between some value 0 to 1. For example if a person's skill index is 4.8 and we have two group which contains skill 3.4 to 5 and another group which contains 4.6 to 6. Member is part of both the groups with varying degrees, in this case. This shows the Fuzziness. This needs to understood that this is not probability but certainty with varying degree of membership in a group.

In Fuzzy logic the problem description in language is translated to the mathematical solution determined by Fuzzy rules. Fuzzy logic model use Fuzziness of data (overlapping values) in software problems, and mathematical/stochastic relationship determine the solution. Monte Carlo simulation is partially based on Fuzzy logic concepts.

G) Monte Carlo simulation

Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling to obtain numerical results. Monte Carlo Simulation allow, modelling of variables that are uncertain in a range of values. The model is useful in predicting outcomes and is increasingly used.

- Monte Carlo simulation enables a detailed calculation of the potential business results for candidate improvement proposals. Since only the identified areas are calculated in detail, the lead time toward a business case is shortened significantly. This helps a company to make quicker decisions when improving quality.
- This helps in more accurate sensitivity analysis and enables analysis in simultaneous effects of many different uncertain variables.
- User-provided values are included in analysis for uncertain variables.
- The above results in confidence in a model output as it supports risk management.

Monte Carlo simulation shows the distribution of the outcome factors (child nodes) based on the certain and uncertain parent nodes. Monte Carlo simulation tools use random number generators to select a value for each parent node from the parent node's distribution and then compute the outcome child node. This step is repeated thousands, if not hundreds of thousands, of times so that a distribution of values for the child node is formed. The distribution for each child node can be used in Monte Carlo simulation for any successive child nodes of the child node. Additionally, the Monte Carlo simulation provides distributions for all child nodes such that a confidence level can be established for any given value of the child node. This proves quite helpful in using results from Monte Carlo simulation in project forecasting, business case development for process improvement, and risk management.

VI. Conclusion

The solution to any problem is inherent in its definition and parameters governing the solution. A manufacturing engineer knows the quality of product produced at the shop depends not only on the design but also production process available at the site. Similarly in software solution it is possible to think parallel that efficient solutions will emerge based on the process selected. The selection of a particular project process model depends upon the nature of problem and quantum of data available. A model must yield actual results as closely as possible to the predicted result. In our approach to paper, suggestions have been given of various process models. The ability of engineer to pick up the best suitable model for the defined problem will lead to desired

efficient solution. Regression model thus will be applicable to problems with a defined relationship between variables impacting the solution in larger proportion. Bayesian Model will be appropriate where prior information of data related to a variable that impacts significantly on the posterior information which is basis of our solution.

References

- [1]. CMMI v1.3 Nov 2010, Technical Report
- [2]. Process performance models – statistical , probabilistic & Simulation–Vishnu Varthanan Moorthy (2014)
- [3]. Radllinski Lukasz. “A Survey of Bayesian Net Models for Software Development Effort Prediction” pp87-101 J S E T Volume 5, No.2. July-December 2013
- [4]. Simulation and Monte Carlo ‘Some General Principles’ - James C. Spall (2007)
- [5]. A New Measure for the Accuracy of a Bayesian Network: Alexandros Pappas, Duncan Gillies
- [6]. Introduction To Neural Networks : Prof. George Papadourakis, Ph.D. Monte Carlo Simulation : Fawaz hrahsheh , Dr. A. Obeidat
- [7]. Fuzzy Logic - Shane Warren, Brittney Ballard (2010)
- [8]. Assessment of the Modelling Abilities of Neural Networks by Alvin Ramsey (1994)
- [9]. Schneider, Henry. “ How to build Process Performance Models” North American SEPG Albuquerque 2012