

Future Risk Analysis Using Bayesian Network

Bansri Sorathia¹, Ompriya kale²

¹*Pursuing Master of Engineering in Information Technology.*

²*Faculty in Department of Computer Science and Engineering.*

L.J.I.E.T., Ahmedabad, India

Abstract- Now Days Company wants to deliver better, cheaper, efficient and reliable software products to the customer. And also customer's requirement becomes high with the high technology environment so probably company needs to launch high-tech software as well as complex project and services. But to develop efficient software there are so much risk arises during the software development process. By risk assessment we can identify and evaluate the risk at prior stage and also filter the risk based on their priorities. And by using CMMI we can get the guidelines for improvement of processes from that we can also identify the risks & its factor and the probability of that risks will be count with the use of Bayesian Network. Here in this paper we discuss risk assessment in detail and studied the models of risk assessment and to find out the probability of risk Bayesian network has been studied and its advantages and disadvantages.

IndexTerms—Risk assessment, CMMI, Bayesian Network

I. INTRODUCTION

Software project is full of high risk and many uncertainties. Because of the high rate of failure rate of software product there is growing concern about software project risk management. Risk factors can be predicted in advance through risk management, therefore such action like risk preventing, risk avoiding, risk transferring and risk sharing can be taken before the risk actually occurs to reduce software development risk and improve the success rate of software projects.

In software projects Risk is defined as a state of uncertainty where some of the probability of harm or loss in software. It can be negative and positive. Negative risk will specify loss or harm while positive risk represent a new opportunity for the project. To handle these kinds of risks, risk assessment models and techniques are there. [4]

There are certain questions related to risk management .the most important question is why we should have to manage risk? The answer of this question is that some level of risk is associated with them. Even if the product is under development is simply another version of existing system or product.

Risk has mainly 10 critical elements: Product requirement, Software complexity, Project staff, Software

Usability, Targeted reliability, Estimation method, Monitoring method, Development process adopted and tools, over the budget, beyond the schedule. This type of uncertainty arises during the development of software. Because of these uncertainties and others, project managers have to decide how to recover the project plan if above risk arises.

To handle the risk in organization they have risk management which is continuous and systematic process and it consists mainly two parts: risk assessment and risk control.

Now, risk management will applies to particular development process of software. And for particular process specific goal has been define in process area of CMMI model.CMMI model provides guidelines for applying best practices in a development organization. Best practices in the model focus on a activities for developing quality products and services which meet the need of customer.

To count probability of each risk factor Bayesian network can be used. And the risk factor derived from the particular process area by CMMI model. Based on graph theory and probability theory, Bayesian Network can describe the relationship between variables reasoning from incomplete, imprecise or uncertain information.BNs are casual networks that consist of a set of variables and provide the means to structure a situation for reasoning under certainty. In this paper we have discuss different type of models for risk assessment and among those model why Bayesian network is best to do risk assessment. The paper is organized as follows. In section-2 we present the background and the related work of risk management and risk assessment. In section-3 we introduce the different types of models for risk assessment .In section-4 we finally conclude the Bayesian network is best for the risk assessment with its advantages and disadvantages.

II. BACKGROUND THEORY

To handle the risk in organization they have risk management which is continuous and systematic process and it consists mainly two parts: risk assessment and risk control. As shown in fig.1 [2] .We will focus on risk assessment, in this fig risk assessment has 3 other subparts: risk identification, risk Analysis, risk prioritization.

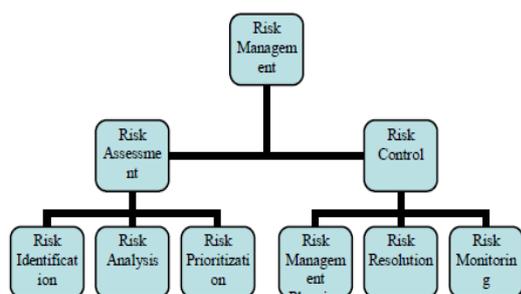


Figure 1. Risk Management [2]

Risk identification produces lists of the projects specific risk items likely to compromise a project success. Where Risk analysis assesses the loss probability and loss magnitude for each identified risk item and it assesses compound risk in the risk item interactions. And Risk prioritization produces a ranked ordering of the risk items identified and analyzed.

The second primary step is risk control and it involves risk management planning, risk resolution, and risk monitoring. Risk management planning helps prepare you to address each risk item including the coordination of individual risk item plans with each other and with the overall project plan. Where Risk resolution produces a situation in which the risk items are eliminated or otherwise resolved (for example risk avoidance via relaxation of requirements).and Risk monitoring involves tracking the project's progress toward resolving its risk items and taking corrective action where appropriate. [2]

III. RISK ASSESSMENT MODELS

To handle the risk assessment there is a various type of models which are Non-AI based (Classical Models), AI-based models and Hybrid Models.

A. Classical Models (Non-AI based Models)

1. Risk assessment based on questionnaire:

In this model project manager and team member prepare a set of questions and then they provide some answers to those questions. Then they compute the level of each risk element to predict the possible outcomes of software project.

2. Risk assessment and estimation based on software metrics:

In this model they defined specific quality attributes and goals based on their importance in developing software projects and their capability to be quantified. A core set of metrics was defined which relate to the software development process and products.

3. Risk assessment models based on classification and Prioritization:

In this model qualitative and quantitative risk assessment of software projects can be done. This model cannot classify the

risk events based on their statistical (in) dependence. Instead, it can do so by using the decision tree to help in risk event classification based on their dependence.

4. Risk assessment based on estimation tools:

This model focuses on technical, cost, and schedule risks. The problem with this model is that it does not consider the requirements and complexity issues. [6]

5. Decision Tree

Decision Tree has a good interpretability and predictive accuracy when used in the domain of risk analysis modeling as it can provide objective facts of risk management based on statistics without having to add prior knowledge. [7]

B. Artificial Intelligence based Models:

1. Artificial Neural Networks based (ANN):

Artificial neural network is machine learning technique that has learning abilities and it is helpful in solving the problems with uncertain and poorly understood conditions. It can be helpful in designing of algorithms. ANN is constructed from a neuron processing element. Processing elements are connected by a network of connections where each connection is weighted and it builds the network knowledge.

2. Bayesian Belief Networks (BBNs):

Bayesian Belief Networks represent probabilistic relationships between variables. BBNs enable reasoning under uncertainty and it is possible to articulate expert beliefs about the dependencies between different variables. They propagate the impact of evidence on the probabilities of uncertain outcomes, such as future system reliability. [6]

Bayesian Network combines the accurate probability distribution and prior knowledge of experts, which are adequate for uncertainty modeling. It is relatively easy to build and the visual modeling can help identify the source of risks.

C. Hybrid Models

This method integrates software metrics with questionnaires. This method focuses on primary facts which are obtained by analyzing, automatically, the source code of software with code metrics.

IV. METHODOLOGY

Bayesian network is a model for the risk assessment it is also known as a Belief networks or Bayesian Belief Networks is one of the most effective theoretical models in the field of uncertain knowledge expressing and reasoning.

Based on graph theory and probability theory, Bayesian Network can describe the relationship between

variables reasoning from incomplete, imprecise or uncertain information.

BNs are casual networks that consist of a set of variables and provide the means to structure a situation for reasoning under certainty. [8]

Bayesian Network consists of two parts:

1. A Directed Acyclic Graph (DAG), also known as Bayesian Network structure, which consists of nodes and directed links between these nodes. Each node is corresponding to a variable, while the directed links represent the correlation or causality between nodes.
2. And a set of Conditional Probability Tables (CPTs), which are a set of local probability distribution (also known as probability parameters) that reflect the relationship between the variables. [7]

Causal networks can be used in order to follow how a change of certainty in one variable may change the certainty for other variables. Formally, the relation between the two nodes is based on Baye’s rule (1).

$$\Pr (B|A) = \frac{\Pr (A|B) \Pr (B)}{\Pr (A)} \quad (1)$$

For example, if any person suffering from shortness of breath (dysponea) and visits the doctor and worried about the lung cancer. The doctor knows two other possible diseases, such as tuberculosis and bronchitis, are possible causes, as well as lung cancer. Doctor also knows that other relevant information or factor of evidence are patient can be smoker or short of air pollution he has been exposed to. A positive X-ray would indicate either TB or lung cancer.

Note that all these nodes have only two values, which keep the model simple, but in general there is no limit to the number of discrete values.

The structure should capture qualitative relationship between variables. In particular, two nodes should be connected directly if one affects or causes the other, with the arc indicating the direction of the effect. So, in this medical diagnosis example, we might ask what factors affect a patient’s chance of having cancer? If the answer is “Pollution and smoking,” then we should add arcs from Pollution and Smoker to Cancer. Similarly, having cancer will affect the patient’s breathing and the chances of having a positive X-ray result. So we add arcs from Cancer to Dyspnoea and X-ray. The resultant structure is shown in Figure 2. [11]

Once the topology of BN is specified, the next step is to quantify the relationship between connected nodes, this is done by conditional probability distribution for each node. As we are considering discrete variables at this stage, this takes the form of a conditional probability table (CPT).

It is important to note that this is just one possible structure for the problem.

An important practical strength of Bayesian networks is that they can be constructed automatically from databases (i.e. learning). Finally, Bayesian networks are able to deal with issues like data over-dispersion (by adding another node representing an additional error term to mean of every observation), relationship between coefficients (representing the coefficients as nodes in the graph), missing data (each missing observation is represented as a node in the graph), measurement errors on observables, or further sources of complexity.

Bayesian networks can also be used as influence diagrams instead of decision trees. Compared to decision trees, Bayesian networks are usually more compact, easier to build, and easier to modify. Unlike decision trees, Bayesian networks may use direct. Each parameter appears only once in a Bayesian network and in case of need, the network may transform into a decision tree, while the reverse is not always possible.

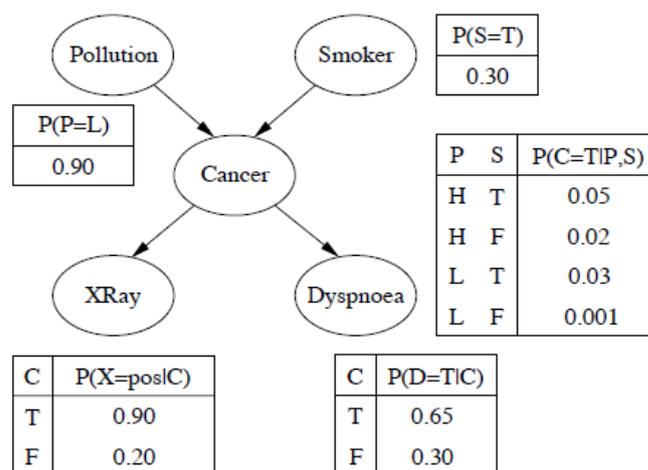


Fig 4.1 A BN For lung Cancer Problem [11]

Nowadays, Bayesian Network has been successfully applied to a wide variety of fields including medical diagnosis, statistical decision making, expert systems, Profit maximization, Outcomes monitoring, Mapping of measures, Economic evaluation, Detecting errors, Decision making and prediction and so on.

V. CONCLUSION

From this survey I have observed that for the huge software project, risk assessment is basic need to launch the reliable, efficient, and low cost product without time delay. And it can be done with using various models but for effective and knowledge based risk assessment Bayesian network is good to use. In this paper, we have also discussed advantages and disadvantages of Bayesian network in real world.

REFERENCES

- [1] Zijian Zhang, Guozheng Rao, Jing Cao, Li Zhang, “*Software Process Risk Measurement Model based on Bayesian Network*”, IEEE 2014.
- [2] Daya Gupta, Mohd. Sadiq, “*Software Risk Assessment and Estimation Model*”, International Conference on Computer Science and Information Technology, IEEE 2008.
- [3] Mohd. Sadiq¹, Mohd. Wazih Ahmad², Md. Khalid Imam Rahmani and SherJung, “*Software Risk Assessment and Evaluation Process (SRAEP) using Model Based Approach*”, International Conference on Networking and Information Technology, IEEE 2010.
- [4] Ying Qu^{1, 2}, Xin-Liang Tang³, “*Software Project Risk Assessing Model Based On Credal Networks*”, Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, IEEE 2010.
- [5] Ahmed Nagy, Mercy Njima and Lusine Mkrtchyan, “*A Bayesian Based Method for Agile Software Development Release Planning and Project Health Monitoring*”, International Conference on Intelligent Networking and Collaborative Systems, IEEE 2010.
- [6] Taher Ahmed Ghaleb, Abdulrhman Ahmed Alsri, Loay Shabaneh, Mahmood Niazi, “*A Survey of Project Risk Assessment and Estimation Models*”, Proceedings of the World Congress on Engineering 2014 Vol I, WCE 2014, July 2 - 4, 2014, London, U.K.
- [7] Yong Hu, Xizhu Mo, Xiangzhou Zhang, Yuran Zeng, Jianfeng Du and Kang Xie, “*Intelligent Analysis Model for Outsourced Software Project Risk Using Constraint-based Bayesian Network*”, JOURNAL OF SOFTWARE, VOL. 7, NO. 2, FEBRUARY 2012
- [8] 1,2Feng XU, 3Guijie QI, 4Yanan Sun, “*The risk analysis of software projects based on Bayesian Network*”, Journal of Convergence Information Technology (JCIT) Volume7, Number5, March 2012.
- [9] Carmel A. Pollino¹ and Barry T. Hart², “*Developing Bayesian network models within a Risk Assessment framework*”, International Environmental Modelling and Software Society (iEMSs), 2008.
- [10]Dimitrios Settas, Stamatia Bibi, Panagiotis Sfetsos, Ioannis Stamelos and Vassilis Gerogiannis, “*Using Bayesian Belief Networks to Model Software Project Management Antipatterns*”
- [11] Introducing Bayesian Networks, Bayesian Artificial Intelligence, Second Edition, modified version of the “Asia” problem Lauritzen and Spiegelhalter, 1988
- [12] <http://www.sei.cmu.edu/cmml/>