USING BAYESIAN BELIEF NETWORK TO MODEL RISK AND IMPACT IN SOFTWARE DEVELOPMENT

Anthony Hui, College of Engineering, California State University, 1250 Bellflower Blvd., Long Beach, CA 90840, 626-823-2016, anthonykhui@yahoo.com
Dar-Biau Liu, Department of Computer Engineering and Computer Science, California State University, 1250 Bellflower Blvd., Long Beach, CA 90840, 562-985-1594, dbliu@verizon.net

ABSTARCT

A successful software development project relies on many factors; it is not an easy task to control all of them going well together. The goal of this paper is to introduce the Bayesian Belief Network (BBN), a mathematical model that can accurately predict risks and quantifies their impacts in a typical development project. Appling this model to a software project, development team can identifies risks in advance, mitigates them by appropriate actions, and gain better control of the resources, budget, and development time frame.

INTRODUCTION

A survey on 8,380 software development projects by the Standish Group in 2003 revealed that 53 % of the projects were behind schedule, or delivered fewer features than originally specified. 52.7% of them overran their budget by 189% or more than their original estimates. Most of the Risk Management tools used today in the software development and technology industry do not scientifically derives risks and their impacts. The BBN model, as described in this paper, can mathematically assess risks and their impacts with pinpoint accuracy. The model first identifies critical risks in a project, then evaluates their probability of occurrence, and finally associates each with an influence rate in order to derive their impact [6]. In 2000, the Arizona State University research concluded 24 most common risks in a software development project and developed a set of causal diagrams for each of them. Our BBN model adapts the structures of all 24 risks as the base [7], associated each with an occurrence probability by a joint study with a group of 29 software development professionals, assigned rate, and then mathematically derives their impacts.

THE RISK AND IMPACT MODEL

The authors of this paper converted the Arizona State University's 24 causal diagrams into 24 risk and impact BBN; in the format that is ready for mathematical analysis. They also conducted a survey in 2004 to estimate the occurrence and transition probabilities of these 24 risk factors. Table 1 shows the result of the survey. Below is an example of one of the 24 risk factors: Staff Experience Shortage, its impacts, weights, and effects expressed in a vector format [7].

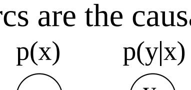
staff_experience_shortage >+ staff_training >++ project_schedule
staff_experience_shortage >+ untrained_staff
keys: risk factor is staff_experience_shortage
    impacts are staff_training, project_schedule and untrained staff
    weight level: >+ influence weight of ONE level effect
        >++ influence weight of TWO level effect

Table 1- Survey result of occurrence probability of the 24 Risk Factors

Problem Group	Risk Factor	Result(0-5)	=Probability
Resource Problem	Staff Experience Shortage	2	0.3
Resource Problem	Reliance on a few Key Person	4	0.75
Resource Problem	Schedule Pressure	3.8	0.7
Human Problem	Low Productivity	1.6	2.2
Human Problem	Lack of Staff Commitment	1.5	2.0
Client Problem	Lack of Client Support	2.25	0.35
Client Problem	Lack of Contact Person Competence	1.25	0.15
Research Data Problem	Lack of Quantity Historical Data	3	0.5
Research Data Problem	Inaccurate Cost Estimating	3	0.5
System Problem	Large and Complex External Interface	2.5	0.4
System Problem	Large and Complex Project	2.75	0.45
System Problem	Unnecessary Features	2	0.3
System Problem	Creeping User Requirement	4	0.75
System Problem	Unreliable Subproject	2.75	0.45
Management Problem	Incapable Project Management	3.3	0.58
Management Problem	Lack of Senior Management Commitment	3	0.5
Management Problem	Lack of Organization maturity	1.75	0.25
Technology Problem	Immature Technology	2.8	0.46
Technology Problem	Inadequate Configuration Control	2.75	0.45
Technology Problem	Excessive Paperwork	2	3
Technology Problem	Inaccurate Metrics	3	0.5
Technology Problem	Excessive Reliance on a Single Process	3	0.5
Experience Problem	Lack of Project Environment Experience	3.5	0.625
Experience Problem	Lack of Project Software Experience	2.6	0.42

THE BAYESIAN BELIEF NETWORK (BNN)

Bayesian Belief Network (BBN) is a type of diagrams that work together with an associate set of probability tables. The graph is made up of nodes and arcs, where the nodes represent uncertain variables and the arcs are the causal/relevance relationships between the variables [2][3][11].

Bayesian network: 

if x causally affect y.

In general, we write $P(A|B)$ to represent a belief in A under the assumption that B is known. The traditional approach to define conditional probabilities is via joint probabilities. We have:

$P(A|B) = P(A, B) / P(B)$ or (1)

$P(A|B) = P(B|A) P(A) / P(B)$ (2)

We can rearrange the formula for conditional probability to get the so-called product rule:

$P(A,B) = P(A|B) P(B)$ (3)

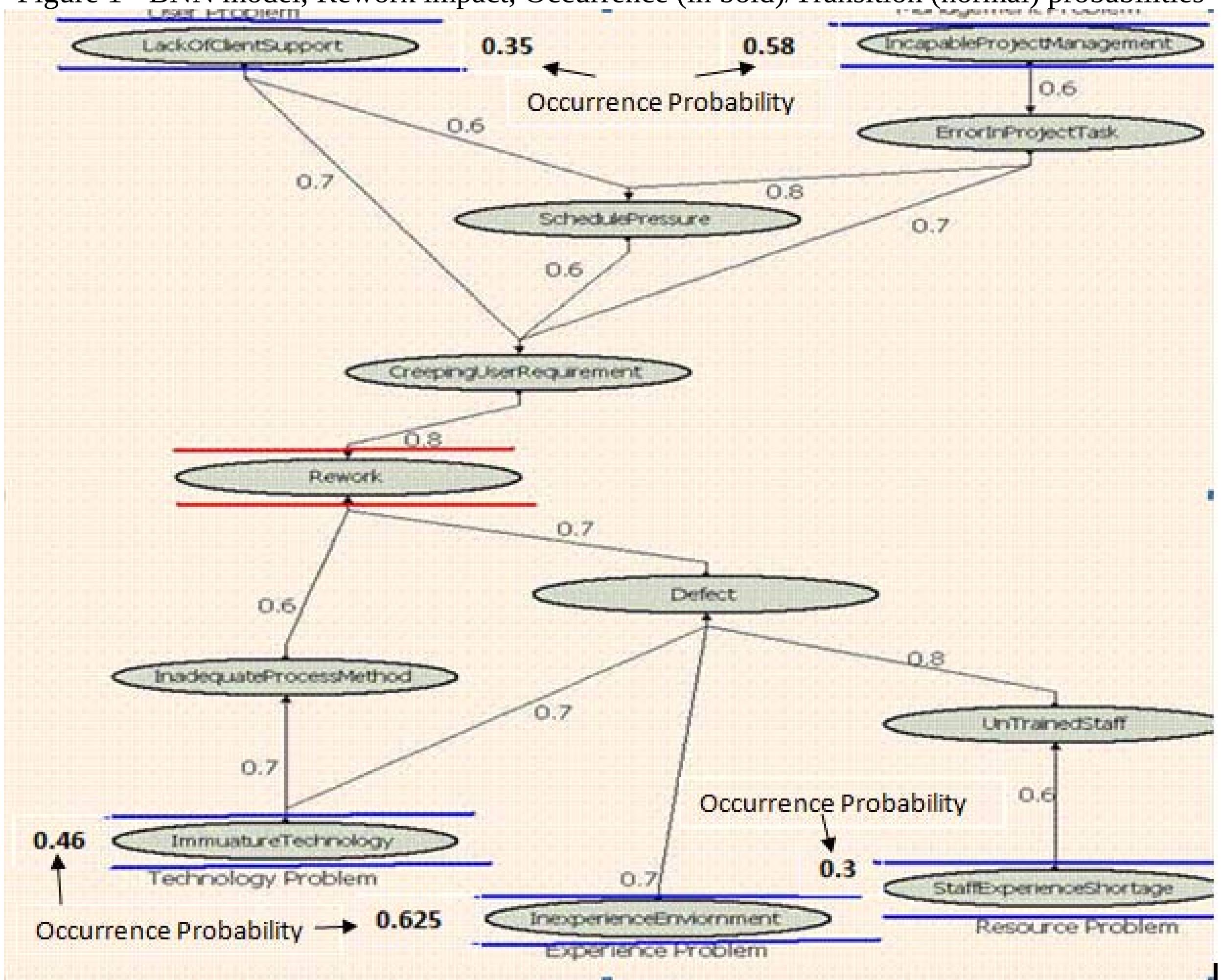
In general to n variables, we refer to this as the “Chain Rule”:

$P(A1, A2, \ldots, An) = P(A1|A2, \ldots , An) P(A2|A3, \ldots , An) P(An\text{-}1|An)P(An)$ (4)

USING BBN TO EVALUATE THE “REWORK” IMPACT

“Rework” is an impact caused by various risk factors in a software development project; not only it will delay the delivery time; it could also require additional resources and overrun development budget. Below illustrates the BNN model graphically and mathematically derives the impact value from 5 of the 24 risk factors defined in the model - Lack of Client Support, Incapable Project Management, Immature Technology, Inexperience Environment, and Staff Experience Shortage.

Figure 1 - BNN model, Rework impact, Occurrence (in bold)/Transition (normal) probabilities



In order to calculate the impact of rework from the 5 risk factors, the model uses their occurrence and transition probability values. The mathematical formulas used by the model are listed below:

Impact:
$$\begin{aligned}F0(x) = {} & F1(x)*F2(x)*F3(x)*0.5+F1(x)*F2(x)*(1-F3(x))*0.5+\\ & F1(x)*(1-F2(x))*F3(x)*0.5+F1(x)*(1-F2(x))*(1-F3(x))*0.8+\\ & (1-F1(x))*F2(x)*F3(x)*0.5+(1-F1(x))*F2(x)*(1-F3(x))*0.7+\\ & (1-F1(x))*(1-F2(x))*F3(x)*0.6+(1-F1(x))*(1-F2(x))*(1-F3(x))*0.5 \qquad (5)\end{aligned}$$

Subject to the Equality Constraint Functions:

$$\begin{aligned}F1(x) = {} & F4(x)*F5(x)*F6(x)*0.5+F4(x)*F5(x)*(1-F6(x))*0.5+\\ & F4(x)*(1-F5(x))*F6(x)*0.5+F4(x)*(1-F5(x))*(1-F6(x))*0.7+\\ & (1-F4(x))* F5(x)*F6(x)*0.5+(1-F4(x))* F5(x)*(1-F6(x))*0.6 +\\ & (1-F4(x))*(1-F5(x))*F6(x)*0.7+(1-F4(x))*(1-F5(x))*(1-F6(x))*0.5 \qquad (6)\end{aligned}$$

$$\begin{aligned}F2(x) = {} & F7(x)*F8(x)*F9(x)*0.5+F7(x)* F8(x)*(1-F9(x))*0.5+\\ & F7(x)*(1-F8(x))*F9(x)*0.5+F7(x)*(1-F8(x))*(1-F9(x))*0.8+\\ & (1-F7(x))*F8(x)*F9(x)*0.5+(1-F7(x))*F8(x)*(1-F9(x))*0.7+\\ & (1-F7(x))*(1-F8(x))*F9(x)*0.7+(1-F7(x))*(1-F8(x))*(1-F9(x))*0.5 \qquad (7)\end{aligned}$$

$$F3(x) = F9(x)*0.7+(1-F9(x))*0.5 \qquad (8)$$

$$F7(x) = F10(x)*0.6+(1-F10(x))*0.5 \qquad (9)$$

$$F5(x) = F4(x)*F6(x)*0.5+F4(x)*(1-F6(x))*0.6+ (1-F4(x))*F6(x)*0.8+(1-F4(x))*(1-F6(x))*0.5 \quad (10)$$

$$F6(x) = F11(x)*0.6+(1-F11(x))*0.5 \quad (11)$$

Source Risks:

$$F4(x) = x1 \qquad 0 \le x1 \le 1 \quad (12)$$

$$F11(x) = x2 \qquad 0 \le x2 \le 1 \quad (13)$$

$$F9(x) = x3 \qquad 0 \le x3 \le 1 \quad (14)$$

$$F8(x) = x4 \qquad 0 \le x4 \le 1 \quad (15)$$

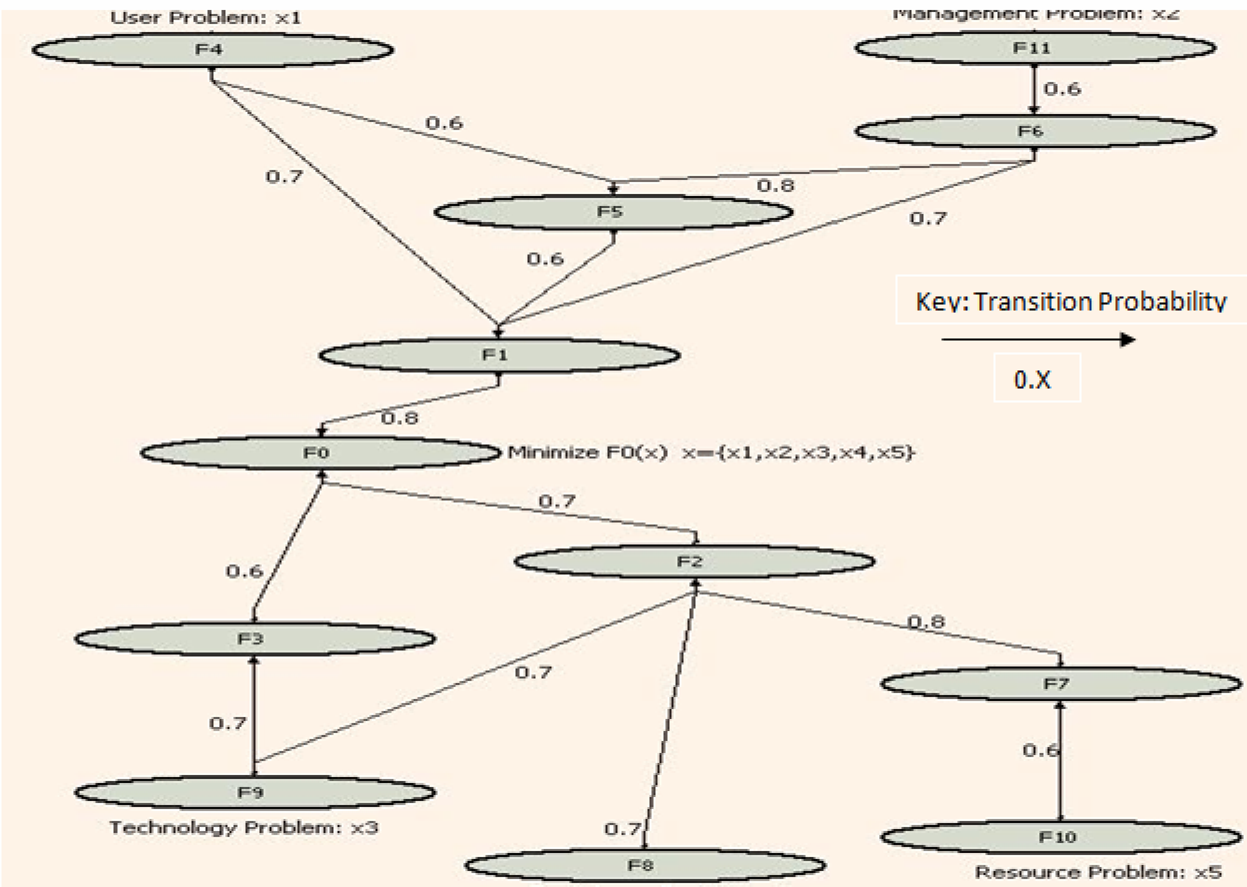
$$F10(x) = x5 \qquad 0 \le x5 \le 1 \quad (16)$$

As shown in Figure 1 above, the occurrence probabilities of the 5 risk factors in our analysis are: $Xn = [X_1=0.35, X_2=0.58, X_3=0.46, X_4=0.625, X_5=0.3]$. They are based on a 2006 in-depth study and survey on typical projects in the software development industry. By inputting these values, the BNN model calculates the Rework impact and generates a minimum value of 0.5608.

OPTIMIZING AND LEVELING THE BNN RESULTS

However, the value of 0.5608 calculated by the BNN model may not be the most minimal value. We are also interested in, if there are, what the occurrence probability values of the 5 risk factors would be, for the model to produce the most minimal result.

Figure 2 – BBN model, with Transition probabilities for Non Linear Programming optimization



As shown in Figure 2 above, applying non-linear programming theory to the model, we want to:

1. Optimize the vector x = (x1, x2, x3, x4, x5) and its variables.
2. Achieve the objective of the most minimal F0(x) value.

There are several ways to solve this non-linear problem, algorithmic method is a mathematical approach that can find the most optimal solution with a high degree of accuracy [4][5][8][10][12]. Using medium-scale minimization method together with numerical gradients calculated by finite-difference approximation, the most minimal impact value for the 5 risk factors could be at 0.5488, when their occurrence probability are x = [0.2715, 0.5715, 0.99, 0.01, 0.2406].

By repeatedly applying different combinations of occurrence probability values of the risk factors, the BNN model shows that low probability values do not always generated low impact values, and vice versa. This phenomenon can be explained that the occurrence probability value is offset by its complement component value (1 - occurrence probability) defined in the model. Therefore, it is very hard to predict the result level of impact just by looking at the occurrence probabilities.

The BNN model also shows that transition probabilities of risk factors interact differently among each others. They affect more on the result's range while occurrence probabilities affect on the minimum value. In Figure 2, the User, Technology, and Experience problem have higher transition probability value (0.7) than the Management and Resources problem (0.6). If we input just the 2 with lower transition probability, the model will calculate an impact with a minimum value of 0.5489 and a maximum value of 0.5692, a deviation of 0.0203. But, if we input just the other 3, the minimum value will be 0.5488 and the maximum value will be 0.5699, an increased deviation of 0.0211. We conclude that lower transition probability values always yield a smaller deviation.

CONCLUSION

To analyze and mitigate risks in a development project, we can minimize their impacts by first prioritizing them based on their most optimal combination of occurrence probabilities as generated by the model, and then applying resources to control them. Further, we can ensure that impacts are predictably more manageable by controlling risks with higher transition probabilities [1][9].

It is the author's believe that if we can identify and control the risk factors at early stages, we will be able to significantly increase the chance of success in a development project. The BNN model, when used properly, will provide details in mathematical values on what, when, and how things may go wrong at initial phases of a project. This methodology could be an integral part of the project management tools for the software development industry.

REFERENCES

[1] Bertsekas, Dimitri P., *Nonlinear Programming,* Athena Scientific, 1999.

[2] Fenton, Norman., *Why do we need a BBN for the probability computations*, Queen Mary and Westfield College, July 2001.

[3] Gilchrist, Warren, *Statistical Modeling*, John Wiley & Sons, 1984.

[4] Gill, Philip., Murray, Walter., Wright, M., *Practical Optimization,* Academic Press, 1981.

[5] Gottfried, Byran S., Weisman, Joel., *Introduction to Optimization,* Prentice-Hill Inc., 1973.

[6] Higuera, Ronald P., Haimes, Yacov Y., *Software Risk Managemen,* CMU/SEI, June 1996.

[7] Houston, Dan., *Survey on Potential Effects of Major Development Risk Factors*. Arizona State University Research Project, February 2000.

[8] Kirkpatrick, S., Gelatt, C.D., *Simulated Annealing*. Wikimedia Foundation, Inc., June 2007.

[9] Lootsma F. A., *Numerical Methods for Non-linear Optimization,* Academic Press, 1972.

[10] Moore, Andrew W., *Iterative Improvement Search – Hill Climbing, Simulated Annealing, WALKSAT and Genetic Algorithms*, Carnegie Mellon University, 2007.

[11] Neil, Martin., *Introduction to Bayesian Networks*, Technical Report of Agnea Limited, 2006.

[12] Zhu, Xiaojin., *Heuristic Search,* University of Wisconsin, Madison, 2006.