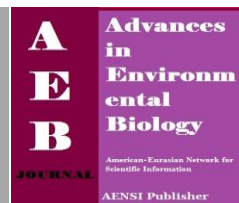




AENSI Journals

Advances in Environmental Biology

ISSN-1995-0756 EISSN-1998-1066

Journal home page: <http://www.aensiweb.com/aeb.html>

A Bayesian Networks Approach for Root Cause Analysis of Manufacturing Systems

Ruhollah Noorian

Instructor of Ilam University, Ilam, Iran.

ARTICLE INFO

Article history:

Received 25 March 2014

Received in revised form 20 April 2014

Accepted 15 May 2014

Available online 5 June 2014

Keywords:

Manufacturing systems; Root cause analysis (RCA); Bayesian networks (BNs)

ABSTRACT

In today's manufacturing systems, due to high equipment cost and capacity limitations, different products with their own recipes are being processed on various types of tools. This brings an extremely complex operation condition to the tools and makes the equipment degradation highly unpredictable. Although sophisticated control programs are designed to ensure product specification, fault still occurs during the operation and subsequently it causes tremendous loss to manufacturing systems. Therefore, occurred failure in production process should be detected as soon as possible, and root cause need to be identified so that corrections can be made in time to avoid further loss. In this paper, a cause and effect model is developed for root cause analysis of manufacturing systems, using Bayesian networks (BNs) -a state-of-the-art technique in modeling uncertainty. An actual test problem was selected to show the applicability of the proposed approach. The results of the Bayesian model showed the weak and strong points of under study system concerning the potential causes. Furthermore, it identifies which causes have the major impacts on the overall performance of manufacturing system. The modeling approach of this paper could be easily utilized for root cause analysis of other systems. The results of such studies would help top managers better understand and improve existing systems.

© 2014 AENSI Publisher All rights reserved.

To Cite This Article: Ruhollah Noorian, A Bayesian networks approach for root cause analysis of manufacturing systems, *Adv. Environ. Biol.*, 8(9), 109-116, 2014

INTRODUCTION

In today's manufacturing industry, due to high equipment cost and capacity limitations, different products with their own recipes are being processed on various types of tools [19]. This brings an extremely complex operation condition to the tools and makes the equipment degradation highly unpredictable. Although sophisticated control programs are designed to ensure product specification, fault still occurs during the operation and subsequently it causes tremendous loss to manufacturer because in most cases rework is infeasible for the mis-processed products. Failure that happens at any time during the process should be detected promptly so that engineer can stop the operation to avoid further loss, and make corrections based on the diagnosis before process can be resumed.

Root cause analysis (RCA) is targeting at identifying the causes of problems in processes for directing counteractive actions [21]. Control charts typically do not have this feature; however, non-random patterns on the chart can be used as a source for RCA [8,17,22]. However, large number of possible relations among patterns and causes makes the process of cause/s identification difficult. Certain information from the process (at the time of change) can be used in accompany with chart patterns to simplify this process. As a simple example, if we know from pattern analysis that either machine condition or the quality of input-material has caused an out of control situation, when the process data shows that the operating machine has not been serviced for a while but the material has recently been tested showing no problem, there is a high chance that the bad condition of the operating machine has caused the problem.

As the number of process sensors grows, a huge amount of data are collected from the production; and yet, the relations among these control variables and their effects on finished products are to be fully understood for both equipment monitoring and quality assurance. Meanwhile, as the product goes through multiple periods with different recipes, failure that occurs during the process can both cause tremendous loss to manufacturer and compromise product quality. Therefore, occurred failure should be detected as soon as possible, and root cause need to be identified so that corrections can be made in time to avoid further loss. In this paper, using an advanced risk modeling approach called Bayesian Networks (BNs) [1], a probabilistic model is proposed to

investigate the causal relationship among process variables and evaluate their influence on product quality. The proposed model can combine subjective information and expert's judgment. BNs are considered as a powerful approach to risk analysis and uncertainty management. This approach provides a decision support system in a wide spectrum of issues including uncertainty and probabilistic reasoning. A Bayesian Network is a means for clear and concise expression of combined probability distribution among variables and interrelated assumptions. In this approach, both main sources of information: "expert opinion" and "previous records" can be used.

The remaining of this paper is organized as follows. Section 2 reviews different techniques of RCA in the literature. In section 3, an overview of Bayesian networks is presented. Section 4 explains the structure of the proposed BN and discusses the model's capabilities of advanced analysis in manufacturing environment. A numerical example is provided in this section. The paper ends with the conclusion and possible extensions to the approach, in section 5.

Root cause analysis (RCA) techniques:

Montgomery [17], Doty [8] and Smith [22] divide the assignable causes into six categories of Man, Machine, Method, Materials, Measure, and Environment. They suggest employing control chart patterns to determine one or more of above categories caused the out-of-control situation. Demirli and Vijayakumar [5] develop a fuzzy rule based system for X hart, based on control patterns-cause relationship network, to resolve the uncertainties in identifying the (real) chart patterns and relating them to assignable causes. They also discuss that categorizing the out of control states into isolated shifts, sustained shifts and gradual shifts can hasten the RCA process.

Dassau and Lewin [4] use optimization as a mean of automating RCA. They formulate the problem as a mixed-integer nonlinear program, whose system variables include the possible perturbations that affect low quality and low yield, with possible process improvements as decision variables. Motschman and Moore [16] discuss the process of RCA as well as corrective action in transfusion medicine. Dhafr *et al.* [7] develop a methodology for identifying various sources of quality defects on the product. Leonhardt and Ayoubi [15] present a summary of methods that can be applied to automatic fault diagnosis with a focus on classification and fuzzy based techniques. Mo *et al.* [14] suggest a methodology based on clustered symptom tree, which utilizes the advantage of the signed digraphs to represent the causal relationship between process variables and/or the propagation paths of faults in a simple and graphical way. Ge *et al.* [10] use Hidden Markov Models for metal stamping process monitoring and fault diagnosis. They use a number of autoregressive models to model the monitoring signal in different time periods of a stamping operation and uses the residues as the features. Then, they employ a Hidden Markov Model (HMM) for classification. Widodo and Yang [23] present a survey of machine condition monitoring and fault diagnosis using support vector machine. Lunze and Schiller [13] provide an example of fault diagnosis by means of probabilistic logic reasoning. Dey and Stori [6] use data from multiple sensors on sequential machining operations through a causal belief network framework to provide a probabilistic diagnosis of the root cause of the process variation. Chang and Ho [2] apply neural network monitoring techniques to process control in an integrated monitoring/diagnosis scheme. The proposed technique contains a modified cause/effect diagram including process and part information that speeds up the diagnosis process.

Bayesian networks:

BNs (also known as Belief Networks, Causal Probabilistic Networks, Graphical Probabilistic Networks and Probabilistic Cause-effect Networks) are powerful tools for knowledge representation and reasoning under uncertainty [11]. A BN consists of a set of nodes (representing variables) and a set of directed edges (representing causal influences between variables) between variables. Each variable has a finite set of mutually exclusive states. The variables together with the edges form a directed acyclic graph (DAG) (a directed graph is acyclic if there is no directed path $A_1 \rightarrow \dots \rightarrow A_n$ such that $A_1 = A_n$).

To each variable 'A' with parents B_1, \dots, B_n , a conditional probability table $P(A|B_1, \dots, B_n)$ is assigned. If the variable has no parents then the table reduces to the unconditional probabilities $P(A)$ (i.e. prior probability).

One important property of BNs is their ability to represent the joint probability distribution $P(A_1, \dots, A_n)$ for all the variables A_1, \dots, A_n in a compact form. This is done by use of the 'chain rule', which says in a BN the full joint probability distribution is the product of all conditional probabilities specified in the BN [1]:

$$P(A_1, \dots, A_n) = \prod_i P(A_i | A_1, \dots, A_n) \quad (1)$$

The more compact representation of the joint probability makes the probability calculation easier. If we have access to the joint probability distribution, then we can calculate the marginal probability for any variable,

$P(A_i)$, and also the conditional probability of $P(A_j|A_i = a_i)$. BNs address the problems of storing and representing the joint probability distribution of a large number of random variables and doing Bayesian inference with these variables. The predictive capability of a BN enables us to infer from cause to effect (from parent to child). The diagnostic capability of the model enables us to infer from effect to cause (from child to parent). During development of a BN, variables (nodes) can be easily added or modified. The graphical nature of BNs allows variables to be conveniently added or removed without significantly affecting the remainder of the network.

The advantages of BNs are now widely recognized and they are being successfully applied in diverse fields [9]. During the last decade, researchers have incorporated BN techniques into easy-to-use toolsets, which in turn have enabled the development of decision support systems in a diverse set of application domains.

As the number of network nodes and edges increase, the volume of calculation of combined distribution of the nodes increases exponentially. Developing an algorithm for calculating conditional probabilities and inference in a BN is an NP-hard problem. However, several accurate (for specific networks) and approximate algorithms have been presented and several software tools for developing BN models are available. These tools provide a graphical editor for building the BN and a runtime module, which takes care of probabilistic calculation and evidence transmission. With such tools, it is possible to build a BN and perform the propagation algorithm in a reasonable amount of time. In this paper, AgenaRisk software has been used.

Proposed approach:

This section explains a new methodology using BNs to root cause analysis in manufacturing systems. First the framework of the proposed method is described. Then real world data are used to validate and verify developed model.

Research framework:

First step is to identify all the potential causes and variables might affect performance of the manufacturing system or production process (e.g. Man, Machine, Material, etc.). This step is essentially a qualitative analysis that can be performed using published literature, brainstorming sessions and conducting interviews with an expert group.

Next step is to assign a probability variable to each cause and estimate them. In this study, a “ranked” variable with three-point Likert scale (Low, Medium, High) is assigned to each identified cause. Then, the prior probability of cause items should be generated. In this study, the new systematic approach in determining the probabilities of Bayesian network, proposed by Chin *et al.* [3] is used to generate prior probabilities of common causes.

After the probabilities are estimated, a ranked node is created in the BN for each cause item. In order to integrate all the information about identified causes, a result node "Overall Risk" is created in the BN model, as portrayed by Fig. 2. After the model is configured, it can show the most important cause, which has greater effect on the manufacturing system. Furthermore, it is possible to see how the overall risk of the system can change by applying different scenario analysis.

Fig. 1 presents the steps required to accomplish the framework of this study.

Generation of prior probabilities:

Suppose there are n states S_1, S_2, \dots, S_n of a node N which has no parent, and the probability of each state S_i , i.e., $P(S_i)$ need to be specified.

Traditionally, $P(S_i)$ is specified directly by experts, using their knowledge and experiences. When the number of states is small, such a method may be feasible. With the increase in states of a variable, estimating probabilities directly to all states at one time may inevitably involve biases and inaccuracies.

An alternative way is to perform pair-wise comparisons between states for generating their probabilities. Since there are only two instead of n states considered at one time in a pair-wise comparison, it should be much easier to provide judgments by pairwise comparisons than the direct estimation of probabilities. In the new approach, the prior probability of each state of a node can be determined by the following pair-wise comparison matrix:

	S_1	S_2	\dots	S_n	ω
S_1	a_{11}	a_{12}	\dots	a_{1n}	ω_1
S_2	a_{21}	a_{22}	\dots	a_{2n}	ω_2
\dots	\dots	\dots	\dots	\dots	\dots
S_n	a_{n1}	a_{n2}	\dots	a_{nn}	ω_n
$\lambda_{\max} =$		CI=		CR=	

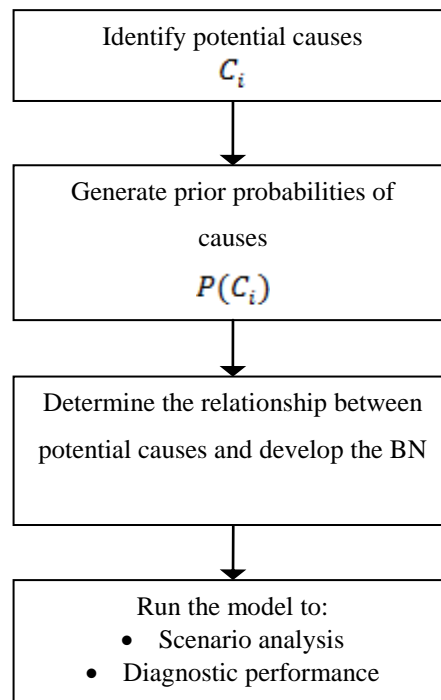


Fig. 1: The proposed approach for assessment of root causes of a manufacturing system.

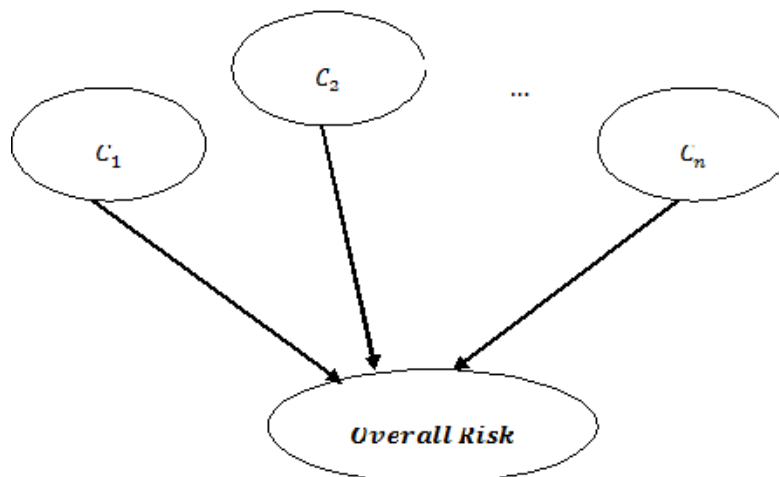


Fig. 2: The proposed BN model.

In the above matrix, a_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, n$) can be specified by questions like "comparing the state S_i with S_j , which one is more likely to occur and how much more likely?" and the value of a_{ij} represents the multiple of the likelihood of the presence of S_i over that of S_j . Note that from the meaning of a_{ij} , we can find that $a_{ji} = 1/a_{ij}$ and $a_{ii}=1$, so there are $n(n-1)$ different comparisons in the above pair-wise comparison matrix. However, it is sufficient to provide $(n-1)$ inter-related comparisons rather than all the $n(n-1)$ different comparisons, although it is useful to have more comparisons for checking consistency.

Similar to Saaty's AHP, the relative priorities of S_i can be generated from the maximum eigenvector $\omega = (\omega_1, \dots, \omega_n)^T$ of the matrix S_i and the consistency of the pair-comparison matrix can be checked by the consistency ratio $CR = CI/RI$, where CI is the consistency index, which is defined by $(\lambda_{\max} - n)/(n-1)$

(λ_{\max} is the maximum eigenvalue corresponding to ω), and RI is a random index related to n as shown in Table 2. A pair-wise comparison matrix with CR less than 0.10 is considered acceptable.

Table 1: Random consistency index.

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

Since the sum of all the elements in x is 1 and its i th element ω_i represents the relative importance of the state S_i among all the states, it is natural to interpret ω_i as the prior probability of state S_i . In other words, we have

$$P(S_i) = \omega_i \quad (2)$$

Test problem:

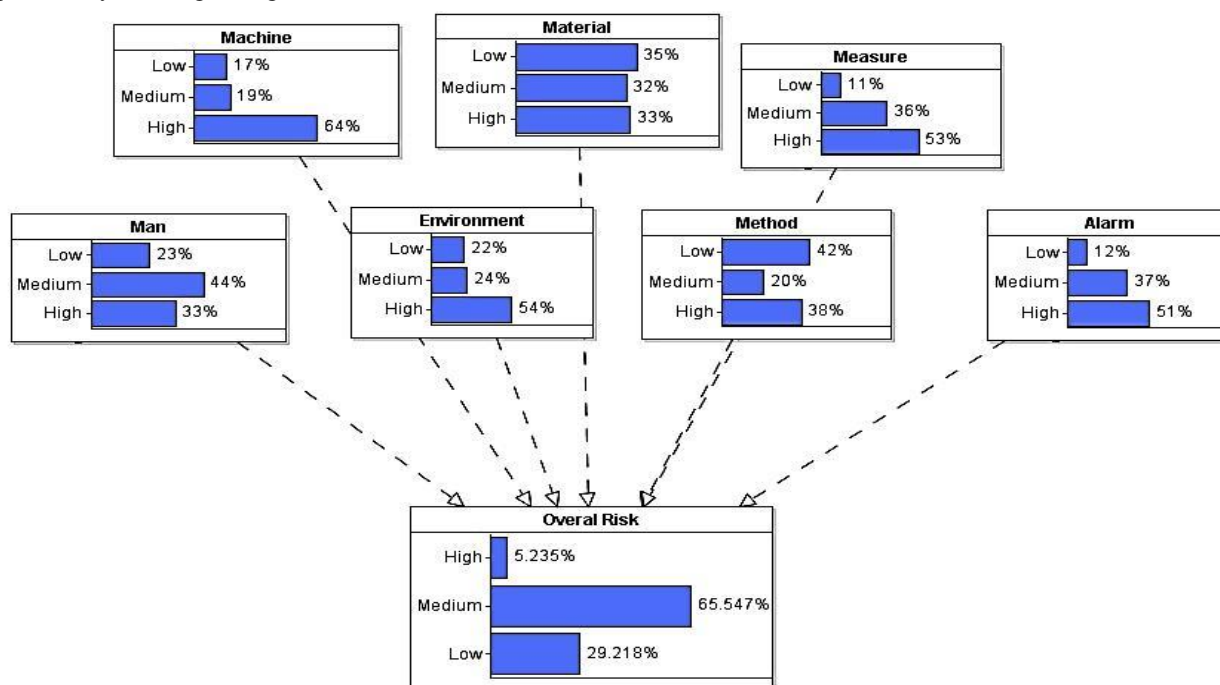
The test problem considers a manufacture of electronic valves and tubes and other electronic components, which refers to telecommunication sector. To achieve the objectives of this study, a study was conducted to locate all common causes, which influence performance of this kind of industry. Using 17 experts, the potential causes were identified and prior probabilities were generated using methods described in section 4.2.

The root causes considered here are: Man, Machine, Material, Method, Measure and Environment, which are the six most commonly considered root causes plus false alarm. False alarm is included because not every out-of-control signal from the control chart is true. Table 1 shows the information gathered related to potential causes.

Table 1: Potential causes and related probabilities.

Causes	Probabilities (Low, Medium, High)
Man	(0.23, 0.44, 0.33)
Machine	(0.17, 0.19, 0.64)
Material	(0.35, 0.32, 0.33)
Method	(0.42, 0.20, 0.38)
Measure	(0.11, 0.36, 0.53)
Environment	(0.22, 0.24, 0.54)
Alarm	(0.12, 0.37, 0.51)

Fig. 3 illustrates the result of running model based on aforementioned variables and probabilities. As shown in the model, the probability distribution of “Overall Risk” is skewed toward “Low”. It can be seen that the probability of being at High level of risk is 0.0523%.

**Fig. 3:** The BN model for root cause analysis of example manufacturing system.

Analysis:

In addition to the Baseline Scenario, which is illustrated in Fig 4, the power of proposed model is better understood by showing the results of running it under various scenarios. It is possible to enter observations

anywhere in the model to perform not just predictions but also many types of trade-off and explanatory analysis. This section explores the capability of the model in performing different ‘what if?’ type analysis. This is one of main advantages of BN that updates the probability of desired variables based on observation made on other variables (evidences). Here we can analyze the Overall Risk from different aspects. For example, Fig. 4 shows how the distribution of the Overall Risk changes when the performance of “Machine” and “Environment” goes to “Low” (Scenario 1).

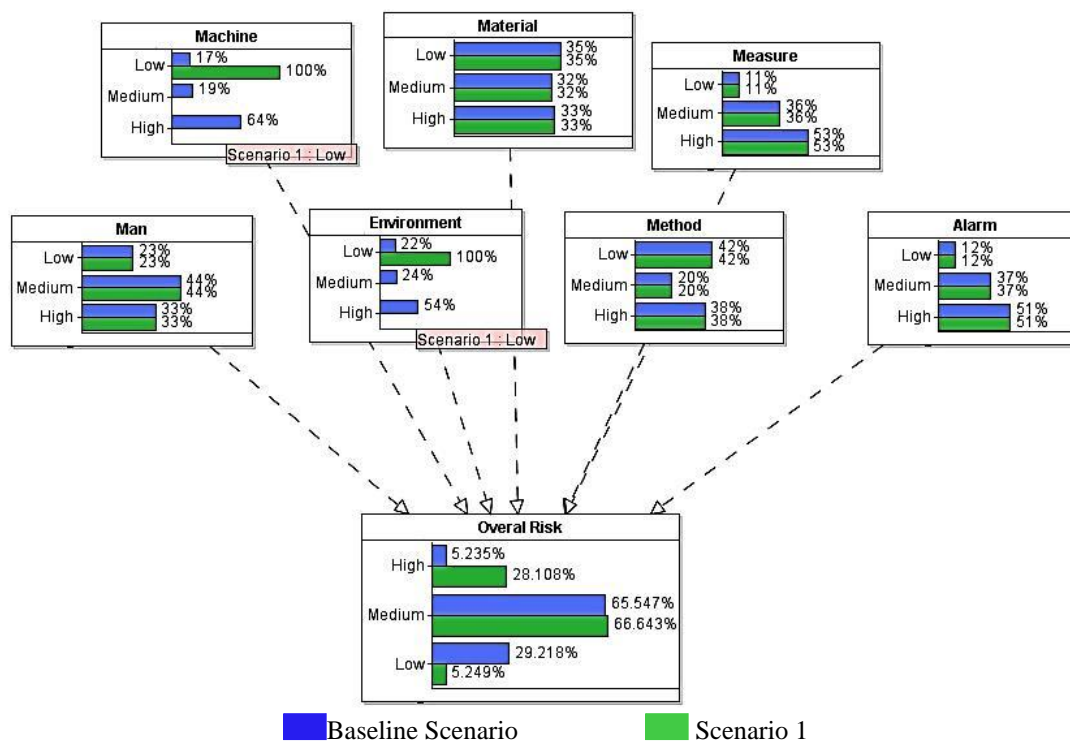


Fig. 4: Change in Overall Risk when level of common causes changes.

Another possible analysis in this model is the trade-off analysis between the Overall Risk and root causes when there is a need to have a perfect performance (i.e. without risk) and it is interesting to know about the required level of different causes. For example, Fig. 5 shows the probability distribution of required quality of “Material” to meet this scenario. Note that how the distribution of Material is skewed toward High (Scenario 2).

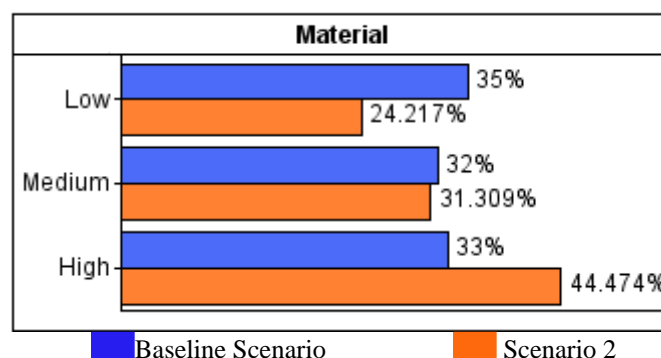


Fig. 5: Scenario 2: P(Material|Overall Risk = Low).

Furthermore, the model can rank the causes to show the most important causes using sensitivity analysis (Fig. 6). The result of sensitivity analysis shows that in order to improve the performance of the manufacturing system, management should concentrate on Machine, Environment, ..., respectively.

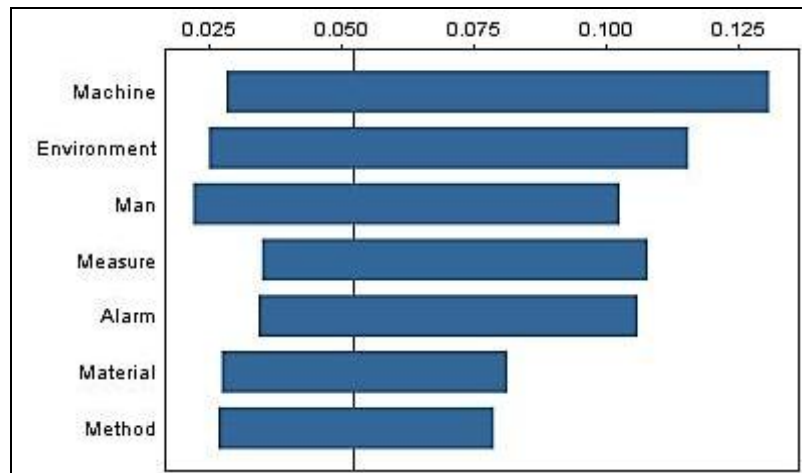


Fig. 6: The most important causes in order to improve the performance of the system.

Conclusion:

Over the years, methodology for fault detection, diagnosis and prognosis has evolved from univariate statistical process control (SPC) to multivariate system analysis. Traditional quality control methods such as Shewhart, CUSUM chart [18] have encountered great challenges as the amount of data and number of variables increase. Since control charts are normally created for critical quality parameters of the process, when out of control alarm comes one need to identify its cause such that correction can be made. However, root cause identification is not trivial without engineer expertise as there are many factors that can affect the process. Furthermore, different control charts are needed for different problems, and to monitor them simultaneously would be impractical when the number of charts is large. More advanced techniques are required to capture different aspects of uncertainty and common causal factors in manufacturing systems.

This paper has proposed a new approach that makes it possible to incorporate uncertainty and causality in root cause analysis of production processes. The approach brings the full weight and power of BN analysis to bear on the problem of root cause analysis. This makes it possible to:

- Express uncertainty about each potential cause with full probability distributions,
- Model the dependency and signal effects among various causes using the causal relationships,
- Find the main causes of low performance,
- Perform complex “what-if?” analysis, and
- Learn from data so that the predictions become more relevant and accurate.

The application of the approach was explained by use of a simple example. The aim of this paper is to demonstrate how advanced artificial intelligence tools such as BN can be employed to capture complex issues such as uncertainty in production systems. The model proposed here is not the ultimate model. In fact, one of the advantages of BN modeling is that the network can be easily (although it needs some knowledge of probability) tailored and modified to fit to the problem. Also when is required the model can be expanded by adding more variables to quantify decision-making scenarios such as management reactions to high risk. However, it has a main shortcoming. In big size systems with several hundred variables and many common causes, the BN becomes too complex in term of efficiency of inference process as well as constructing the model. For our future work, the so-called Object Oriented Bayesian Network (OOBN) approach would be sought. By using OOBNs, complex models can be constructed easily using inter-related objects. This not only facilitates the process of building the model but also can speed up the inference process [20].

REFERENCES

1. Ben-Gal, I., 2007. “Bayesian Networks”, in Ruggeri, F., Faltin, F. and Kenett, R., Encyclopedia of Statistics in Quality and Reliability. John Wiley & Sons.
2. Chang, S.I., E.S. Ho, 1999. A two-stage neural network approach for process variance change detection and classification. International Journal of Production Research, 37(7): 1581-1599.
3. Chin, K.S., D.W. Tang, J.B. Yang, Sh. Y. Wang and H. Wang, 2009. Assessing new product development project risk by Bayesian network. Expert Systems with Applications, 36: 9879-9890.
4. Dassau, E., D. Lewin, 2006. Optimization-based root cause analysis. In 16th European symposium on computer aided process engineering, pp: 943-948.

5. Demirli, K., S. Vijayakumar, 2008. Fuzzy assignable cause diagnosis of control chart patterns. In Annual meeting of the North American fuzzy information processing society (NAFIPS), pp: 1-6.
6. Dey, S., J.A. Stori, 2005. A Bayesian network approach to root cause diagnosis of process variations. *International Journal of Machine Tools & Manufacture*, 45(1): 75-91.
7. Dhafr, N., M. Ahmad, B. Burgess, S. Canagassababady, 2006. Improvement of quality performance in manufacturing organizations by minimization of production defects. *Robotics and Computer-Integrated Manufacturing*, 22(5-6): 536-542.
8. Doty, L.A., 1996. *Statistical process control* (2nd ed.). NY, USA: Industrial Press.
9. Fenton, N. and M. Neil, 2007. "Managing risk in the modern world: Applications of Bayesian networks". Knowledge transfer report, London Mathematical Society.
10. Ge, M., R. Du, G. Zhang, Y. Xu, 2004. Fault diagnosis using support vector machine with an application in sheet metal stamping operations. *Mechanical Systems and Signal Processing*, 18: 143-159.
11. Hecherman, D., A. Mamdani and M. Wellman, 1995. Real-world application of Bayesian networks. *Comm ACM*, 3: 25-26.
12. Pitchforth, J., K. Mengersen, 2013. A proposed validation framework for expert elicited Bayesian Networks with Applications, 40: 162-167.
13. Lunze, J., F. Schiller, 1999. An example of fault diagnosis by means of probabilistic logic reasoning. *Control Engineering Practice*, 7: 271-278.
14. Mo, K.J., G. Lee, D.S. Nam, Y.H. Yoon, E.S. Yoon, 1997. Robust fault diagnosis based on clustered system trees. *Control Engineering Practice*, 5(2): 199-208.
15. Leonhardt, S., M. Ayoubi, 1997. Methods of fault diagnosis. *Control Engineering Practice*, 5(5): 683-692.
16. Motschman, T.L., S.B. Moore, 1999. Corrective and preventive action. *Transfusion Science*, 21: 163-178.
17. Montgomery, D.C., 2005. *Introduction to statistical quality control* (5th ed.). New York, NY: Wiley.
18. Montgomery, D., 1997. *Introduction to statistical process control*. John Wiley & Sons, Inc.
19. Pasadyn, A., T. Edgar, 2005. Observability and state estimation for multiple product control in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, 18(4): 592-604.
20. Pitchforth, J., K. Mengersen, 2013. A proposed validation framework for expert elicited Bayesian Networks, *Expert Systems with Applications*, 40: 162-167.
21. Rooney, J.J., L.N.V. Heuvel, 2004. Root cause analysis for beginners. *Quality Progress*, 45: 53.
22. Smith, G.M., 2004. *Statistical process control and quality improvement* (5th ed.). NJ, USA: Pearson/Prentice Hall.
23. Widodo, A., B.S. Yang, 2007. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 21(6): 2560-2574.
24. www.agenarisk.com, 2014.