IIE Transactions (2010) **42**, 564–576

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170903232597

Optimal sensor allocation by integrating causal models and set-covering algorithms

JING LI[1] and JIONGHUA (JUDY) JIN[2,*]

[1] *Department of Industrial Engineering, Arizona State University, Tempe, AZ 85287-5906, USA*
[2] *Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109-2117, USA*
E-mail: jhin@umich.edu

Received February 2009 and accepted July 2009

Massive amounts of data are generated in Distributed Sensor Networks (DSNs), posing challenges to effective and efficient detection of system abnormality through data analysis. This article proposes a new method for optimal sensor allocation in a DSN with the objective of timely detection of the abnormalities in a underlying physical system. This method involves two steps: first, a Bayesian Network (BN) is built to represent the causal relationships among the physical variables in the system; second, an integrated algorithm by combining the BN and a set-covering algorithm is developed to determine which physical variables should be sensed, in order to minimize the total sensing cost as well as satisfy a prescribed detectability requirement. Case studies are performed on a hot forming process and a large-scale cap alignment process, showing that the developed algorithm satisfies both the cost and detectability requirements.

Keywords: Bayesian networks, causal models, sensor allocation, set-covering algorithm

1. Introduction

Advances in sensing and computing technologies have lead to the widespread use of Distributed Sensor Networks (DSNs) in many areas, such as manufacturing industry, environmental protection, and homeland security (Ding *et al.*, 2006). One of the major challenges in the use of DSNs is to effectively and efficiently analyze the massive amount of data created by the large number of sensors, so that useful knowledge about the underlying physical system can be discovered and timely decisions can be made. Since the total amount of data collected by the sensors in a DSN, even within a relatively short time period, can be overwhelming, attempting to analyze all the data simultaneously in a data fusion center may not be appropriate. This is especially true when a real-time decision is required in response to a catastrophic event. In Wireless Sensor Networks (WSNs), not only is analyzing the large amount of data a serious concern, communicating the data to the fusion center is also a major issue due to a limited communication bandwidth (Akyildiz *et al.*, 2002). Increasing the bandwidth results in more power being consumed, thus reducing the lifetime of a WSN since the sensors in the WSN are usually powered by batteries that have a limited lifespan.

While sensing every physical parameter of a system can minimize information loss, the resulting sensor network may generate an overwhelming amount of data, leading to tremendous difficulty in data communication and analysis. Moreover, it is quite common that the data contains redundant information, such as data collected on two highly correlated physical parameters. Therefore, it is desirable to study how to optimally allocate sensors in a sensor network; i.e., how to collect sufficient information from a system at the minimum sensing cost. This is equivalent to selecting the minimum number of sensors, if the sensing cost of each sensor is assumed to be the same.

The problem of sensor allocation has been increasingly investigated in recent years. There are two common research issues: one is to decide where to physically install the sensors, and the other is to decide the physical parameters that are to be measured by the sensors. A typical application area is computer vision, in which the physical locations and settings of the sensors are selected to achieve the best quality in object recognition (Tarabanis *et al.*, 1995). Another application area is WSNs, which focuses on where to put sensors so that every physical spot in the investigated area is within the sensing range of at least one sensor, and every sensor is able to deliver its data to a base station within its lifespan at the minimum cost (Dhillon and Chakrabarty, 2003; Mhatre *et al.*, 2005; Xu *et al.*, 2005). Recently, optimal sensor allocation for system monitoring, diagnosis,

*Corresponding author

and control has become a promising research area in quality engineering (Shi, 2006). In this area, research has been focused on sensor allocation, as well as data modeling and analysis of systems where there is sufficient prior knowledge to be able to quantitatively describe the relationships among physical parameters (Khan *et al.*, 1998; Ceglarek *et al.*, 1999; Wang and Nagarkar, 1999; Ceglarek and Khan, 2000; Ding *et al.*, 2003; Liu *et al.*, 2005; Shi, 2006). There have been few attempts to create general formulations and solutions for data-driven relationship modeling of physical parameters, based on which optimal sensor allocation is discussed.

Generally speaking, optimal sensor allocation is "mission specific"; that is, where to allocate sensors is closely related to specific objectives of knowledge discovery and decision making. In this article, we focus on the important topic of abnormality detection in a physical system; i.e., how to track overall system performance and quickly generate alarms when there is any abnormality. To accomplish this mission, it is reasonable to believe that not all physical parameters in a system need to be sensed by considering their dependence/correlation relationships. For example, if there is a causal relationship between variables X_i and X_j, e.g., X_i is a cause of X_j, and a sensor has been installed on X_j, then it may not be necessary to sense X_i because any abnormality in X_i will propagate to X_j and may incur alarms in X_j.

An effective way to monitor the performance of a physical system and detect abnormalities is to build monitoring control charts on sensor outputs. In a system with p physical variables $\{X_1, \ldots, X_p\}$, letting $\mathbf{\Omega} = \{\Omega_1, \ldots, \Omega_q\} \subseteq \{X_1, \ldots, X_p\} (q \leq p)$ be the subset of variables on which sensors are installed, then control charts can be constructed on the outputs of these sensors and out-of-control signals can be interpreted as signs of system abnormalities. Here, two types of strategies may be employed to construct the control charts: one is to build a multivariate control chart that combines the information from all the sensors, and the other is to build a control chart for each sensor that results in a total of q control charts. The second strategy has good interpretability since it not only flags system abnormalities but also facilitates diagnosis as to where (i.e., in which specific variables) the abnormalities occur. Therefore, the second strategy is adopted in this article.

Specifically, we propose a method for optimal sensor allocation in a sensor network with the objective of timely detecting the abnormalities occurring in the underlying physical system. The method involves two steps: first, a Bayesian network is built to represent the causal relationships among the physical variables in the system; second, an algorithm is developed to determine which physical variables should be sensed, in order to minimize the total sensing cost as well as satisfy a prescribed detectability requirement (i.e., how soon an alarm should be generated after the abnormality occurs). In developing this method, we focus on system abnormalities that are single mean shifts.

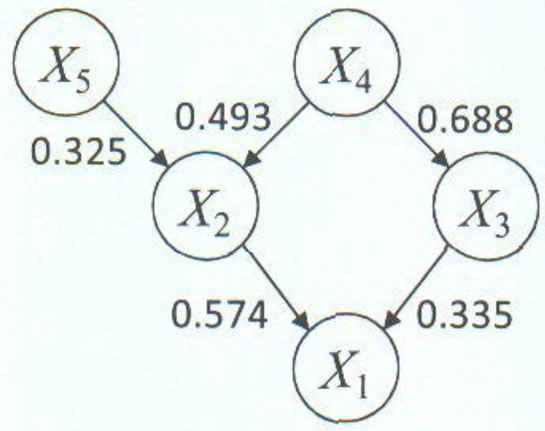

Fig. 1. BN structure of a hot forming process.

2. Bayesian networks in causal relationship representation

A Bayesian Network (BN) can be used to represent the causal relationships among the physical variables $\mathbf{X} = \{X_1, \ldots, X_p\}$ in a system. A BN has two components: structure and parameters (see Fig. 1 for an example). The structure of a BN is a directed acyclic graph; i.e., a set of nodes, each corresponding to one physical variable, connected by directed arcs. If there is a directed arc from X_i to X_j, i.e., $X_i \rightarrow X_j$, then X_i is a direct cause (called a parent) of X_j and X_j is a direct effect (called a child) of X_i, where *direct* means that the causal influence from X_i to X_j is not mediated through any other variables in the BN. In other words, the lack of a directed arc between any two variables means that these two variables do not have a direct cause–effect relationship; however, they may have one of the following relationships:

(a) they are dependent/correlated by sharing a common cause (e.g., X_2 and X_3 in Fig. 1);
(b) one variable is an indirect cause of the other (e.g., X_5 and X_1);
(c) they are independent (e.g., X_5 and X_4).

Furthermore, if there is a directed path from X_i to X_j, i.e., $X_i \rightarrow \cdots \rightarrow X_j$, then X_i is a direct or indirect cause (called an ancestor) of X_j and X_j is a direct or indirect effect (called a descendant) of X_i. In this article, the sets of parents, children, descendents and ancestors of a variable X_j are denoted by $\mathbf{PA}(X_j)$, $\mathbf{CH}(X_j)$, $\mathbf{DE}(X_j)$, and $\mathbf{AN}(X_j)$, respectively; and a component of each set is denoted by $\mathrm{PA}_k(X_j)$, $\mathrm{CH}_k(X_j)$, $\mathrm{DE}_k(X_j)$, and $\mathrm{AN}_k(X_j)$, respectively.

The parameters of a BN are a set of conditional probability distributions, $P(X_j|\mathbf{PA}(X_j))$, $j = 1, \ldots, p$. In this article, it is assumed that all variables, when the system runs under the normal condition, follow the standard normal distribution, and a linear Gaussian parameterization of the BN is used (Lauritzen and Wermuth, 1989; Korb and Nicholson, 2003); that is,

$$X_j = \sum_{k=1}^{m_j} p(PA_k(X_j), X_j) PA_k(X_j) + V_j. \qquad (1)$$

In Equation (1), $p(\cdot, \cdot)$ is a linear coefficient, called a *path coefficient*; $p(\cdot, \cdot) \in (0, 1)$ because all variables

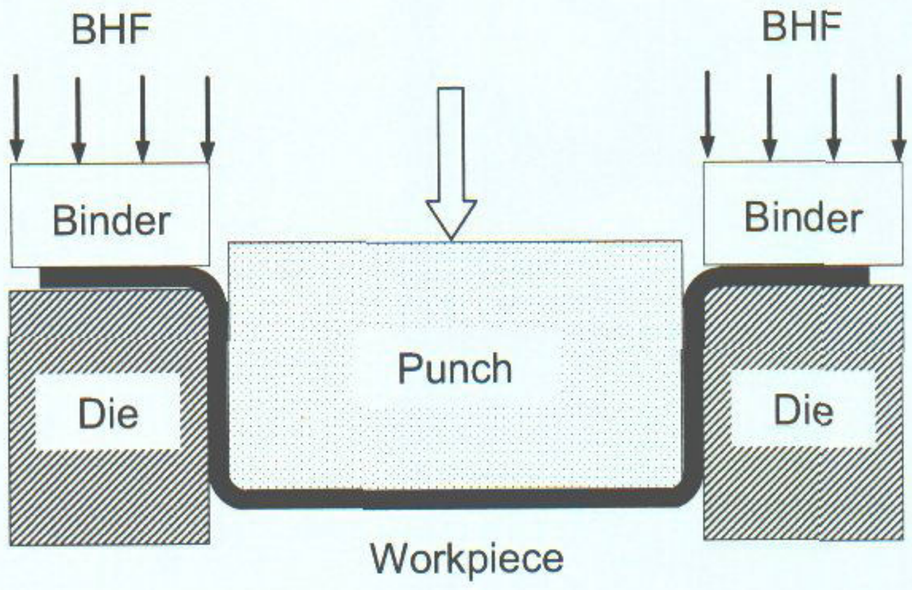

Fig. 2. 2-D illustration of the hot forming process.

follow the standard normal distribution. V_j is independent of $PA_k(X_j)$, $V_j \sim N(0, \sigma_j^2)$; here, $\sigma_j^2 = 1 - \sum_{k=1}^{m_j} p(PA_k(X_j), X_j)\rho(PA_k(X_j), X_j)$ in order to maintain unit variance of each variable given the Pearson correlation, denoted by $\rho(\cdot, \cdot)$, along the incoming paths. The parameterization in Equation (1) is also adopted in path modeling (Wright, 1921), which has been extensively employed in social and behavioral sciences.

As an example, Fig. 1 shows a linear Gaussian BN for a hot forming process with one quality variable (X_1: final dimension of workpiece) and four process variables (X_4: temperature, X_3: material flow stress, X_2: tension in workpiece and X_5: Blank Holding Force (BHF; Li *et al.*, 2008). A two-dimensional (2-D) physical illustration of the hot forming process is given in Fig. 2. The numbers beside the directed arcs in Fig. 1 are the parameters (i.e., the path coefficients).

Generally speaking, the structure of a BN can be obtained from engineering knowledge or learned from data. One commonly used data-driven algorithm for structure learning is called the PC (Peter and Clark) alogrithm (Spirtes *et al.*, 1993), which uses a series of statistical significance tests of conditional independence. For linear Gaussian BNs, PC uses partial correlation tests. Furthermore, the parameters (i.e., the path coefficients) in a linear Gaussian BN can be estimated based on the sample correlation matrix of all variables through multiple regressions (Kline, 2005) or Maximum Likelihood Estimation (MLE). An advantage of MLE is that it can be integrated with Bayesian estimation, resulting in a Bayesian maximum a posterior approach that augments the likelihood with a prior that gives the initial belief about the parameters before seeing any data (Buntine, 1996).

Under the linear Gaussian parameterization in Equation (1), we define, in this article, the component effects and total effect between two variables X_i and X_j. Specifically, a component effect of X_i on X_j is the effect due to a specific directed path from X_i to X_j. It is easy to derive that the component effect is equal to the product of all path coefficients on the corresponding directed path. The total effect of X_i on X_j is the sum of all component effects and is denoted by $\tilde{\gamma}(X_i, X_j)$. Note that:

$$\tilde{\gamma}(X_i, X_j) = \begin{cases} \in (0, 1) \text{ if } X_i \in \mathbf{AN}(X_j), \\ 1 \qquad \text{if } i = j, \\ 0 \qquad \text{otherwise} \end{cases} \tag{2}$$

Based on Equation (2), it is possible to trace how the fault in one variable propagates to other variables for a given BN. Specifically, a single mean shift in X_i, i.e., $E(X_i) = \Delta_i \neq 0$, will incur a mean shift in X_j as follows:

$$E(X_j) = \tilde{\gamma}(X_i, X_j)\Delta_i. \tag{3}$$

Consequently, the total effect of an ancestor of a variable on this variable can be computed. For example, in Fig. 1, because there are two different directed path from X_4 to X_1, i.e., $X_4 \to X_2 \to X_1$ *and* $X_4 \to X_3 \to X_1$, the total effect of X_4 on X_1 is the sum of two component effects; i.e., $\tilde{\gamma}(X_4, X_1) = 0.493 \times 0.574 + 0.688 \times 0.335 \approx 0.513$. If there is a mean shift in X_4, i.e., $E(X_4) = \Delta_4 \neq 0$, it will incur a mean shift of $0.513\Delta_4$ in X_1.

3. Problem formulation of optimal sensor allocation for system abnormality detection

If the causal relationships among the variables can be known and represented by a BN, it is possible to trace how the mean shift Δ_i in a variable X_i propagates to any variable X_j in the BN, $i, j \in \{1, \ldots, p\}$. Specifically, given Δ_i, the mean in X_j is given in Equation (3). Supposing that there is a sensor on X_j, i.e., $X_j \in \mathbf{\Omega}$, then the average run length of a Shewhart control chart on X_j in detecting the mean shift in X_i is (Montgomery, 2001)

$$ARL_1^{X_j}(\Delta_i) = \frac{1}{1 - \phi(z_{\alpha/2} - E(X_j)) + \phi(-z_{\alpha/2} - E(X_j))}, \tag{4}$$

where $\phi(\cdot)$ and $z_{\alpha/2}$ are the cumulative distribution function and the upper $\alpha/2$ percentile of the standard normal distribution, respectively, and α is a Type-I error of the control chart. If X_i is an ancestor of X_j, or $i = j$, then $\tilde{\gamma}(X_i, X_j) \neq 0$ and it can be assumed without loss of generality that $\tilde{\gamma}(X_i, X_j) > 0$. Also, let δ_i be the magnitude of the mean shift Δ_i; i.e., $\delta_i > 0$. Thus, for a positive mean shift $\Delta_i = \delta_i$, Equation (4) becomes

$$ARL_1^{X_j}(\delta_i) \approx \frac{1}{1 - \phi\left(z_{\alpha/2} - \tilde{\gamma}\left(X_i, X_j\right)\delta_i\right)}, \tag{5}$$

for a negative mean shift $\Delta_i = -\delta_i$, Equation (4) becomes:

$$\begin{aligned} ARL_1^{X_j}(\delta_i) &\approx \frac{1}{\phi(-z_{\alpha/2} - \tilde{\gamma}(X_i, X_j)(-\delta_i))} \\ &= \frac{1}{1 - \phi(z_{\alpha/2} - \tilde{\gamma}(X_i, X_j)\delta_i)}, \end{aligned}$$

which is the same as Equation (5). In other words, Equation (5) holds for both positive and negative mean shifts.

Furthermore, it can be shown that $ARL_1^{X_j}(\delta_i)$ is monotonically decreasing with respect to $\tilde{\gamma}(X_i, X_j)$, for given α and δ_i.

Let $\Omega_{i_0} \in \mathbf{\Omega}$ be the sensor such that the total effect of X_i on Ω_{i_0} is the largest among all the sensors in $\mathbf{\Omega}$; i.e., $\Omega_{i_0} = \text{argmax}_{\Omega_j \in \mathbf{\Omega}} \tilde{\gamma}(X_i, \Omega_j)$. Then, the control chart on Ω_{i_0} has the minimum average run length in detecting δ_i among all control charts. Furthermore, in order for the mean shift δ_i to be detectable, $ARL_1^{\Omega_{i_0}}(\delta_i)$ must not exceed an upper bound, $ARL_{1U}(\delta_i)$, which is set according to specific domain standards; i.e., $ARL_1^{\Omega_{i_0}}(\delta_i) \leq ARL_{1U}(\delta_i)$. It can be shown that this inequality holds if and only if $\tilde{\gamma}(X_i, \Omega_{i_0}) \geq \eta_i$, where:

$$\eta_i = \frac{\left(z_{\alpha/2} - \phi^{-1}\left(1 - \frac{1}{ARL_{1U}(\delta_i)}\right)\right)}{\delta_i}. \tag{6}$$

Therefore, the following definitions can be given under a BN framework.

Definition 1. A mean shift δ_i in $X_i \in \mathbf{X}$ is *detectable* by a set of sensors, $\mathbf{\Omega}$, if the largest total effect of X_i on the sensors in $\mathbf{\Omega}$, $\tilde{\gamma}(X_i, \Omega_{i_0})$, is no less than η_i; i.e.; $\tilde{\gamma}(X_i, \Omega_{i_0}) \geq \eta_i$.

Definition 2. $\mathbf{\Omega} = \{\Omega_1, \ldots, \Omega_q\}$ is a feasible solution to the problem of sensor allocation, if the mean shift δ_i in any $X_i \in \mathbf{X}$ is detectable by $\mathbf{\Omega}$; i.e., $\tilde{\gamma}(X_i, \Omega_{i_0}) \geq \eta_i$ for $\forall i \in \{1, \ldots, p\}$.

Furthermore, the problem of optimal sensor allocation can be defined as finding a set of sensors, $\mathbf{\Omega}$, in order to

$$\min W = \sum_{j=1,}^{q} w(\Omega_j),$$

subject to

$$\tilde{\gamma}(X_i, \Omega_{i_0}) \geq \eta_i, \quad \forall i \in \{1, \ldots, p\}, \tag{7}$$

where $w(\Omega_j)$ is the cost of sensor Ω_j.

Taking the hot forming process in Fig. 1 as an example, given that $\alpha = 0.05$, $\delta_i = 3$ $(i \in \{1, \ldots, 5\})$, and $ARL_{1U}(\delta_i) = 5$, then $\eta_i = 0.373$ by Equation (6). If $\mathbf{\Omega} = \{X_1, X_3, X_5\}$, then $\tilde{\gamma}(X_4, \Omega_{4_0}) = \tilde{\gamma}(X_4, X_3) = 0.688 > \eta_i$. Thus, by Definition 1, a mean shift of three in X_4 is detectable by $\mathbf{\Omega}$. Similarly, it can be shown that a mean shift of three in X_i, $i = 1, 2, 3, 5$, is detectable by $\mathbf{\Omega}$. Therefore, according to Definition 2, $\mathbf{\Omega}$ is a feasible solution to the problem of sensor allocation in detecting the mean shift of three at any node in the process. Also, the total sensing cost is $W = w(X_3) + w(X_5) + w(X_1)$.

4. Solution to optimal sensor allocation for system abnormality detection

4.1. *Solution based on set-covering algorithms*

The set-covering problem is a classic optimization problem in computer science and complexity theory for resource selection (Cormen *et al.*, 2001). The input to a set-covering problem is a finite set $\mathbf{Y} = \{Y_1, \ldots, Y_p\}$; m subsets of $\mathbf{Y}$, whose union is $\mathbf{Y}$, i.e., $\bigcup_{k=1}^{m} \mathbf{S}_k = \mathbf{Y}$; and a cost for each subset, $w(\mathbf{S}_k)$. The goal is to find a few subsets, with minimum total cost, whose union is $\mathbf{Y}$; namely, the goal is to find a set $\mathbf{L} \subseteq \{1, \ldots, m\}$ that:

$$\min W = \sum_{l \in \mathbf{L}} w(S_l),$$

subject to

$$\bigcup_{l \in \mathbf{L}} S_l = Y. \tag{8}$$

This section will show that the problem of optimal sensor allocation, as defined in Equation (7), can be translated into a set-covering problem. To facilitate this translation, Definition 3 and Proposition 1 below are needed.

We can aggregate all the variables whose mean shifts are detectable by $\mathbf{\Omega}$ into a set, $\mathbf{C}^{\mathbf{\Omega}}$, called a *duty set* of $\mathbf{\Omega}$.

Definition 3. The duty set of $\mathbf{\Omega}$ is $\mathbf{C}^{\mathbf{\Omega}} = \{X_i | \tilde{\gamma}(X_i, \Omega_{i_0}) \geq \eta_i, i = 1, \ldots, p\}$.

In a special case when only one variable, X_k, is sensed, i.e., $\mathbf{\Omega} = \{X_k\}$, the duty set of $\{X_k\}$ is

$$\mathbf{C}^{\{X_k\}} = \{X_i | \tilde{\gamma}(X_i, X_k) \geq \eta_i, i = 1, \ldots, p\}. \tag{9}$$

Furthermore, a sufficient and necessary condition for $\mathbf{\Omega}$ to be a feasible solution is given in Proposition 1.

Proposition 1. *$\mathbf{\Omega} = \{\Omega_1, \ldots, \Omega_q\} \subseteq \mathbf{X}$ is a feasible solution to the problem of sensor allocation, if and only if the union of the duty sets $\mathbf{C}^{\{\Omega_j\}}$, $j \in \{1, \ldots, q\}$, is $\mathbf{X}$; that is,*

$$\bigcup_{j=1}^{q} \mathbf{C}^{\{\Omega_j\}} = \mathbf{X}.$$

Proof. See Appendix 1. ■

Based on Proposition 1, the problem of optimal sensor allocation in Equation (7) can be redefined as finding a set of sensors, $\mathbf{\Omega}$, in order to

$$\min W = \sum_{j=1,}^{q} w(\Omega_j),$$

subject to

$$\bigcup_{j=1}^{q} \mathbf{C}^{\{\Omega_j\}} = \mathbf{X}. \tag{10}$$

The input and goal of the general set-covering problem and those of the problem of optimal sensor allocation in Equation (10), are compared in Table 1. It is clear from this table that the problem of optimal sensor allocation is a set-covering problem.

Table 1. Translation of optimal sensor allocation into a set-covering problem

	General set-covering problem	*Optimal sensor allocation*
Input	$\mathbf{Y} = \{Y_1, \ldots, Y_p\}$ $\mathbf{S}_k \subseteq \mathbf{Y}, \quad k = 1, \ldots, m$ such that $\bigcup_{k=1}^{m} \mathbf{S}_k = \mathbf{Y}$ $w(S_k)$	$\mathbf{X} = \{X_1, \ldots, X_p\}$ $\mathbf{C}^{\{X_k\}} \subseteq \mathbf{X}, \quad k = 1, \ldots, p$, naturally $\bigcup_{k=1}^{p} \mathbf{C}^{\{X_k\}} = \mathbf{X}$ $w(\mathbf{C}^{\{X_k\}})$ equivalent to sensing cost $w(X_k)$
Goal	Find $\mathbf{L} \subseteq \{1, \ldots, m\}$ in order to min $W = \sum_{l \in \mathbf{L}} w(\mathbf{S}_l)$ subject to $\bigcup_{l \in \mathbf{L}} \mathbf{S}_l = \mathbf{Y}$	Find $\mathbf{L} \subseteq \{1, \ldots, p\}$ in order to min $W = \sum_{l \in \mathbf{L}} w(X_l)$ subject to $\bigcup_{l \in \mathbf{L}} \mathbf{C}^{\{X_l\}} = \mathbf{X}$

It is known that the set-covering problem is NP-hard; that is, an algorithm to obtain the optimal solution in polynomial-time has not been found. Several approximation algorithms exist, which run in polynomial-time and can produce a solution that is within a factor v of the optimal solution, called v-approximation algorithms. Among them, one of the most popular algorithms is an $H(\max_{k=1,\ldots,m}|\mathbf{S}_k|)$—approximation greedy algorithm (Cormen *et al.*, 2001), where $|\cdot|$ denotes the cardinality of a set, and $H(d)$ is the dth harmonic number; that is,

$$H(d) = \sum_{i=1}^{d} 1/i. \tag{11}$$

The greedy algorithm can be applied to search for the solution to the problem of optimal sensor allocation, as illustrated in the steps of Fig. 3, where A/B denotes the complement of set B with respect to set A. By applying these steps, a feasible solution, denoted by $\mathbf{L}_G \subseteq \{1, \ldots, p\}$, can be found, which leads to a total sensing cost W_G no greater than $H(\max_{k=1,\ldots,p})|\mathbf{C}^{\{x_k\}}|)$ times of the optimal (i.e., minimum) total sensing cost W_{Opt}; that is,

$$W_{\text{G}} \leq H(\max_{k=1,\ldots,p}|\mathbf{C}^{\{X_k\}}|)\, W_{\text{Opt}}. \tag{12}$$

For example, for the hot forming process in Fig. 1, assume that $\alpha = 0.05$, $\delta_i = 3$ $(i \in \{1, \ldots, 5\})$, and $ARL_{1\text{U}}(\delta_i) = 5$, then $\eta_i = 0.373$ by Equation (6). By Definition 3, $\mathbf{C}^{\{X_5\}} = \{X_5\}$, $\mathbf{C}^{\{X_4\}} = \{X_4\}$, $\mathbf{C}^{\{X_3\}} = \{X_4, X_3\}$, $\mathbf{C}^{\{X_2\}} = \{X_4, X_2\}$, and $\mathbf{C}^{\{X_1\}} = \{X_4, X_2, X_1\}$. In addition, assume that the sensing cost is the same across different sensors; i.e., $w(X_k) = w, k = 1, \ldots, 5$. Therefore, by applying the greedy algorithm in Fig. 3, $\mathbf{L}_G = \{1, 3, 5\}$; that is, variables X_1, X_3, and X_5 should be sensed, and $W_G = 3w$.

$\mathbf{U} \leftarrow \mathbf{X}$
$\mathbf{L}_G \leftarrow \phi$
While $\mathbf{U} \neq \phi$, do:
  Select an $\mathbf{C}^{\{X_{k_0}\}}$ that minimizes $w(X_{k_0}) \big/ \left|\mathbf{C}^{\{X_{k_0}\}} \cap \mathbf{U}\right|$
  $\mathbf{U} \leftarrow \mathbf{U} \setminus \left(\mathbf{C}^{\{X_{k_0}\}} \cap \mathbf{U}\right)$
  $\mathbf{L}_G \leftarrow \mathbf{L}_G \cup \{k_0\}$

Fig. 3. A greedy algorithm for optimal sensor allocation.

While the greedy algorithm does not guarantee to find the optimal solution, it guarantees to find a feasible solution that is bounded by a factor $H(\max_{k=1,\ldots,p}|\mathbf{C}^{\{X_k\}}|) = H(3) = 1.8$ of the optimal solution; e.g., $W_{\text{G}} \leq 1.8\, W_{\text{Opt}}$. Because this example involves only five variables, it is easy to perform an exhaustive search to identify the optimal solution, $\mathbf{L}_{\text{opt}} = \{1, 3, 5\}$, although the exhaustive search runs in exponential-time. Therefore, in this example, the greedy algorithm happens to find the optimal solution.

4.2. *An improved solution procedure by integrating BN and the set-covering greedy algorithm*

It is obvious from Equation (12) that the smaller the $H(\max_{k=1,\ldots,p}|\mathbf{C}^{\{X_k\}}|)$, the closer the total sensing cost by the greedy algorithm, W_{G}, is to the optimal total sensing cost, W_{Opt}. Due to the definition of the function H in Equation (11), $H(\max_{k=1,\ldots,p}|\mathbf{C}^{\{X_k\}}|)$ is monotonically increasing with respect to $\max_{k=1,\ldots,p}|\mathbf{C}^{\{X_k\}}|$. Therefore, an effective way for bringing W_G close to the optimal total sensing cost W_{Opt} is to reduce $\max_{k=1,\ldots,p}|\mathbf{C}^{\{X_k\}}|$. This motivates us to develop a pre-processing algorithm, which recursively runs two rules (see Fig. 4) before the greedy algorithm is executed, with the purpose of potentially reducing $\max_{k=1,\ldots,p}|\mathbf{C}^{\{X_k\}}|$. Proofs of the validity of the two rules are provided in Appendix 2. As a result, an integrated algorithm, which combines the pre-processing algorithm and the greedy algorithm in Fig. 3, can be developed, as shown in Fig. 4.

The integrated algorithm in Fig. 4 is applied to the previous example. In the pre-processing stage, because X_2 and X_4 satisfy that any duty set including X_2 must include X_4, X_4 can be eliminated from $\mathbf{U}$ and every duty set that includes it, according to Rule 1. As a result, $\mathbf{U} = \{X_1, X_2, X_2, X_5\}$, and the duty sets become $\mathbf{C}^{\{X_5\}} = \{X_5\}$, $\mathbf{C}^{\{X_4\}} = \phi$, $\mathbf{C}^{\{X_3\}} = \{X_2\}$, $\mathbf{C}^{\{X_2\}} = \{X_2\}$, and $\mathbf{C}^{\{X_1\}} = \{X_2, X_1\}$. Furthermore, by applying Rule 2, because X_1 is only included in $\mathbf{C}^{\{X_1\}}$, thus X_1 must be sensed, i.e., $\mathbf{L}_G = \{1\}$; $\mathbf{U} = \{X_2, X_5\}$; and the duty sets become $\mathbf{C}^{\{X_5\}} = \{X_5\}$, $\mathbf{C}^{\{X_4\}} = \phi$, $\mathbf{C}^{\{X_3\}} = \{X_2\}$, $\mathbf{C}^{\{X_2\}} = \phi$, and $\mathbf{C}^{\{X_1\}} = \phi$. By applying Rule 2 again, X_2 and X_5 must be sensed, i.e., $\mathbf{L}_G = \{1, 3, 5\}$; $\mathbf{U} = \phi$; and all duty sets become empty sets. Because $\mathbf{U} = \phi$, the second part of the integrated algorithm, i.e., the greedy algorithm,

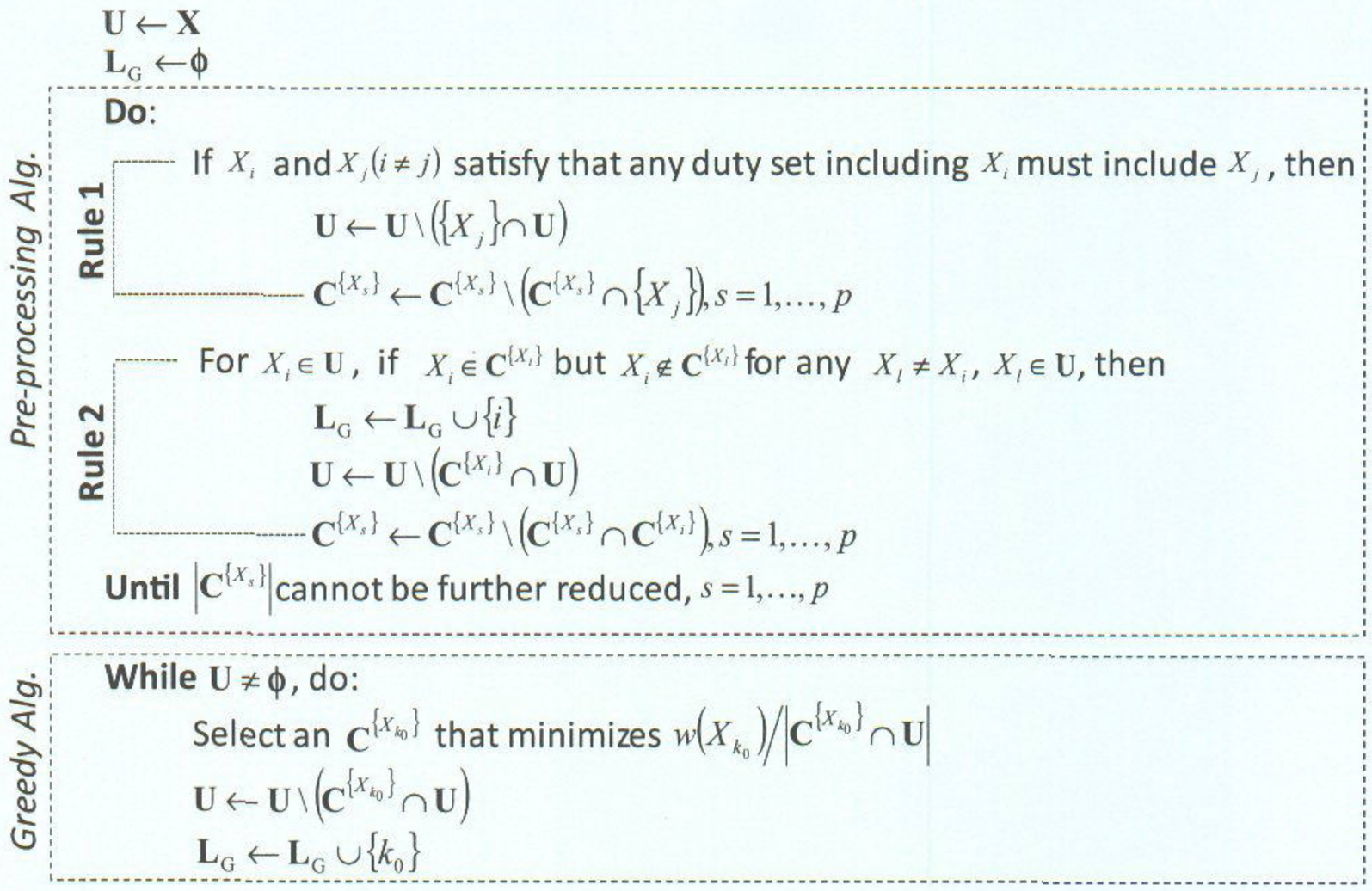

Fig. 4. An integrated algorithm for optimal sensor allocation by combining a pre-processing algorithm and the greedy algorithm in Fig. 3.

does not need to be run; that is, the pre-processing algorithm identifies a solution, in fact the optimal solution, to the problem of sensor allocation. This example shows that the pre-processing algorithm has at least two advantages: first, it helps reduce the cardinalities of the duty sets; second, it expedites identifying the optimal solution.

4.3. *A practical procedure to guide the decision making in optimal sensor allocation*

Note that the pre-processing algorithm proposed in the previous section can be combined with either the greedy algorithm or the exhaustive search to improve search efficiency. This section aims to identify practical considerations that may influence a decision maker to choose between the greedy algorithm and the exhaustive search for optimal sensor allocation.

It is known from the previous section that the greedy algorithm does not guarantee to find the optimal solution to minimize the total sensing cost. Therefore, a decision maker faces the choice of whether to run the greedy algorithm to get a feasible solution, whose corresponding total sensing cost can be as high as $H(\max_{k=1,\ldots,p}|\mathbf{C}^{\{X_k\}}|)$ times the minimum total sensing cost; or run an exhaustive search to find the optimal solution, which is computationally very expensive, if not impossible. To choose one over the other, the decision maker needs to be informed of two sets of information in comparing the greedy algorithm against the exhaustive search; i.e., how much computational resource can be saved and how much more total sensing cost has to be spent.

The computational resources needed by the greedy algorithm and the exhaustive search are primarily determined by their respective run time complexities. Given a system of p physical variables, the run time complexity of the exhaustive search is an exponential function of p; i.e., $O(e^p)$. Although the greedy algorithm is known to be executed in polynomial-time, to find the exact order of the polynomial requires a detailed analysis of the algorithm. Specifically, we show in Appendix 3 that the run time complexity of the greedy algorithm is a quadratic function of p; i.e., $O(p^2)$. Therefore, the greedy algorithm can save $(1 - p^2/e^p) \times 100\%$ of the computational resource used by the exhaustive search. Denote this percentage as λ_p; that is,

$$\lambda_p \equiv \left(1 - p^2/e^p\right) \times 100\%. \tag{13}$$

The total sensing costs using the greedy algorithm and the exhaustive search can be compared based on Equation (11), which, through some simple algebra, gives:

$$\frac{(W_G - W_{Opt})}{W_{Opt}} \leq \left(H\left(\max_{k=1,\ldots,p}|\mathbf{C}^{\{X_k\}}|\right) - 1\right) \times 100\%.$$

This indicates that the greedy algorithm may lead to a total sensing cost *at most* $(H(\max_{k=1,\ldots,p}|\mathbf{C}^{\{X_k\}}|) - 1) \times 100\%$ more than the minimum total sensing cost. Denote this percentage as λ_H, then

$$\lambda_H \equiv \left(H\left(\max_{k=1,\ldots,p}|\mathbf{C}^{\{X_k\}}|\right) - 1\right) \times 100\%. \tag{14}$$

Figure 5 plots λ_p with respect to different numbers of variables and λ_H with respect to different values for the maximum cardinalities of duty sets.

Considering the example in the previous section, because $p = 5$, $\lambda_p = 83\%$; because $H(\max_{k=1,\ldots,5}|\mathbf{C}^{\{X_k\}}|) = H(3) =$

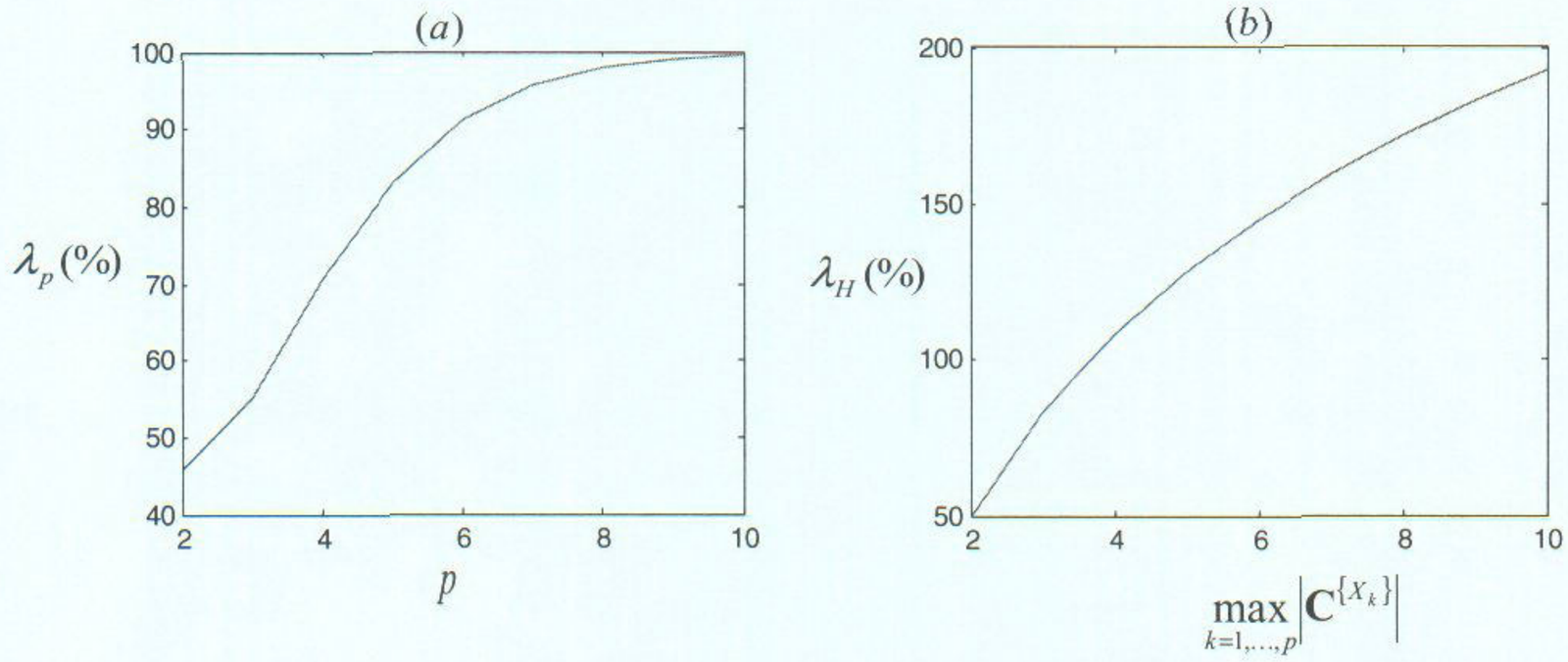

Fig. 5. (a) Relation between λ_p and number of variables and (b) relation between λ_H and maximum cardinality of duty sets.

1.8, $\lambda_H = 80\%$. Note that this value of λ_H corresponds to a *particular* mean shift of interest, i.e., $\delta = 3$, under certain requirements on the control chart performance including Type-I error $\alpha = 0.05$ and $ARL_{1U}(\delta) = 5$, which is directly linked to Type-II error. As mean shifts with various magnitudes may occur in practice, it is important to know how the value of λ_H changes with respect to different mean shift magnitudes. Toward this end, the function between λ_H and the mean shift δ is identified through the following proposition.

Proposition 2. *For a BN whose structure and path coefficients are known, λ_H is an increasing function of the mean shift of interest, δ, under given requirements on the maximum allowable Type-I error of individual control charts, α, and the maximum allowable Type-II error of individual control charts or, equivalently, the maximum allowable average run length until the mean shift δ is detected, $ARL_{1U}(\delta)$. Denote this function as $\lambda_H = \varphi_{\mathrm{BN},\alpha,\, ARL_{1U}(\delta)}(\delta)$.*

Proof. See Appendix 4. The expression of $\varphi_{\mathrm{BN},\alpha,\, ARL_{1U}(\delta)}$ is complicated and thus represented in an algorithmic format in Fig. 6. ■

In summary, in comparing the greedy algorithm against the exhaustive search option the decision maker needs to evaluate two sets of information: how much computational resource can be saved, reflected by λ_p; and at most how much more total sensing cost has to be spent with respect to different magnitudes of the mean shift δ, reflected by λ_H linked to δ through the function $\varphi_{\mathrm{BN},\alpha,\, ARL_{1U}}(\delta)$. Specifically, if the potential saving in total sensing cost overrules the additional spending in computational resource, then the exhaustive search is suggested for optimally allocating sensors; otherwise, the greedy algorithm is preferred. Figure 7 gives a procedure suggested for optimal sensor allocation in practice.

5. Examples

5.1. *Hot forming process*

Although exhaustive search can be adopted for optimal sensor allocation in this simple system, we use this system to illustrate the procedure in Fig. 7 and justify the integrated algorithm in Fig. 4.

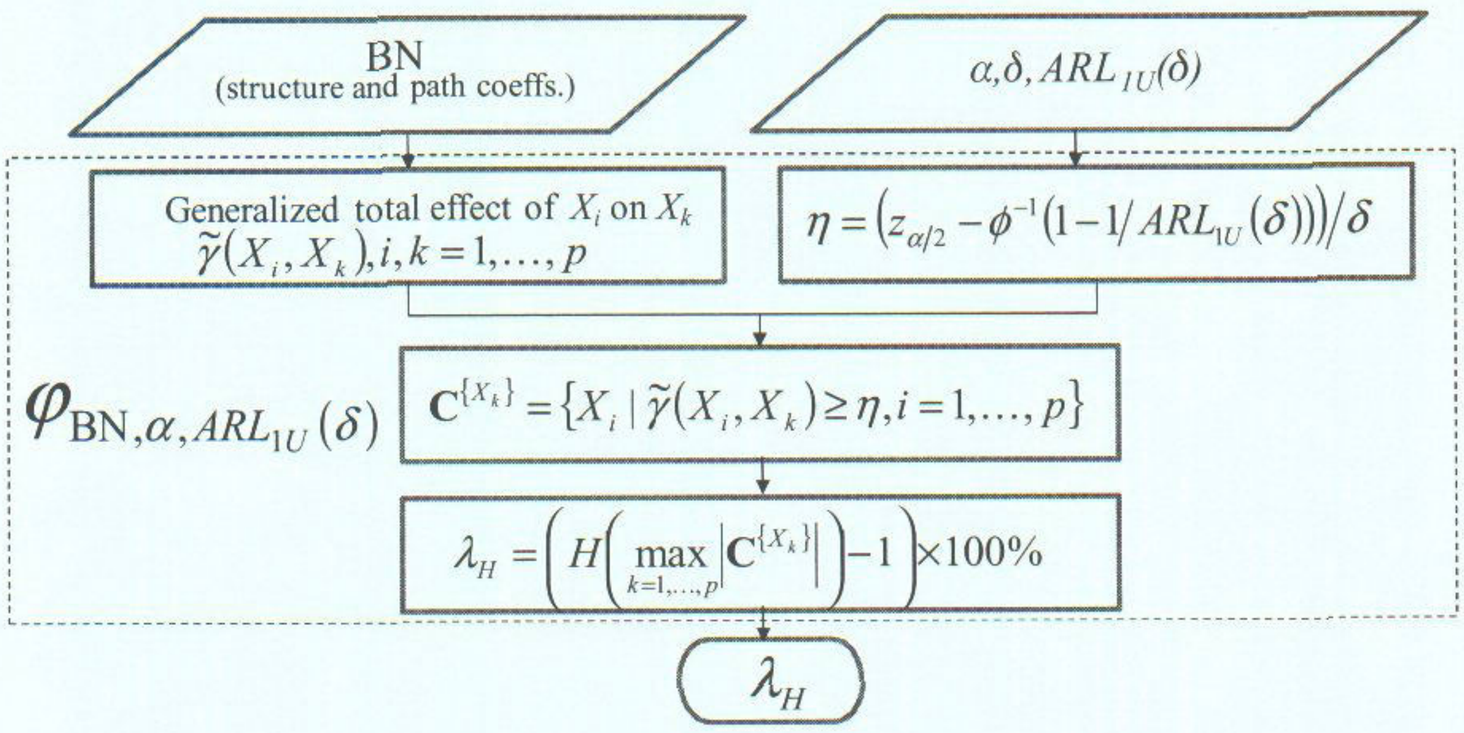

Fig. 6. Function $\lambda_H = \varphi_{\mathrm{BN},\alpha,\, ARL_{1U}(\delta)}(\delta)$ in algorithmic format.

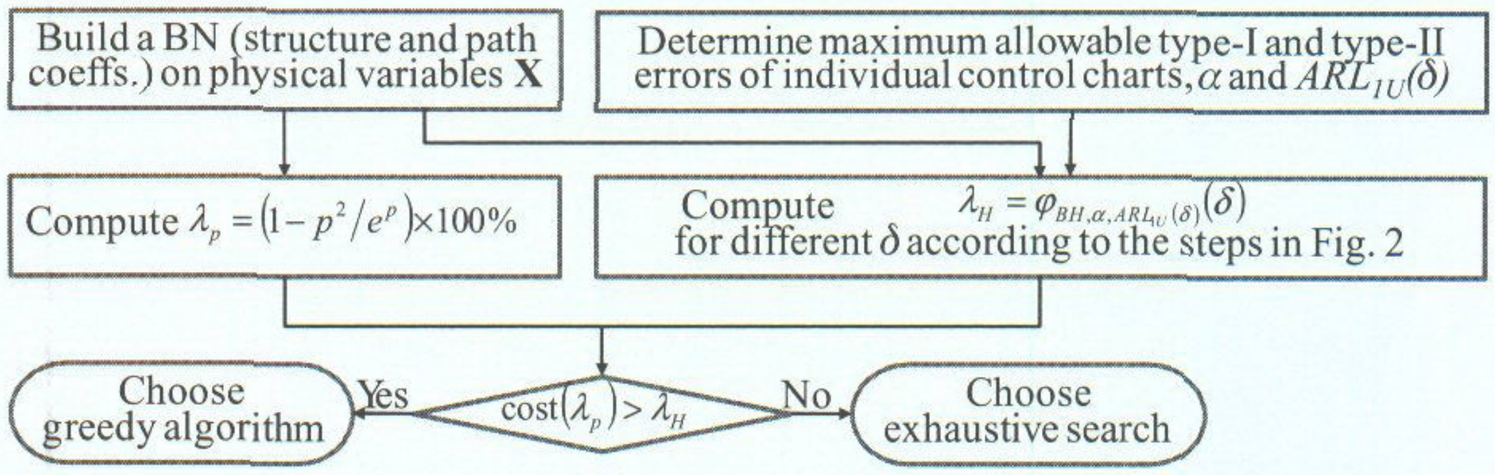

Fig. 7. A practical procedure to guide the decision maker in optimal sensor allocation.

5.1.1. *Demonstration of the practical procedure (Fig. 7) for optimal sensor allocation*

The procedure in Fig. 7 is followed for optimal sensor allocation in the hot forming process of Figs. 1 and 2. Because the process includes five variables, i.e., $p = 5$, $\lambda_p = 83\%$, indicating that the greedy algorithm can save 83% of the computational resource used by the exhaustive search. Furthermore, Fig. 8 plots λ_H against δ for different combinations of the maximum allowable Type-I error α and the average run length in mean shift detection, ARL_{1U}. It indicates that, for example, if $\alpha = 0.05$ and $ARL_{1U}(\delta) = 10$, then *at most* 83% and 108% more total sensing cost than the minimum total sensing cost will be spent, for detecting mean shifts $\delta \leq 2$ and $\delta > 2$, respectively. The decision maker can weigh the computed λ_p and the λ_H against δ plot in Fig. 8 in order to determine if he/she wants to run the greedy algorithm or the exhaustive search approach for optimal sensor allocation.

5.1.2. *Sensor allocation by the integrated algorithm*

Suppose that the greedy algorithm is chosen, it can be combined with the proposed pre-processing rules, resulting in the integrated algorithm shown in Fig. 4. The integrated algorithm was applied to the hot forming process. As a result, Table 2 shows the feasible solution L_G, i.e., the indices of the variables that should be sensed, for different combinations of δ, $ARL_{1U}(\delta)$ and α, assuming the same sensing cost across different sensors. A general pattern in the tables is that the larger the α, the $ARL_{1U}(\delta)$ or the δ, the fewer the number of variables that will be sensed, which agrees well with intuition.

5.1.3. *Comparison between the integrated algorithm solution and optimal solution*

It is of interest to compare how the solution from the integrated algorithm is different from the optimal solution, in terms of which variables to be sensed as well as the total sensing cost. Therefore, the exhaustive search was run for different combinations of δ, $ARL_{1U}(\delta)$ and α. As a result, a similar table for the optimal solution L_{opt}, to Table 2 for the integrated algorithm solution L_G, could be generated. Comparison between L_{opt} and L_G shows that L_G is the same as L_{opt} for any combination of δ, $ARL_{1U}(\delta)$ and α, indicating that the integrated algorithm performs as well as the exhaustive search for solving the problem of optimal sensor allocation in the hot forming process.

Furthermore, because the path coefficients of the hot forming process are estimated from a specific sample

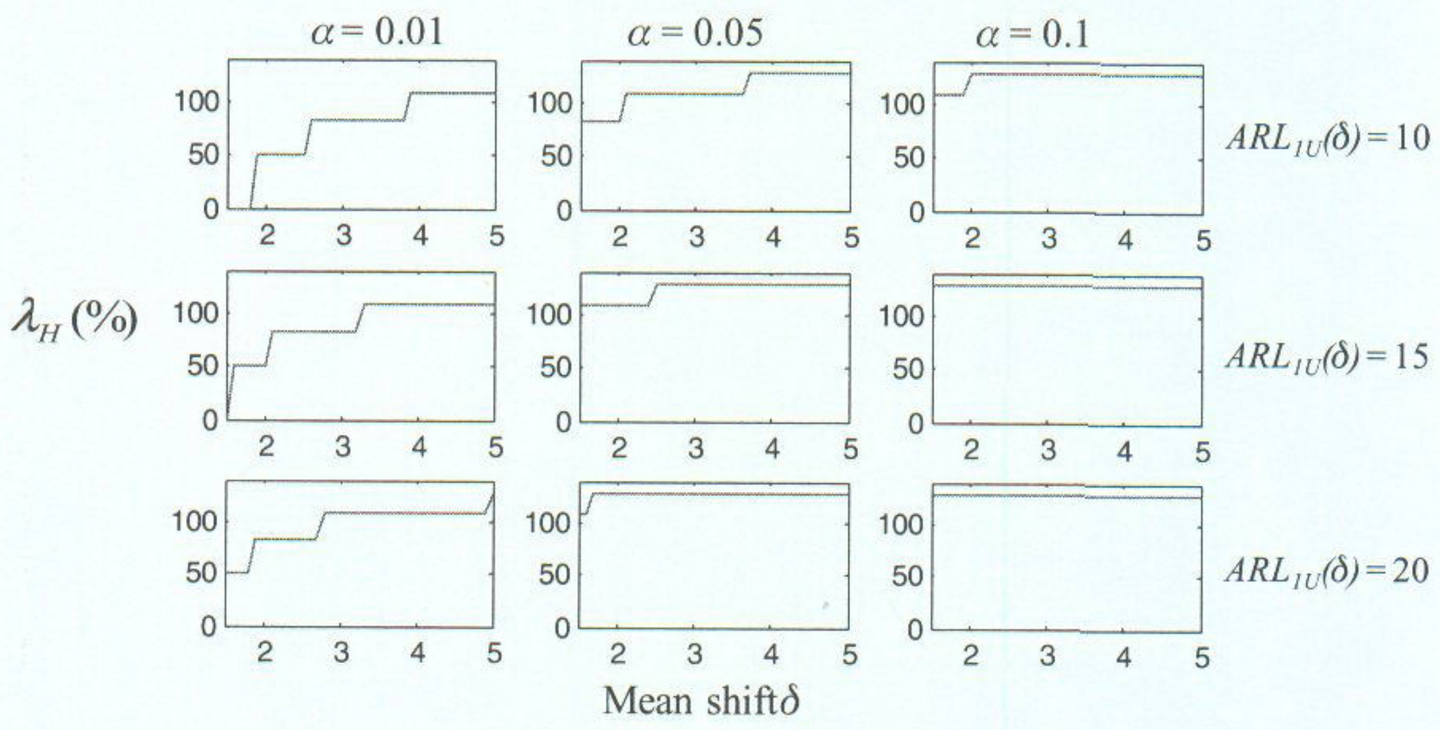

Fig. 8. Percentages of largest possible additional spending in total sensing cost by greedy algorithm, compared with minimum total sensing cost, as a function of mean shifts of interest, under different combinations of α and $ARL_{1U}(\delta)$ for the hot forming process in Figs. 1 and 2.

Table 2. Integrated algorithm solution for the hot forming process, $\mathbf{L}_G$ ($\alpha = 0.01, 0.05$ and 0.1)

		δ			
α	$ARL_{IU}(\delta)$	*1.5*	*2*	*2.5*	*3*
0.01	10	{1,2,3,4,5}	{1,2,3,5}	{1,3,5}	{1,3,5}
	15	{1,2,3,4,5}	{1,3,5}	{1,3,5}	{1,3,5}
	20	{1,2,3,5}	{1,3,5}	{1,3,5}	{1,2}
0.05	10	{1,3,5}	{1,3,5}	{1,2}	{1,2}
	15	{1,2}	{1,2}	{1}	{1}
	20	{1,2}	{1}	{1}	{1}
0.1	10	{1,2}	{1}	{1}	{1}
	15	{1}	{1}	{1}	{1}
	20	{1}	{1}	{1}	{1}

dataset, it is of interest to study how different estimates of the path coefficients may impact the performance of the integrated algorithm. Toward this end, another simulation study was performed which included the following steps.

1. The BN structure is randomly parameterized; that is, with the fixed structure (i.e., nodes and arcs), the parameters (i.e., path coefficients of the arcs) take values of randomly generated numbers following a uniform distribution between zero and one.
2. The integrated algorithm and the exhaustive search are run on the parameterized BN in Equation (1); the solutions, i.e., the indices of the variables to be sensed, are saved in L_G and L_{Opt}, respectively; the total sensing costs are saved in W_G and W_{Opt}, respectively.
3. If L_G contains the same indices as L_{Opt}, then $diff_{Sol} = 0$; otherwise, $diff_{Sol} = 1$. Similarly, $diff_{Cost} = (W_G - W_{Opt})/W_{Opt}$.
4. Steps 1 to 3 are repeated N times, and the average $diff_{Sol}$ and $diff_{Cost}$ are computed and saved into $diff_{Sol}$ and $diff_{Cost}$, respectively.

Steps 1 to 4 were performed under different combinations of δ, $ARL_{1U}(\delta)$ and α. It turns out that $diff_{Sol} = 0$ and $diff_{Cost} = 0$, indicating that the integrated algorithm performs as well as the exhaustive search, for different parameterizations of the BN.

5.1.4. *System abnormality detection using the sensors allocated by the integrated algorithm*

This study will evaluate the performance of the sensors allocated by the integrated algorithm in detecting different kinds of system abnormalities. The system abnormalities under consideration include mean shifts of magnitude δ in single variables or in multiple variables simultaneously. Specifically, single-variable mean shifts are considered, because they represent weaker system abnormalities than multiple-variable mean shifts and thus are more difficult to detect. In other words, if the allocated sensors perform well in detecting single-variable mean shifts, they should perform better in detecting multiple-variable mean shifts, which are more noticeable.

It is known that the sensors allocated by the integrated algorithm may be different for each combination of δ, $ARL_{1U}(\delta)$ and α. In particular, when $\delta = 3$, $ARL_{1U}(\delta) = 10$, and $\alpha = 0.05$, sensors will be allocated to variables X_1 and X_2; i.e., $L_G = \{1, 2\}$ (see Table 2). In what follows, this particular sensor allocation scheme is evaluated in detecting single-variable mean shifts. Other sensor allocation schemes, corresponding to other combinations of δ, $ARL_{1U}(\delta)$, and α, can be evaluated in a similar manner.

Specifically, the evaluation involves four steps.

1. A mean shift of three standard deviations (i.e., $\delta = 3$) is introduced to a variable X_i, $i \in \{1, \ldots, 5\}$.
2. One dataset of $\mathbf{X}$ with M samples is generated, in which the introduced mean shift occurs at the first sample of X_i. Samples of X_1 and X_2 (i.e., the two variables to which sensors have been allocated) are plotted on Shewhart control charts, respectively, with control limits being $\pm z_{\alpha'/2}$. Here, the Bonferroni method is used to set $\alpha' = \alpha/2 = 0.025$ (i.e., $z_{\alpha'/2} = 2.24$) in order to control Type-I error at the system level.
3. Indices of the first out-of-control samples on the two control charts are recorded as RL_1 and RL_2, respectively. Furthermore, $RL_{min} = \min\{RL_1, RL_2\}$ and $I_{min} = \text{argmin}_{i\in\{1,2\}}\{RL_i\}$ are computed.
4. Steps 1 to 3 are repeated N times. The average RL_{min}, $\overline{RL}_{min}$, is computed; and the mode of I_{min}, mode (I_{min}) is also computed.

Following the above four steps, $\overline{RL}_{min}$ and mode (I_{min}) were obtained for each single-variable mean shift, as shown in Table 3.

It can be seen from Table 3 that the estimated average run length $\overline{RL}_{min}$ for sensors on X_1 and X_2 to detect any single-variable mean shift is below the maximum allowable average run length $ARL_{1U}(\delta) = 10$, indicating that the sensors allocated by the integrated algorithm can successfully detect system abnormalities. Furthermore, the values of mode (I_{min}) imply that the mean shift in a variable is signaled by its descendent at a majority of the times. This makes it possible to trace backward from the signaling sensor to locate the variable that initiates the mean shift and, therefore, facilitate root cause diagnosis after the abnormality detection. Similar results to those in Table 3 can be

Table 3. Estimated average run length (over 5000 simulation runs; i.e., $N = 5000$) and the primary responsible sensor in detecting single-variable mean shifts in the hot forming process

	Mean shift variable				
	X_1	X_2	X_3	X_4	X_5
$\overline{RL}_{min}$	1.34	1.20	5.94	3.36	7.50
mode(I_{min})	1	2	1	1	2

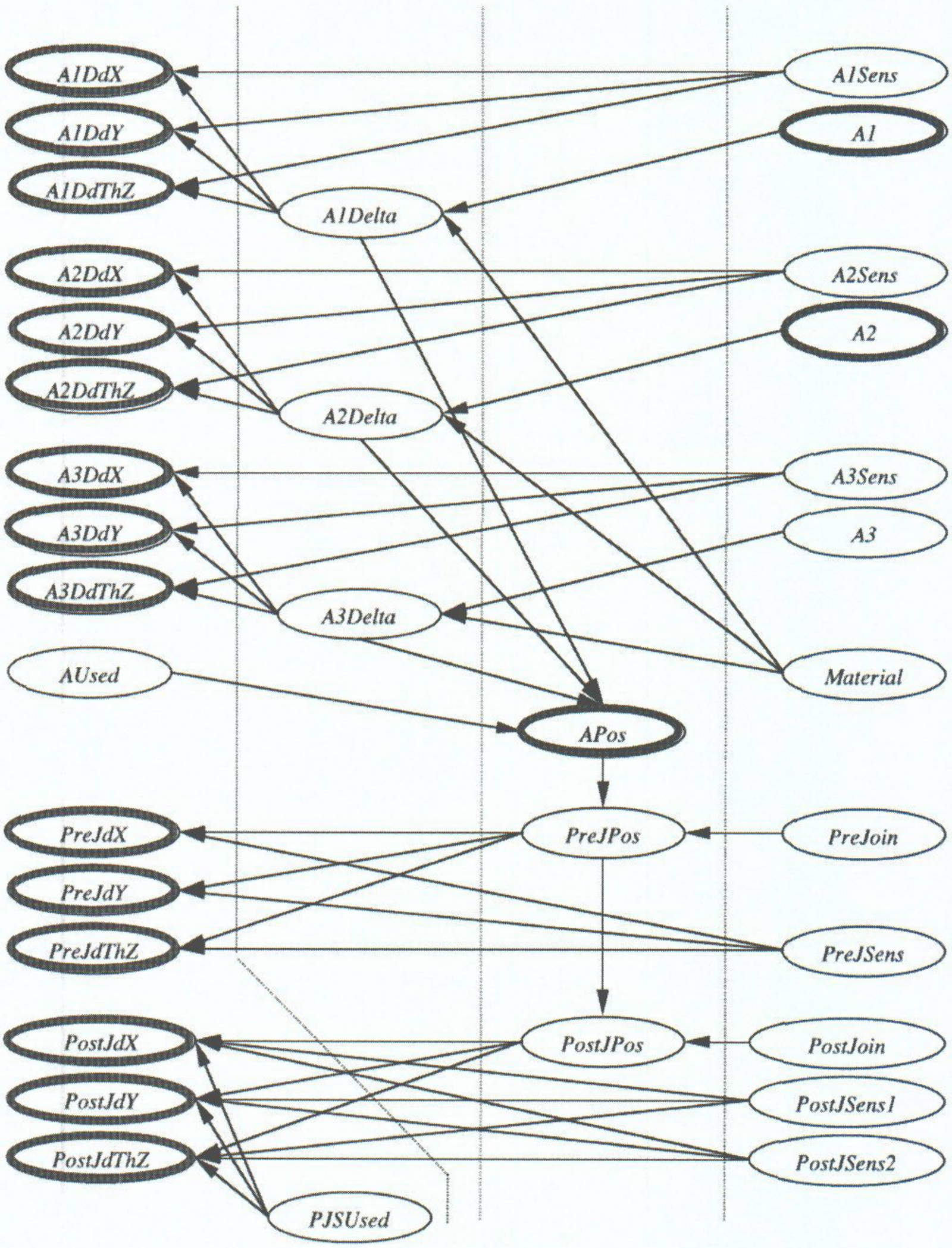

Fig. 9. BN of a cap alignment process (Wolbrecht *et al.*, 2000); highlighted nodes are variables to which sensors are allocated by the integrated algorithm.

observed for other sensor allocation schemes, demonstrating the effectiveness of sensor allocation by the integrated algorithm in the hot forming process.

5.2. *A large-scale manufacturing process*

In large-scale manufacturing processes involving numerous variables, optimal sensor allocation by exhaustive search is impossible. For example, we ran an exhaustive search on BNs of various sizes and found that when BNs include more than 15 variables, the exhaustive search is very likely to fail due to memory overflow errors (Intel Xeon 1.86 GHz dual core processor, 2 GB RAM). Therefore, the greedy algorithm must be applied for sensor allocation in large-scale manufacturing processes.

In this section, we applied the integrated algorithm in Fig. 4, i.e., the pre-processing algorithm followed by the greedy algorithm, to a cap alignment process, a high-speed automated assembly process essential for the production of some precision products. Monitoring and abnormality detection of the cap alignment process are very important because the quality of the precision product depends on the positional accuracy of the cap on the base part. Wolbrecht *et al.* (2000) developed a BN to represent the causal relationships among 35 key variables involved in a cap

Table 4. Estimated average run length (over 5000 simulation runs) and the primary responsible sensor in detecting single-variable mean shifts in the cap alignment process

Mean	A1DdXdX	A1DdY	A1DdThZ	A2DdX	A2DdY	A2DdThZ	A3DdX	A3DdY	A3DdThZ
$\overline{RL}_{\min}$	1.91	1.96	1.99	2.13	1.24	1.8	1.8	1.97	1.91
mode($I_{\min}$)	A1DdXdX	A1DdY	A1DdThZ	A2DdX	A2DdY	A2DdThZ	A3DdX	A3DdY	A3DdThZ
Mean	PreJdX	PreJdY	PreJdThZh	PostJdX	PostJdYYhYh	PostJdThZ	A1Sens	A2Sens	A3Sens
$\overline{RL}_{\min}$	1.89	1.88	1.93	2.16	2.01	2.03	3.26	2.97	2.3
mode($I_{\min}$)	PreJdX	PreJdY	PreJdThZh	PostJdX	PostJdYYhYh	PostJdThZ	A1DdXdX	A2DdThZ	A3DdX
Mean	PreJSens	PostJSens1	PostJSens2	PJSUsed	PostJPos	PostJoin	PreJPos	PreJoin	Apos
$\overline{RL}_{\min}$	2.77	2.49	1.76	3.11	2.4	3.5	1.56	1.96	1.82
mode($I_{\min}$)	PreJdY	PostJdY	PostJdX	PostJdX	PostJdY	PostJdY	PreJdY	PreJdY	Apos
Mean	AUsed	A3Delta	A3	A2Delta	A2	A1Delta	A1	Material	
$\overline{RL}_{\min}$	5.62	3.08	5.9	6.73	1.95	4.25	1.83	7.17	
mode($I_{\min}$)	Apos	A3DdThZ	A3DdThZ	A2DdX Z	A2	A1DdY	A1	A3DdThZ	

alignment process at Hewlett Packard. The BN is shown in Fig. 9 and physical interpretation of the variables can be found in their paper.

Given that $\delta = 3$, $ARL_{1U}(\delta) = 10$, and $\alpha = 0.05$, the integrated algorithm generates a solution according to which sensors are allocated. The variables to which sensors are allocated are denoted by nodes with thick-lined circles in Fig. 9. This particular sensor allocation scheme was evaluated for its ability to detect single-variable mean shifts, and results are shown in Table 4 by following similar steps to steps 1 to 4 in the previous section. It can be seen from Table 4 that the estimated average run length, $\overline{RL}_{\min}$, for the sensors allocated by the integrated algorithm to detect any single-variable mean shift is well below the maximum allowable average run length $ARL_{1U}(\delta) = 10$. Also, the values of mode ($I_{\min}$) imply that the mean shift in a variable is signaled by its decrease in a majority of cases. Therefore, the integrated algorithm facilitates not only abnormality detection but also root cause diagnosis. Similar results to those in Table 4 can be observed for other sensor allocation schemes, demonstrating the effectiveness of sensor allocation by the integrated algorithm in the cap alignment process.

6. Conclusions

Optimal sensor allocation for system abnormality detection is an important research issue in quality engineering. This article proposed to represent the causal relationships among physical variables in the system by using a BN, based on which optimal sensor allocation can be formulated into a set-covering problem. Furthermore, an integrated algorithm, created by combining the BN and a set-covering greedy algorithm, was developed to determine which physical variables should be sensed, in order to minimize the total sensing cost as well as satisfy a prescribed detectability requirement. In addition, a practical procedure was developed to guide decision makers to choose between the greedy algorithm and exhaustive search, for the case where the greedy algorithm cannot find the optimal solution, by considering the trade-off between sensing and computational costs. Finally, two case studies were conducted: one on a hot forming process, showing that the integrated algorithm can find the optimal solution for sensor allocation and the control chart system based on the allocated sensors satisfies the prescribed detectability requirement; the other on a large-scale cap alignment process, showing that only half of the total number of variables need to be sensed by the integrated algorithm and the detectablity requirement is also satisfied.

Note that although the proposed method was intended for abnormality detection, it can also help other important goals in quality control to be better realized, such as root cause identification. For example, assuming that sensors are allocated following the proposed method and that a mean shift is detected by the sensor on X_i, it is immediately known that root cause variables should be included in the upstream variables of X_i and X_i but definitely not others. Then, further inspection or mobile sensing can be conducted only on the upstream variables to identify the root cause variables. This avoids installing sensors on all physical variables at once but enables a new adaptive sensor allocation strategy that advocates allocating mobile sensing efforts progressively and on-demand to fulfill the quality control objectives of detection and root cause identification. This strategy leads to cost reduction and eases massive data communication and analysis in DSNs.

It is also worth mentioning that this article offers only one formulation for the problem of optimal sensor allocation; i.e., to formulate it into an optimization problem that treats detectability as a constraint and the sensing cost as the objective function. Two other potential formulations include: (i) treating the sensing cost as a constraint and detectability as the objective function; and (ii) combining the sensing cost and detectability into the objective function, if a cost function for detectability can be known. All these three formulations are useful, depending on what information is available and what objective is concerned in a specific application. Therefore, while this article focuses on the first formulation, the other two formulations will be thoroughly

investigated in future research. Another future research direction is to conduct sensitivity studies to see how variability around the sharp cut-off for detectability will be translated into the variability in the minimum sensing cost. Furthermore, multivariate control charting techniques may be compared with the currently used single-variable control charts, to see what competitive advantages they may offer.

References

Akyildiz, I.F., Su, W., Sankarasubramaniam, Y. and Cayirci, E. (2002) Wireless sensor networks: a survey. *Computer Networks*, **38**(4), 393–422.

Buntine, W. (1996) A guide to the literature on learning probabilistic networks from data. *IEEE Transactions on Knowledge and Data Engineering*, **8**, 195–210.

Ceglarek, D. and Khan, A. (2000) Sensor optimization for fault diagnosis in multi-fixture assembly systems with distributed sensing. *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **122**(1), 215–226.

Ceglarek, D., Khan, A., Shi, J., Ni, J. and Woo, T.C. (1999) Sensor optimization for fault diagnosis in single fixture systems: a methodology. *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **121**(4), 771–777.

Cormen, T.H., Leiserson, C.E, Rivest, R.L. and Stein, C. (2001) *Introduction to Algorithms*, MIT Press, Cambridge, MA.

Dhillon, S.S. and Chakrabarty, K. (2003) Sensor placement for effective coverage and surveillance in distributed sensor networks. in *Proceedings of the 2003 IEEE Wireless Communications and Networking Conference, Volume 3*, pp. 1609–1614.

Ding, Y., Elsayed, E.A., Kumara, S., Lu, J.C., Niu, F. and Shi, J. (2006) Distributed sensing for quality and productivity improvements. *IEEE Transactions on Automation Science and Engineering*, **3**(4), 344–358.

Ding, Y., Kim, P., Ceglarek, D. and Jin J. (2003) Optimal sensor distribution for variation diagnosis in multistation assembly processes. *IEEE Transactions on Robotics and Automation*, **19**(4), 543–556.

Khan, A. and Ceglarek, D. (1998) Sensor location optimization for fault diagnosis in multi-fixture assembly systems. *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **120**(4), 781–791.

Kline, R.B. (2005) *Principles and Practice of Structural Equation Modeling*. The Guilford Press, New York.

Korb, K.B. and Nicholson, A.E. (2003) *Bayesian Artificial Intelligence*, Chapman & Hall/CRC, London.

Lauritzen, S.L. and Wermuth, N. (1989) Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics*, **17**, 31–57.

Li, J., Jin, J. and Shi, J. (2008) Causation-based T^2 decomposition for multivariate process monitoring and diagnosis. *Journal of Quality Technology*, **40**(1), 46–58.

Liu, C.Q., Ding, Y. and Chen, Y. (2005) Optimal coordinate sensor placements for estimating mean and variance components of variation sources. *IIE Transactions*, **37**, 877–889.

Mhatre, V.P., Rosenberg, C., Kofman, D., Mazumdar, R. and Shroff, N. (2005) A minimum cost heterogeneous sensor network with a lifetime constraint. *IEEE Transactions on Mobile Computing*, **1**, 4–15.

Montgomery, D.C. (2001) *Introduction to Statistical Quality Control*, Wiley, New York.

Shi, J. (2006) *Stream of Variation Modeling and Analysis for Multistage Manufacturing Processes*, CRC Press, Boca Raton, FL.

Spirtes, P., Glymour, C. and Scheines, R. (1993) *Causation, Prediction, and Search*, Springer-Verlag, New York.

Tarabanis, K.A., Allen, P.K. and Tsai, R.Y. (1995) A survey of sensor planning in computer vision. *IEEE Transactions on Robotics and Automation*, **11**(1), 86–104.

Wang, Y. and Nagarkar, S.R. (1999) Locator and sensor placement for automated coordinate checking fixtures. *Transactions of the ASME, Journal of Manufacturing Science and Engineering*, **121**, 709–719.

Wolbrecht, E., D'Ambrosio, B., Paasch, R. and Kirby, D. (2000) Monitoring and diagnosis of a multistage manufacturing process using Bayesian networks. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, **14**(1), 53–67.

Wright, S. (1921) Correlation and causation. *Journal of Agricultural Research*, **20**, 557–585.

Xu, K., Wang, Q., Hassanein, H. and Takahara, G. (2005) Optimal wireless sensor networks (WSNs) deployment: minimum cost with lifetime constraint. *2005 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, Volume 3*, pp. 454–461.

Appendix

Appendix 1: Proof of Proposition 1.

1. Proof of sufficiency.
Because $\bigcup_{j=1}^{q} \mathbf{C}^{\{\Omega_j\}} = \mathbf{X}$, any $X_i \in \mathbf{X}$ must be included in at least one $\mathbf{C}^{\{\Omega_j\}}$, $j \in \{1, \ldots, q\}$. Denoting a $\mathbf{C}^{\{\Omega_j\}}$ that includes X_i by $\mathbf{C}^{\{\Omega_{i^*}\}}$, then the mean shift in X_i is detectable by $\{\Omega_{i^*}\}$ according to Definition 3 of the duty set. Furthermore, because $\Omega_{i^*} \in \mathbf{\Omega}$, the mean shift in X_i is detectable by $\mathbf{\Omega}$. Therefore, according to Definition 2, $\mathbf{\Omega}$ is a feasible solution. ■

2. Proof of necessity.
If the mean shift in any $X_i \in \mathbf{X}$ is detectable by $\mathbf{\Omega}$, then there must exist a $\Omega_{i^*} \in \mathbf{\Omega}$ such that the mean shift in X_i is detectable by $\{\Omega_{i^*}\}$; i.e., $X_i \in \mathbf{C}^{\{\Omega_{i^*}\}}$. Thus,

$$\mathbf{X} = \bigcup_{i=1}^{p} \mathbf{C}^{\{\Omega_{i^*}\}} = \bigcup_{j=1}^{q} \mathbf{C}^{\{\Omega_j\}}.$$

■

Appendix 2: Proofs of Rule 1 and Rule 2

Proof of Rule 1. Let $\mathbf{C}^{\{X_k\}}$ be a duty set that is included in the solution by the greedy algorithm, i.e., $k \in \mathbf{L}_G$, and that includes X_i; i.e., $X_i \in \mathbf{C}^{\{X_k\}}$. Then, $X_j \in \mathbf{C}^{\{X_k\}}$ by the condition in Rule 1. This means that X_j will become detectable as long as X_i is detectable. Thus, X_j can be eliminated from $\mathbf{X}$. As a result, if any of these variables are included in a duty set, they should be eliminated from that duty set. ■

Proof of Rule 2. If X_i belongs to only one duty set $\mathbf{C}^{\{X_h\}}$, then X_h must be sensed in order to detect the mean shift in X_i. In this case, X_h is actually just X_i because X_i must belong to the duty set of itself. Furthermore, because X_i is sensed, the mean shifts in all the variables included in the duty set $\mathbf{C}^{\{X_i\}}$ will become detectable. Therefore, the variables included in $\mathbf{C}^{\{X_i\}}$ can be eliminated from $\mathbf{X}$. ■

Appendix 3: Analysis of the greedy algorithm in optimal sensor allocation.

To analyze the greedy algorithm, the abstract steps of the algorithm given in Fig. 3 have to be expanded into the pseudocode presented in Table A1, with each line i taking a constant amount of run time, t_i, when executed.

Table A1. Analysis of the greedy algorithm in optimal sensor allocation

Pseudocode of greedy algorithm input: $\mathbf{X}, \mathbf{C}^{\{X_k\}}, w(X_k), k = 1, \ldots, p$	*Run time for one execution*	*Times of executions (worst-case scenario)*
$\mathbf{U} \leftarrow \mathbf{X}$	t_1	1
$\mathbf{L}_G \leftarrow \emptyset$	t_2	1
While $\mathbf{U} \neq \emptyset$	t_3	$p+1$
do $\zeta_{\min} \leftarrow$ *infinity*	t_4	p
for $k \leftarrow 1$ **to** p	t_5	$p(p+1)$
$\zeta(k) \leftarrow w(X_k)/\lvert\mathbf{C}^{\{X_k\}} \cap \mathbf{U}\rvert$	t_6	p^2
if $\zeta(k) < \zeta_{\min}$	t_7	p^2
$\zeta_{\min} \leftarrow \zeta(k)$	t_8	p^2
$k_0 \leftarrow k$	t_9	p^2
$\mathbf{U} \leftarrow \mathbf{U} \setminus \left(\mathbf{C}^{\{X_{k_0}\}} \cap \mathbf{U}\right)$	t_{10}	p
$\mathbf{L}_G \leftarrow \mathbf{L}_G \cup \{k_0\}$	t_{11}	p

Therefore, the run time of the greedy algorithm is the sum of run times for each line executed; that is,

$$t_{\text{total}} = t_1 + t_2 + t_3(p+1) + t_4 p + t_5 p(p+1) + (t_6 + t_7 + t_8 + t_9)p^2 + (t_{10} + t_{11})p \quad \text{(A1)}$$

Because the highest order of the polynomial in Equation (A1) is two, the run time complexity of the greedy algorithm is $O(p^2)$.

Appendix 4: Proof of Proposition 2

Let $B = f(A_1 \uparrow, A_2 \downarrow)$ denote that B is an increasing function of A_1 with other arguments (i.e., A_2) held constant and a decreasing function of A_2 with other arguments (i.e., A_1) held constant. Then,

$$\lambda_H = f(H(\max_{k=1,\ldots,p} \lvert\mathbf{C}^{\{X_k\}}\rvert) \uparrow).$$

Because

$$H(\max_{k=1,\ldots,p} \lvert\mathbf{C}^{\{X_k\}}\rvert) = f(\lvert\mathbf{C}^{\{X_k\}}\rvert \uparrow),$$
$$\lambda_H = f(\lvert\mathbf{C}^{\{X_k\}}\rvert \uparrow).$$

According to the definition of $\mathbf{C}^{\{X_k\}}$ in (9), $\lvert\mathbf{C}^{\{X_k\}}\rvert = f(\eta_i \downarrow)$, then $\lambda_H = f(\eta_i \downarrow)$. Furthermore, according to the definition of η_i in Equation (6), $\eta_i = f(\alpha \downarrow, \delta_i \downarrow, ARL_{1U}(\delta_i) \downarrow)$. Thus,

$$\lambda_H = f(\alpha \uparrow, \delta_i \uparrow, ARL_{1U}(\delta_i) \uparrow). \quad \text{(A2)}$$

If the minimum interested mean shift to be detected for each variable can be assumed to be the same, i.e., $\delta_i = \delta$, Equation (A2) can be simplified into

$$\lambda_H = \varphi_{BN,\alpha,ARL_{1U}(\delta)}(\delta). \quad \blacksquare$$

Biographies

Jing Li received her B.S. from Tsinghua University in China and an M.A. in Statistics, and a Ph.D. in Industrial and Operations Engineering from the University of Michigan in 2005 and 2007, respectively. She is currently an Assistant Professor in the Department of Industrial, Systems, and Operations Engineering at Arizona State University. Her research interests are in the general area of applied statistics, quality engineering, and system engineering, with applications in manufacturing and health informatics. She is a member of IIE, INFORMS, and ASQ.

Jionghua (Judy) Jin is an Associate Professor in the Department of Industrial and Operations Engineering at the University of Michigan. She received her Ph.D. degree from the University of Michigan in 1999. Her recent research focuses on data fusion for complex system improvement with the goal of developing novel methodologies for variation reduction, condition monitoring and fault diagnosis, process control, knowledge discovery, and decision making. Her research emphasizes a multidisciplinary approach by integrating applied statistics, signal processing, reliability engineering, system control, and decision making theory. She has received a number of awards including the NSF CAREER Award in 2002, the PECASE Award in 2004, and six Best Paper Awards during 2005–2009. She is a member of ASME, ASQC, IEEE, IIE, INFORMS, and SME.