A Novel Qualitative Maximum a Posteriori Estimation for Bayesian Network Parameters Based on Computing the Center Point of Constrained Parameter Regions

RUOHAI DI\textsuperscript{1}, PENG WANG\textsuperscript{1}, JIAO WU\textsuperscript{1}, AND ZHIGAO GUO\textsuperscript{2}

\textsuperscript{1}School of Electronics and Information Engineering, Xi'an Technological University, Xi'an 710021, China
\textsuperscript{2}School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K.

Corresponding author: Zhigao Guo (guozhigao2004@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61573285, in part by the State Key Laboratory Fund under Grant CEMEE2020Z0202B, and in part by the Basic Research Plan of Natural Science in Shaanxi Province under Grant 2020Q-816 and Grant 2020KJRC0033.

\textbf{ABSTRACT} Introducing parameter constraints has become a mainstream approach for learning Bayesian network parameters with small datasets. The QMAP (Qualitative Maximum a Posteriori) estimation has produced the best learning accuracy among existing learning approaches. However, the rejection-acceptance sampling employed in the QMAP algorithm for determining average BN (Bayesian Network) parameter values is time-consuming, particularly when the number of parameter constraints is large. This paper proposes a new analytical approach that enhances the learning efficiency of the QMAP algorithm without reducing its learning accuracy by treating the average value of the parameters as the center point of the constrained parameter region, which is a much more efficient method than the rejection-acceptance sampling method employed in the traditional QMAP algorithm. First, a novel objective function is designed and a constrained objective optimization model is constructed based on parameter constraints. Second, the constructed model is employed to obtain the center point of the constrained parameter region based on its boundary points, and the average parameter value is the average of all boundary points. The results obtained from a large number of simulation experiments with four benchmark Bayesian networks demonstrate that the parameter learning accuracy of the proposed algorithm is slightly better than that of the original QMAP algorithm under specific conditions, and the computational efficiency is substantially increased under all conditions.

\textbf{INDEX TERMS} Bayesian networks, small datasets, qualitative maximum a posteriori, constrained regions.

\section{I. INTRODUCTION}

A Bayesian network (BN) is a statistical model that represents a group of variables as nodes in a directed acyclic graph, and denotes the conditional dependencies between variables as the graph edges. As such, a BN contains a directed acyclic graph structure and network parameters reflecting the conditional dependency distributions between variables. The BN concept was first proposed by Professor Judea Pearl of UCLA in 1986 [1]. A BN can adopt multi-informational variables, and can infer probable outcomes via data fusion and bi-directional reasoning. Moreover, a BN can colligate prior information with current information to increase the accuracy and believability of inferences. As such, BNs have become powerful tools for addressing statistical problems, and they have been wildly applied in many fields, such as geological hazard prediction [2], gene analysis [3], medical diagnosis [4]–[6], reliability analysis [7], pattern classification [8], fault diagnosis [9], language recognition [10], [11], and action recognition [12].

The construction of an accurate BN model is the core task in BN-related research. Similar to the neural networks [13], [14], the learning of BN model parameters based on existing data is very important. This parameter learning task is generally easy when sufficient data samples are available. However, sufficient data samples can be difficult to obtain due to many reasons such as measurement equipment failure or a natural scarcity of data due to rare conditions, such as...
when the statistical problem involves rare geological hazards or rare diseases, or when the data acquisition cost is quite expensive. This condition can severely limit the feasibility of BNs. Therefore, numerous BN modeling methods based on small datasets have been developed, and the introduction of parameter constraints has become a mainstream approach applied to address the condition. The expert knowledge employed in BN parameter learning can be expressed by the following eight types of parameter constraints [32]: axiomatic constraints, range constraints, intra-distribution constraints, cross-distribution constraints, inter-distribution constraints, approximate-equality constraints, additive synergy constraints, and product synergy constraints. These constraints are defined in detail in Subsection 2.4. The methods developed for BN parameter learning with small datasets can be divided into two types: parameter learning methods with a single type of constraint and parameter learning methods with multiple types of constraints.

Among the single constraint type methods, Wittig and Jameson [15] proposed a constrained BN parameter learning algorithm that employed cross-distribution parameter constraints. This algorithm first employs the cross-distribution parameter constraints to build a penalty function. Second, the penalty function is combined with an entropy function to obtain an optimization model. Finally, the optimization model is solved using the adaptive probability networks method. Feelders and Gaag [16], [17] regarded parameter learning as an isotonic regression problem. This problem first obtains the initial BN parameters via the maximum likelihood (ML) estimation. Then, the parameters are regulated by the minimum lower sets algorithm to ensure that the regulated parameters satisfy the cross-distribution constraints. Altendorf et al. [18] also suggested a constrained BN parameter learning algorithm employing cross-distribution constraints. Here, an objective function is defined by integrating the cross-distribution constraints into an entropy function that is solved using the gradient-descent algorithm. Zhou et al. [19] proposed a constrained BN parameter learning algorithm called constrained optimization with a flat prior. The method improves the accuracy of the BN parameters obtained via the ML estimation using a flat prior and cross-distribution constraints, and applies sequential quadratic programming to solve the improved ML function. Di et al. [20], [21] proposed a constrained BN parameter learning algorithm that employed intra-distribution constraints. Here, the intra-distribution constraints are translated into the hyper-parameters of a Dirichlet distribution or a beta distribution, and a maximum a posteriori algorithm is applied to obtain the BN parameters. Ren et al. [22] proposed a constrained BN parameter learning algorithm that employed range constraints. This algorithm converts the range constraints into the hyper-parameters of a beta distribution, and applies a maximum a posteriori algorithm to obtain the BN parameters.

While reasonable BN parameter learning has been obtained using single-type constraints, these constraints can only represent a single type of expert knowledge. However, many different types of expert knowledge can be used in the BN parameter learning process. As a result, multiple constraint type methods have gained increasing attention in the last few years. For example, Niculescu et al. [23] and Campos and Qiang [24] proposed BN parameter learning algorithms based on convex optimization, in which the ML function is treated as the objective function and multiple types of parameter constraints are regarded as the feasible solution region. Liao and Qiang [25] introduced multiple types of parameter constraints into the BN parameter learning process under the condition of missing data using the expectation-maximization algorithm. Rui et al. [26] proposed the qualitative maximum a posteriori (QMAP) algorithm for BN parameter learning. The QMAP algorithm is an approximate Bayesian learning method that incorporates both quantitative and qualitative constraints. This approach requires that a sufficient number of parameters be sampled from within the constrained parameter region using a rejection-acceptance sampling strategy to obtain the mean values of the parameters. In addition, the hyper-parameters of the Dirichlet distribution are determined as the products of an equivalent sample size and the mean values of the sampled parameters. The optimal parameters are then computed as interpolations of the sample observations and the obtained hyper-parameters. Guo et al. [27] found that the BN parameters learned by the QMAP algorithm could violate the parameter constraints, and accordingly proposed the further constrained QMAP algorithm. The algorithm regulates the QMAP estimation by replacing data estimation with a further constrained estimation conducted via convex optimization. Zhou et al. [28], [29] proposed the multinomial parameter learning with constraints algorithm. First, the frequency of the configuration states of particular child and parent nodes is counted. Second, an auxiliary BN model is built by integrating both the sample data and the parameter constraints. Finally, the optimal parameters are computed as the mean values of the probability distribution. Guo et al. [30] proposed a dual constrained estimation method employing both intra-distribution parameter constraints and cross-distribution constraints. Compared with existing BN parameter learning algorithms, the above-mentioned QMAP algorithms have been demonstrated to provide the best BN parameter learning accuracy with small datasets [30]. However, the rejection-acceptance sampling strategy used in these QMAP algorithms is time-consuming, particularly when the number of constraints is large, which greatly decreases the computational efficiency of the algorithms. This paper addresses this issue by proposing a new analytical approach that enhances the learning efficiency of the QMAP algorithm without reducing its learning accuracy. The proposed approach treats the average value of the parameters as the center point of the constrained parameter region, which is a much more efficient method than the rejection-acceptance sampling method employed in the QMAP algorithm. First, a novel objective function is designed and a constrained objective optimization model is constructed based on parameter constraints.
Second, the constructed model is employed to obtain the center point of the constrained parameter region based on its boundary points, and the average parameter value is determined as the average of the boundary points. The results obtained for a large number of simulation experiments with four benchmark networks demonstrate that the parameter learning accuracy of the proposed algorithm is slightly better than that of the original QMAP algorithm under specific conditions, and the computational efficiency is substantially increased under all conditions.

II. PRELIMINARIES

A. BAYESIAN NETWORKS
For each node \(X_i\), where \(i = 1, 2, \ldots, n\), of a BN composed of \(n\) nodes, we denote the parents of \(X_i\) as \(\pi(X_i)\). A parameter represents the conditional probability \(P(X_i|\pi(X_i))\) reflecting the conditional dependencies between \(\pi(X_i)\) and \(\pi(X_i)\). If the conditional distributions are specified for all \(n\) nodes of the BN, the joint probability can be calculated as follows:

\[
P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\pi(X_i))
\]

Decomposing the joint probability according to (1) provides not only a compact representation of the joint probability but also provides an efficient implementation of reasoning. Node \(X_i\) could represent a discrete variable or a continuous variable. This paper considers only discrete variables.

B. PARAMETER LEARNING IN A BAYESIAN NETWORK
The discussion in the above subsection indicates that the BN parameter learning process involves estimating conditional probability distributions from available data pertinent to the problem to which the BN is applied. Two classical methods are typically applied for BN parameter learning, which include ML estimation and Bayesian estimation. This parameter estimation process can be presented in terms of the ML estimation algorithm as an example. Here, when the dataset is complete (i.e., it has no missing data), parameter estimation can be expressed as a maximization of the log-likelihood function \(\ell(\theta|D)\), as follows:

\[
\ell(\theta|D) = \sum_{i} \sum_{j} q_i \sum_{k} r_i \sum_{k} N_{ijk} \log(\theta_{ijk})
\]

where \(q_i\) is the number of parent node states, \(r_i\) is the number of child node states, \(N_{ijk}\) is the number of nodal configurations that satisfy \(X_i = k\) and \(\pi(X_i) = j\), and \(\theta_{ijk}\) is the maximum likelihood estimation of the BN parameter for the \(i\)th node when it is in the \(k\)-th state and its parent nodes are in the \(j\)th state. \(\theta_{ijk}\) is computed as follows:

\[
\theta_{ijk} = \frac{N_{ijk}}{N_j}
\]

where \(N_j\) is the number of nodal configurations of the \(i\)th node that satisfy \(\pi(X_i) = j\). The estimation of network parameters can be decomposed into the product of independent estimations for individual nodes using the decomposability property of a BN illustrated in (1).

C. SAMPLE COMPLEXITY OF BN PARAMETER LEARNING
The ML estimation algorithm is computationally efficient, and is therefore ideal for accurately learning BN parameters if the dataset is sufficiently large. Unfortunately, datasets are often small, which substantially decreases the accuracy of the ML estimation algorithm. However, determining the dataset size that can ensure the accurate estimation of BN parameters using the ML estimation algorithm is a problem. In this regard, Dasgupta [31] provided a lower dataset size limit for a known BN structure. For a BN composed of \(n\) Boolean nodes, where no node has more than \(k\) parents, the lower dataset size limit is calculated with confidence \((1 - \delta)\) as follows:

\[
\text{Number} \geq \frac{288 \times n^2 \times 2^k}{\varepsilon^2} \times \ln^2(1 + \frac{3n}{\varepsilon}) \times \ln\left(\frac{1 + 3n/\varepsilon}{\varepsilon \delta}\right)
\]

where \(\varepsilon\) is the error rate, which is often computed as \(\varepsilon = n\sigma\), for a small arbitrary constant \(\sigma\).

D. COMMON PARAMETER CONSTRAINTS
The eight common types of parameter constraints can be formally defined as follows.

(1) Axiomatic constraints

\[
\sum_{k=1}^{r_i} \theta_{ijk} = 1, \quad 0 \leq \theta_{ijk} \leq 1, \quad \forall i, j, k
\]

These constraints require that the sum of the BN parameters obtained from the same distribution is 1, which is a basic law from probability theory.

(2) Range constraints

\[
0 \leq \alpha_{ijk} \leq \theta_{ijk} \leq \beta_{ijk} \leq 1
\]

Here, \(\alpha_{ijk}\) is a lower bound and \(\beta_{ijk}\) is an upper bound. These constraints restrict BN parameters to specified intervals.

(3) Constraints

\[
\theta_{ijk} \leq \theta_{ij'k}, \quad \forall k \neq k'
\]

As discussed above, these constraints describe relationships between parameters obtained from the same distribution, where the parameters have the same parent node configuration state \(j\) and different child node states \(k\) and \(k'\).

Cross-distribution (4) constraints

\[
\theta_{ijk} \leq \theta_{ij'k}, \quad \forall j \neq j'
\]

As discussed above, these constraints describe relationships between two parameters with the same child node state \(k\) and different parent node configuration states \(j\) and \(j'\).

Inter-distribution (5) constraints

\[
\theta_{ijk} \leq \theta_{ij'k'}, \quad \forall i \neq i', j \neq j', k \neq k'
\]
These constraints describe relationships between the parameters of different nodes.

Approximate-equality constraints

\[ |\theta_{ijk} - \theta_{ij'k'}| \leq \varepsilon, \quad \forall i \neq i', \ j \neq j', \ k \neq k' \]  

(10)

These constraints define close relationships between any two parameters.

(7) Additive synergy constraints

\[ \theta_{ijk} + \theta_{ij'k} \leq \theta_{ij'k} + \theta_{ijk}, \quad \forall i, k \]  

(11)

These constraints describe the comparative relationships between the sums of two parameters with different parent node configuration states.

(8) Product synergy constraints

\[ \theta_{ijk} \theta_{ij'k} \leq \theta_{ij'k} \theta_{ijk}, \quad \forall i, k \]  

(12)

These constraints describe the comparative relationships between the products of two parameters with different parent node configuration states.

E. HOW TO OBTAIN THE PARAMETER CONSTRAINTS

The parameter constraints come from expert knowledge or domain knowledge. For example, smokers are more likely to get cancer than nonsmokers. If \( C = 1 \) means that a person has cancer, \( C = 0 \) means that the person does not have cancer. If \( S = 1 \) means that a person smokes, \( S = 0 \) means that the person does not smoke. This domain knowledge can be expressed as \( P(C = 1 | S = 1) > P(C = 1 | S = 0) \), which can be treated as a parameter constrain for Bayesian network parameter learning. There are many such examples in different fields. For example, Europeans are more likely to be white, and smokers who have the novel coronavirus pneumonia have a greater probability of the virus developing into more severe diseases.

III. CENTER POINT OF A CONSTRAINED PARAMETER REGION

A. ANALYSIS OF LOW-DIMENSIONAL CONSTRAINED PARAMETER REGIONS

We visualize two-dimensional (2D) and three-dimensional (3D) constrained parameter regions to analyze the functionality of constrained parameter regions in the BN parameter learning process.

For the 2D constrained parameter condition, it is assumed that the BN parameters \( \theta_i (i = 1, 2) \) satisfy the following constraints.

\[
\Omega = \begin{cases} 
\theta_1 + \theta_2 = 1 \\
\theta_1 > \theta_2 \\
0 \leq \theta_1 \leq 1 \\
0 \leq \theta_2 \leq 1 
\end{cases} 
\]  

(13)

The constraints in (13) can be illustrated as shown in Figure 1, where the line segment \( AC \) is the constrained region.

To determine the center point of line segment \( AC \), the coordinates of points \( A \) and \( C \) must first be obtained, and the center point is established as the average value of the two end points. We note from Figure 1 that the actual task of obtaining the coordinates of points \( A \) and \( C \) is to acquire the boundary points of the constrained region, which is a complex problem, particularly for high-dimensional constrained parameter regions. We address this problem by modeling the acquisition of the boundary points of a constrained region as a constrained optimization problem. Under the 2D constrained parameter condition, the constrained optimization models to obtain the coordinates of points \( A \) and \( C \) are respectively given as follows.

\[
\min \left[ (\theta_1 - 0)^2 + (\theta_2 - 1)^2 \right] \\
\text{s.t.} \quad \begin{cases} 
\theta_1 + \theta_2 = 1 \\
\theta_1 > \theta_2 \\
0 \leq \theta_1 \leq 1 \\
0 \leq \theta_2 \leq 1 
\end{cases} 
\]  

(14)

\[
\min \left[ (\theta_1 - 1)^2 + (\theta_2 - 0)^2 \right] \\
\text{s.t.} \quad \begin{cases} 
\theta_1 + \theta_2 = 1 \\
\theta_1 > \theta_2 \\
0 \leq \theta_1 \leq 1 \\
0 \leq \theta_2 \leq 1 
\end{cases} 
\]  

(15)

The basic process treats the boundary points as the nearest points to the vertices of the constrained parameter region. As shown in (14) and (15), the vertices are the points (0,1) and (1,0).

Similarly, for the 3D constrained parameter condition, it is assumed that the BN parameters \( \theta_i (i = 1, 2, 3) \) satisfy the following constraints.

\[
\Omega = \begin{cases} 
\theta_1 + \theta_2 + \theta_3 = 1 \\
\theta_1 > \theta_2 \\
\theta_1 < \theta_3 \\
0 \leq \theta_1 \leq 1 \\
0 \leq \theta_2 \leq 1 \\
0 \leq \theta_3 \leq 1 
\end{cases} 
\]  

(16)
where the triangle $CDE$ is the constrained region.

The constraints in (16) can be illustrated as shown in Figure 2, where the triangle $CDE$ is the constrained region.

The center point of the triangle $CDE$ is obtained by acquiring the coordinates of points $C$, $D$, and $E$. As was conducted for the 2D case, the coordinates of points $C$, $D$, and $E$ can be respectively obtained as follows.

$$
\min \left[ (\theta_1 - 0)^2 + (\theta_2 - 2)^2 + (\theta_3 - 3)^2 \right]
\begin{align*}
\theta_1 + \theta_2 + \theta_3 &= 1 \\
\theta_1 > \theta_2 \\
\theta_1 < \theta_3 \\
0 \leq \theta_1 \leq 1 \\
0 \leq \theta_2 \leq 1 \\
0 \leq \theta_3 \leq 1
\end{align*}
$$

(17)

Then, the center point is established as the average value of the three boundary points.

**B. APPROXIMATE CENTER POINT CALCULATION FOR A HIGH-DIMENSIONAL CONSTRAINED PARAMETER REGIONS**

The method proposed in Subsection 3.1 for acquiring the boundary points of a constrained region are now employed to obtain an approximate calculation of the center point of a high-dimensional constrained parameter region. The method is illustrated using the simple BN in Figure 3 with BN parameters $i, \theta_j$ ($i = 1, 2, 3, 4$). The corresponding parameter constraints $D$ are given as follows.

$$
D = \begin{cases} 
\theta_1 + \theta_2 = 1 \\
\theta_3 + \theta_4 = 1 \\
\theta_1 > \theta_2 \\
\theta_3 < \theta_4 \\
\theta_1 > \theta_3 \\
0 \leq \theta_1 \leq 1 \\
0 \leq \theta_2 \leq 1 \\
0 \leq \theta_3 \leq 1 \\
0 \leq \theta_4 \leq 1 \\
a_2 \leq \theta_1 \leq b_2 \\
a_3 \leq \theta_3 \leq b_3
\end{cases} 
$$

(20)

The main task is constructing the objective function to obtain the boundary points of the constrained parameter region for the BN in Figure 3. Two methods are proposed in this paper. The first method constructs the objective function with four-dimensional (4D) coordinates based on standard basis vectors, such as $[1 0 0 0]^T$, $[0 1 0 0]^T$, $[0 0 1 0]^T$, and $[0 0 0 1]^T$. The other method constructs the objective function with 4D coordinates based on improved vectors that introduce the axiomatic constraint. Examples of such vectors can be given as $[0 1 0 1]^T$, $[0 1 1 0]^T$, $[1 0 0 1]^T$, and $[1 0 1 0]^T$. The objective functions obtained using these two methods are given in detail as follows.

1. Objective function constructed with standard basis vectors $[1 0 0 0]^T$, $[0 1 0 0]^T$, $[0 0 1 0]^T$, and $[0 0 0 1]^T$

$$
\begin{align*}
\theta_1, \theta_2, \theta_3, \theta_4 &= \arg \min_{\theta_j \in D} (\theta_1 - 1)^2 + (\theta_2 - 0)^2 \\
&+ (\theta_3 - 0)^2 + (\theta_4 - 0)^2 \\
\theta_1, \theta_2, \theta_3, \theta_4 &= \arg \min_{\theta_j \in D} (\theta_1 - 0)^2 + (\theta_2 - 1)^2 \\
&+ (\theta_3 - 0)^2 + (\theta_4 - 0)^2 \\
\theta_1, \theta_2, \theta_3, \theta_4 &= \arg \min_{\theta_j \in D} (\theta_1 - 0)^2 + (\theta_2 - 0)^2 \\
&+ (\theta_3 - 1)^2 + (\theta_4 - 0)^2 \\
\theta_1, \theta_2, \theta_3, \theta_4 &= \arg \min_{\theta_j \in D} (\theta_1 - 0)^2 + (\theta_2 - 0)^2 \\
&+ (\theta_3 - 0)^2 + (\theta_4 - 1)^2
\end{align*}
$$

(21)–(24)

After obtaining the boundary points of the constrained parameter region using (21)–(24), the center point is defined as the average value of the four boundary points.
(2) Objective function constructed with improved vectors considering the axiomatic constraint

\[ \theta_1, \theta_2, \theta_3, \theta_4 = \arg \min_{\theta_i \in D} [(\theta_i - 1)^2 + (\theta_i - 1)^2] \]

Again, the center point is defined as the average value of the four boundary points obtained from (25)–(28). It is noteworthy that, if the cross-distribution constraint does not exist, (35) can be converted to the following form.

\[ \theta_{\text{rang}} = \arg \min_{\theta} (\theta - e_i)(\theta - e_i)^T \]

\[ \Omega = \begin{bmatrix} \theta_1 > \theta_2 \\ 0 \leq \theta_1 \leq 1 \\ 0 \leq \theta_2 \leq 1 \\ a_1 \leq \theta_1k_1 \leq b_1 \\ a_2 \leq \theta_2k_2 \leq b_2 \\ \vdots \end{bmatrix} \]

For the method employing improved vectors considering the axiomatic constraint, we first obtain the improved vectors, as illustrated in Figure 4. Here, we define a set of fundamental vectors \( f_1, f_2, \ldots, f_m, f_{m+1}, \ldots, f_{2m}, \ldots, f_{(q-1)m-1}, \ldots, f_{qm} \)

\[ f_1 = \left( \begin{array}{c} 1 \\ \vdots \\ f_m \\ \vdots \\ f_{2m} \\ \vdots \\ f_{(q-1)m-1} \\ \vdots \\ f_{qm} \end{array} \right), \quad f_{2m} = \left( \begin{array}{c} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{array} \right) \]

where \( M_1, M_2, \ldots, M_q \).

The obtained improved vectors are then employed to construct the following objective function.

We note from Figure 4 that \( m^q \) improved vectors are employed to construct the objective function. As such, \( m^q \)

\[ \theta_{\text{rang}} = \arg \min_{\theta \in \Omega} (\theta - T_1)(\theta - T_1)^T \]

We note from Figure 4 that \( m^q \) improved vectors are employed to construct the objective function. As such, \( m^q \)

\[ \theta_{\text{rang}} = \arg \min_{\theta \in \Omega} (\theta - T_1)(\theta - T_1)^T \]

We note from Figure 4 that \( m^q \) improved vectors are employed to construct the objective function. As such, \( m^q \)

\[ \theta_{\text{rang}} = \arg \min_{\theta \in \Omega} (\theta - T_1)(\theta - T_1)^T \]

We note from Figure 4 that \( m^q \) improved vectors are employed to construct the objective function. As such, \( m^q \)

\[ \theta_{\text{rang}} = \arg \min_{\theta \in \Omega} (\theta - T_1)(\theta - T_1)^T \]
In general, the parameter learning process of the CMAP algorithm is similar to that of the QMAP algorithm. First, the log-form score function of the QMAP algorithm is decomposed into a prior likelihood \(P(\theta | G, G)\), where \(G\) denotes the known structure of the BN. The data likelihood \(P(D | \theta, G)\) is calculated as follows:

\[
\log P(\theta | G, D, \Omega) = \log P(D | \theta, G) + \log P(\theta | G, D) - c \tag{39}
\]

where \(c = P(D | \Omega, G)\). The prior likelihood is defined according to the parameter constraints. When sampling prior instances are obtained from the parameter constraints using the rejection-acceptance sampling method, the log prior likelihood is expressed as

\[
\log P(\theta | \Omega, G) = \sum_{i} \sum_{j} \sum_{k} M_{ijk} \log \theta_{ijk} \tag{40}
\]

Then, combining (40) and (41) in (39) yields the following.

\[
\log P(D | \theta, G) = \sum_{i} \sum_{j} \sum_{k} N_{ijk} \log \theta_{ijk} \tag{41}
\]

Finally, the maximum estimation of the QMAP score function is obtained as

\[
\log \theta_{ijk} = \frac{N_{ijk} + M_{ijk}}{\sum_{k} (N_{ijk} + M_{ijk})} \tag{43}
\]

Here \(M_{ijk} = A \cdot P(X_i = k, \pi(X_i) = j | \Omega)\), where the coefficient \(A\) can be determined from expert knowledge or via the cross-validation method. \(P(X_i = k, \pi(X_i) = j | \Omega)\) is the average value of the parameters sampled from the parameter constraints using the rejection-acceptance sampling strategy, which is computed as follows:

\[
P(X_i = k, \pi(X_i) = j | \Omega) = \frac{\sum_{i=1}^{S} P_i(X_i = k, \pi(X_i) = j | \Omega)}{S} \tag{44}
\]

where \(S\) is the number of sampled parameters that satisfy the parameter constraints. \(P(X_i = k, \pi(X_i) = j | \Omega)\) can also be treated as the center point of the parameter constraints. Therefore, the CMAP-I and CMAP-II algorithms compute \(P(X_i = k, \pi(X_i) = j | \Omega)\) using the proposed methods presented in Subsection 3.2.

The process flow of the CMAP algorithm can be summarized as follows.

**Step 1:** Count the numbers of observations \(N_{ijk}\) and \(N_{ij}\) from the dataset.

**Step 2:** Construct the objective function to obtain the boundary points of the constrained parameter region, and
TABLE 1. Details of the Bayesian network in the experiments.

| Size       | Nodes | Edges | Parameters |
|------------|-------|-------|------------|
| small      | 8     | 8     | 18         |
| medium     | 8     | 46    | 509        |
| large      | 8     | 46    | 574        |
| very large | 8     | 46    | 1157       |

obtain the center point of the region from the boundary points using the method presented in Subsection 3.2.

Step 3: Compute the maximum a posteriori (MAP) estimation using (43) with the best value of \( A \) obtained using the cross-validation method, and adopt the center point of the parameter constraints as the average parameter value.

V. EXPERIMENTS

In the following experiments, we evaluate the performances of eight BN learning algorithms for four typical benchmark BNs varying in size from small to medium, large, and very large. The BN learning algorithms include the ML algorithm, the constrained ML (CML) algorithm, the maximum entropy (ME) algorithm, the constrained ME (CME) algorithm, the MAP algorithm, the QMAP algorithm, and the proposed CMAP-I and CMAP-II algorithms. It is noted that the flat prior was used in the experiments for the MAP algorithm, and the best value of the coefficient \( A \) was initially set between 1 and 20 and determined by the cross-validation method for the QMAP and CMAP algorithms. The BNs used in the experiments are summarized in Table 1. The learning accuracy and learning efficiency performances of the different algorithms are evaluated. The learning accuracy is evaluated using the Kullback-Leibler (KL) divergence, which expresses the divergence between the true parameters and the learned parameters. Therefore, the learning accuracy increases as the KL divergence decreases. The learning efficiency is evaluated according to the required computation time.

A. EXPERIMENTAL SETTINGS

The experimental datasets were composed of from 20 to 200 entries in increments of 20 to determine the effect of the dataset size on the performance of the eight algorithms. The parameter constraints were generated from the true parameters of the BNs, and the maximum number of constraints for each node was 30. We employed two experimental groups, denoted as constraint groups A and B. Constraint group A conducted BN parameter learning using a set of parameter constraints that included axiomatic constraints, range constraints, intra-distribution constraints, and cross-distribution constraints while constraint group B employed all the same types of parameter constraints except for the cross-distribution constraints.

The average KL divergence values and computation times obtained by the eight algorithms for the four BNs under different dataset sizes with constraint group A are listed in Tables 2 and 3, respectively, where the best results are highlighted in bold.

B. LEARNING ACCURACY ANALYSIS

We note from Table 2 that the learning accuracy of all algorithms increased by different degrees for all BNs as the dataset size increased. In most cases, the QMAP algorithm provided the highest learning accuracy of all algorithms considered. However, the KL divergence values of the CMAP-I and CMAP-II algorithms were less than twice the magnitude of those of the QMAP algorithm, but they were less than the KL divergence values of the other algorithms. We also

| Dataset size | ML | ME | MAP | CML | CME | QMAP | CMAP-I | CMAP-II |
|--------------|----|----|-----|-----|-----|------|--------|---------|
| Asia         | 40 | 15.52±3.50 | 3.90±0.04 | 3.36±0.18 | 3.60±0.57 | 1.82±0.07 | 0.82±0.14 | 1.59±0.04 | 0.81±0.02 |
| 80           | 14.35±3.64 | 3.73±0.06 | 2.62±0.44 | 2.76±0.34 | 1.52±0.02 | 0.60±0.06 | 1.16±0.07 | 0.61±0.04 |
| 120          | 10.17±2.93 | 2.81±0.24 | 1.98±0.11 | 1.76±0.35 | 1.15±0.13 | 0.45±0.04 | 0.69±0.09 | 0.45±0.04 |
| 160          | 4.86±2.94 | 2.59±0.28 | 1.56±0.45 | 1.36±0.18 | 1.04±0.07 | 0.42±0.04 | 0.84±0.04 | 0.50±0.02 |
| 200          | 2.56±1.58 | 2.28±0.41 | 1.05±0.54 | 0.79±0.18 | 0.84±0.09 | 0.36±0.06 | 0.71±0.04 | 0.38±0.04 |
| Alarm        | 40 | 273.94±22.9 | 139.84±2.8 | 118.75±7.3 | 95.88±6.34 | 60.46±1.72 | 33.78±1.17 | 58.14±1.87 | 50.69±1.46 |
| 80           | 275.49±25.36 | 124.45±1.75 | 95.27±3.90 | 102.91±5.69 | 52.45±1.48 | 29.78±0.75 | 49.39±1.59 | 43.75±1.03 |
| 120          | 262.14±22.68 | 111.18±5.14 | 89.46±5.95 | 96.93±5.30 | 64.78±2.90 | 27.55±1.19 | 45.77±3.07 | 40.89±2.18 |
| 160          | 243.75±19.58 | 106.12±5.25 | 85.46±3.50 | 97.37±4.52 | 45.97±2.84 | 27.00±1.20 | 43.36±1.10 | 39.34±1.84 |
| 200          | 229.07±23.50 | 101.64±3.93 | 80.39±5.20 | 94.01±6.24 | 43.96±2.09 | 25.67±1.66 | 42.07±1.80 | 37.53±1.58 |
| Win59Pts     | 40 | 269.71±22.86 | 152.27±1.38 | 134.34±2.66 | 136.24±6.01 | 96.66±0.74 | 57.98±1.51 | 89.79±1.64 | 78.75±1.60 |
| 80           | 283.48±41.42 | 142.88±1.39 | 124.96±5.31 | 127.38±5.00 | 92.80±1.11 | 55.06±0.66 | 82.96±1.03 | 74.58±1.20 |
| 120          | 308.45±42.10 | 136.53±1.45 | 121.07±6.28 | 118.95±7.32 | 83.52±0.88 | 53.44±0.36 | 79.86±0.72 | 72.52±0.11 |
| 160          | 305.57±15.75 | 131.81±1.66 | 114.93±3.14 | 98.69±6.61 | 82.41±0.71 | 51.36±0.19 | 76.86±0.57 | 76.08±0.65 |
| 200          | 304.90±36.19 | 127.80±1.49 | 114.46±2.38 | 98.19±2.66 | 80.01±0.12 | 50.33±0.60 | 74.99±0.60 | 68.88±1.76 |
| Andes        | 40 | 1252.65±37.19 | 271.36±1.96 | 228.02±6.60 | 744.61±30.50 | 119.50±1.62 | 43.24±0.19 | 121.32±3.84 | 90.50±3.82 |
| 80           | 1046.26±44.69 | 215.98±1.05 | 187.57±6.33 | 636.29±33.50 | 93.18±1.02 | 36.40±0.33 | 92.15±1.75 | 73.70±2.55 |
| 120          | 892.49±31.50 | 185.69±1.70 | 160.63±5.56 | 566.77±15.83 | 78.12±1.53 | 31.11±0.38 | 75.45±4.80 | 60.72±2.16 |
| 160          | 868.22±48.74 | 165.74±2.35 | 140.11±5.37 | 542.73±2.77 | 68.62±1.30 | 28.80±0.38 | 67.43±3.73 | 55.05±4.10 |
| 200          | 780.94±52.41 | 150.57±1.53 | 135.50±7.81 | 515.63±7.81 | 61.99±1.67 | 26.44±0.36 | 60.84±1.17 | 49.82±0.73 |
TABLE 3. Computation times of the different BN learning algorithms for the four BNs under different dataset sizes with constraint group A.

| Dataset size | ML          | ME          | MAP          | CML          | CME          | QMAP         | CMAP-I       | CMAP-II       |
|--------------|-------------|-------------|--------------|--------------|--------------|--------------|---------------|---------------|
| Asia         | 40          | 0.0011±1.0000 | 0.283±1.0000 | 0.002±1.0000 | 0.231±1.0000 | 0.218±1.0000 | 0.178±1.0000 | 0.303±1.0000 | 0.331±1.0000 |
| Attr         | 80          | 0.002±1.0000  | 0.283±1.0000 | 0.002±1.0000 | 0.231±1.0000 | 0.218±1.0000 | 0.178±1.0000 | 0.303±1.0000 | 0.331±1.0000 |
| Alarm        | 120         | 0.004±1.0000  | 0.283±1.0000 | 0.002±1.0000 | 0.231±1.0000 | 0.218±1.0000 | 0.178±1.0000 | 0.303±1.0000 | 0.331±1.0000 |
| Win9SPTs     | 160         | 0.005±1.0000  | 0.283±1.0000 | 0.002±1.0000 | 0.231±1.0000 | 0.218±1.0000 | 0.178±1.0000 | 0.303±1.0000 | 0.331±1.0000 |

TABLE 4. KL divergence values obtained by the different BN learning algorithms for the four BNs under different dataset sizes with constraint group B.

| Dataset size | ML          | ME          | MAP          | CML          | CME          | QMAP         | CMAP-I       | CMAP-II       |
|--------------|-------------|-------------|--------------|--------------|--------------|--------------|---------------|---------------|
| Asia         | 40          | 7.58±1.59   | 3.85±0.28    | 2.28±0.90    | 5.71±1.57    | 1.91±0.35    | 0.83±0.16     | 1.49±0.13     | 0.82±0.12     |
| Attr         | 80          | 8.6±2.52    | 3.11±0.11    | 2.16±0.87    | 6.99±1.41    | 1.63±0.41    | 0.73±0.06     | 1.31±0.27     | 0.80±0.17     |
| Alarm        | 120         | 9.02±2.26   | 2.82±0.57    | 1.63±0.27    | 7.62±1.35    | 1.36±0.13    | 0.70±0.08     | 1.07±0.10     | 0.68±0.10     |
| Win9SPTs     | 160         | 7.39±2.62   | 2.46±0.14    | 1.50±0.30    | 6.54±1.18    | 1.26±0.07    | 0.62±0.12     | 1.02±0.09     | 0.61±0.14     |
| Andes        | 200         | 6.40±2.58   | 2.35±0.10    | 1.47±0.29    | 6.50±1.18    | 1.22±0.06    | 0.58±0.06     | 1.01±0.14     | 0.55±0.04     |

We note that the CMAP-II algorithm uniformly obtained higher learning accuracies than the CMAP-I algorithm.

C. LEARNING EFFICIENCY ANALYSIS

We note from Table 3 that the computation time of all algorithms increased as the dataset size increased. The ML and MAP algorithms uniformly required the least amount of computation time. However, the proposed CMAP-I and CMAP-II algorithms provided better learning efficiency than all other algorithms considered. Meanwhile, the CMAP-I and CMAP-II algorithms required considerably less computation time than the QMAP algorithm. In fact, the QMAP algorithm required more computation time than the CMAP-I algorithm by a factor greater than three.

The KL divergence values and computation times obtained by the eight algorithms for the four BNs under the different dataset sizes with constraint group B are listed in Tables 4 and 5, respectively, where the best results are highlighted in bold. Moreover, the KL divergence values and computation times obtained by the eight algorithms for the Alarm BN are given in Figures 5 and 7 with respect to dataset size, respectively. We note from Figure 5 that the KL divergence values
TABLE 5. Computation times of the different BN learning algorithms for the four BNs under different dataset sizes with constraint group B.

| Data size | ML       | ME       | MAP     | CML     | CME     | QMAP    | CMAP-I   | CMAP-II   |
|-----------|----------|----------|---------|---------|---------|---------|----------|-----------|
| Asia      | 0.00023±0.0000 | 0.1920±0.0052 | 0.0002±0.0000 | 0.2115±0.0090 | 0.2099±0.0034 | 0.1823±0.0006 | 0.0329±0.0002 | 0.0297±0.0003 |
| 80        | 0.00031±0.0000 | 0.1885±0.0057 | 0.0003±0.0000 | 0.2146±0.0091 | 0.2086±0.0040 | 0.2014±0.0004 | 0.0519±0.0002 | 0.0498±0.0003 |
| 120       | 0.00046±0.0000 | 0.1926±0.0044 | 0.0004±0.0000 | 0.2231±0.0045 | 0.2095±0.0035 | 0.2236±0.0006 | 0.0726±0.0002 | 0.0701±0.0004 |
| 160       | 0.00085±0.0000 | 0.1916±0.0052 | 0.0005±0.0000 | 0.2301±0.0059 | 0.2094±0.0028 | 0.2422±0.0007 | 0.0976±0.0006 | 0.0902±0.0005 |
| 200       | 0.00086±0.0000 | 0.1927±0.0025 | 0.0006±0.0000 | 0.2357±0.0068 | 0.2107±0.0029 | 0.2624±0.0014 | 0.1130±0.0009 | 0.1107±0.0008 |

FIGURE 5. KL divergence values of the different BN learning algorithms with respect to the dataset size.

We note from Table 4 and Figure 5 that the learning accuracy of all algorithms considered increased by different degrees for all BNs as the dataset size increased. In addition, the CMAP-II algorithm generally provided the highest learning accuracy of all algorithms considered. The KL divergence values obtained by the CMAP-II algorithm shown in Figure 5 were very close to those obtained by the QMAP algorithm for the Alarm BN, but closer inspection (Figure 6) indicates that these values were always slightly less than those obtained by the QMAP algorithm, indicating that the CMAP-II algorithm provided uniformly higher learning accuracy than the QMAP algorithm for the Alarm BN under all dataset sizes. Finally, we note that the learning accuracy of the CMAP-I algorithm was slightly inferior to that of the QMAP algorithm.

D. LEARNING ACCURACY ANALYSIS

FIGURE 6. KL divergence values of the QMAP and CMAP-II algorithms with respect to the dataset size.

obtained by the QMAP and CMAP-II algorithms are very close. Therefore, these results are reproduced in Figure 6 to more adequately compare these two results.

E. LEARNING EFFICIENCY ANALYSIS

We note from Table 5 and Figure 7 that the computation time required by all of the algorithms considered increased as the dataset size increased. Again, the ML and ME algorithms uniformly required the least amount of computation time. In addition, we also note that the proposed CMAP-II algorithm generally required greater computation time than all other algorithms considered, except for the QMAP algorithm, under the condition of constraint group B. This was also often the case for the proposed CMAP-I algorithm as dataset

FIGURE 7. Computation times of the different BN learning algorithms for the four BNs under different dataset sizes with constraint group B.
size increased. The CMAP-I and CMAP-II algorithms again required considerably less computation time than the QMAP algorithm. (3) The computational time required by the QMAP algorithm was greater than that of the CMAP-II algorithm by a factor greater than five times.

The above results indicate that the proposed CMAP-I and CMAP-II algorithms have higher computational efficiencies than the QMAP algorithm under all conditions considered, and the learning accuracy of the CMAP-II algorithm is only slightly less than that of the QMAP algorithm when cross-distribution constraints are included within the constraint set. However, the CMAP-II algorithm has the highest learning accuracy of all the existing algorithms when cross-distribution constraints are not included within the constraint set.

VI. CONCLUSION

The rejection-acceptance sampling method employed in the QMAP algorithm for determining average BN parameter values is time-consuming, particularly when the number of parameter constraints is large. Therefore, this paper proposed a new analytical method for determining the average value of BN parameters by applying the center point of the constrained parameter region, which is a far more computationally efficient method. Experiments involving four benchmark BNs demonstrated that the proposed CMAP-I and CMAP-II algorithms have higher computational efficiency than the QMAP algorithm under all experimental conditions considered, and the learning accuracy of the CMAP-II algorithm is only slightly less than that of the QMAP algorithm when cross-distribution constraints are included within the constraint set. The CMAP-II algorithm has the highest learning accuracy of all existing algorithms when cross-distribution constraints are not included within the constraint set. This includes the following points: (1) Develop a more efficient solution engine for solving the objective function of the boundary points; (2) Design a more appropriate vector set to obtain the objective function of the boundary points; and (3) Introduce more types of parameter constraints in the proposed algorithm. The other part is to apply the proposed algorithm to solve the practical problems that can be modeled with Bayesian networks, such as fault diagnosis, genetic engineering, image processing, intelligent decision and other problems.

ACKNOWLEDGMENT

The authors would like to thank LetPub for its linguistic assistance during the preparation of this manuscript.

REFERENCES

[1] J. Pearl, “Fusion, propagation, and structuring in belief networks,” Artif. Intell., vol. 29, no. 3, pp. 241–288, Mar. 1986, doi: 10.1016/0004-3702(86)90072-X.
[2] T. Hincks, W. Aspinall, R. Cooke, and T. Gerton, “Oklahoma’s induced seismicity strongly linked to wastewater injection depth,” Science, vol. 359, no. 6381, pp. 7911–7924, Feb. 2018, doi: 10.1126/science.aap7911.
[3] N. Friedman, “Inferring cellular networks using probabilistic graphical models,” Science, vol. 303, no. 5659, pp. 799–805, Feb. 2004, doi: 10.1126/science.1094068.
[4] M. Velikova, J. T. van Scheltinga, P. J. F. Lucas, and M. Spaanderman, “Exploiting causal functional relationships in Bayesian network modelling for personalised healthcare,” Int. J. Approx. Reasoning, vol. 55, no. 1, pp. 59–73, Jan. 2014, doi: 10.1016/j.ijar.2013.05.016.
[5] S. Lax, N. Sangwan, D. Smith, P. Larsen, K. M. Handley, M. Richardson, K. Guyton, M. Krezalek, B. D. Shogan, J. Defazio, I. Fleming, B. Shakhsher, S. Weber, E. Landon, S. Garcia-Houchins, J. Siegel, J. Alverdy, R. King, B. Stephens, and J. A. Gilbert, “Colonization and succession in a newly opened hospital,” Sci. Transl. Med., vol. 9, no. 391, pp. 6500–6513, May 2017, doi: 10.1126/scitranslmed.aah6500.
[6] M. L. P. Bueno, A. Hommersom, P. J. F. Lucas, M. Lappenschaar, and J. G. E. Janzing, “Understanding disease processes by partitioned dynamic Bayesian networks,” J. Biomed. Informat., vol. 61, no. 5, pp. 283–297, Jun. 2016, doi: 10.1016/j.jbi.2016.05.003.
[7] I. Tien and A. D. Kiureghian, “Algorithms for Bayesian network modeling and reliability assessment of infrastructure systems,” Rel. Eng. Syst. Saf., vol. 156, no. 7, pp. 134–147, Dec. 2016, doi: 10.1016/j.ress.2016.07.022.
[8] F. Pernkopf, M. Wohlmayr, and S. Tschiatschek, “Maximum margin Bayesian network classifiers,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 3, pp. 521–532, Mar. 2012, doi: 10.1109/TPAMI.2011.149.
[9] Z. Wang, Z. Wang, S. He, X. Gu, and Z. F. Yan, “Fault detection and diagnosis of chillers using Bayesian network merged distance rejection and multi-source non-sensor information,” Appl. Energy, vol. 188, no. 10, pp. 200–214, Feb. 2017, doi: 10.1016/j.apenergy.2016.11.130.
[10] Q. Xiao, M. Qin, P. Guo, and Y. Zhao, “Multimodal fusion based on LSTM and a couple conditional hidden Markov model for Chinese sign language recognition.” Neurocomputing, vol. 434, no. 28, pp. 45–54, May 2020, doi: 10.1109/ACCESS.2019.2925654.
[11] Q. Xiao, Y. Zhao, and W. Huan, “Multi-sensor data fusion for sign language recognition based on dynamic Bayesian network and convolutional neural network,” Multimedia Tools Appl., vol. 78, no. 11, pp. 15335–15352, Jun. 2019, doi: 10.1007/s11042-018-6939-8.
[12] Q. Xiao and R. Song, “Action recognition based on hierarchical dynamic Bayesian network,” Multimedia Tools Appl., vol. 77, no. 6, pp. 6955–6968, Mar. 2018, doi: 10.1007/s11042-017-4614-0.
[13] J. Long, S. Zhang, and C. Li, “Evolving deep echo state networks for intelligent fault diagnosis,” IEEE Trans. Ind. Informat., vol. 16, no. 7, pp. 4928–4937, Jul. 2020, doi: 10.1109/TII.2019.2938884.
[14] J. Long, Z. Sun, C. Li, Y. Hong, Y. Bai, and S. Zhang, “A novel sparse echo autoencoder network for data-driven fault diagnosis of delta 3-D printers,” IEEE Trans. Instrum. Meas., vol. 69, no. 3, pp. 683–692, Mar. 2020, doi: 10.1109/TIM.2019.2905752.
[15] F. Wittig and A. Jameson, “Exploiting qualitative knowledge in the learning of conditional probabilities of Bayesian networks,” in Proc. 16th Int. Conf. Uncertainty Artif. Intell., San Francisco, CA, USA, 2000, pp. 644–652.

[16] A. Feelders and L. C. van der Gaag, “Learning Bayesian network parameters under order constraints,” Int. J. Approx. Reasoning, vol. 42, nos. 1–2, pp. 37–53, May 2006, doi: 10.1016/j.ijar.2005.10.003.

[17] A. R. Masegosa, A. J. Feelders, and L. C. van der Gaag, “Learning from incomplete data in Bayesian networks with qualitative influences,” Int. J. Approx. Reasoning, vol. 69, pp. 18–34, Feb. 2016, doi: 10.1016/j.ijar.2015.11.004.

[18] E. E. Altendorf, A. C. Restificar, and T. G. Dietterich, “Learning from sparse data by exploiting monotonicity constraints,” in Proc. 21st Int. Conf. Uncertainty Artif. Intell., Edinburgh, U.K., 2005, pp. 18–26.

[19] Y. Zhou, N. Fenton, and C. Zhu, “An empirical study of Bayesian network parameter learning with monotonic influence constraints,” Decis. Support Syst., vol. 87, pp. 69–79, Jul. 2016, doi: 10.1016/j.dss.2016.05.001.

[20] R. Di, X. Gao, and Z. Guo, “Discrete Bayesian network parameter learning based on monotonic constraint,” J. Syst. Eng. Electron., vol. 36, no. 2, pp. 272–277, Jan. 2014, doi: 10.3969/j.issn.1001-500X.2014.02.12.

[21] R. Di, X. Gao, and Z. Guo, “Learning Bayesian networks parameters under new monotonic constraints,” J. Syst. Eng. Electron., vol. 28, no. 6, pp. 1248–1255, Jun. 2017.

[22] J. Ren, X. Gao, and Y. Bai, “Discrete dynamic BN parameter learning under small sample and incomplete information,” J. Syst. Eng. Electron., vol. 34, no. 8, pp. 1723–1728, Aug. 2012, doi: 10.3906/jissn.1001-500X.2012.08.33.

[23] R. Niculescu, T. Mitchell, and B. Rao, “Bayesian network learning with parameter constraints,” J. Mach. Learn. Res., vol. 7, no. 1, pp. 1357–1383, Jul. 2006, doi: 10.1007/s10846-006-9065-1.

[24] C. Campos and Q. Ji, “Improving Bayesian network parameter learning using constraints,” in Proc. 19th Int. Conf. Pattern Recognit., Tampa, FL, USA, 2009, pp. 1–4.

[25] W. Liao and Q. Ji, “Learning Bayesian network parameters under incomplete data with domain knowledge,” Pattern Recognit., vol. 42, no. 11, pp. 3046–3056, Nov. 2009, doi: 10.1016/j.patcog.2009.04.006.

[26] R. Chang, R. Shoemaker, and W. Wang, “A novel knowledge-driven systems biology approach for phenotype prediction upon genetic intervention,” IEEE Trans. Comput. Bioinform., vol. 3, no. 8, pp. 683–697, Sep. 2011, doi: 10.1109/TCBB.2011.18.

[27] Z.-G. Guo, X.-G. Gao, H. Ren, Y. Yang, R.-H. Di, and D.-Q. Chen, “Learning Bayesian network parameters from small data sets: A further constrained qualitatively maximum a posteriori method,” Int. J. Approx. Reasoning, vol. 91, no. 9, pp. 22–35, Dec. 2017, doi: 10.1016/j.ijar.2017.08.009.

[28] Y. Zhou, N. Fenton, and M. Neil, “Bayesian network approach to multinomial parameter learning using data and expert judgments,” Int. J. Approx. Reasoning, vol. 55, no. 5, pp. 1252–1268, Jul. 2014, doi: 10.1016/j.ijar.2014.02.008.

[29] Y. Zhou, N. Fenton, T. Hospedales, and M. Neil, “Probabilistic graphical models parameter learning with transferred prior and constraints,” in Proc. 31st Int. Conf. Uncertainty Artif. Intell., Amsterdam, The Netherlands, 2015, pp. 34–41.

[30] Z.-G. Guo, X.-G. Gao, and R.-H. Di, “Learning Bayesian network parameters under dual constraints from small data sets,” Acta Autom. Sinica, vol. 40, no. 7, pp. 1509–1516, Jan. 2014, doi: 10.3724/SP.J.1004.2014.01509.

[31] S. Dasgupta, “The sample complexity of learning fixed-structure Bayesian networks,” in Proc. 14th Conf. Mach. Learn., Nov. 1997, pp. 165–180.

[32] M. P. Wellman, “Fundamental concepts of qualitative probabilistic networks,” Artif. Intell., vol. 44, no. 3, pp. 257–303, Aug. 1990, doi: 10.1016/0004-3702(90)90026-V.

RUOHAI DI received the M.S. and Ph.D. degrees in systems engineering from Northwestern Polytechnical University, in 2011 and 2016, respectively. From 2012 to 2017, he worked as a Senior Engineer with the Huawei Laboratory. In 2019, he completed his postdoctoral research with the Electronic Science and Technology Mobile Station, Northwestern Polytechnical University. He currently works as an Associate Professor with Xi’an Technological University. His research interests include Bayesian networks (structure and parameter learning algorithms) and data mining.

PENG WANG received the M.A. degree from Xi’an Technological University, in 2003. He is currently a Professor with the Department of Electronic Information Engineering, Xi’an Technological University. His research interests include homeland security, sensors and sensor networks, and artificial intelligence.

JIAO WU is currently pursuing the B.S. degree with the Department of Communication Engineering, Xi’an Technological University. Her main research interests include machine vision and deep learning.

ZHI GAO GUO received the B.S. degree from Hainan University, in 2008, the M.S. degree from Shenyang Aerospace University, in 2010, and the Ph.D. degree from Northwestern Polytechnical University, in 2019. He currently works as a Postdoctoral Research Associate with the Queen Mary University of London. His research interest includes Bayesian networks, especially parameter estimation from insufficient data and structure learning in high-dimensional problems.

***