Extended Learning Algorithm for Constrained Maximum Entropy Model of BN Parameters (Postprint)

Authors: Guo Wenqiang, Li Ran, Hou Yongyan, Gao Wenqiang

Date: 2018-05-20T00:00:00+00:00

Abstract

In parameter modeling for many intelligent systems, users often face the challenge of scarce training samples. To address the problem of Bayesian network (BN) parameter modeling under small dataset conditions, a Constrained Data Maximum Entropy BN Parameter Learning Algorithm (CDME) is proposed. First, BN parameters are estimated using the small dataset, then qualitative expert experience is converted into inequality constraints, and the Bootstrap algorithm is utilized to generate a set of parameter candidates satisfying the constraints, after which BN parameters are calculated through weighted averaging based on the maximum entropy principle. Experimental results demonstrate that when the data volume is sufficient, the learning accuracy of the CDME parameter learning algorithm is comparable to that of the classical MLE algorithm, indicating the correctness of the algorithm; under small dataset conditions, the CDME algorithm can be used for BN parameter modeling, with learning accuracy superior to both the MLE algorithm and the QMAP algorithm. The CDME algorithm obtains diagnostic BN model parameters under conditions of relatively scarce actual fault diagnosis sample data, and the diagnostic inference results completed on this basis also confirm the effectiveness of the algorithm, providing a new approach for parameter modeling under small dataset conditions.

Full Text

Extension of BN Parameter Learning Algorithm Based on Constrained Maximum Entropy Model

Guo Wenqiang†, Li Ran, Hou Yongyan, Gao Wenqiang
(School of Electrical & Information Engineering, Shaanxi University of Science & Technology, Xi' an 710021, China)

Abstract: In parameter modeling for many intelligent systems, users often face the dilemma of scarce modeling samples. To address the problem of Bayesian network (BN) parameter modeling under small datasets, this paper proposes a constrained data maximum entropy (CDME) algorithm for BN parameter learning. The algorithm first estimates BN parameters using a small dataset, then converts qualitative expert experience into inequality constraints. It generates a set of parameter candidates satisfying these constraints using the Bootstrap algorithm, and finally computes the BN parameters through maximum entropy weighting. Experimental results demonstrate that when data is sufficient, the learning accuracy of the CDME algorithm approximates that of the classical MLE algorithm, confirming its correctness. Under small dataset conditions, the CDME algorithm enables BN parameter modeling with learning accuracy superior to both MLE and QMAP algorithms. When applied to real-world fault diagnosis with relatively scarce sample data, the diagnostic reasoning results obtained from the learned BN model parameters further validate the algorithm' s effectiveness, providing a novel approach for parameter modeling under small dataset conditions.

Keywords: Bayesian network; small dataset; parameter learning; maximum entropy model

0 Introduction

Bayesian networks (BN) demonstrate powerful adaptability in solving uncertainty and incompleteness problems in complex systems and have been successfully applied in numerous domains related to intelligent systems [1~3]. BN modeling forms the foundation for BN inference, and the results of BN reasoning can often validate the effectiveness of modeling approaches. BN modeling comprises both structural modeling and parameter modeling [4~5]. While domain experts can often provide qualitative knowledge to complete BN structural modeling, they typically struggle to directly determine accurate quantitative parameters for BN parameter modeling—that is, the problem of estimating BN parameters. Under sufficient sample data conditions, the classical maximum likelihood estimation (MLE) algorithm represents one of the more effective data-driven learning methods. However, in practice, obtaining large amounts of effective sample data from certain systems is extremely difficult and expensive, such as in earthquake prediction or aero-engine fault diagnosis systems. This results in relatively limited system characteristic data that cannot accurately estimate BN parameters, consequently affecting system model inference and even leading to results that contradict objective laws [6~8].

In research on BN parameter modeling under small dataset conditions, reference [9] utilized qualitative constraints to propose a gradient-descent estimation (GDE) method, employing adaptive probabilistic networks (APN) algorithms for solution. Reference [10] proposed an isotonic regression estimation (IRE) method for qualitative influence constraints, using isotonic regression algorithms to correct initial parameters to satisfy order constraints. Reference [11]

introduced a minimum free energy (MFE) estimation method for normative constraints, which first constructs an objective optimization function containing both KL divergence and entropy functions, then combines it with normative constraints using Lagrange multipliers, and finally obtains parameters through extremum methods. Reference [2] proposed a constrained expectation maximum (CEM) method for BN parameter learning using qualitative influence constraints. However, GDE, IRE, MFE, and CEM methods employ single constraint types with relatively weak constraint strength, limiting parameters to broad feasible regions and resulting in insufficiently precise learning outcomes. Reference [8] proposed a dual constrained estimation (DCE) method based on double constraints, using Beta distribution fitting and isotonic regression for parameter learning under small datasets. This algorithm is limited to BN models with binary-valued nodes. In real-world systems, such as multi-mode classification problems in fault diagnosis, network node states are often multi-valued, restricting the applicability of the DCE algorithm. References [12,13] proposed a constrained maximum likelihood (CML) method, which first constructs non-monotonic constraint models and then solves them by combining the maximum entropy function with non-monotonic constraints to form convex optimization problems. Reference [14] introduced a qualitative maximum a posteriori (QMAP) method, which combines real statistical data with virtual small data using Bayesian estimation to learn BN parameters.

Analysis reveals that while the CML method imposes no special requirements on parameter constraint types, it exhibits low utilization of parameter constraints and often yields inferior learning performance compared to QMAP [12]. The QMAP method represents a relatively effective new approach under small dataset conditions [8]. QMAP uses qualitative constraints to perform virtual sampling of parameter sets satisfying constraints through rejection-acceptance algorithms. However, since rejection-acceptance algorithms are based on probability distribution function estimation, and the probability distribution functions of BN parameters are often difficult to estimate in practice, the practicality of the QMAP method faces challenges.

To address BN parameter estimation problems under small dataset conditions, this paper proposes a constrained data maximum entropy (CDME) parameter learning algorithm. It organically combines small dataset parameter learning results with parameter candidate sets generated from expert experience constraints, estimating BN parameters based on the principle of maximum information entropy.

1 Background on BN Parameter Learning

BN parameter learning aims to obtain the conditional probability distribution of all network nodes through sample data and prior knowledge when the network structure is known. A BN consists of a directed acyclic graph (DAG) G and a set of probability distributions associated with each variable. Equation (1) represents the joint probability distribution of a BN [1]:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Pa(X_i))$$

where $Pa(X_i)$ denotes the parent node set of X_i in G, and $P(X_i|Pa(X_i))$ represents the conditional probability distribution of each variable value given its parent node values.

Let $\theta_{ijk} = P(X_i = k|Pa(X_i) = j)$ denote the k-th probability value in the conditional probability table for node X_i, where $\theta_{ijk} \in \theta$, $1 \leq i \leq n$, $1 \leq j \leq q_i$, and $1 \leq k \leq r_i$. Here, r_i represents the number of states for node X_i, and q_i represents the number of combined states for the parent nodes of X_i. Clearly, node X_i has a total of $r_i \times q_i$ parameters, which form an $r_i \times q_i$ dimensional matrix called the conditional probability table (CPT) for node X_i.

1.1 Maximum Likelihood Estimation Algorithm

Maximum likelihood estimation (MLE) is expressed as:

$$L(\theta) = \prod_{i=1}^{m} f(e_i; \theta) = f(e_1, e_2, \ldots, e_m; \theta)$$

where $e_1, e_2, \ldots, e_m$ represents the sample set and θ denotes the parameters of the distribution. $f(e_i; \theta)$ is the probability density function [7]. In practice, each parameter can be estimated using Equation (3):

$$\theta_{ijk}^{MLE} = \frac{N_{ijk}}{N_{ij}}$$

where N_{ij} represents the frequency of parent node $Pa(X_i)$ taking the j-th combined state in sample D, and N_{ijk} represents the frequency of node X_i being in state k when its parent nodes are in the j-th combined state.

MLE is commonly used for BN parameter estimation with relatively sufficient datasets. However, Equation (3) has a major drawback: it cannot estimate θ_{ijk} when $N_{ij} = 0$. When the training set is limited, observations with zero values frequently occur, especially with small-scale datasets. To address this issue, a slight modification is made to Equation (3), resulting in Equation (4):

$$\theta_{ijk}^{*} = \frac{N_{ijk} + \mu}{N_{ij} + \varepsilon}$$

where μ and ε are constants close to zero but not equal to zero. To satisfy experimental requirements, ε is set to 0.00001 in Equation (4).

1.2 Common Constraint Sets

In the real world, although domain experts find it difficult to provide quantitative expressions for BN model parameters, they can often supply qualitative information between nodes—that is, constraint information. Equation (5) reflects the relationship between expert prior knowledge (Ω and G), data (D), and BN parameters [11,12]:

$$\log P(\theta|D,G,\Omega) = \log P(\theta|G,\Omega) + \log P(D|\theta,G) - \log P(D|G,\Omega)$$

where θ represents the learned BN parameters, G is the BN structure, D is the sample data, and Ω is the constraint information formed from expert experience knowledge.

Common constraint sets in BNs can be derived from qualitative domain knowledge, which can be defined as various constraints on conditional probabilities. These constraints can be expressed in the following forms [13,15]:

a) **Range constraints** define the upper and lower bounds for any parameter in the BN: $0 < \alpha_{ijk} \leq \theta_{ijk} \leq \beta_{ijk} \leq 1$.

b) **Cross-distribution constraints** define the relative relationship between two parameters from different conditional distributions. If two constrained parameter nodes have the same index i and state value k but different parent node combinations j, this constraint is called a cross-distribution constraint, typically obtained from qualitative causal relationships: $\theta_{ijk} \leq \theta_{ijk'}$ or $\theta_{ijk} \geq \theta_{ijk'}$, $\forall j \neq j'$.

c) **Intra-distribution constraints** define the relative relationship between two parameters sharing the same node index i and parent node combination j but having different state values k: $\theta_{ijk} \leq \theta_{ijk'}$ or $\theta_{ijk} \geq \theta_{ijk'}$, $\forall k \neq k'$.

d) **Inter-distribution constraints** define the relative relationship between two parameters that do not share any node index i, parent node combination j, or state value k: $\theta_{ijk} \leq \theta_{i'j'k'}$ or $\theta_{ijk} \geq \theta_{i'j'k'}$, $\forall i \neq i', j \neq j', k \neq k'$.

e) **Axiom constraints** describe relationships between parameters referring to fixed parent node configurations. These are fundamental constraints that domain experts need not provide: $\sum_{k=1}^{r_i} \theta_{ijk} = 1$, $0 \leq \theta_{ijk} \leq 1$, $\forall i, j, k$.

f) **Approximate equality constraints** describe close relationships between any two parameters. For computational convenience, they can be written as: $|\theta_{ijk} - \theta_{i'j'k'}| \leq \sigma$, $\forall i \neq i', j \neq j', k \neq k'$.

g) **Additive synergy constraints** describes comparative relationships between sums of two parameters under different distributions: $\theta_{i_1 j_1 k_1} + \theta_{i_2 j_2 k_2} \leq \theta_{i_3 j_3 k_3} + \theta_{i_4 j_4 k_4}$, $\forall i_1, i_2, i_3, i_4, j_1, j_2, j_3, j_4, k_1, k_2, k_3, k_4$.

h) **Product synergy constraint** describes comparative relationships between products of two parameters under different distributions: $\theta_{i_1 j_1 k_1} \times \theta_{i_2 j_2 k_2} \leq \theta_{i_3 j_3 k_3} \times \theta_{i_4 j_4 k_4}$, $\forall i_1, i_2, i_3, i_4, j_1, j_2, j_3, j_4, k_1, k_2, k_3, k_4$.

1.3 Maximum Entropy Principle

The fundamental idea of maximum entropy is to make the most objective and uniform estimates of unknown events while satisfying all known events. To select a model from an allowable probability distribution set C, the model with maximum entropy H(p) is chosen, formally expressed as [16]:

$$p^* = \arg\max_{p \in C} H(p) = \arg\max_{p \in C} \sum_{x,y} p(x,y) \log \frac{p(y|x)}{p(y)}$$

where C is the set of probability distributions satisfying constraints, and x, y are random variable pairs corresponding to events. The mathematical measurement of conditional distribution is provided by conditional entropy.

This paper selects the maximum entropy statistical model as the estimation model bridging the original domain (BN knowledge and sample data) and target domain (BN parameter learning) based on the following considerations: the BN distribution parameters employed by maximum entropy impose no requirements on the prior distribution of data and can be arbitrarily combined without compromising the consistency of the entire estimation model. Simultaneously, it naturally addresses parameter smoothing in statistical models. Assuming L parameter candidate sets satisfy the constraint knowledge Ω, according to the maximum entropy principle, they have equal probability of approximating the true BN parameters. Therefore, even if some features in the target domain do not appear in the original domain, satisfactory parameter estimation results can still be obtained due to the support of constraint knowledge.

1.4 KL Distance

KL distance is commonly used to measure the distance between two probability distributions. This paper employs KL distance to measure the accuracy of BN parameter estimation methods, as shown in Equation (7) [17]:

$$KL(\theta || \theta') = \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} \theta_{ijk} \ln \frac{\theta_{ijk}}{\theta'_{ijk}}$$

where θ represents the true parameters and θ' represents the estimated parameters. Clearly, the closer the estimated BN parameters are to the true BN parameters, the smaller the KL value.

2 Constrained Data Maximum Entropy Parameter Learning Algorithm

The flowchart of the proposed constrained data maximum entropy BN parameter learning algorithm is shown in [Figure 1: see original paper]. The main steps of the CDME algorithm are as follows:

a) Set the parameter weight factor α ($0 \leq \alpha \leq 1$), the number of candidate parameter sets L, the constraint set Ω, and initialize the intermediate variable parameter set θ as empty.

b) Calculate an initial parameter set $\theta^*(S_{ijk})$ from the initial small sample set using Equation (4) and insert it into parameter set θ.

c) Generate the $(L-1)$-th parameter candidate set $\theta^{(L-1)}(S_{ijk})$ using the Bootstrap method [18].

d) Check whether the $(L-1)$-th parameter candidate set $\theta^{(L-1)}(S_{ijk})$ satisfies constraints Ω. If satisfied, insert $\theta^{(L-1)}(S_{ijk})$ into the intermediate variable parameter set θ and set $L = L-1$; otherwise, return to step c).

e) Check whether $L < 1$:

- (a) If true, output and extract the candidate parameter set θ. Calculate the final learning parameters θ_{ijk}^{CDME} according to the maximum entropy principle using Equation (8):

$$\theta_{ijk}^{CDME} = \alpha\theta_{ijk}^{*} + \frac{(1-\alpha)}{L}\sum_{l=1}^{L-1}\theta_{ijk}^{(l)}$$

  where $\sum_{l=1}^{L-1}\theta_{ijk}^{(l)}$ is the sum of $(L-1)$ constraint-satisfying parameter candidate sets obtained through parameter expansion, θ_{ijk}^{*} is the initial parameter set calculated from the small sample set, and α is the weight factor.

- (b) Otherwise, jump to step c).

Note that in the parameter model of Equation (8), when $\alpha = 0$ and $L = 1$, the form is essentially the MLE result when calculating BN parameters under sufficient data conditions. Therefore, this parameter model is a generalization of the MLE algorithm; MLE is a special case of this parameter model.

The candidate parameters generated according to qualitative constraints have equal weight, reflecting the maximum entropy principle. Additionally, when α takes a smaller value, the decisive role of sample data in parameter learning becomes prominent, while the influence of qualitative constraint expert experience on learning results decreases. Conversely, when α takes a larger value, the influence of qualitative constraint expert experience on learning results increases, compensating for the impact of insufficient sample data on learning results.

The constrained data maximum entropy BN parameter learning algorithm establishes a framework for estimating BN parameters using constraint information and small datasets. It can integrate different sources of information, such as sample data and expert qualitative experience, into a unified framework for comprehensive consideration, demonstrating good formal consistency in solving BN parameter problems.

3 Experiments and Analysis

This paper conducts two sets of experimental analyses to validate the effectiveness of the proposed method. The first set of experiments employs the classic Bayesian network model—the lawn wetness model [19]—to compare BN parameter learning using MLE, QMAP, and the proposed method under simulated data. The second set of experiments applies the proposed method to BN parameter learning and inference diagnosis under real fault diagnosis data to further validate the superiority and practicality of the BN parameter learning method. All experiments in this paper were conducted on a PC with a 2GHz processor and 6GB of memory, with programs implemented in MATLAB 2014a.

3.1 Simulated Data BN Parameter Modeling Experiment

The lawn wetness model is a classic experimental model for validating BN learning algorithms. [Figure 2: see original paper] shows the structure of this BN model. The network contains four nodes: C, S, R, and W. Parameter learning is performed on node W.

3.1.1 Experimental Setup The parameter weight factor α is set to 0.3, and the number of candidate parameter sets L is 500. The parameter event set is shown in , where parameters in node W are numbered. Event value "1" indicates the event is "false," and "2" indicates the event is "true" . shows the constraint set Ω satisfied by parameter $P(W|S,R)$.

In , $P1 > P5$ indicates that under the condition of "sprinkler (S) not watering and weather (R) not raining," the probability of lawn event W being "not wet" is greater than the probability of "wet lawn." This is referred to as constraint condition "C1," and similarly for the other eight constraints. Clearly, these constraint sets align with common knowledge and are readily accepted by domain experts.

3.1.2 Experimental Results Analysis Under the condition of 35 sample groups, the parameters of node W were learned using MLE, QMAP, and CDME methods respectively. [Figure 3: see original paper] shows the comparison of average KL distances between estimated parameters and true parameters across 15 learning runs. [Figure 4: see original paper] shows the corresponding box plots of KL distances.

When sample data is relatively sufficient, 300 sample groups were used in the experiment. [Figure 5: see original paper] shows the KL distance comparison among the three methods for learning node W parameters.

Through comparative experiments on BN parameter learning, the following observations can be made: Under small sample sizes, the KL distance of the MLE method is relatively large, while QMAP and CDME methods show clear advantages, with CDME achieving slightly higher learning precision than QMAP. As sample size increases, the KL distance of the MLE method gradually decreases, intersecting with those of QMAP and CDME methods at sample sizes of 30 and 31 respectively. However, the KL distance of the QMAP method shows an insignificant downward trend. Unlike QMAP, the KL distance of the CDME method exhibits an overall decreasing trend as sample size increases.

Analysis reveals that under small dataset conditions, the CDME method can fully utilize parameter constraints to estimate BN parameters with superior precision compared to MLE and QMAP methods. When sample data is sufficient, the proposed method remains feasible and can fully leverage the information contained in the obtained datasets to compensate for uncertainties in BN modeling, continuously refining and improving the precision of learned parameters.

3.2 Real Fault Diagnosis Data Experiment

The bearing fault diagnosis model and 469 feature datasets used in this experiment are derived from reference [20]. The original experimental data comes from the Bearing Fault Simulation Laboratory at Case Western Reserve University, where rich monitoring data for typical fault types were obtained through destructive testing. In actual production processes, however, equipment fault data available to practitioners is often limited. Therefore, this paper attempts to perform BN parameter learning using the CDME algorithm under both small data (35 groups) and sufficient data (300 groups) conditions. Subsequently, diagnostic reasoning experiments were conducted using the remaining 169 groups of feature datasets to validate the effectiveness of the CDME algorithm. The bearing fault diagnosis BN model structure is shown in [Figure 6: see original paper].

In [Figure 6: see original paper], the BN model contains 9 nodes: 1 parent node and 8 child nodes representing 8 feature vectors D0~D7. The parent node has four states: normal condition, inner ring fault, outer ring fault, and rolling element fault, corresponding to the four diagnostic outputs of the BN diagnosis model with diagnostic confidence $\gamma = 0.70$. To diagnose bearing faults, the BN model structure was first completed through domain expert experience. Second, feature vectors were obtained by applying feature extraction functions and discretization to fault data. The CDME method was then applied for fault diagnosis BN parameter modeling to obtain the BN model required for fault diagnosis reasoning.

3.2.1 Experimental Parameter Settings The number of candidate parameter sets L is set to 500. The parameter constraint set Ω contains 8 constraints derived from expert experience. α is set to 0.3. The classical junction tree algorithm is selected for BN inference.

3.2.2 Fault Diagnosis Reasoning Results Comparison and present the diagnostic reasoning results using 169 groups of feature data, based on models trained with 35 groups and 300 groups of feature data respectively.

**** CDME learning method reasoning results under 35 groups of data
| State Type | Test Count | Correct Count | Correct Rate | |———|———|———|———-| | Normal condition | 169 | 169 | 100% | | Inner ring fault | 169 | 169 | 100% | | Rolling element fault | 169 | 154 | 91.12% | | Outer ring fault | 169 | 139 | 82.25% |

**** CDME learning method reasoning results under 300 groups of data
| State Type | Test Count | Correct Count | Correct Rate | |———|———|———|———-| | Normal condition | 169 | 169 | 100% | | Inner ring fault | 169 | 169 | 100% | | Rolling element fault | 169 | 162 | 95.86% | | Outer ring fault | 169 | 157 | 92.90% |

Experimental results demonstrate that under small dataset conditions, reasoning after modeling with the CDME algorithm achieves high overall correct rates with strong inference capability (only slightly lower for outer ring fault diagnosis). This further validates the effectiveness of using the CDME algorithm for parameter modeling under small dataset conditions. When sample data is relatively sufficient, the CDME algorithm can utilize dataset information to refine BN learning parameters, improving BN model learning precision.

4 Conclusion

The CDME algorithm proposed in this paper estimates BN parameters using small datasets while converting qualitative expert experience into inequality constraints. It generates constraint-satisfying parameter candidate sets using the Bootstrap algorithm and computes BN parameters according to the principle of maximum information entropy. Experiments demonstrate that the proposed CDME method exhibits certain effectiveness and superiority in learning BN parameters under small datasets. Its application in fault diagnosis modeling confirms the algorithm' s practicality through diagnostic reasoning results. When sample data is sufficient, the proposed algorithm remains feasible and can fully leverage information contained in the obtained datasets to reduce uncertainties in BN modeling and continuously refine and improve learning parameter precision. The CDME algorithm studied in this paper shows promising practical prospects for intelligent system modeling under data-scarce conditions, providing a novel approach for BN parameter modeling under small dataset conditions.

References

[1] Dojer N. Learning Bayesian networks from datasets joining continuous and discrete variables [J]. International Journal of Approximate Reasoning, 2016, 78(C): 116-124.

[2] Liao Wenhui, Ji Qiang. Learning Bayesian network parameters under incomplete data with domain knowledge [J]. Pattern Recognition, 2009, 42(11): 3046-3056.

[3] Chen Wei, Zhu Biao, Zhang Hongxin. BN-Mapping: Visual analysis of geospatial data based on Bayesian networks [J]. Chinese Journal of Computers, 2016, 39(07): 1281-1293.

[4] Li Shuohao, Zhang Jun. Survey on Bayesian network structure learning [J]. Application Research of Computers, 2015, 32(3): 641-646.

[5] Chen Jing, Jiang Zhengkai, Fu Jingqi. Construction of self-learning Bayesian network based on Netica [J]. Journal of Electronic Measurement and Instrumentation, 2016, 30(11): 1687-1693.

[6] Xiao Meng, Zhang Youpeng. Bayesian network parameter learning for multi-state systems under small dataset conditions [J]. Computer Science, 2015, 42(4): 253-257.

[7] Li Zida, Liao Shizhong. Bayesian network parameter learning method for small samples [J]. Computer Engineering, 2016, 42(8): 153-159+165.

[8] Guo Zhigao, Gao Xiaoguang, Di Ruohai. BN parameter learning based on dual constraints under small dataset conditions [J]. Acta Automatica Sinica, 2014, 40(7): 1509-1516.

[9] Altendorf E E, Restificar A C, Dietterich T G. Learning from Sparse Data by Exploiting Monotonicity Constraints [C]// Proc of the 21st Conference on Uncertainty in Artificial Intelligence. 2005: 18-25.

[10] Feelders A, Gaagl. Learning Bayesian networks parameters under order constraints [J]. International Journal of Approximate Reasoning, 2006, 42(1//2): 37-53.

[11] Isozaki T, Kato N, Ueno M. "Data temperature"in minimum free energies for parameter learning of Bayesian networks [J]. International Journal on Artificial Intelligence Tools, 2009, 18(5): 653-671.

[12] Niculescu R, Mitchell M, Rao B. Bayesian network learning with parameter constraints [J]. Journal of Machine Learning Research, 2006, 7(1): 1357-3046.

[13] Campos C, Tong Yang, Ji Qiang. Constrained maximum likelihood learning of bayesian networks for facial action recognition [C]// Computer Vision. 2008: 168-181.

[14] Chang Rui, Shoemaker R, Wang Wei. A novel knowledge-driven systems biology approach for phenotype prediction upon genetic intervention [J]. Transactions on Computational Biology and Bioinformatics, 2011, 1(8): 603-618.

[15] Guo Zhigao, Gao Xiaoguang, Di Ruohai, et al. Learning Bayesian network parameters from small data set: a spatially maximum a posteriori method [C]// Proc of International Workshop on Advanced Methodologies for Bayesian Networks. 2015: 32-45.

[16] Berger A L, Pietra S, Pietra V. A Maximum Entropy approach to Natural Language Processing [J]. Computational Linguistics, 1996, 22(1): 38-73.

[17] Kullback S, Leibler R A. On information and sufficiency [J]. Annals of Mathematical Statistics, 1951, 22(1): 79-86.

[18] Sun Huiling, Hu Weiwen, Liu Haitao. Improved method for parameter interval estimation under small sample conditions [J]. Journal of Harbin University of Science and Technology, 2017, 22(01): 109-113.

[19] Stuart J. Russell, Peter N. Artificial Intelligence: A Modern Approach [M]. Translated by Yin Jianping. 3rd ed. Beijing: Tsinghua University Press, 2013.

[20] Guo Wenqiang, Zhang Baorong, Peng Cheng, et al. Deep groove ball bearing fault diagnosis based on wavelet packet and BN model [J]. Bearing, 2016, 59(3): 48-52.

Note: Figure translations are in progress. See original paper for figures.