

小数据集条件下贝叶斯网络自适应参数学习方法

郭志高, 高晓光[†], 邸若海

(西北工业大学 电子信息学院, 陕西 西安 710129)

摘要: 针对小数据集条件下贝叶斯网络参数学习问题, 约束最大似然(CML)和定性最大后验概率(QMAP)方法是两种约束适用性较好的方法. 当样本数量、约束数量、参数位置不同时, 上述两种方法互有优劣, 进而导致方法上的难以选择. 因此, 本文提出一种自适应参数学习方法: 首先, 利用CML和QMAP方法学习得到两组参数; 然后, 基于拒绝-接受采样和空间最大后验概率思想自定义计算得到样本权重、约束权重、参数位置权重; 最后, 基于上述参数和权重计算得到新的参数解. 实验表明: 在任何条件下, 本文方法计算得到参数的精度接近甚至优于CML和QMAP方法的最优解.

关键词: 贝叶斯网络; 小数据集; 参数估计; 凸优化; 自适应方法

中图分类号: TP18 文献标识码: A

Learning Bayesian network parameters from small data set: an adaptive method

GUO Zhi-gao, GAO Xiao-guang[†], DI Ruo-hai

(School of Electronics and Information, Northwestern Polytechnical University, Xi'an Shaanxi 710129, China)

Abstract: For parameter learning of Bayesian networks from small data set, constrained maximum likelihood (CML) method and qualitative maximum a posterior (QMAP) method are two approaches, which suit all types of existing parameter constraints. However, those two approaches dominate each other when samples size, constraint number or true-parameter location varies. That makes it tough to choose between those two methods. For that reason, a novel adaptive parameter learning method is proposed in this paper. First, CML method and QMAP method are employed to learn BN parameters. Then, sample weight, constraint weight, and parameter-location weight are defined and calculated based on rejection-acceptance sampling and spatial maximum a posterior analysis. Finally, a new set of parameters are calculated as the weighted values of CML and QMAP solutions. Furthermore, simulation results reveal that precision of parameters learnt by the proposed method, in any cases, approaches and even outperforms those of CML method and QMAP method.

Key words: Bayesian networks; small data set; parameter estimation; convex optimization; adaptive method

1 引言(Introduction)

贝叶斯网络(Bayesian network, BN)是表示随机变量间依赖和独立关系的网络模型^[1], 由节点集、有向边集、条件概率分布集组成. 其中, 节点表示随机变量, 有向边表示变量间的依赖关系, 条件概率分布表示变量间的依赖程度. 现实生活中的许多实例如医疗诊断^[2]、故障诊断^[3]、智能控制^[4]、目标跟踪^[5]、信号处理^[6]、生态系统建模^[7]等都可以通过BN进行建模. 通常, BN的构建需要咨询领域专家, 以获取该领域的经验知识. 对于专家而言, 他们能够可靠地给出BN的结构^[8], 却很难指定条件概率值即参数^[9]. 如果能够获得充足的样本数据, 则利用最大似然估计(maximum

likelihood estimation, MLE)方法^[10]可以准确地估计出BN参数. 但是, 从某些系统中获取大量样本数据非常困难或者代价昂贵, 如地震预测系统中的地震数据、空战态势评估系统中的空战数据、航空发动机故障诊断系统中的发动机故障数据等. 在上述情况下, 由于样本数据较少, 很难精确估计出BN参数. 而不精确的参数往往导致BN的推理结果违背常理. 所以, 一直以来, 提高小数据集条件下BN参数学习的精度是一个备受关注的问题.

虽然专家很难给出网络参数, 却可以给出网络节点间的关系信息即约束信息^[11]. 而事实上, 约束信息比参数信息具有更好的鲁棒性^[12]. 迄今为止, 小数据

收稿日期: 2015-06-02; 录用日期: 2016-04-28.

[†]通信作者. E-mail: cxg2012@nwpu.edu.cn; Tel.: +86 29-88431266.

本文责任编辑: 刘淑君.

国家自然科学基金面上项目(61573285), 中央高校基本科研业务费专项资金(3102015BJ(II)GH01)资助.

Supported by National Natural Science Foundation of China (61573285) and Fundamental Research Funds for the Central Universities (3102015BJ(II)GH01).

集条件下基于约束信息的BN参数学习研究经历了两个阶段: 基于单一约束的参数学习和基于复合约束的参数学习. 在基于单一约束的参数学习研究阶段: 文献[12]针对定性影响约束提出了约束自适应概率网络(constrained adaptive probabilistic networks, CAPN)参数学习方法. 文献[13]针对定性影响约束提出了梯度下降估计(gradient-descent estimation, GDE)方法. 文献[14]针对定性影响约束提出了保序回归估计(isotonic regression estimation, IRE)方法. 文献[15]针对规范性约束提出了最小自由能(minimum free energy, MFE)估计方法. 由于单一约束对参数的约束力较小, 只能将参数限制到较为宽泛的可行域内, 导致参数学习结果不够精确. 而且, 针对单一特定类型约束的参数学习方法约束适用性较差, 很难适用于其他类型的约束.

因此, 针对小数据集条件下BN参数学习研究逐渐过渡到基于复合约束的参数学习阶段. 在基于复合约束的参数学习阶段: 文献[16–17]提出了约束最大似然(constrained maximum likelihood, CML)方法. 文献[18]提出了约束最大期望(constrained expectation maximum, CEM)方法. 文献[19]提出了定性最大后验概率(qualitative maximum a posterior, QMAP)方法. 文献[20]提出了基于双重约束(dually constrained estimation, DCE)的参数学习方法. 文献[21]提出了约束多项式参数学习(multinomial parameter learning with constraints, MPL-C)方法.

分析可以发现: CEM, DCE, MPL-C方法仅适用于几种特定类型的参数约束, 即约束适用性有限. 而CML和QMAP方法对参数约束类型没有特殊要求, 具有很好的约束适用性. CML和QMAP方法具有以下特点: 1) CML方法对样本数据利用率较高, 而对参数约束利用率较低, 学习结果往往位于参数可行域的边缘

位置. 2) QMAP方法对样本数据利用率较低, 而对参数约束利用率较高, 学习结果往往位于参数可行域的中心位置. 因此, 上述两种方法在样本量、约束量、真实参数位置不同的情况下互有优劣. 然而, 在样本数量大小、约束数量多少、真实参数位于可行域的边缘、中心或是两者之间等因素无法确定的情况下, CML和QMAP孰优孰劣? 笔者更无法确定. 为此, 本文提出一种自适应参数学习方法: 首先, 利用CML和QMAP方法进行参数学习得到两组参数. 然后, 自定义计算得到样本权重、约束权重、参数位置权重. 最后, 在两组参数和三组权重的基础上计算得到新的自适应参数解. 该方法通过合理分配对CML和QMAP方法学习结果的信任度, 使得在任何条件下, 都可以学习得到更加精确的参数.

2 研究问题描述(Problem description)

根据马尔科夫独立性, BN网络全局参数学习问题可以分解为每个节点与其父节点之间的局部参数学习问题. 因此, 以图1所示二值网络^[22]中参数 $P(W | S, R)$ 的学习为例阐明本文的研究问题.

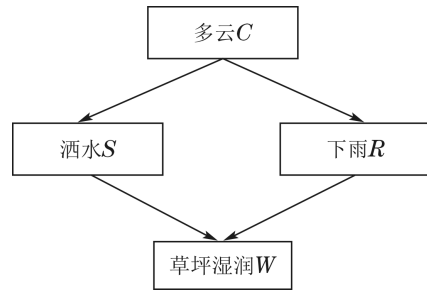


图1 草坪湿润推理贝叶斯网络
Fig. 1 Lawn wet inference Bayesian network

首先, 为了获取数据以进行参数学习, 将 $P(C)$, $P(S|C)$, $P(R|C)$ 参数值设置如表1所示.

表1 部分网络真实参数
Table 1 Part of the true parameters

$P(C = 1) = 0.50$	$P(C = 2) = 0.50$
$P(S = 1 C = 1) = 0.10, P(R = 1 C = 1) = 0.99$	$P(S = 1 C = 2) = 0.20, P(R = 1 C = 2) = 0.20$
$P(S = 2 C = 1) = 0.90, P(R = 2 C = 1) = 0.01$	$P(S = 2 C = 2) = 0.80, P(R = 2 C = 2) = 0.80$

然后, 分析参数 $P(W|S, R)$ 所满足的具体约束并建立如表2所示复合参数约束库. 最后, 通过3组

实验引出所要研究的问题. 其中, 为便于表示, 将表2中参数进行编号, 具体编号如表3所示.

表2 参数 $P(W|S, R)$ 所满足的约束
Table 2 Constraints satisfied by parameter $P(W | S, R)$

$P(W = 1 S = 1, R = 1) > P(W = 2 S = 1, R = 1)$
$P(W = 2 S = 2, R = 1) > P(W = 1 S = 2, R = 1), P(W = 2 S = 2, R = 2) > P(W = 2 S = 1, R = 1)$
$P(W = 2 S = 1, R = 2) > P(W = 2 S = 1, R = 1), P(W = 2 S = 2, R = 2) > P(W = 1 S = 2, R = 2)$
$P(W = 2 S = 1, R = 2) > P(W = 1 S = 1, R = 2), P(W = 2 S = 2, R = 2) > P(W = 2 S = 2, R = 1)$
$P(W = 2 S = 2, R = 1) > P(W = 2 S = 1, R = 1), P(W = 2 S = 2, R = 2) > P(W = 2 S = 1, R = 2)$

表 3 参数编号

Table 3 Parameter index

序号	参数
P_1	$P(W = 1 S = 1, R = 1)$
P_2	$P(W = 1 S = 1, R = 2)$
P_3	$P(W = 1 S = 2, R = 1)$
P_4	$P(W = 1 S = 2, R = 2)$
P_5	$P(W = 2 S = 1, R = 1)$
P_6	$P(W = 2 S = 1, R = 2)$
P_7	$P(W = 2 S = 2, R = 1)$
P_8	$P(W = 2 S = 2, R = 2)$

1) 约束数量对CML和QMAP参数学习的影响分析.

为公平比较两种方法, 进行如下实验设置:

① 样本数量设定为介于1和35之间的18组.

② 真实参数在参数约束确定后设定在介于可行域中心和边缘之间的中间位置. 在约束量为1条和9条的情况下分别利用CML和QMAP方法进行参数学习, 学习结果如表4和表5所示.

注 1 本文采用如式(1)所示的KL散度^[23]来度量CML方法和QMAP方法学习结果的精度. KL值越小表明学习结果越精确. 其中: θ_{ijk} 代表真实参数, θ'_{ijk} 代表学习得到的参数.

$$KL(\theta_i | \theta'_i) = \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} \theta_{ijk} \ln \frac{\theta_{ijk}}{\theta'_{ijk}}. \quad (1)$$

表 4 约束数量为1时学习结果对比

Table 4 Comparison of parameter learning with 1 constraint

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	KL距离
真实参数	0.50	0.20	0.75	0.78	0.50	0.80	0.25	0.22	0
CML估计	1	0.72	0	1	0	0.28	1	0	0.33
QMAP估计	0.68	0.36	0.50	0.50	0.32	0.64	0.50	0.50	0.62

表 5 约束数量为9时学习结果对比

Table 5 Comparison of parameter learning with 9 constraints

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	KL距离
真实参数	0.77	0.40	0.30	0.20	0.23	0.60	0.70	0.80	0
CML估计	0.50	0.40	0.50	0	0.50	0.60	0.50	1	0.95
QMAP估计	0.74	0.31	0.32	0.12	0.26	0.69	0.68	0.88	0.07

注 2 实验设置①中样本数量设定在1组和35组之间的原因是, 精确学习图1网络参数所需最小样本量为35^[24].

分析可以发现: 在样本量和真实参数位置相同的情况下, 当约束较少时, CML方法学习结果更加靠近真实参数; 当约束较多时, QMAP方法学习结果更加靠近真实参数.

2) 样本数量对CML和QMAP参数学习的影响

分析.

为公平比较两种方法, 进行如下实验设置:

① 约束数量设定为介于1和9之间的5条.

② 真实参数在约束确定后设定在可行域的中间位置. 在样本量为1组和35组的情况下, 分别利用CML和QMAP方法进行参数学习, 学习结果如表6和表7所示.

表 6 样本量为1时学习结果对比

Table 6 Comparison of parameter learning when sample size is 1

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	KL距离
真实参数	0.93	0.80	0.45	0.40	0.07	0.20	0.55	0.60	0
CML估计	0.50	0	0	0	0.50	1	1	1	5.08
QMAP估计	0.78	0.40	0.32	0.18	0.22	0.60	0.68	0.82	0.85

表7 样本量为35时学习结果对比

Table 7 Comparison of parameter learning when sample size is 35

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	KL距离
真实参数	0.88	0.62	0.42	0.32	0.12	0.38	0.58	0.68	0
CML估计	0.86	0.86	0.39	0.39	0.14	0.14	0.61	0.61	0.07
QMAP估计	0.78	0.40	0.32	0.18	0.22	0.60	0.68	0.82	0.85

分析可以发现: 在约束量和真实参数位置相同的情况下, 当样本较少时, QMAP方法学习结果更加靠近真实参数; 而当样本较多时, CML方法学习结果更加靠近真实参数.

3) 真实参数位置对CML和QMAP参数学习的影响分析.

为公平比较两种方法, 进行如下实验设置:

① 样本数量设定为18组.

② 约束数量设定为5条. 在真实参数位于可行域的中心和边缘位置的情况下, 分别利用CML方法和QMAP方法进行参数学习, 学习结果如表8和表9所示.

表8 真实参数位于可行域中心位置时学习结果对比

Table 8 Comparison of parameter learning when true parameters locate in center of constrained region

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	KL距离
真实参数	0.81	0.39	0.25	0.29	0.19	0.61	0.75	0.71	0
CML估计	1	0.90	0.50	0.45	0	0.10	0.50	0.55	1.95
QMAP估计	0.78	0.39	0.25	0.22	0.22	0.61	0.75	0.78	0.02

表9 真实参数位于可行域边缘位置时学习结果对比

Table 9 Comparison of parameter learning when true parameters locate on border of constrained region

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	KL距离
真实参数	0.98	1	0.48	0.50	0.02	0	0.52	0.50	0
CML估计	1	0.90	0.50	0.45	0	0.10	0.50	0.55	0.12
QMAP估计	0.78	0.39	0.25	0.22	0.22	0.61	0.75	0.78	2.01

分析可以发现: 在样本量和约束量相同的情况下, 当真实参数位于可行域的中心位置时, QMAP方法学习结果更加靠近真实参数; 而当真实参数位于可行域的边缘位置时, CML方法学习结果更加靠近真实参数.

总结可以发现: CML和QMAP方法在样本量、约束量、真实参数位置不同的情况下学习结果互有优劣, 导致方法的难以选择. 为此, 本文提出一种小数据集条件下BN自适应参数学习方法.

3 BN自适应参数学习方法(BN adaptive parameter learning method)

首先, 给出用于融合CML和QMAP方法优点的3个权重的定义、计算方法、特征分析. 然后, 介绍CML和QMAP方法的核心参数优化模型. 最后, 详

细描述自适应参数学习方法的步骤.

3.1 权重定义(Weight definitions)

3.1.1 约束权重(Constraint weight)

约束权重是评估约束量多少或约束力强弱的权重因子, 是分配对CML和QMAP方法学习结果信任度的第一个指标. 约束权重越小表明约束量越多或约束力越强. 该权重计算方法为:

首先, 随机生成 N_1 组满足如下约束的基本参数组 $(\theta_{i1k}, \theta_{i2k}, \dots, \theta_{iqik})$:

$$\theta_{ijk} \in (\Omega_1 \cap \Omega_2), j = 1, 2, \dots, q_i, \quad (2)$$

其中: θ_{ijk} 代表节点 i 取第 j 个状态值且其父节点取第 k 个组合状态值时的参数值, Ω_1, Ω_2 代表如下约束:

$$\theta_{ijk} \in [0, 1], j = 1, 2, \dots, q_i, \quad (3)$$

$$(\theta_{i1k} + \theta_{i2k} + \dots + \theta_{iq_i k}) = 1. \quad (4)$$

然后, 统计进一步满足如下约束的参数的数量并记为 N_2 . 其中: 约束 Ω_C 代表复合参数约束,

$$\theta_{ijk} \in (\Omega_1 \cap \Omega_2 \cap \Omega_C), j = 1, 2, \dots, q_i. \quad (5)$$

最后, 通过式(6)计算得到约束权重

$$w_C = \frac{N_2}{N_1}. \quad (6)$$

以表2中参数约束为例, 分析约束权重与约束数量之间的关系, 具体关系如图2所示.

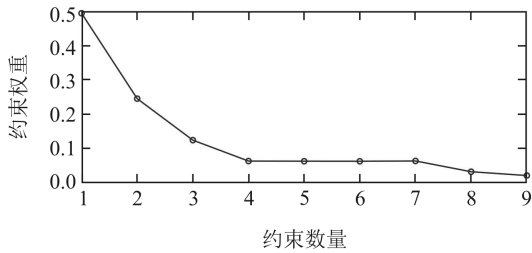


图 2 约束权重与约束数量之间的关系

Fig. 2 Relation between constraint weight and number

分析可以发现: 约束权重随约束量的增加而减小, 实现了对CML和QMAP学习结果信任度的分配.

3.1.2 样本权重(Sample weight)

样本权重是评估样本量多少的权重因子, 是分配对CML和QMAP方法学习结果信任度的第2个指标. 样本权重越大表明样本量越多. 该权重计算方法为:

首先, 根据式(7)^[23]计算得到学习网络参数所需最小样本量 N_3 . 其中: n 为全局网络节点数量, m 为全局网络中最大父节点数,

$$N_3 = n^2 \log n^2 \log n^{(m+1)}. \quad (7)$$

然后, 根据式(8)计算得到样本权重

$$w_S = \min\{1, \frac{N_4}{N_3}\}, \quad (8)$$

其中 N_4 为已知样本量. 以图1网络参数为例, 分析样本权重与样本量之间的关系, 具体关系如图3所示.

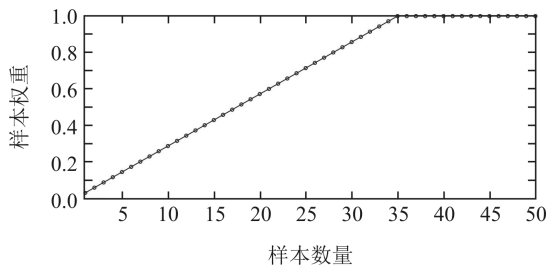


图 3 样本权重与样本数量之间的关系

Fig. 3 Relation between sample weight and size

分析可以发现: 样本权重随样本数量的增大而

增大, 实现了对CML和QMAP学习结果信任度的分配.

3.1.3 参数位置权重(Parameter location weight)

真实参数位置权重是评估真实参数距离参数可行域边界远近的权重因子, 是分配对CML和QMAP方法学习结果信任度的第3个指标. 真实参数位置权重越大表明真实参数距离可行域边界越远. 由于真实参数位置无法确定, 所以将真实参数位置权重设定为0.5, 即

$$w_L = 0.5. \quad (9)$$

上述设定依据为: 从统计和几何的角度分析, 真实参数位于参数可行域中心和边界位置的概率小于位于中间位置的概率(如图4, 点散落于BC间的概率(0.5)大于散落于AB和CD间的概率(0.25)). 因此, 将真实参数位置权重设定为0.5.

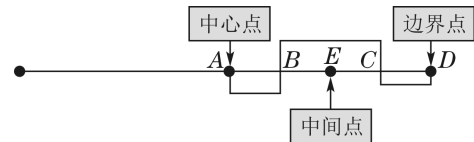


图 4 真实参数位置权重分析示意图

Fig. 4 Analysis of true parameter location weight

3.2 参数优化模型(Parameter optimization models)

3.2.1 有约束参数优化模型(Constrained parameter optimization model)

CML方法将BN参数学习问题视为有约束优化问题, 其参数优化模型^[16-17]为

$$\max \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} N_{ijk} \log \theta_{ijk}, \quad (10)$$

$$\text{s.t. } \Omega(\theta_{ijk}) \leq 0, \quad (11)$$

其中 N_{ijk} 代表样本数据中第 i 个节点取第 j 个状态值且其父节点取第 k 个组合状态值的样本数量.

3.2.2 无约束参数优化模型(Unconstrained parameter optimization model)

QMAP方法将BN参数学习问题视为无约束优化问题, 其参数优化模型^[19]为

$$\max \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} (N_{ijk} + M_{ijk}) \log \theta_{ijk}, \quad (12)$$

其中 M_{ijk} 代表第 i 个节点取第 j 个状态值且其父节点取第 k 个组合状态值的虚拟样本数量, 大小为

$$M_{ijk} = A * P(X_i = k, \Pi(i) = j | \Omega), \quad (13)$$

其中 A 代表虚拟采样次数.

3.3 算法描述(Algorithm description)

基于上述权重和模型, 小数据集条件下BN自适应参数学习方法的具体步骤如下:

步骤1 统计样本量 N_{ijk} , 然后根据已知参数约束构建有约束参数优化模型即式(10)和(11);

步骤2 利用凸优化方法^[25]求解上述优化模型, 得到参数值 $\theta_{ijk}^{\text{CML}}$;

步骤3 利用拒绝-接受采样法随机抽取满足约束 Ω 的参数 $P(X_i = j, \Pi(i) = k | \Omega)$, 然后根据式(13)计算得到参数 M_{ijk} ;

步骤4 利用 N_{ijk} 和 M_{ijk} 构建无约束参数优化模型即式(12), 求解得到如下参数:

$$\theta_{ijk}^* = \frac{N_{ijk} + M_{ijk}}{\sum_{k=1}^{r_i} (N_{ijk} + M_{ijk})}. \quad (14)$$

步骤5 将步骤3和4重复进行 L 次, 然后将 L 次结果平均化得到参数值 $\theta_{ijk}^{\text{QMAP}}$;

步骤6 根据式(2)-(9)计算约束权重、样本权重、参数位置权重, 然后根据式(15)-(17), 最终得到自适应参数值 $\theta_{ijk}^{\text{Adaptive}}$:

$$\theta_{ijk}^{\text{Adaptive}} = w_1 * \theta_{ijk}^{\text{CML}} + w_2 * \theta_{ijk}^{\text{QMAP}}, \quad (15)$$

$$w_1 = \frac{2 - w_C + w_S - w_L}{3}, \quad (16)$$

$$w_2 = \frac{1 + w_C - w_S + w_L}{3}. \quad (17)$$

4 复杂度分析(Complexity analysis)

MLE, CML, QMAP和本文方法复杂度为

$$T_{\text{MLE}} = N_4, \quad (18)$$

$$T_{\text{CML}} = \frac{1}{\varepsilon} + N_4, \quad (19)$$

$$T_{\text{QMAP}} = L * N_s + N_4, \quad (20)$$

$$T_{\text{Adaptive}} = \frac{1}{\varepsilon} + L * N_s + N_1 + 2 * N_4. \quad (21)$$

注3 ε 代表计算精度, 其默认值为 $2.22 * 10^{-16}$, N_1 取值为 10^3 左右, N_s 取值为 10^2 左右, L 取值为 10^2 左右.

结论 小数据集条件下, MLE, CML, QMAP, 本文方法复杂度之间的大小关系为

$$T_{\text{Adaptive}} > T_{\text{CML}} > T_{\text{QMAP}} > T_{\text{MLE}}, \quad (22)$$

且本文方法和CML方法的复杂度相近, 即

$$T_{\text{Adaptive}} \approx T_{\text{CML}}. \quad (23)$$

证 由于 N_4 , $L * N_s$ 和 N_1 远小于 $\frac{1}{\varepsilon}$, 故根据式(18)-(21)可知MLE, CML, QMAP, 本文方法复杂度满足如下大小关系:

$$T_{\text{Adaptive}} > T_{\text{CML}} > T_{\text{QMAP}} > T_{\text{MLE}}. \quad (24)$$

而小数据集条件下 N_4 往往很小, 且 $L * N_s$, N_1 远小

于 $\frac{1}{\varepsilon}$, 故保留算法复杂度主项后, 可得

$$T_{\text{MLE}} \approx O(1), \quad (25)$$

$$T_{\text{CML}} \approx O\left(\frac{1}{\varepsilon}\right), \quad (26)$$

$$T_{\text{Adaptive}} \approx O\left(\frac{1}{\varepsilon}\right), \quad (27)$$

$$T_{\text{QMAP}} \approx O(L * N_s), \quad (28)$$

故可得 $T_{\text{Adaptive}} \approx T_{\text{CML}}$.

5 仿真实验(Simulations)

实验平台为MATLAB, 实验网络为图1网络, 实验约束为表2约束, 实验方法为MLE, CML, QMAP, 本文方法, 具体实验如下.

5.1 有效性分析(Effective analysis)

5.1.1 约束数量变化实验(Learning with different constraint number)

实验设置:

1) 样本量设为18组.

2) 约束量由1条依次递增至9条.

3) 真实参数在实验约束确定后设定在参数可行域的中间地带.

实验内容: 首先, 按照上述设置利用不同方法进行单次学习实验, 实验结果如图5(a)和5(b)所示. 然后, 按照上述实验设置重复进行50次学习实验, 平均实验结果如图5(c)和5(d)所示.

实验分析:

1) 约束较少时, CML学习精度优于QMAP; 约束较多时, QMAP学习精度优于CML.

2) 本文方法学习精度在多数情况下优于CML和QMAP方法.

5.1.2 样本数量变化实验 (Learning with different sample size)

实验设置:

1) 约束量设为5条.

2) 样本量由1组递增至35组.

3) 真实参数在实验约束确定以后设定在参数可行域的中间地带.

实验内容: 首先, 按照上述设置利用不同方法进行单次学习实验, 结果如图6(a)和6(b)所示. 然后, 按照上述实验设置重复进行50次学习实验, 平均实验结果如图6(c)和6(d)所示.

实验分析:

1) 样本较少时, QMAP学习结果优于CML; 样本较多时, CML学习结果优于QMAP.

2) 本文方法学习精度在多数情况下优于CML和QMAP方法.

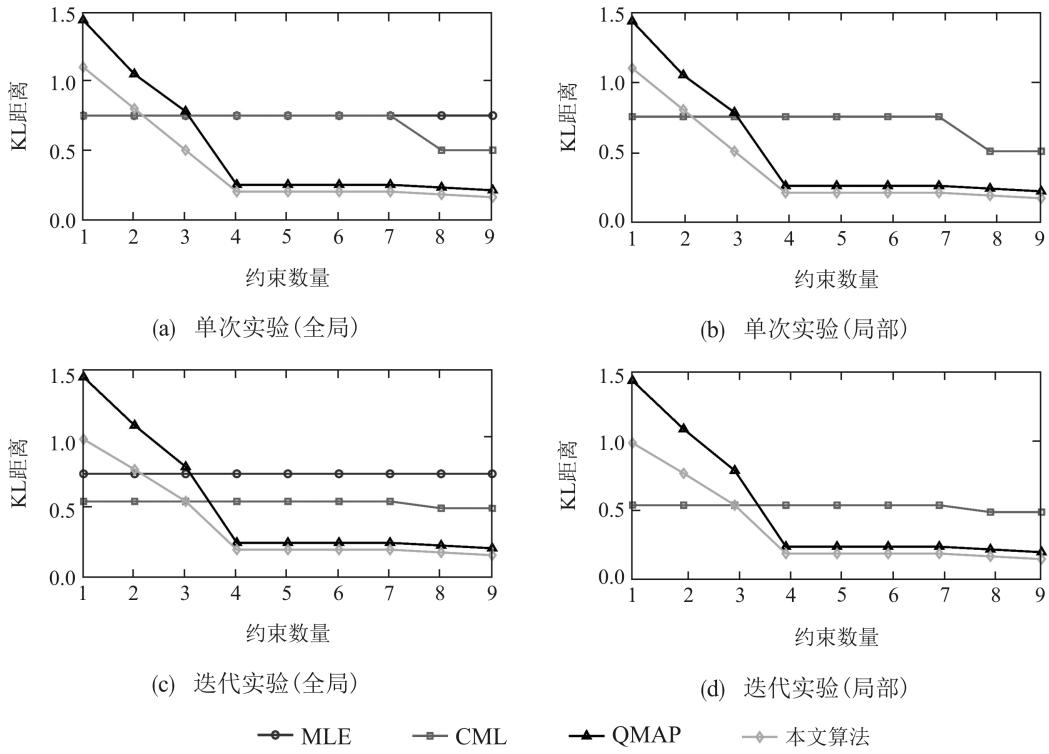


图 5 约束数量变化实验

Fig. 5 Effectiveness analysis-I

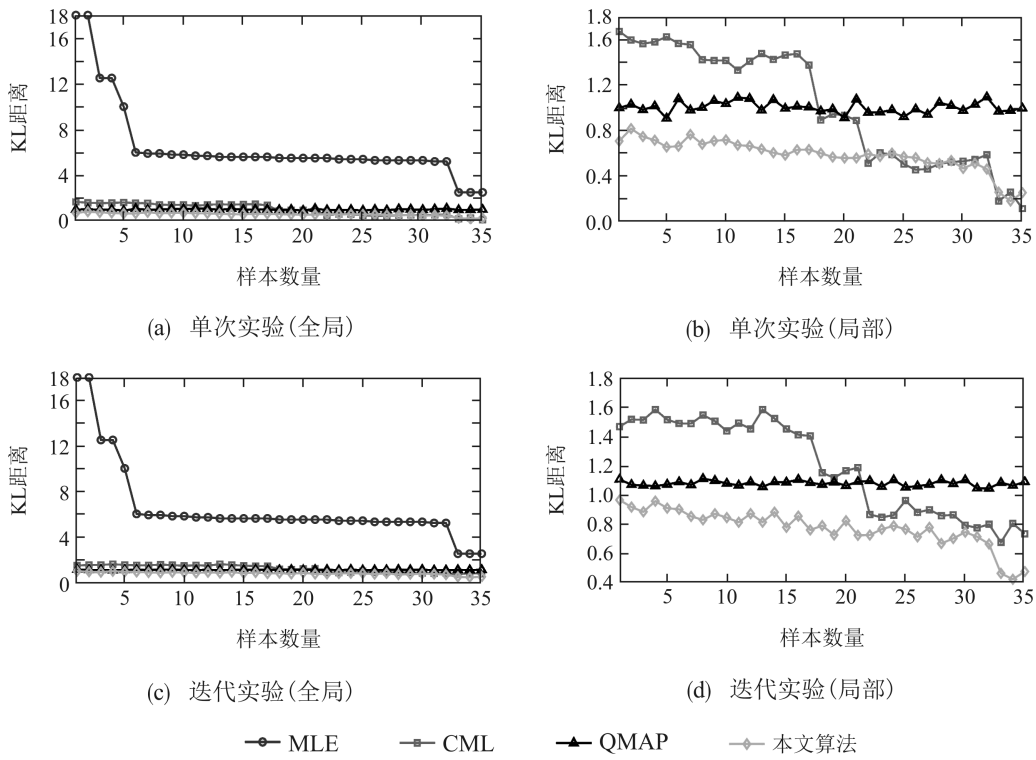


图 6 样本数量变化实验

Fig. 6 Effectiveness analysis-II

5.1.3 真实参数位置变化实验(Learning with different true parameter location)

实验设置:

- 1) 约束量设为5条.
- 2) 样本量设为18组.

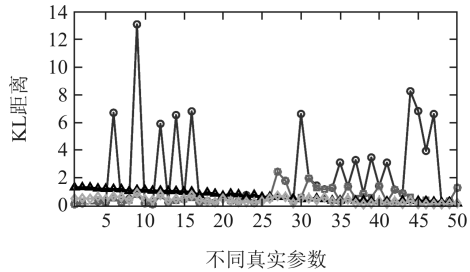
3) 真实参数在实验约束确定以后由参数可行域的边缘位置向中心位置逐渐变化.

实验内容: 首先, 按照上述设置利用不同方法进行单次学习实验, 结果如图7(a)和7(b)所示. 然后, 按照上述实验设置重复进行50次学习实验, 平均实验结果如图7(c)和7(d)所示.

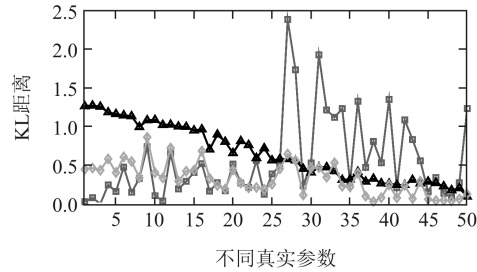
实验分析:

1) 当真实参数位于参数可行域边缘位置时, CML学习结果优于QMAP; 当真实参数位于参数可行域中心位置时, QMAP学习结果优于CML.

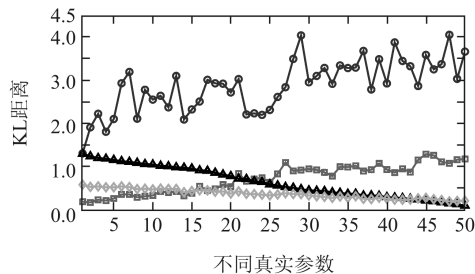
2) 本文方法学习精度在多数情况下优于CML和QMAP方法.



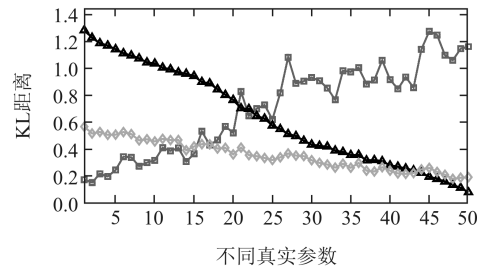
(a) 单次实验(全局)



(b) 单次实验(局部)



(c) 迭代实验(全局)

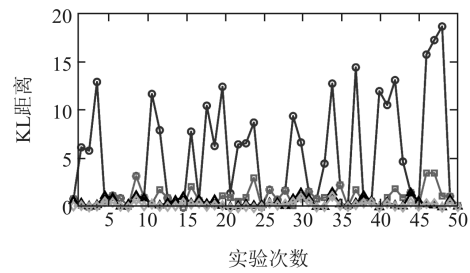


(d) 迭代实验(局部)

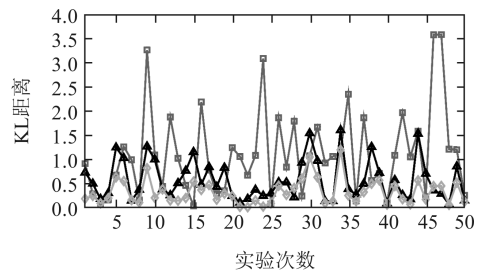
—○— MLE —□— CML —▲— QMAP —◇— 本文算法

图7 真实参数位置变化实验

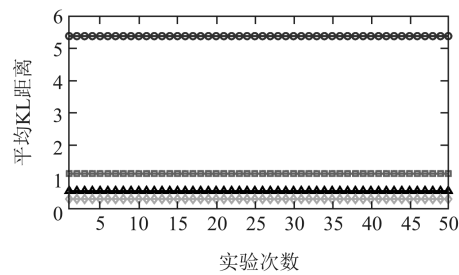
Fig. 7 Effectiveness analysis-III



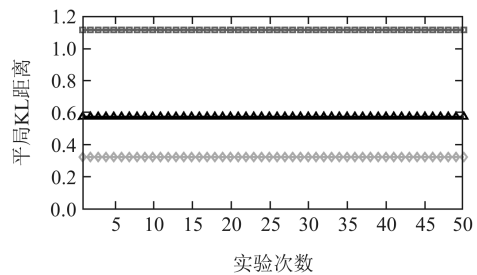
(a) 全局结果



(b) 局部结果



(c) 全局结果



(d) 局部结果

—○— MLE —□— CML —▲— QMAP —◇— 本文算法

图8 任意条件变化实验

Fig. 8 Effectiveness analysis-IV

5.1.4 随机条件变化实验(Learning under any conditions)

实验设置: 1) 约束量随机生成(介于1和9之间)并从约束库中随机抽取.

2) 真实参数在约束确定后从参数可行域中随机选定.

3) 样本量随机生成(介于1和35之间)并从网络中随机抽取.

实验内容: 首先, 按照上述设置利用不同方法重复进行50次实验, 结果如图8(a)和8(b)所示. 然后, 将学习结果平均化, 结果如图8(c)和8(d)所示.

实验分析: 1) 本文方法学习精度在多数情况下优于CML和QMAP方法.

2) 本文方法学习结果平均KL值(0.33)低于CML(1.12)和QMAP(0.58).

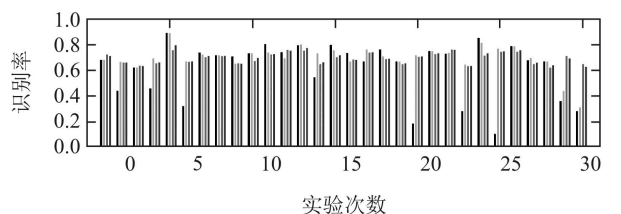
5.2 实用性分析(Practicability analysis)

实验设置: 1) 约束量随机生成(介于1和9之间)并从约束库中随机抽取.

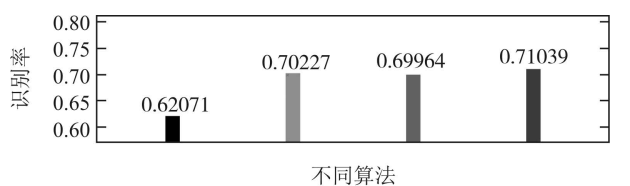
2) 真实参数在约束确定后从参数可行域中随机选定.

3) 样本量随机生成(介于1和35之间)并从网络中随机抽取.

实验内容: 首先, 抽取两份样本数据: 训练样本和测试样本. 其中, 训练样本量为实验设置中的随机样本数量, 测试样本量为100. 然后, 基于训练样本和参数约束利用不同方法进行参数学习. 最后, 基于学习结果和测试样本进行推理并记录推理结果即识别率. 实验重复30次并将实验结果平均化, 结果分别如图9(a)和9(b)所示.



(a) 随机实验结果



(b) 随机实验结果平均值

图 9 实用性分析
Fig. 9 Practicability analysis

实验分析:

1) 基于本文方法学习结果的推理精度接近于MLE, CML, QMAP三者的最优值.

2) 基于本文方法学习结果的平均推理精度(71.039%)高于其他3种方法(65.072%, 70.227%, 69.964%).

5.3 实时性分析(Real-time analysis)

实验设置:

1) 约束量随机生成(介于1和9之间)并从约束库中随机抽取.

2) 真实参数在约束确定后从参数可行域中随机选定.

3) 样本量随机生成(介于1和35之间)并从网络中随机抽取.

实验内容: 首先, 按照上述设置利用不同方法重复进行50次学习实验并记录时间消耗, 统计结果如图10(a)所示. 然后, 将统计结果平均化, 平均结果如图10(b)所示.

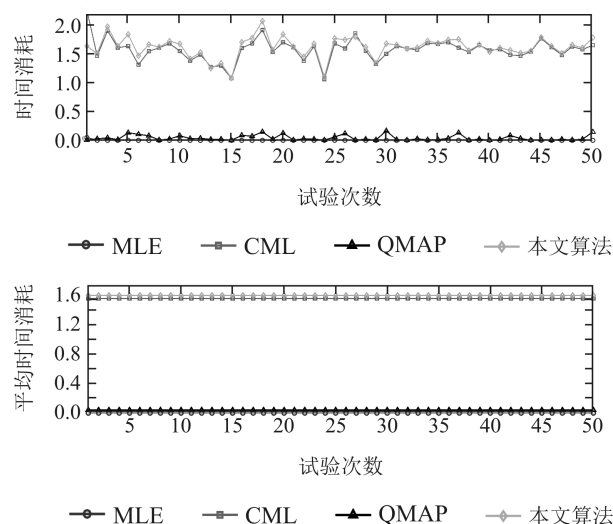


图 10 实时性分析

Fig. 10 Real-time analysis

实验分析:

1) MLE和QMAP方法学习时间较短, 计算复杂度相对较低; CML和本文方法学习时间较长, 计算复杂度相对较高.

2) 本文方法的学习时间即计算复杂度与CML方法接近.

5.4 稳定性分析(Stability analysis)

实验设置:

1) 约束量随机生成(介于1和9之间)并从约束库中随机抽取.

2) 真实参数在约束确定后从参数可行域中随机选定.

3) 样本量随机生成(介于1和35之间)并从网络中随机抽取.

实验内容: 首先, 按照上述设置利用不同方法重复进行100次学习实验. 然后, 计算不同方法学习结果波动率, 结果如图11所示. 最后, 将不同方法波动

率进行平均化, 结果如图11中黑色虚线所示.

注4 参数学习结果波动率计算方法定义如下:

$$\phi_i = \frac{|KL_i - KL_{average}|}{KL_{max}}, 1 \leq i \leq r, \quad (29)$$

其中: KL_i 代表第*i*次学习的KL值, $KL_{average}$ 代表*r*次学习的平均KL值, KL_{max} 代表*r*次学习的最大KL值.

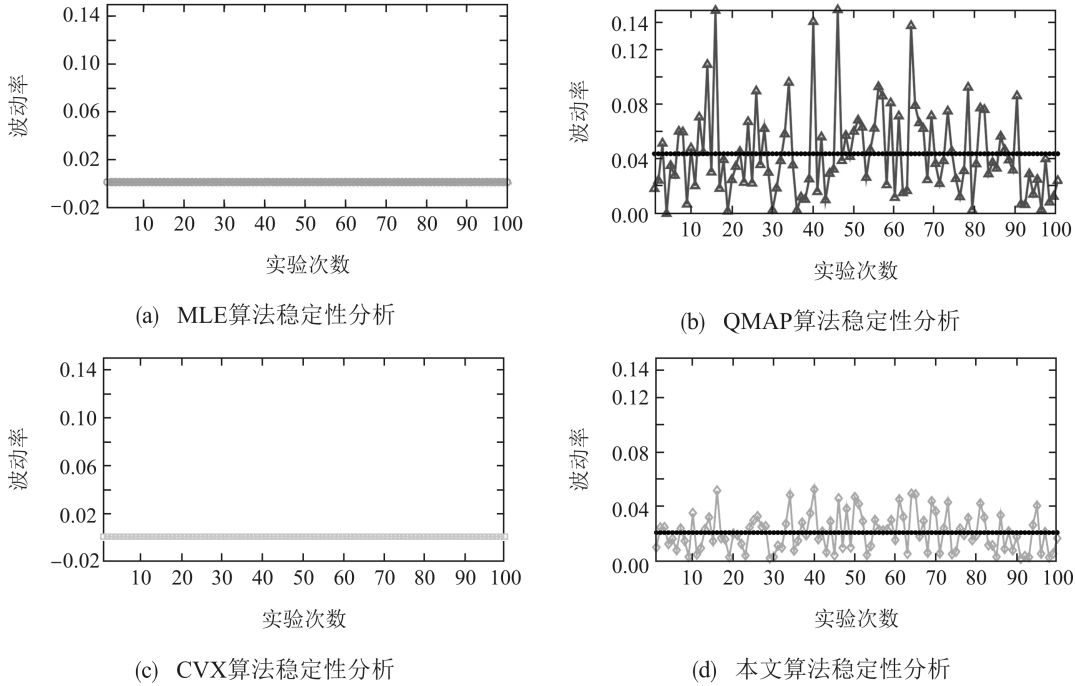


图 11 稳定性分析

Fig. 11 Stability analysis

实验分析:

1) MLE和CML方法学习结果波动率为0, 为稳定性学习方法; QMAP和本文方法学习结果波动率在0~10%之间, 为不稳定性学习方法.

2) 本文方法学习结果平均波动率(1.0%) 低于QMAP方法(3.0%), 稳定性更好.

6 结论(Conclusions)

针对小数据集条件下BN参数学习问题, 提出一种适用于任何类型参数约束的BN自适应参数学习方法. 本文贡献概括如下:

1) 指出了两种约束适用性较好的参数学习方法—CML和QMAP方法各自的优缺点, 为提出更好的参数学习方法指明了方向.

2) 自定义了3种权重: 约束权重、样本权重、参数位置权重, 为融合上述两种方法的优点提供了良好的手段.

3) 给出了1种BN自适应参数学习方法, 该方法参数学习精度接近甚至优于CML和QMAP方法中的最优者.

针对小数据集条件下BN参数学习问题, 已有研

究方法将上述问题视为确定性优化问题, 进而将唯一最优解作为最优参数. 然而, 由于小数据集条件下样本数据较少, 利用样本数据构造的优化函数(如似然函数、熵函数)往往并不精确. 因此, 优化函数的最优解往往无法逼近真实参数, 而某些次优解却能较好地逼近真实参数. 因此, 寻找合理的次优解将是值得尝试的研究思路.

参考文献(References):

[1] PEARL J. *Probabilistic Reasoning in Intelligent Systems* [M]. Massachusetts: Morgan Kaufmann, 1988.

[2] TAMDA Y, IMOTO S, ARAKI H. Estimating genome-wide gene networks using nonparametric Bayesian network models on massively parallel computers [J]. *IEEE Transactions on Computational Biology and Bioinformatics*, 2011, 3(8): 683 – 697.

[3] IBRAHIM W, BEIU V. Using bayesian networks to accurately calculate the reliability of complementary metal oxide semiconductor gates [J]. *IEEE Transactions on Reliability*, 2011, 60(3): 538 – 549.

[4] DUAN Yong, CUI Baoxia, XU Xinhe. Multi-agent reinforcement learning and its application to role assignment of robot soccer [J]. *Control Theory & Applications*, 2009, 26(4): 371 – 376. (段勇, 崔宝侠, 徐心和. 多智能体强化学习及其在足球机器人角色分配中的应用 [J]. *控制理论与应用*, 2009, 26(4): 371 – 376.)

- [5] STEVEN M, ANN N, KEVIN K. Anomaly detection in vessel tracks using Bayesian networks [J]. *International Journal of Approximate Reasoning*, 2014, 55(1): 84 – 98.
- [6] NEIL W, MAHMOOD R. Detection and classification of non-stationary transient signals using sparse approximations and Bayesian networks [J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2014, 22(12): 1750 – 1764.
- [7] DRIES L, STEVEN B, ROB D, et al. A review of Bayesian belief networks in ecosystem service modeling [J]. *Environmental Modelling and Software*, 2013, 46(1): 1 – 11.
- [8] DRUZDZEL M. *Probabilistic reasoning in decision support systems: from computation to common sense* [D]. Pittsburgh: Carnegie Mellon University, 1995.
- [9] DRUZDZEL M, GAAG L. Building probabilistic networks: where do the numbers come from? [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2000, 12(4): 481 – 486.
- [10] REDNER R, WALKER H. Mixture densities, maximum likelihood and the EM algorithm [J]. *SIAM Review*, 1984, 26(2): 195 – 239.
- [11] HELSPER E, GAAG L, GROENENDAL F. Designing a procedure for the acquisition of probability constraints for Bayesian networks [C] // *Proceedings of the Fourteenth Conference on Engineering Knowledge in the Age of the Semantic Web*. Northampton: IEEE, 2004, 8: 280 – 292.
- [12] WITTIG F, JAMESON A. Exploiting qualitative knowledge in the learning of conditional probabilities of Bayesian networks [C] // *Proceedings of the Sixteenth International Conference on Uncertainty in Artificial Intelligence*. California: IEEE, 2000, 7: 644 – 652.
- [13] ALTENDORF E, RESTIFICAR A, DIETTERICH T. Learning from sparse data by exploiting monotonicity constraints [C] // *Proceedings of the Twenty First International Conference on Uncertainty in Artificial Intelligence*. Edinburgh: IEEE, 2005, 7: 18 – 26.
- [14] FEELDERS A, GAAG L. Learning Bayesian networks parameters under order constraints [J]. *International Journal of Approximate Reasoning*, 2006, 42(1/2): 37 – 53.
- [15] 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.
- [16] NICULESCU R, MITCHELL M, RAO B. Bayesian network learning with parameter constraints [J]. *Journal of Machine Learning Research*, 2006, 7(1): 1357 – 1383.
- [17] CAMPOS C, QIANG J. Constrained maximum likelihood learning of Bayesian networks for facial action recognition [C] // *Proceedings of the Tenth European Conference on Computer Vision*. Marseille: IEEE, 2008, 8: 168 – 181.
- [18] WEIHUA L, QIANG J. Learning Bayesian network parameters under incomplete data with domain knowledge [J]. *Pattern Recognition*, 2009, 42(11): 3046 – 3056.
- [19] RUI C, SHOEMAKER R, WEI W. A novel knowledge-driven systems biology approach for phenotype prediction upon genetic intervention [J]. *IEEE Transactions on Computational Biology and Bioinformatics*, 2011, 1(8): 1170 – 1181.
- [20] GUO Zhigao, GAO Xiaoguang, DI Ruohai. Learning Bayesian network parameters under dual constraints from small data set [J]. *Acta Automatica Sinica*, 2014, 40(7): 1509 – 1516.
(郭志高, 高晓光, 邸若海. 小数据集条件下基于双重约束的贝叶斯网络参数学习 [J]. *自动化学报*, 2014, 40(7): 1509 – 1516.)
- [21] YUN Z, FENTON M, NEIL M. Bayesian network approach to multinomial parameter learning using data and expert judgments [J]. *International Journal of Approximate Reasoning*, 2014, 40(7): 1509 – 1516.
- [22] MURPHY K. *Dynamic Bayesian networks: representation, inference and learning* [D]. California: University of California, Berkeley, 2002.
- [23] KULLBACK S, LEIBLER R. On information and sufficiency: annals of mathematical statistics [J]. *Annals of Mathematical Statistics*, 1951, 22(1): 79 – 86.
- [24] HOFFGEN K. Learning and robust learning of product distributions [C] // *Proceedings of the Sixth Annual Conference on Computational Learning Theory*. New York: IEEE, 1993, 8: 77 – 83.
- [25] BOYD S, VANDENBERGHE L. *Convex Optimization* [M]. Cambridge: Cambridge University Press, 2004.

作者简介:

郭志高 (1986–), 男, 博士研究生, 目前研究方向为小数据集条件下贝叶斯网络参数学习, E-mail: guozhigao2004@163.com;

高晓光 (1957–), 女, 教授, 博士生导师, 目前研究方向为智能决策、复杂系统建模与效能分析, E-mail: cxg2012@nwpu.edu.cn;

邸若海 (1986–), 男, 博士研究生, 目前研究方向为小数据集条件下贝叶斯网络结构和参数学习, E-mail: xfwtdrh@163.com.