

文章编号: 1001-0920(2009)04-0598-03

可分解马尔可夫网络参数学习中的噪声平滑

王双成^{1a}, 李小琳², 刘凤霞^{1b}

(1. 上海立信会计学院 a. 信息科学系, b. 工商管理系, 上海 201600; 2. 南京大学 商学院, 南京 210093)

摘要: 对于存在噪声的可分解马尔可夫网络参数学习, 目前可通过调整可分解马尔可夫网络的结构来增强其抗噪声能力, 但调整后的结构往往会降低网络的有效性和可靠性, 而且不能从源头上减少噪声对参数的影响. 针对该问题, 首先建立基于可分解马尔可夫网络理论的邻域最优预测; 然后以变量为基本单位, 使用预测值平滑噪声. 实验结果显示, 使用平滑后的数据进行参数学习, 其可靠性能得到显著的改进.

关键词: 可分解马尔可夫网络; 参数学习; 邻域预测; 噪声平滑

中图分类号: TP181 **文献标识码:** A

Noise smoothing in learning the parameters of decomposable Markov net works

WANG Shuang-cheng^{1a}, LI Xiaolin², LIU Feng-xia^{1b}

(1a. Department of Information Science, 1b. Department of Business Administration, Shanghai Lixin University of Commerce, Shanghai 201600, China; 2. School of Business, Nanjing University, Nanjing 210093, China. Correspondent: WANG Shuang-cheng, E-mail: wangsc@lixin.edu.cn)

Abstract: At present, the influence of noise can be reduced by simplifying the structure in the parameter learning of decomposable Markov network. But regulated structure will decrease the network reliability and the influence of noise can not be decreased or removed from root. In this paper, an effective method of smoothing noise is presented. Firstly, optimal neighborhood prediction is built based on the theory of decomposable Markov network. Then, a variable is taken as a basic unit and the noise can be smoothed by using prediction value to revise corresponding data. Finally, experimental results show that the reliability of learning parameters from smoothed data is greatly improved.

Key words: Decomposable Markov network; Parameter learning; Neighborhood prediction; Noise smoothing

1 引言

可分解马尔可夫网络^[1] (DMN) 是用于描述随机变量之间依赖关系的无向图形模式. 因为联合概率能够按其结构进行边缘分解, 并可避免进行复杂的势函数^[1] 计算, 所以, 在多 agents 之间的信息结构研究、图像和声音处理、多源信息融合、系统监测与控制、web 图的宏观与微观结构研究、变量之间的一般证据传递等^[2,3] 方面都有着广泛的应用.

对于具有噪声的 DMN 参数学习, 目前主要采用简化 DMN 结构的方法来提高 DMN 参数的抗噪声能力^[3-5] (如使用最大似然树替代 DMN). 当噪声比例较小时, 能够收到一定的效果, 但可能会将一些重要的依赖关系排除在外, 破坏 DMN 的完整性, 从而降低 DMN 模型的可靠性. 以上方法使 DMN 模

型去适应具有噪声的数据, 作法被动, 而且新噪声的出现会进一步降低模型的可靠性, 无法从数据源头上排除或减少噪声的影响.

基于 EM 算法^[6] 的思想, 首先建立基于可分解马尔可夫网络理论的邻域最优预测; 然后以变量为基本单位, 使用最优预测值平滑噪声 (最大似然平滑), 可从源头上减少噪声的影响. 实验结果显示, 使用平滑后的数据进行参数学习, 其可靠性得到了显著的改进.

2 邻域预测及最优性

本文作如下规定: 用 X_1, X_2, \dots, X_n 表示离散随机变量, 简称为变量, x_1, x_2, \dots, x_n 为其值; 已知数据集 D 中具有 N 个例子, 它们随机产生于概率分布 P , 对一个变量值的平滑简称为对这个变量的平滑;

收稿日期: 2008-02-26; 修回日期: 2008-05-06.

基金项目: 国家自然科学基金项目 (60675036); 上海市教委重点学科和上海市教委科研创新重点项目 (09ZZ202).

作者简介: 王双成 (1958 →), 男, 吉林省吉林人, 教授, 博士, 从事人工智能、机器学习等研究; 李小琳 (1978 →), 女, 长春人, 博士, 从事机器学习的研究.

在概率模式中的变量,以及表示概率模式的图形模式中的结点也不加以区分.

一般马尔可夫网络不能直接进行联合概率的边缘分解,需要三角剖分转化弦图^[1](将一般马尔可夫网络转化为 DMN),才可进行联合概率的边缘分解,转化与分解方法参见文献[1].

定义 1 在由 X_1, X_2, \dots, X_n 构成的 DMN 中,由与 X_i 直接相连接的结点构成的结点集称为 X_i 的邻域(用 $BL(X_i)$ 表示),它们与 X_i 构成的局部无向图称为 X_i 的邻域结构.

Pearl^[1] 建立了如下两个定理,为基于 DMN 的联合概率边缘分解奠定了基础.

定理 1 P 是可分解概率模式, G 是 P 的 DMN, Q_1, \dots, Q_k, \dots 为 G 中 Clique(最大完全子图)的顺序, q_1, \dots, q_k, \dots 为其配置,则有

$$p(x_1, \dots, x_n) = \prod_k \frac{p(q_k | G)}{p(q_k | q_{j(k)} | G)},$$

$Q_{j(k)}$ 是 Clique 结合树^[1] T 中 Q_k 的父结点 Clique.

定理 2 P 是可分解概率模式, G 是 P 的 DMN, 存在一棵 Clique 结合树 T , 对 G 中任意的结点 X_i , 使含有 X_i 的 Cliques 构成 T 的一棵子树 T_{X_i} .

理论上应该使用所有的变量平滑一个变量,即采用满条件分布 $p(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ 来平滑变量 X_i , 但满条件分布的计算复杂程度随变量增加指数增长, 对此, 基于 DMN 的联合概率分解可通过转化满条件分布为邻域边缘分布来解决这一问题. Q_1, \dots, Q_k, \dots 分成两组, 一组为 $Q_1^i, \dots, Q_k^i, \dots$ 都含有 X_i ; 另一组为 $Q_1^{n(i)}, \dots, Q_k^{n(i)}, \dots$ 都不含有 X_i .

$$p(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \frac{p(x_1, \dots, x_n)}{p(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)} = \frac{\prod_{Q_k^i} \frac{p(q_k^i)}{p(q_k^i | q_{j(k)}^i)}}{\prod_{Q_k^{n(i)}} \frac{p(q_k^{n(i)})}{p(q_k^{n(i)} | q_{j(k)}^{n(i)})}} = \frac{\prod_{Q_k^i} \frac{p(q_k^i)}{p(q_k^i | q_{j(k)}^i)}}{\prod_{Q_k^{n(i)}} \frac{p(q_k^{n(i)})}{p(q_k^{n(i)} | q_{j(k)}^{n(i)})}}.$$

其中: $bl(x_i)$ 和 q_k^i 分别为 $BL(X_i)$ 和 Q_k^i 的配置; $Q_1^i, \dots, Q_k^i, \dots$ 为 Clique 结合树 T 中 Clique 的顺序; $Q_{j(k)}^i$ 为 T 中 Q_k^i 的父结点 Clique.

对例子 $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ (简记为 $x - \{x_i\}$), 使用预测 F 对类变量 X_i 进行预测, 记 $X_i^F(x - \{x_i\})$ 为预测的结果, $X_i(x - \{x_i\})$ 为真正的结果.

定义 2 对预测 F , 称

$$LR(F) = \prod_{x - \{x_i\}} p(x -$$

$$\{x_i\}) p(X_i^F(x - \{x_i\}) | X_i(x - \{x_i\}))$$

为 F 平均 0-1 损失(或风险), 具有最小平均 0-1 损失的预测称为最优预测. 在最优预测中, 具有最少变量(或属性)的预测称为最小预测.

定义 3 在 DMN 中, 称使用

$$\arg \max_{x_i^{bl(x_i)}} \left\{ \prod_k \frac{p(q_k^i)}{p(q_k^i | q_{j(k)}^i)} \right\}$$

对变量 X_i 进行的预测为邻域预测.

下面给出两个定理, 限于篇幅省略证明.

定理 3 P 是可分解概率模式, 在 0-1 损失下, P 的 DMN 中对 X_i 的邻域预测是最优预测, 当 P 存在完全图^[1] 时, 还是最小预测.

定理 4 设 T_{X_i} 是依据 Pearl 方法建立的 X_i 邻域 Clique 子结合树, Q_1, \dots, Q_h, \dots 是 Clique 的顺序, 则有

$$\arg \max_{x_i^{bl(x_i)}} \left\{ \prod_k \frac{p(q_k^i)}{p(q_k^i | q_{j(k)}^i)} \right\} = \arg \max_{x_i^{bl(x_i)}} \left\{ \prod_h \frac{p(q_h)}{p(q_h | q_{j(h)})} \right\}.$$

定理 4 保证了可对邻域进行局部三角剖分, 以及使用邻域 Clique 子结合树建立邻域预测, 这将使邻域预测的构建更加简单化, 而且易于程序实现.

3 基于邻域预测的噪声平滑

基于 EM 算法的思想, 需使用具有最大似然的值去修正一个数据(平滑噪声), 而邻域的预测值就是具有最大似然的取值, 因此可以用来平滑噪声.

3.1 噪声平滑机制与过程

使用预测值修正待平滑的数据是基于邻域预测噪声平滑的基本机制. 用 x_{mi} 表示变量 X_i 在第 m 行的取值, 平滑 x_{mi} 即是用 \hat{x}_{mi} 修正 x_{mi} , 其中

$$\hat{x}_{mi} = \arg \max_{x_i^{bl(x_i)}} \left\{ \prod_{h=1}^{r(q_i)} \frac{p(q_h^i)}{p(q_h^i | q_{j(h)}^i)} \mid x - \{x_i\} \right\}.$$

推论 1 基于邻域最优预测的噪声平滑具有最小风险.

对变量 X_i 的平滑是一个迭代过程, 当平滑完毕所有的数据 $x_{1i}, x_{2i}, \dots, x_{Ni}$ 便实现一次迭代, 在此基础上再进入下一次迭代, 直到收敛. 当需要平滑多个变量或全部变量时, 只需依次平滑即可.

3.2 迭代终止检验

在对一个变量进行平滑迭代的过程中, 根据相邻两次平滑与该变量有关的参数平方误差和的大小来检验其收敛性. 设包含 X_i 的所有 Cliques 为 $Q_1^i, Q_2^i, \dots, Q_i^{r(q_i)}$, 对给定的小正数 ϵ , 如果满足

$$\frac{1}{M(Q_i)} \prod_{h=1}^{r(q_i)} (p(q_h^{(v)}) - p(q_h^{(v+1)}))^2 < \epsilon,$$

就认为收敛. 其中: $M(Q_i)$ 是与 X_i 有关的所有参数

的数量, $p(q_{hi}^{(v)})$ 和 $p(q_{hi}^{(v+1)})$ 是经过 v 次和 $v+1$ 次平滑后关于 Clique $Q_i^h(1 \leq h \leq r(q_i))$ 的参数集,

$(p(q_{hi}^{(v)}) - p(q_{hi}^{(v+1)}))^2$ 表示相邻两次平滑关于 Q_i^h 的对应参数差的平方和.

4 实验

根据 Alarm 网(一个国际标准贝叶斯网络, 具有 37 个变量, 变量最多取值为 4) 概率分布表生成具有 6000 个例子的模拟数据集, 复制 9 个数据集, 分别为 $D_2, D_4, D_6, D_8, D_{10}, D_{20}, D_{30}, D_{40}, D_{50}$. 使用随机初始化的方法引入噪声, 在 9 个数据集中分别产生具有 2%, 4%, 6%, 8%, 10%, 20%, 30%, 40% 和 50% 的噪声. 以没有引入噪声的数据集所估计的参数为标准参数, 利用被平滑数据集估计的参数与标准参数之间的平方误差来衡量平滑效果, 使用迭代终止标准检验平滑迭代的收敛性, 取 $\epsilon = 0.001$, 分别从确定变量的邻域结构、平滑噪声迭代的收敛性, 以及平滑效果 3 个方面进行实验与分析.

(1) 邻域结构

将 Alarm 网的道德图^[1] (moral graph) 进行三角剖分, 得到弦图, 其中结点 X_{35} 和 X_{29} 的邻域结构如图 1 所示.

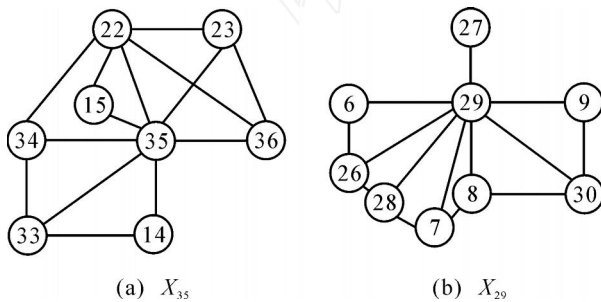


图 1 结点的邻域结构

(2) 迭代收敛性

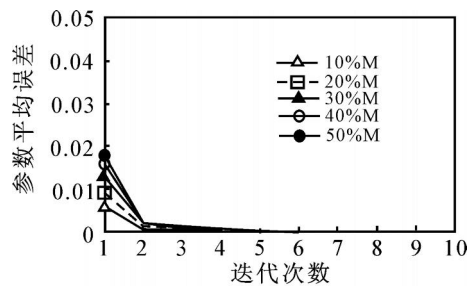
首先基于已知的 DMN 结构建立邻域最优预测, 然后使用预测值对数据集中的对应数据进行迭代平滑. 图 2(a) 显示了具有不同的噪声比例, 对变量 X_{35} 平滑迭代的收敛情况; 图 2(b) 给出了噪声比例为 40%, 对 5 个变量 ($X_{35}, X_{31}, X_{29}, X_{27}$ 和 X_{16}) 平滑迭代的收敛情况.

从图 1 可以看出, 在两种情况的迭代中, 经过 4 次迭代后均收敛, 没有波动现象, 这显示了基于邻域预测噪声平滑具有较高的效率和良好的稳定性.

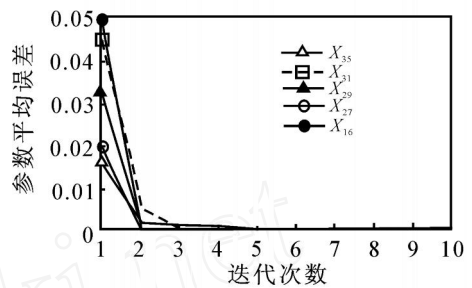
(3) 平滑效果与比较

选择 10 个变量 (6 个内部, 4 个边缘), 噪声比例为 40%, 平滑方法分别采用邻域预测、Gibbs 抽样和梯度 (grads) 优化. 由于梯度优化方法对边缘结点需要先验知识, 用 Gibbs 抽样来代替. 平滑前以及使用

3 种平滑方法平滑后参数平方误差的比较如表 1 所示.



(a) 不同比例的噪声平滑



(b) 不同变量的噪声平滑

图 2 平滑噪声迭代收敛情况

表 1 不同方法的平滑效果比较

方法	X_3	X_{10}	X_{16}	X_{25}	X_{27}
平滑前	0.1710	0.1746	0.2563	0.4221	0.1205
邻域预测	0.0869	0.0605	0.1113	0.1526	0.0447
Gibbs	0.1646	0.1668	0.2548	0.3659	0.1087
Grads	0.1646	0.1668	0.2548	0.4398	0.1255
方法	X_{28}	X_{29}	X_{31}	X_{33}	X_{35}
平滑前	0.1498	0.4489	0.3818	0.3377	0.9049
邻域预测	0.1094	0.2359	0.2544	0.1719	0.5718
Gibbs	0.1455	0.3385	0.3894	0.3458	0.8554
Grads	0.1455	0.3743	0.2638	0.2629	0.6484

从表 1 可以看出, 基于邻域预测的噪声平滑方法明显优于其他两种方法, 其主要原因是: Gibbs 抽样需要根据一个变量的概率分布情况确定平滑值, 而由于噪声的存在, 实际分布与真实分布可能具有较大的差距 (概率分布对噪声比较敏感), 导致相对较差的平滑效果; 而基于邻域预测的平滑使用一个变量的条件 (邻域) 概率分布中, 具有最大概率的取值作为平滑值, 一方面可使平滑风险最小化, 另一方面条件概率分布中概率大小的顺序关系具有很好的稳定性, 能够得到良好的平滑效果; 基于梯度优化的平滑方法对只利用邻域中部分结点的信息进行预测, 平滑易于走向极端, 从而影响平滑效果.

(下转第 605 页)

优化,从而可以得到一条最优或基本最优的全局路径.由于蚂蚁寻食目标明确,且不允许重复搜索,加上两组蚂蚁搜索的并行性,规划效率极高.仿真实验结果表明,本算法比已有蚁群算法等的收敛速度提高显著,效果好.

本文算法作为一种新的基础算法,若用于未知环境下的动态规划,可满足实时规划的需求.

参考文献(References)

- [1] Chunmiao Wang, Soh Y C, Han Wang, et al. A hierarchical genetic algorithm for path planning in a static environment with obstacles [C]. *Electrical and Computer Engineering. Canadian*, 2002, 3: 1652-1657.
- [2] Kuffner J J, LaValle Jr S M. RRT-connect: An efficient approach to single-query path planning, robotics and automation [C]. *IEEE Int Conf on ICRA. San Francisco*, 2000, 2: 995-1001.
- [3] Fan Xiaoping, Luo Xiong, Yi Sheng, et al. Optimal path planning for mobile robots based on intensified ant colony optimization algorithm[C]. *Robotics, Intelligent Systems and Signal Processing. 2003*: 131-136.
- [4] Gengqian Liu, Tiejun Li, Yuqing Peng, et al. The ant algorithm for solving robot path planning problem[C]. *3rd Int Conf on Information Technology and Applications. 2005*, 2: 25-27.
- [5] Shirong Liu, Linbo Mao, Jinshou Yu. Path planning based on ant colony algorithm and distributed local navigation for multi-robot systems, mechatronics and automation[C]. *Proc of the 2006 IEEE Int Conf on. 2006*: 1733-1738.
- [6] 朱庆保. 动态复杂环境下的机器人路径规划蚂蚁预测算法[J]. *计算机学报*, 2005, 28(11): 1898-1906. (Zhu Q B. Ants predictive algorithm for path planing of robot in a complex dynamic environment [J]. *Chinese J of Computers*, 2005, 28(11): 1898-1906.)
- [7] 朱庆保. 复杂环境下的机器人路径规划蚂蚁算法[J]. *自动化学报*, 2006, 32(4): 586-593. (Zhu Q B. Ant algorithm for path planning of mobile robot in a complex environment [J]. *Acta Automatica Sinica*, 2006, 32(4): 586-593.)
- [8] Michael J Greene, Deborah M Gordon. Cuticular hydrocarbons inform task decisions[J]. *Nature*, 2003, 423(6935): 35-36.
- [9] Jackson D E, Holcombe W L M, Ratnieks F L W. Knowing which way to go - trail geometry gives polarity to ant foraging trails [J]. *Nature*, 2004, 432(7019): 907-909.
- [10] 张美玉, 黄翰, 郝志峰, 等. 基于蚁群算法的机器人路径规划[J]. *计算机工程与应用*, 2005, 25: 34-37. (Zhang M Y, Huang H, Hao Z F. Motion planning of autonomous mobile robot based on ant colony algorithm [J]. *Computer Engineering and Application*, 2005, 25: 34-37.)

(上接第 600 页)

5 结 论

本文基于 DMN 的邻域预测建立了一种有效的离散数据噪声平滑方法,不需要改变 DMN 结构,在理论上具有最小的平滑风险,可从源头上减少噪声对参数的影响.相对于现有减小噪声影响的方法更加可靠和实用,并且以变量为基本单位的噪声平滑使得该方法更具灵活性(既可局部平滑,也可全局平滑),这种方法还具有一般性,可广泛用于离散数据的噪声平滑.

参考文献(References)

- [1] Pearl J. Probabilistic reasoning in intelligent systems: Networks of plausible inference [M]. San Mateo: Morgan Kaufmann, 1988: 383-408.
- [2] 曹建农. 基于可分解马尔可夫网的图像分割方法研究[D]. 武汉: 武汉大学, 2005. (Cao J N. Study on approach of image segmentation based on decomposable Markov networks[D]. Wuhan: University of Wuhan, 2005.)
- [3] Zhao B Y, Joseph A D A, Ludwig R. Markov-based channel model algorithm for wireless networks [J]. *Wireless Networks*, 2003, 9(3): 189-199.
- [4] Wong S K M, Xiang Y. Construction of a Markov network from data for probabilistic inference[C]. *Proc of the 3th Int Conf on Rough Sets and Soft Computing. San Jose*, 1994: 562-569.
- [5] Deshpande A, Garofalakis M N, Jordan M I. Efficient stepwise selection in decomposable models[C]. *Proc of the 17th Conf on Uncertainty in Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers*, 2001: 128-135.
- [6] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm [J]. *J of the Royal Statistical Society*, 1977, 39(1): 1-38.