

基于条件互信息和概率突跳机制的贝叶斯网络结构学习算法

魏中强 徐宏喆 李 文 桂小林

(西安交通大学陕西省计算机网络重点实验室 西安 710049)

摘 要 贝叶斯网络分类器的精确构造是 NP 难问题,使用 K2 算法可以有效地缩减搜索空间,提高学习效率。然而 K2 算法需要初始的节点次序作为输入,这在缺少先验信息的情况下很难确定;另一方面,K2 算法采用贪婪的搜索策略,容易陷入局部最优解。提出了一种基于条件互信息和概率突跳机制的贝叶斯网络结构学习算法(CMI-PK2 算法),该算法首先利用条件互信息生成有效的节点次序作为 K2 算法的输入,然后利用概率突跳机制改进 K2 算法的搜索过程来提高算法的全局寻优能力,学习较为理想的网络结构。在两个基准网络 Asia 和 Alarm 上进行了实验验证,结果表明 CMI-PK2 算法具有更高的分类精度和数据拟合程度。

关键词 贝叶斯网络分类器,结构学习,条件互信息,概率突跳,K2 算法

中图分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.3.044

Bayesian Network Structure Learning Algorithm Based on Conditional Mutual Information and Probabilistic Jumping Mechanism

WEI Zhong-qiang XU Hong-zhe LI Wen GUI Xiao-lin

(Shaanxi Key Lab of Computer Network, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract Precise construction of Bayesian network classifier is an NP-hard problem. K2 algorithm can reduce search space effectively and improve learning efficiency, but it requires the initial node ordering as input, which is very limited by the absence of the priori information. On the other hand, K2 algorithm uses a greedy search strategy and is easy to fall into local optimization solutions. This paper presented a new Bayesian network structure learning algorithm based on conditional mutual information and probabilistic jumping mechanism. Firstly, conditional mutual information is used to determine the initial node ordering as input of K2 algorithm. Then probabilistic jumping mechanism is introduced into K2 algorithm to improve the search process and the ability of global optimization, and learn more reasonable network structure. Experimental results over two benchmark networks Asia and Alarm show that this new improved algorithm has higher classification accuracy and better degree of data matching.

Keywords Bayesian network classifier, Structure learning, Conditional mutual information, Probabilistic jumping, K2 algorithm

1 引言

贝叶斯网络是一种基于图理论和概率理论的不确定知识表示和推理模型^[1],已被广泛应用于数据挖掘分类过程中^[2,3]。贝叶斯网络分类器的构造主要分为 3 个过程:(1)从训练数据中学习表示变量之间依赖关系的网络拓扑结构;(2)根据学习到的拓扑结构构造条件概率表;(3)根据最大后验概率分类并制定决策。其中,第二个过程是构造贝叶斯网络分类器的关键,也是贝叶斯网络研究的热点问题。

贝叶斯网络结构学习算法主要分为两类:基于约束的算法^[4-6]和基于评分搜索的算法^[7-9]。基于约束的算法根据随机变量之间的条件依赖关系构造贝叶斯网络结构,这种方法较

为直观,但是对条件独立性测试的误差较为敏感。基于评分搜索的算法将学习看作优化问题,选择具有最高评分值的网络结构作为最优的网络结构。通常这种算法学习的网络结构具有较高的分类精度,但算法的时间复杂度较高,而且容易陷入局部最优解。

K2 算法^[9]是基于评分搜索的经典算法,该算法利用爬山搜索算法搜索网络结构空间,并利用贝叶斯评分方法评估网络结构的优劣。在给定初始节点次序的条件下,K2 算法可以有效地缩减搜索空间,降低计算复杂性,然而初始的节点次序在实际应用中很难获得,而且不合理的节点次序将使得到的网络结构与最优解相差较大。因此,为 K2 算法提供有效的节点次序十分重要。为了解决这个问题,研究人员已经提出

到稿日期:2014-04-17 返修日期:2014-07-20 本文受国家自然科学基金(61172090)资助。

魏中强(1989—),男,硕士生,主要研究方向为数据挖掘,E-mail:xjtu201305@163.com;徐宏喆(1961—),女,博士,教授,CCF 会员,主要研究方向为人工智能、数据挖掘;李文(1966—),女,硕士,讲师,主要研究方向为计算机智能软件;桂小林(1966—),男,博士,教授,CCF 会员,主要研究方向为可信计算、云计算。

了一些解决方法^[10-12],然而这些方法在实际情况下的应用效果并不理想。另一方面,K2算法采用贪婪的搜索策略,使得结果容易陷入局部最优,为了提高算法的全局寻优能力,有必要改进算法的搜索过程。

针对上述情况,本文提出了一种基于条件互信息和概率突跳机制的贝叶斯网络结构学习算法(CMI-PK2算法),该算法分为两个部分,首先利用条件互信息^[13]生成初始的节点次序作为K2算法的输入,然后利用概率突跳机制^[14]改进K2算法搜索过程,增强算法的全局寻优能力,学习全局最优的网络结构。

本文第2节详细描述基于条件互信息和概率突跳机制的贝叶斯网络结构学习算法CMI-PK2,并给出CMI-PK2算法的详细过程;第3节给出实验验证及分析;最后总结及展望未来的研究工作。

2 基于条件互信息和概率突跳机制的结构学习算法CMI-PK2

2.1 随机变量间依赖关系的确定

图结构是表示数据集中随机变量(或节点,下文统一采用“节点”一词)间依赖关系的有力工具,要构建一个图结构,需要确定节点间的连接性(即边)以及依赖关系的强弱(即边的权值)。

任意两个节点间依赖关系的大小可以通过节点间的条件互信息(Conditional Mutual Information,CMI)值表示。在给定节点 Z 的条件下,如果节点 X_i 和节点 X_j ($1 \leq i, j \leq n, n$ 是数据集中节点的个数)的条件互信息值较大,说明在给定节点 Z 的条件下,节点 X_i 和节点 X_j 具有较强的相互依赖关系,这种依赖关系可以用来确定节点间的连接性。

在给定节点 Z 的条件下,节点 X_i 和节点 X_j 的条件互信息如式(1)所示:

$$I(X_i; X_j | Z) = \sum_k \sum_l \sum_m \sum_n P(x_l, y_m | z_k) \log \frac{P(x_l, y_m | z_k)}{P(x_l | z_k) P(y_m | z_k)}, i \neq j \quad (1)$$

其中, q_z 表示节点 Z 所有可能取值的个数, z_k 表示节点 Z 取第 k 个值, q_i 表示节点 X_i 所有可能取值的个数, x_l 表示节点 X_i 取第 l 个值, q_j 表示节点 X_j 所有可能取值的个数, y_m 表示节点 X_j 取第 m 个值, $i \neq j$ 表示节点 X_i 和 X_j 不能是同一个变量。从条件互信息的定义可以看出,两个节点间的条件互信息是对称的,即 $I(X_i; X_j | Z) = I(X_j; X_i | Z)$ 。

假定已知节点集合 $X = \{X_1, X_2, \dots, X_n\}$ 及类属性 C ,则任意两个节点 X_i 和 X_j 之间的条件互信息可以计算为 $I(X_i; X_j | C)$,其中 $1 \leq i, j \leq n, i \neq j$ 。在这两个节点 X_i 和 X_j 间添加一条边,边的权值被设置为它们的条件互信息 $I(X_i; X_j | C)$,根据条件互信息的对称性,添加的边是无向的。利用条件互信息确定节点间的依赖关系和强度,最终生成表示节点间依赖关系的带权无向图。

2.2 基于有向图的最大权生成树的获取

为了获得初始的节点次序,需要确定带权无向图中边的方向,从而确定节点的先后顺序。采用条件相对平均熵^[15](Conditional Relative Average Entropy,CRAE)确定无向边的方向。CRAE能够表示节点之间的依赖倾向,对于任意两个 X_i 和 X_j ,其条件相对平均熵定义为:

$$CRAE(X_j \rightarrow X_i) = \frac{H(X_i | X_j)}{H(X_i) \cdot |X_i|}, i \neq j \quad (2)$$

其中, $H(X_i) = -\sum_{x_l} P(x_l) \log P(x_l)$ 表示节点 X_i 的信息熵, $H(X_i | X_j) = -\sum_{x_l, x_m} P(x_l, x_m) \log P(x_l | x_m)$ 表示在给定节点 X_j 的情况下节点 X_i 的条件信息熵, $|X_i|$ 表示节点 X_i 所有可能取值的个数, x_l 表示节点 X_i 所有可能的取值, x_m 表示节点 X_j 所有可能的取值。

对节点 X_i 和 X_j 而言,如果 $CRAE(X_j \rightarrow X_i) > CRAE(X_i \rightarrow X_j)$,则边的方向被设置为从节点 X_j 指向节点 X_i ,即 $X_j \rightarrow X_i$ 。如果 $CRAE(X_i \rightarrow X_j) > CRAE(X_j \rightarrow X_i)$,则边的方向被设置为从 X_i 指向节点 X_j ,即 $X_i \rightarrow X_j$ 。如果 $CRAE(X_i \rightarrow X_j) = CRAE(X_j \rightarrow X_i)$,采用贝叶斯评分方法,穷尽搜索边的最优方向。经过上述步骤,可以得到一个带权的有向图。

经过上述步骤得到的有向图中每个节点几乎都有入度和出度,很难确定节点间的顺序关系,而基于该带权有向图构造的生成树可以很好地确定节点间的顺序关系。假定初始图为 $G(V, E)$,构建的最大权生成树为 $T(U, D)$,则利用带权有向图构建最大权生成树的过程如下:

- 初始化节点集合 U 和边集合 D 为空集,即 $U \leftarrow \emptyset, D \leftarrow \emptyset$ 。
- 从图 G 的节点集合 V 中随机选择一个节点 u_0 ,添加到生成树 T 的节点集合 U 中,即 $U = \{u_0\}$ 。选择图 G 中与节点 u_0 相连的,且权值最大的一条有向边添加到生成树 T 的边集合 D 中,与这条有向边相连的另外一个节点(假定为 v)添加到节点集合 U 中,即 $U = \{u_0, v\}$ 。
- 在图 G 的边集合 E 中选择一条满足如下条件的有向边 $\langle x, y \rangle$:1)有向边的一个节点在集合 U ,另一个节点不在集合 U 中;2)该有向边弧头 y 的入度为0;3)在满足条件1)和2)的情况下具有最大的权值。满足条件的有向边添加到集合 D 中,不在集合 U 中的节点添加到 U 中。
- 重复第三步直到节点集合 U 包含所有的节点。

经过上述步骤,可以得到表示节点间关系的最大权生成树,图1所示为根据有向图构造最大权生成树的过程,首先随机选择节点 A ,由于边 $\langle A, B \rangle$ 满足条件,在生成树中加入这条边;与节点 A 或 B 相邻且权值最大的边 $\langle A, C \rangle$ 被选中;与节点 A, B 或 C 相邻且权值最大的边为 $\langle D, C \rangle$,但弧头 C 的入度不为0,所以忽略这条有向边并选择边 $\langle D, A \rangle$,得到最大权生成树。

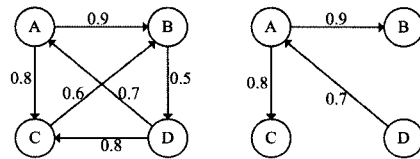


图1 根据有向图生成最大权生成树

根据带权有向图构造最大权生成树,可以利用拓扑排序对生成树中的节点进行排序,确定初始节点次序。此时树中 B 和 C 两个节点的顺序可以随机确定,即 D, A, B, C 或 D, A, C, B 。

2.3 基于概率突跳机制K2搜索过程的优化

K2算法可以有效地缩减搜索空间提高学习效率,但是其贪婪的搜索策略容易陷入局部最优,本节采用概率突跳

(Probability Jumping)机制提高 K2 算法的全局寻优能力,以学习接近全局最优的网络结构。

当采用 K2 算法为每个节点 X_i 搜索最优的父节点集合 $Pa(X_i)$ 时,利用 BDe 评分方法评价当前搜索到的父节点集 $Pa_0(X_i)$ 。假设评分值为 S ,如果此时集合 $Pa(X_i)$ 为空,则令 $Pa(X_i) = Pa_0(X_i)$,作为当前搜索到的最优的父节点集合,并令最优父节点集合的评分值 $S_{max} = S$ 。如果集合 $Pa(X_i)$ 不为空,则比较评分值 S 和最优父节点集合的评分值 S_{max} ,如果 $S > S_{max}$,则接受新的父节点集合作为最优的父节点集合,即 $Pa(X_i) = Pa_0(X_i)$ 。

在集合 $Pa(X_i)$ 不为空且 $S \leq S_{max}$ 的情况下,利用概率突跳机制判断是否接受新的节点集合作为最优的父节点集合,具体方法如下:产生 0 到 1 之间的随机数 γ ,若 $P_{acc} > \gamma$,则接受新的节点集合作为最优的父节点集合,并将评分值更新为 S ,即 $Pa(X_i) = Pa_0(X_i)$, $S_{max} = S$,其中 P_{acc} 表示接受概率,即新解的接收次数与搜索总次数之比;否则保持原来的父节点集合和评分值不变。

原来 K2 算法为每个节点搜索最优父节点集合的终止条件(满足任意一个条件即可)为:1)节点的父节点个数已经达到父节点个数上限。2)节点的所有可能的父节点已经搜索完。3)增加新的节点到父节点集合无法提高评分函数值。引入概率突跳机制后,除了终止条件 1)和 2)保持不变外,终止条件 3)需要采用如下方法:

- 设定最大初始迭代次数 $L_{max} = L_0$ 以及当前迭代次数 $L = 1$ 。

- 每利用概率突跳机制判断一次,如果新解被接收,则 $L_{max} = L_0$,否则令 $L_{max} = L_0 \times C \times r \times P_{acc}$,其中 C 为节点状态的维数,较高的维数表明与该节点相关联的节点个数可能越多; r 为衰减系数,随着搜索的进行,适当地减少迭代次数,加速算法的收敛。无论新解有没有被接收,迭代次数 L 都增 1。

- 如果当前迭代次数 $L \geq L_{max}$,则终止为当前节点搜索最优父节点集合的过程。

本文提出的新的终止条件是一种动态迭代次数控制方法,这种迭代次数控制方法可以从很大程度上跳出陷入局部最优解,并且避免了过多冗余迭代的次数,在一定程度上提高了算法的性能。

2.4 算法详细描述

根据上述算法思路,基于条件互信息和概率突跳机制的贝叶斯网络结构学习算法(Bayesian Network Structure Learning Base-on Conditional Mutual Information and Probability Jumping, CMI-PK2)的详细描述如下所示:

CMI-PK2 算法

输入:数据集 D ,节点集合 $X = \{X_1, X_2, \dots, X_n\}$,类属性节点 C
输出:每个节点的最优父节点集合 $Pa(X_i)$ ($1 \leq i \leq n$)。

- Step 1 采用 CMI 确定任意两个节点间的连接性,构造权值为 $I(X_i; X_j | C)$ 的带权无向图
 - Step 2 采用 CRAE 确定节点间的依赖倾向,得到带权的有向图
 - Step 3 基于上述思路中构造最大权生成树的过程,计算基于带权有向图的最大权生成树
 - Step 4 利用拓扑排序生成初始节点次序
 - Step 5 初始化每个节点的父节点集合 $Pa(X_i) \leftarrow \emptyset$
- While 仍有节点的父节点集合未确定
For 对当前节点 X_i 所有可能的父节点集合 $Pa'(X_i)$

Step 6 计算评分函数值 S

Step 7 如果 $Pa(X_i) = \emptyset$ 或者 $Pa(X_i) \neq \emptyset$ 且 S 大于 S_{max} ,更新节点 X_i 的最优父节点集及其评分函数值,即 $Pa(X_i) = Pa'(X_i)$, $S_{max} = S$

Step 8 如果 $Pa(X_i) \neq \emptyset$, S 不大于 S_{max} 且 $P_{acc} > \gamma$,同样更新节点 X_i 的最优父节点集及其评分函数值

Step 9 如果 Step 7 和 Step 8 的条件均不满足,执行迭代次数衰减操作,加速算法的收敛,即 $L_{max} = L_{max} \times C \times r \times P_{acc}$

Step 10 当前迭代次数加 1,即 $L = L + 1$

Step 11 如果 L 大于等于 L_{max} ,结束为节点 X_i 搜索父节点的过程
End For

Step 12 更新当前节点为下一个节点

End While.

3 实验结果与分析

3.1 实验参数设定

为了验证本文提出算法的有效性,实验采用 10-折交叉验证方法。假定数据集的所有属性都是离散的并且不包含缺失值。在每个数据集上进行 10 次 10-折交叉验证,取最优结果、平均结果以及 10 次交叉验证所得结果的标准差。

本文利用条件互信息确定初始的节点次序,并采用概率突跳机制提高 K2 搜索策略的全局寻优能力,构建较为理想的网络结构,然后进行网络参数条件概率表(Conditional Probability Table, CPT)的学习,构建贝叶斯分类器。实验中所有参数采用 BDe 先验,并采用联结树推理算法。概率突跳机制中迭代次数 L_0 设置为 5 次,衰减系数 r 取值为 0.8。算法学习过程中每个节点的最大父节点个数上限设置为 5。

3.2 对比和分析参数

用来分析和比较算法的参数如下所示:

评分值 $S(\text{Score})$:算法在数据集上学习到的网络结构的评分值。

分类精度 $CA(\text{Classification Accuracy})$:算法在数据集上学习到的网络结构的分类精度。

平均结果 $MR(\text{Mean Result})$:算法在数据集上进行 10 次 10-折交叉验证所得结果的平均值。

标准差 $SD(\text{Standard Deviation})$:算法在数据集上进行 10 次 10-折交叉验证所得结果的标准差。

最优结果 $BR(\text{Best Result})$:算法在数据集上进行 10 次 10-折交叉验证所得结果的最优值。

3.3 实验结果分析

实验分析在两个基准网络 Asia^[16] 和 Alarm^[17] 上进行,并将本文提出的算法与 TAN、Hill Climbing、Random-K2 进行比较。Asia 网络是一个小型的贝叶斯网络,具有 8 个节点和 8 条边,每个节点是离散的且具有两个离散状态(或值)。Alarm 网络是一种中等大小的贝叶斯网络,包含 37 个节点和 46 条边,每个节点都是离散的且可以取 2 个、3 个或 4 个离散状态。

表 1 算法在两种基准网络上学习到的网络结构 CA 比较

算法	MR		SD		BR	
	Asia	Alarm	Asia	Alarm	Asia	Alarm
CMI-PK2	97.55	98.54	16.4	17.4	98.51	98.65
TAN	93.64	94.77	15.8	16.1	94.98	95.73
Hill Climbing	97.42	97.55	19.2	19.0	98.21	97.30
Random-K2	94.32	94.26	22.3	21.1	96.39	96.44

表 1 显示了 CMI-PK2 算法和另外 3 种算法学习到的网络结构分类精度比较,从表中可以看出,本文提出的算法具有更高的分类精度,分类效果更好。

表 2 和表 3 分别显示了算法在 Asia 网络和 Alarm 网络上学习到的网络结构的评分值比较,从表中可以看出,本文提出的算法学习到的网络结构的评分值较高,能够很好地表示节点之间的依赖关系,与数据的拟合程度更好。

表 2 算法在 Asia 网络上学习到的网络结构 S 比较

算法	MR	SD	BR
CMI-PK2	-525.10	17.5	-523.21
TAN	-530.20	17.4	-526.43
Hill Climbing	-531.17	19.2	-526.33
Random-K2	-529.84	22.7	-523.97

表 3 算法在 Alarm 网络上学习到的网络结构 S 比较

算法	MR	SD	BR
CMI-PK2	-526.17	18.6	-524.22
TAN	-555.26	17.8	-553.74
Hill Climbing	-529.63	20.1	-525.72
Random-K2	-532.38	22.6	-526.13

从上面 3 个表中可以看出,本文提出的算法比 TAN、Hill Climbing 以及 Random-K2 算法具有更好的性能。综合表 1 一表 3 可以看出,本文提出的贝叶斯网络结构学习算法学习的网络结构的评分值更高,说明与数据集的拟合程度更好,而且具有更好的分类准确性。

尽管 Hill Climbing 算法也具有较高的性能,但是从图 2 可以看出 Hill Climbing 算法的运行时间较高,而 CMI-PK2 算法运行时间较低,对样本容量的大小不敏感,具有处理较大数据集的能力。

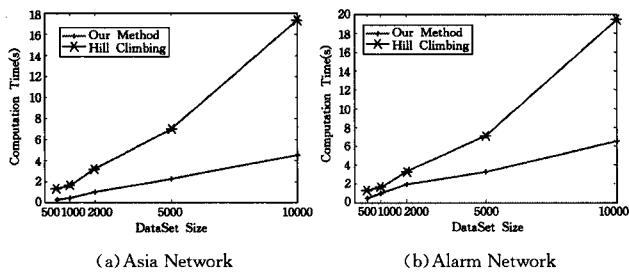


图 2 运行时间比较

从上面的实验结果可以看出:本文提出的算法具有较高的分类精度,可以更好地反映变量之间的实际依赖关系。

结束语 使用 K2 算法学习网络结构需要确定初始的节点次序,而且容易陷入局部最优。为此,本文提出了一种改进的贝叶斯网络结构学习算法,该算法利用条件互信息获得初始的节点次序,然后利用概率突跳机制提高 K2 算法的全局寻优能力。实验结果验证了算法的有效性。在以后的研究中将研究更多的数据挖掘算法,并将算法应用到计算机教育中。

参考文献

[1] Pearl J. Fusion, Propagation, and Structuring in Belief Networks [J]. Artificial Intelligence, 1986, 29: 241-288

[2] 黄建明. 贝叶斯网络在学生成绩预测中的应用[J]. 计算机科学, 2012, 39(11A): 280-282

[3] 李亚飞, 吕强, 苏伟峰, 等. 一种小规模数据集下的贝叶斯网络学习方法及其应用[J]. 计算机科学, 2011, 38(7): 181-184

[4] Wong M L, Leung K S. An efficient data mining method for learning Bayesian networks using an evolutionary algorithm based hybrid approach[J]. IEEE Transactions on Evolutionary Computation, 2004, 8(4): 378-404

[5] Chen X W. Improving Bayesian Network structure learning with mutual information-based node ordering in the K2 algorithm [J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(1): 1-13

[6] 冀俊忠, 张鸿勋, 胡仁兵, 等. 基于独立性测试和蚁群优化的贝叶斯网结构学习算法[J]. 自动化学报, 2009, 35(3): 281-288

[7] Darwiche A. A differential approach to inference in Bayesian networks[J]. Journal of ACM, 2003, 50(3): 280-305

[8] Lerner B, Malka R. Investigation of the K2 algorithm in learning Bayesian network classifiers[J]. Applied Artificial Intelligence, 2011, 25(1): 74-96

[9] Cooper G F, Herskovits E. A Bayesian method for the induction of probabilistic networks from data [J]. Machine Learning, 1992, 9(4): 309-347

[10] Chen X W, Anantha G, Lin X. Improving Bayesian network structure learning with mutual information-based node ordering in the K2 algorithm[J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(5): 628-640

[11] Zhang Y, Zhang W, Xie Y. Improved heuristic equivalent search algorithm based on Maximal Information Coefficient for Bayesian Network Structure Learning [J]. Neurocomputing, 2013, 117(2): 186-195

[12] Chickering D M. Optimal structure identification with greedy search[J]. The Journal of Machine Learning Research, 2002, 3: 507-554

[13] Fleuret F. Fast binary feature selection with conditional mutual information[J]. The Journal of Machine Learning Research, 2004, 5: 1531-1555

[14] Al-Rifaie M M, Blackwell T. Bare bones particle swarms with jumps[M]. Swarm Intelligence. Springer Berlin Heidelberg, 2012: 49-60

[15] Jiang J, Wang J, Yu H, et al. Poison Identification Based on Bayesian Network: A Novel Improvement on K2 Algorithm via Markov Blanket[M]. Advances in Swarm Intelligence. Springer Berlin Heidelberg, 2013: 173-182

[16] Lauritzen S L, Spiegelhalter D J. Local computations with probabilities on graphical structures and their application to expert systems[J]. Journal of the Royal Statistical Society, Series B (Methodological), 1988, 50(2): 157-224

[17] Beinlich I A, Suermondt H J, Chavez R M, et al. The ALARM monitoring system[C]//A case study with two probabilistic inference techniques for belief networks. Springer Berlin Heidelberg, 1989: 247-256