

基于节点输入策略贝叶斯网络的骨盆骨折分型研究

李 清, 苏 强

(同济大学 经济与管理学院, 上海 200092)

摘要: 基于历史数据的统计和收集, 选取骨盆骨折患者存在的 18 个体表特征, 采用基于 K2 算法的贝叶斯网络方法挖掘各体表特征之间和骨盆骨折类型与体表特征间的相互关系; 设计不同的节点输入策略, 分析不同输入策略对算法性能的影响; 基于骨盆稳定性将骨盆骨折分成 A、B、C 三种类型, 分别找到与其直接相关的体表特征, 作为判断骨盆骨折类型的依据. 基于体表特征和骨盆骨折类型的分析结果, 借助早期的观察及简单检查, 对患者进行初步分型.

关键词: 体表特征; 贝叶斯网络; 骨盆分型; K2 算法
中图分类号: TP181 **文献标志码:** A

Pelvic Fracture Classification Based on the Bayesian Network of Node Ordering Strategy

LI Qing, SU Qiang

(School of Economics and Management, Tongji University, Shanghai 200092, China)

Abstract: Based on the statistics and collection of historical data, 18 surface characteristics of patients with pelvic fractures were selected. Bayesian network based on K2 algorithm was used to mine the causal relationship between the 18 surface characteristics, also between the surface characteristics and the pelvic fracture types. Different node ordering strategies were designed to analyze the influence on algorithm performance. Based on the stability of the pelvis, pelvic fracture was divided into A, B and C 3 types. Then found the features associated with each type of pelvic fracture, which was the basis of judgment. Based on the analysis of surface characteristics and pelvic fracture types, preliminary classification were made by means of early observation and simple examination.

Key words: surface characteristic; Bayesian network; pelvic fracture classification; K2 algorithm

骨盆是人体骨骼的重要部分, 骨盆周围分布着重要的器官及血管神经, 是一个复杂且关键的结构. 骨盆骨折是严重的创伤性骨折, 常合并有其他重要器官及系统的损伤. 早期诊断、院前急救和骨折复位固定是控制出血、降低死亡率的关键. 早期诊断中, 骨盆骨折的分型至关重要, 分型结果是后续治疗的基础, 直接决定了治疗方案和手术方法. 基于骨盆的稳定性, 将骨盆骨折分成 A、B、C 三种类型. 其中 A 型为稳定的骨盆环损伤, B 型为部分稳定骨折, C 型为不稳定骨折^[1]. A 型骨折一般采取保守治疗, B 型骨折采用固定前环的措施, 对 C 型骨折进行骨盆环前后联合固定.

目前, 对骨盆骨折的分型诊断主要集中在患者到达医院之后, 医生采用 X 光、CT (computed tomography slices) 片及其他辅助设备, 如图 1 所示. 电子设备的使用能快速准确地定位骨盆损伤, 帮助医生划分骨折类型. 然而, 对设备的依赖也导致时间的浪费: 运送患者从事故现场到医院的时间得不到充分利用, 患者到达医院后需要等待空闲的设备和医生, 等待时间的累积可能导致患者错过救援的“黄金时间”. 为充分利用等待时间, 本文收集患者的体表特征, 设计图 1 的新诊疗路径. 基于统计和贝叶斯网络方法, 分析体表特征间关系、体表特征与骨盆骨折类型间关系, 初步判断骨盆损伤患者的骨折类型, 采取相应的急救措施. 同时将患者的情况及时告知医院, 方便医生提前准备相关的诊断仪器, 制定可能的治疗方案, 为后续治疗节约时间.

朴素贝叶斯分类方法的准确率和分类效率较高, 然而其特征属性需要满足条件独立或完全独立的条件. 实际案例中的特征属性往往具有较强的相关性, 限制了朴素贝叶斯分类的能力. 贝叶斯网络 (Bayesian network, BN) 将有向无环图与概率理论有机结合, 可简洁有效地表示变量间的相互关系, 为

收稿日期: 2016-11-10

基金项目: 国家自然科学基金(71090404, 71072026)

第一作者: 李 清(1991—), 女, 博士生, 主要研究方向为创伤诊断. E-mail: mzsqli26@126.com

通讯作者: 苏 强(1969—), 男, 工学博士, 教授, 博士生导师, 主要研究方向为服务科学、服务质量管理. E-mail: suq@tongji.edu.cn

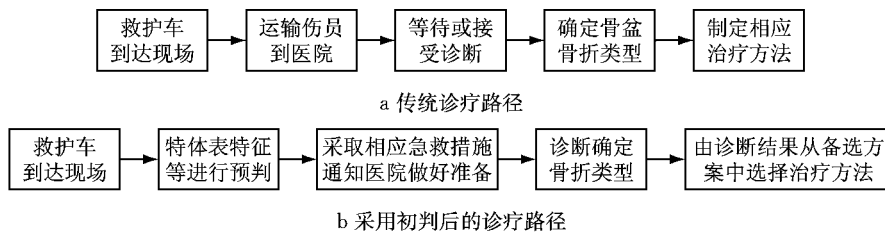


图 1 两种诊疗路径对比

Fig.1 Comparison of two paths of diagnosis and treatment

不确定性推理提供强有力的工具. 贝叶斯网络理论中, 结构学习的问题可以表述为给定节点变量集 X 的一个观测数据集 D , 通过学习找出一个最匹配数据集 D 的网络结构 G , 评分函数表示网络结构与真实网络的匹配程度.

1 文献综述

在不确定决策领域中, 机器学习和决策制定首要采用概率图模型, 贝叶斯网络是其中应用最广泛的一类. 贝叶斯网络中的推断和学习涉及到复杂的优化问题, 其学习包括结构学习和参数学习两部分, 结构学习是最为重要的内容. Larrañaga 等^[2]回顾了通用随机搜索算法在贝叶斯网络推理和学习中求解非确定性多项式 (non-deterministic polynomial, NP) 难题的应用. 基于局部结构的推导, Minn 等^[3]采用基于约束的学习类算法 (accelerated PC) 获得 BN, 该方法继承了 PC (peter & Clark) 算法的高数据效率特性, 节省了计算量. Cooper 等^[4]采用 K2 算法构造 BN, 通过局部贪婪搜索选择最优的网络结构, 并将此方法应用于数据缺失和含有隐藏变量的例子. Gasse 等^[5]提出新的 H2PC (hybrid PC) 混合算法: 重构贝叶斯网络框架, 基于贪婪爬山搜索方法确定边的方向. Gheisari 等^[6]使用基于粒子群算法的贝叶斯网络构造算法 (BNC-PSO), 在算法中增加插入边和删除边的功能, 使得粒子能够达到最优解. 循环删除程序的增加防止了无效解的迭代, 马尔可夫链的引入证明了该算法的全局收敛性. Yang 等^[7]提出细菌觅食优化方法 (bacterial foraging optimization, BFO-B), 算法中的每个细菌对应一个贝叶斯网络结构候选解, BFO 在网络空间中搜索最优网络结构, K2 准则引导搜索进程. Kreimer 等^[8]采用条件熵方法学习最优结构, 该方法在计算时间和计算性能上较之其他方法有显著优势.

网络结构学习包含评分和搜索两部分, 评分准则衡量贝叶斯网络对于给定数据集的质量, 搜索程序在可能的空间中智能移动搜索, 建立合理的网络

结构. K2 算法是贝叶斯网络中最有效的结构学习方法之一, 然而其性能很大程度上依赖节点的排列顺序. 基于当子节点被分配给正确的父节点时, 其条件概率更高的已知, Ko 等^[9]设计了新的评分函数和节点排序方法. 给定两个变量时, 利用评分函数推断出更好的父节点; 考虑所有变量时, 推断出父节点候选节点, 由此确定 K2 算法中节点的输入顺序. 基于隐式推理框架, Bouchaala 等^[10]提出隐式评分法, 采用 K2 和最大生成树 (maximum weight spanning tree, MWST) 算法实现网络结构学习. Hruschka 等^[11]提出特征排序算法, 采用信息增益和卡方检验两种方法得到有效的变量顺序, 大幅降低了 K2 算法的计算复杂度. 以信息理论方法和评分函数方法为基础, Chen 等^[12]采用条件独立测试方法获得节点排序信息. 针对贝叶斯评分准则中难以估计的自由参数, Cano 等^[13]采用局部边际化元参数的方法, 接受更广泛的参数假设, 并基于简单局部平均的方法来评估自由参数. Wang 等^[14]使用带平滑参数的高斯核函数估计属性密度, 基于分类精度准则, 贪婪搜索父节点, 构建扩展朴素贝叶斯分类器.

目前, BN 已在多个领域的决策和诊断中获得成功应用. 范建平^[15]基于聚类的贝叶斯网络模型约简医疗指标, 通过分类及提取主特征和类特征, 研究中医症候分型及动态演变, 为类风湿关节炎病因的研究、诊断提供了重要依据. 针对贪婪贝叶斯模式搜索 (greedy Bayesian pattern search algorithm, GBPS) 算法易陷入局部最优的不足, 王学伟等^[16]在 GBPS 算法的邻域生成过程中引入有向边的变向操作, 并用该算法研究中医临床诊断数据中症状与辨证要素间的复杂关系. 为研究精神分裂症疾病中基因型对外表型的影响, Ouali 等^[17]使用贝叶斯网络方法确定连续变量的最佳分界点. 采用贝叶斯狄利克雷等效, Akaike 信息准则和最小描述长度三种评分标准, 使用三种搜索算法: K2 算法, 爬山算法和模拟退火算法实现.

为充分利用救援时间, 快速准确地诊断骨盆骨折类型, 本文选取了患者的 18 个体表特征, 采用基

于 K2 算法的贝叶斯网络方法挖掘特征间、特征与骨折类型间的关系。为弥补 K2 算法在输入顺序上的不足,设计 4 种节点输入策略,分析不同策略对结果的影响。最后,基于患者体表特征描述和与骨折类型相关的主特征分析,对骨盆骨折患者进行初步诊断分型。

2 贝叶斯网络及构造

2.1 贝叶斯网络定义

贝叶斯网络又称为贝叶斯信念网络,由网络结构 B_S 和条件概率表 B_P 组成,即 BN 为 (B_S, B_P) 。贝叶斯网络结构是一个有向无环图(directed acyclic graph, DAG),其中节点集合为 $X, x_i \in X$ 表示随机变量;节点间的有向边集合为 $A, a_{ij} \in A$ 表示随机变量间的直接依赖关系。 a_{ij} 为 x_i 与 x_j 间的有向连接, $x_i \rightarrow x_j, x_i$ 被称为 x_j 的父节点。使用 $\pi(x_i)$ 表示变量 x_i 的父节点集合,条件概率分布 $P(x_i | \pi(x_i))$ 可定量表示 x_i 与 $\pi(x_i)$ 之间的依赖关系。条件概率表中的元素对应图中的节点,存储该节点与其父节点的联合条件概率。在确定节点及其父节点集后,假设该节点独立于它的所有非父节点,即贝叶斯网络中的条件独立性假设。因此,贝叶斯网络中所有节点的联合概率可以表示为各节点条件概率的乘积,如式(1)。

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | x_1, x_2, \dots, x_{i-1}) = \prod_{i=1}^n P(x_i | \pi(x_i)) \quad (1)$$

给定测试数据集 D ,构造贝叶斯网络的任务即找到最佳匹配数据集 D 的网络结构。网络结构的学习包含结构学习和参数学习,结构学习确定网络的拓扑结构,参数学习定义网络拓扑结构的条件概率。结构学习分为完备数据和不完备数据两种情况,完备数据下结构学习算法有 3 类:基于约束的学习算法(constraint based, CB)、基于评分搜索的学习算法(scoring and searching, SS)和混合学习算法。本文采用基于评分搜索的算法, K2 算法。

2.2 评分函数

贝叶斯网络学习中使用的评分函数主要分为两大类:基于贝叶斯的评分函数和基于信息论的评分函数。Cooper 等^[3]于 1992 年提出了用于计算贝叶斯网络评分的 K2 算法,计算公式如式(2):

$$P(B_S, D) =$$

$$P(B_S) \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \quad (2)$$

式中: D 为包含 m 个样本的数据集; r_i 表示随机变量 x_i 所有可能取值的数量; q_i 为 $\pi(x_i)$ 中包含实例的数量; w_{ij} 为 $\pi(x_i)$ 中第 j 个实例的取值情况, $1 \leq j \leq q_i$; N_{ijk} 表示数据库 D 中变量 x_i 取其第 k 个值 x_{ik} , 且其父代 $\pi(x_i)$ 取其第 j 个值 w_{ij} 的样本数量, 则 N_{ij}

$$= \sum_{k=1}^{r_i} N_{ijk}, N_{ik} = \sum_{j=1}^{q_i} N_{ijk}.$$

可分解是评分函数一个非常重要的特性,利用该特性,网络结构的评分可以转化为各点局部结构的评分之和。当贝叶斯网络中一个点的局部结构发生变化时,只需重新计算该点的局部结构评分,可分解的评分函数能大大降低重复计算的次数。因此,使用式(2)的代数形式,式(3)。其中, $f(x_i, \pi(x_i))$ 表示每个节点的 K2 评分,其具体的计算见式(4),常数 $\log P(B_S)$ 可省去不计。通过取对数的方法,将相乘的评分函数形式转化为各节点结构评分相加的形式,降低了计算的复杂度,减少了计算时间。

$$f(B_S, D) =$$

$$\log(P(B_S, D)) \approx \sum_{i=1}^n f(x_i, \pi(x_i)) \quad (3)$$

$$f(x_i, \pi(x_i)) = \sum_{j=1}^{q_i} \left(\log \left(\frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \right) + \sum_{k=1}^{r_i} \log(N_{ijk}!) \right) \quad (4)$$

每个节点与其父节点的联合概率均小于 1,分解后每个节点的 K2 评分值都是负数。要构造最优的贝叶斯网络结构,就是要寻找最大的 K2 评分值。

2.3 搜索方法

K2 搜索算法在给定输入节点顺序的情况下,依次计算各点的评分值。给定最大父节点数量,选择评分值最高的节点或节点组合作为该点的父节点。K2 算法的算法性能很大程度上取决于输入节点的顺序,输入顺序中排在前面的节点被默认是后续节点的父节点。随机设置的节点顺序或由数据集给定的节点顺序往往导致较差的结果,由领域专家确定节点顺序的方法因专家的难以获得而无法实施。本文中 18 个体表特征间的关系及先后输入顺序未知,因此采用 4 种策略确定节点的输入顺序:① 数据集给定的节点顺序,即样本中各特征的编号顺序;② 采用信息增益方法计算三种骨盆骨折类型下各特征的值,十折交叉验证后按从大到小的顺序排列各特征值,其对应编号作为输入顺序;③ 采用卡方检验方

法计算三种骨盆骨折类型下各特征的值,十折交叉验证后按从大到小的顺序排列各特征值,其对应编号作为输入顺序;④ 对每一个节点,遍历除其之外的所有节点,选择最大值代表的节点或节点组作为该节点的父节点,该策略下节点的输入顺序对算法结果没有影响.为挖掘骨盆骨折类型与体表特征间的关联,分析不同骨盆骨折的主特征,4 种策略下骨折类型均作为最后一个变量输入.每个节点的最大父节点数量一般设定为 $n-1$,本文取 18.

信息增益和卡方检验是有效的特征选择方法,可以度量三种骨盆骨折类型下各特征的重要程度.卡方检验通过一定的测量方法赋予每个特征权重,量化特征与类别间的相关度,关联性越强分值越高;信息增益衡量特征为分类系统带来的信息多少,越重要的特征带来的信息越多.三种骨盆骨折类型下,各特征的信息增益值和卡方检验值在软件 Weka3. 6. 4 中计算得到.

K2 算法的伪代码描述如下:

程序 K2 算法伪代码

输入 n 个节点,包含 m 个样本的数据集 D ,每个节点可拥有的最大父节点个数 u

输出 n 个节点的网络结构图

Steps 1. for $i=1:n$, do
 2. $\pi(x_i) = \emptyset$;
 3. $P_{old} = f(i, \pi(x_i))$;
 4. OKToProceed = true;
 5. While OKToProceed and $|\pi(x_i)| < u$ do
 6. let z be the node in $\text{pred}(x_i) - \pi(x_i)$ that maximize $f(i, \pi(x_i) \cup \{z\})$;

7. $P_{new} = f(i, \pi(x_i) \cup \{z\})$;
 8. if $P_{new} > P_{old}$ then
 9. $P_{old} = P_{new}$;
 10. $\pi(x_i) = \pi(x_i) \cup \{z\}$;
 11. else OKToProceed = false;
 12. end While;
 13. write ('Node: ', x_i , 'Parent of x_i : ', $\pi(x_i)$);
 14. end for;
 15. end K2;

3 数据收集与指标选取

收集 2012 年上海市第一人民医院骨盆骨折患者的电子病历和纸质病历,病历详细记录了患者的基本资料、体检报告数据、入院诊断、手术资料和出院信息等.由于病历中包含的数据过多、过杂,首先筛选出标记骨盆骨折类型的病历,其次提取与本研究相关的患者的体格检查描述.标记骨盆骨折类型的样本共有 172 例,筛选出的 18 个体表特征并依次编号为:①腰部压痛、②臀部压痛、③髌部压痛、④骨盆分离试验阳性、⑤下肢疼痛、⑥纵向叩击痛、⑦腹部压痛、⑧耻骨联合处压痛、⑨皮肤擦伤、⑩肩部压痛、⑪髂区压痛、⑫骨盆压痛、⑬上肢肿胀、⑭骶骨压痛、⑮胸廓挤压试验阳性、⑯头部受伤、⑰4 字征实验阳性和⑱对光反射.设计如表 1 的数据结构表,记录每个患者的骨折类型及存在的体表特征,并将文字描述转化成可以处理的 0-1 数字.所有患者的统

表 1 病例样本整理后的数据结构

Tab.1 Data structure after case arrangement

患者编号	体表特征					骨盆骨折类型		
	腰部压痛	臀部压痛 (省略的体表特征见第 3 节文字描述)	下肢疼痛	光反	A 型	B 型	C 型
1	1	1	0	0	1	0	0
2	0	1	1	1	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
172	0	1	1	0	0	1	0

计情况表中不予列出.

完成病历的逐条筛选后,计算三种骨盆骨折类型的患者数量,统计每种骨盆骨折下各体表特征出现的次数,绘制如表 2 所示的体表特征频数分布表.

4 实验分析

4.1 实验设计

基于第 4 节的资料收集和特征选取,得到实验

需要的数据集. 根据 4 种节点输入顺序策略, 首先计算三种骨折类型下每个特征的信息增益值和卡方检验值; 按照特征值从大到小的顺序进行排列, 返回特征值在数据集中的编号, 如表 3 所示; 依次排列编号作为节点的输入顺序, 进行实验计算. 比较不同输入

策略下, 实验结果和算法性能的差异, 构造骨盆骨折类型与体表特征的关系网络图.

4.2 不同节点输入策略的结果比较

针对三种骨盆骨折类型, 采用 4 种节点输入策略进行计算, 计算结果比较见表 4 和表 5:

表 2 骨盆骨折类型及体表特征频数分布

Tab.2 Classification of pelvic fracture and frequencies of surface characteristics

类型	数量	腰部压痛	臀部压痛	髋部压痛	骨盆分离试验阳性	下肢疼痛	纵向叩击痛	腹部压痛	上肢肿胀	皮肤擦伤
A	9	4	1	4	3	3	2	1	2	0
B	118	24	16	58	63	66	13	13	7	18
C	45	10	9	17	26	28	3	4	9	18
合计	172	38	26	79	92	97	18	18	18	36

类型	数量	肩部压痛	髂区压痛	骨盆压痛	胸廓挤压试验阳性	骶骨压痛	耻骨联合压痛	头部受伤	4 字征	对光反射
A	9	1	1	1	0	0	0	1	2	0
B	118	7	5	10	9	2	11	10	6	2
C	45	1	6	3	4	1	6	7	1	1
合计	172	9	12	14	13	3	17	18	9	3

数据来源: 上海市第一人民医院, 2012 年

表 3 策略 2 和 3 下各体表特征排序

Tab.3 Sorting of each surface characteristic under strategy 2 and 3

骨盆骨折类型	信息增益值从大到小排列的特征编号	卡方检验值从大到小排列的特征编号
A	9,17,8,1,5,15,4,13,6,10,11,14,18,12,2,7,3,16	17,9,1,5,4,13,6,8,10,15,11,12,18,14,3,2,7,16
B	13,9,11,16,3,2,1,10,8,7,17,6,4,5,12,15,14,18	13,9,11,16,3,2,1,10,8,17,7,6,4,5,15,12,14,18
C	9,13,11,3,16,17,10,2,6,5,8,4,14,12,15,7,18,1	9,13,11,3,16,17,2,10,6,8,5,4,14,12,15,7,18,1

表 4 4 种节点输入策略的算法性能比较

Tab.4 Performance comparison of 4 node input strategies

节点输入顺序策略	骨盆骨折类型								
	A			B			C		
	计算时间/s	与①比优化节点数/个	连接边数/条	计算时间/s	与①比优化节点数/个	连接边数/条	计算时间/s	与①比优化节点数/个	连接边数/条
①按数据集的特征编号输入	0.746 5	0	26	0.937 6	0	30	0.878 4	0	29
②特征信息增益值从大到小输入	0.799 7	7	27	0.944 9	5	31	1.249 3	5	31
③特征卡方检验值从大到小输入	0.796 4	7	27	0.962 2	6	31	0.942 4	5	31
④对每一个节点, 遍历其他各节点	2.078 8	11	41	2.197 0	11	41	2.161 8	7	41

表 5 4 种节点输入策略的结果比较

Tab.5 Result comparison of 4 nod input strategies

节点输入顺序策略	骨盆骨折类型					
	A		B		C	
	与①比优化节点数/个 (占百分数/%)	与④比连接边数/条 (占百分数/%)	与①比优化节点数/个 (占百分数/%)	与④比连接边数/条 (占百分数/%)	与①比优化节点数/个 (占百分数/%)	与④比连接边数/条 (占百分数/%)
①按数据集的特征编号输入	0(0)	26(63.41)	0(0)	30(73.17)	0(0)	29(70.73)
②特征信息增益值从大到小输入	7(38.89)	27(65.85)	5(27.78)	31(75.61)	5(27.78)	31(75.61)
③特征卡方检验值从大到小输入	7(38.89)	27(65.85)	6(33.33)	31(75.61)	5(27.78)	31(75.61)
④对每一个节点, 遍历其他各节点	11(61.11)	41(100)	11(61.11)	41(100)	7(38.89)	41(100)

通过比较可以发现:

(1) 策略 1 的计算时间最短, 策略 4 的计算时间最长, 策略 2 和 3 的计算时间与策略 1 的差距较小. A 型骨折下, 策略 2 和 3 分别比策略 1 的计算时间多了 7.13% 和 6.69%; B 型骨折下, 策略 2 和 3 分别

比策略 1 的计算时间多了 0.78% 和 2.62%; C 型骨折下, 策略 2 和 3 分别比策略 1 的计算时间多了 42.22% 和 7.29%.

(2) 对比策略 1, 策略 2~4 对节点的 K2 评分值有不同程度的优化. 策略 4 优化的节点数量最多, 三

种骨折类型中的优化数量分别占有所有节点的 61%、61% 和 39%;策略 2 和 3 优化的节点数量基本一致。

(3)策略 1 下连接的边数最少,即发现的节点间关系的数量最少,而采用策略 4 的遍历方法能找到节点间的尽可能多的关系。

综上,由于 K2 算法中先输入节点被默认是后输入节点的父节点,在给定节点顺序下(前 3 种策略),算法的计算时间相差不大。策略 2 和 3 采用特征选择方法,根据特征重要性制定节点顺序。随机产生的节点顺序或数据集中给定的输入顺序一般只能发现相邻节点间的关系;采用相关特征选择方法,根据特征重要性制定节点顺序时,易于挖掘与骨盆骨折类型高度相关的特征间关系。对节点进行遍历计算,通过计算选择评分函数值较高的节点作为其父节点,虽然导致计算时间增加,但可以发现所有特征间的关系及特征与骨折类型间的关系。

4.3 体表特征间关联分析

运行 K2 算法,采用遍历节点的输入策略,分析体表特征间的相互关系,如图 2 所示。图中节点分别表示:①腰部压痛、②臀部压痛、③髌部压痛、④骨盆分离试验阳性、⑤下肢疼痛、⑥纵向叩击痛、⑦腹部压痛、⑧耻骨联合处压痛、⑨皮肤擦伤、⑩肩部压痛、⑪髌区压痛、⑫骨盆压痛、⑬上肢肿胀、⑭骶骨压痛、⑮胸廓挤压试验阳性、⑯头部受伤、⑰ 4 字征实验阳性和⑱对光反射。由图可知:节点 4 是众多节点的子节点,即存在由一些部位的压痛可以推断出骨盆分离试验阳性的结论。

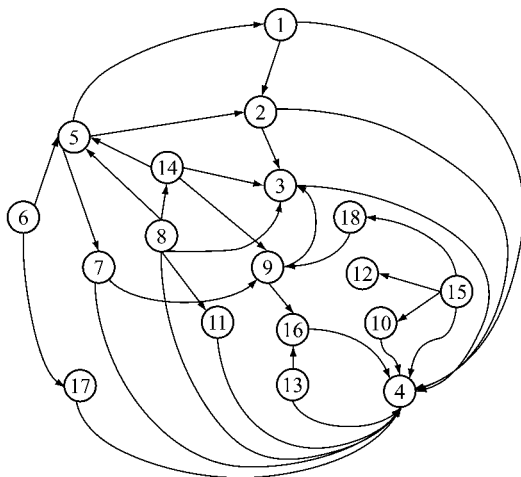


图 2 体表特征间网络关联图

Fig.2 Network diagram of surface characteristics

4.4 骨盆骨折类型与体表特征间关联分析

通过分析骨盆骨折类型与体表特征间关系,分别得到与三种骨折类型相关的体表特征,如图 3~图 5 所示。从图中可以发现:4 字征试验阳性是 A 型骨

折的父节点,而 A 型骨折容易导致腰部压痛、下肢疼痛及上肢肿胀这些体表特征的出现;当存在骨盆分离试验阳性、皮肤擦伤、髌区压痛、肩部压痛和上肢肿胀这些体表特征时,极有可能是 B 型骨盆骨折;骨盆分离试验阳性、皮肤擦伤、髌区压痛和上肢肿胀这些体表特征是发生 C 型骨盆骨折时存在的主要特征。

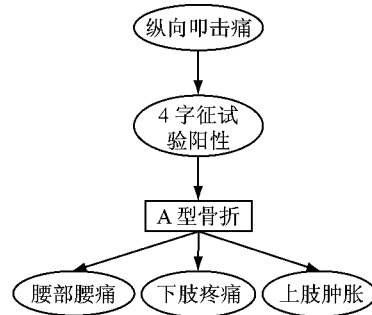


图 3 A 型骨盆骨折相关特征

Fig.3 Features related to type A pelvic fracture

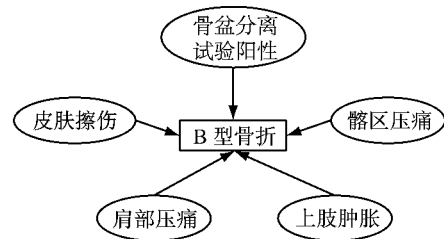


图 4 B 型骨盆骨折相关特征

Fig.4 Features related to type B pelvic fracture

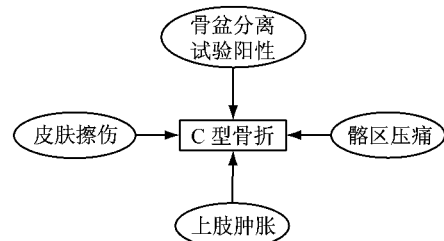


图 5 C 型骨盆骨折相关特征

Fig.5 Features related to type C pelvic fracture

5 结论

本文对收集到的 172 例骨盆骨折患者的病历进行整理,从原始病历的体表特征中提取 18 个指标。根据 Tile 分型方法将患者分成 3 类,其中 A 型骨折 9 例,B 型骨折 118 例,C 型骨折 45 例,统计不同骨折类型下各体表特征的频数。采用基于 K2 算法的贝叶斯网络结构学习算法学习体表特征间关系;设计不同的节点输入策略,分析不同节点输入顺序对算法性能的影响;分别提取与三种骨盆骨折类型直接

相关的特征. 特征间关联分析及特征与骨盆骨折类型的分析可以为骨盆骨折类型的初步判断提供依据,并采取有针对性的急救措施,为后续治疗节约时间. 由于收集到的病历数量有限,特别是 A 型骨盆骨折病历的严重不足,导致对 A 型骨盆骨折的研究不充分,相关特征的挖掘不全面;同时,三种骨盆骨折类型的病历数量相差较大,文章没有考虑样本数据量对计算结果准确性的影响.

进一步的研究包括:收集更多骨盆骨折患者的资料,降低样本量对结果的影响;提取更全面有效的指标,删减冗余、无效的指标,更好挖掘各骨盆骨折类型的主特征;改进结构学习算法或采用更有效的方法,使之更好地适用于医学统计与分析.

参考文献:

- [1] PIZANIS A, POHLEMANN T, BURKHARDT M, *et al.* Emergency stabilization of the pelvic ring; clinical comparison between three different techniques [J]. *Injury*, 2013, 44(12):1760.
- [2] LARRAÑAGA P, KARSHENAS H., BIELZA C., *et al.* A review on evolutionary algorithms in Bayesian network learning and inference tasks [J]. *Information Sciences*, 2013, 233: 109.
- [3] MINN S, 傅顺开. 贝叶斯网络结构加速学习算法[J]. *计算机科学*, 2016, 43(2): 263.
MINN S, FU Shunkai. Bayesian network structure accelerated learning algorithm [J]. *Computer Science*, 2016, 43(3): 263.
- [4] COOPER G F, HERSKOVITS E. A Bayesian method for the induction of probabilistic networks from data [J]. *Machine Learning*, 1992, 9(4): 309.
- [5] GASSE M, AUSSEM A, ELGHAZEL H. A hybrid algorithm for Bayesian network structure learning with application to multi-label learning [J]. *Expert Systems with Applications*, 2014, 41(15): 6755.
- [6] GHEISARI S, MEYBODI M R. BNC-PSO: structure learning of Bayesian networks by Particle Swarm Optimization [J]. *Information Sciences*, 2016, 348(2016): 272.
- [7] YANG C C, JI J Z, LIU J, *et al.* Structural learning of Bayesian networks by bacterial foraging optimization [J]. *International Journal of Approximate Reasoning*, 2016, 69: 147.
- [8] KREIMER A, Herman M. A novel structure learning algorithm for optimal Bayesian network; best parents [J]. *Procedia Computer Science*, 2016,96: 43.
- [9] KO S, Kim D W. An efficient node ordering method using the conditional frequency for the K2 algorithm [J]. *Pattern Recognition Letters*, 2014, 40: 80.
- [10] BOUCHAALA L, Masmoudi A, Gargouri F, *et al.* Improving algorithms for structure learning in Bayesian Networks using a new implicit score [J]. *Expert Systems with Applications*, 2010, 37(7): 5470.
- [11] HRUSCHKA Jr E R, Ebecken N F F. Towards efficient variables ordering for Bayesian networks classifier [J]. *Data & Knowledge Engineering*, 2007, 63(2): 258.
- [12] 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): 1.
- [13] CANO A, GÓMEZ-OLMEDO M, MASEGOSA A R, *et al.* Locally averaged Bayesian Dirichlet metrics for learning the structure and the parameters of Bayesian networks [J]. *International Journal of Approximate Reasoning*, 2013, 54(4): 526.
- [14] WANG S C, GAO R, WANG L M. Bayesian network classifiers based on Gaussian kernel density [J]. *Expert Systems with Applications*, 2016, 51(C): 207.
- [15] 范建平,李常洪,吴美琴,等. 贝叶斯网络在中医诊断中的应用研究 [J]. *管理科学学报*, 2008,11(6): 143.
FAN Jianping, LI Changhong, WU Meiqin, *et al.* Study on application of Bayesian network in Chinese medicine diagnose [J]. *Journal of Management Sciences in China*, 2008, 11(6): 143.
- [16] 王学伟,瞿海斌,刘雪松,等. 贝叶斯网络杂交学习算法及其在中医中的应用 [J]. *浙江大学学报(工学版)*, 2005, 39(7): 948.
WANG Xuewei, ZHAI Haibin, LIU Xuesong, *et al.* Bayesian network approach to knowledge discovery in traditional Chinese medicine [J]. *Journal of Zhejiang University (Engineering Science)*, 2005, 39(7): 948.
- [17] OUALI A, CHERIF A R, KREBS M O. Data mining based Bayesian networks for best classification [J]. *Computational Statistics & Data Analysis*, 2006, 51(2): 1278.