

doi: 10.3969/j.issn.1002-0268.2018.05.013

基于贝叶斯网络模型的道路交通事故链生成与演化研究

熊晓夏, 陈 龙, 梁 军, 陈月霞
(江苏大学 汽车与交通工程学院, 江苏 镇江 212013)

摘要: 为了研究道路交通事故链的生成和演变规律并全面反映道路交通事故的产生机理, 以美国 100-Car 自然驾驶研究数据为基础, 充分考虑了驾驶员状态和驾驶行为对道路交通安全的影响, 构建了事故发生前驾驶员状态和行为特征参数等变量, 并与其他传统驾驶员特征、道路交通特性以及环境特点等影响因素变量共同建立了关于道路交通事故风险类型的贝叶斯网络模型。在贝叶斯网络模型基础上, 引入事故因果链理论, 利用有向无环网络中简单路径搜索算法生成事故链集合, 并采用信息增益特征选择方法识别关键事故链。通过 100-Car 自然驾驶数据得到的关键事故链显示, 单车事故类型更容易在弯道和坡道条件下驾驶员无注意力转移或进行简单非驾驶任务的情况下发生; 而正面、侧面和尾部碰撞事故类型的发生, 往往在水平直线道路条件下, 伴随着驾驶员注意力由前方道路转移至两侧车窗或内后视镜, 以及驾驶员非驾驶任务变复杂的情况。通过改变事故链中各节点的状态概率, 可以探索不同类型事故的演化路径和规律, 克服了传统方法对每个道路交通事故致因因素进行独立分析的局限性, 揭示了事故链中影响因素/事件之间的相关关系, 为更好地掌控道路交通风险状态和实现事故链阻断提供了新的思路。

关键词: 交通工程; 事故链生成与演化; 贝叶斯网络模型; 道路交通事故链; 路径搜索; 特征选择

中图分类号: U491.1

文献标识码: A

文章编号: 1002-0268 (2018) 05-0099-09

Study on Generation and Evolution of Road Traffic Accident Chain Based on Bayesian Network Model

XIONG Xiao-xia, CHEN Long, LIANG Jun, CHEN Yue-xia

(School of Automobile and Traffic Engineering, Jiangsu University, Zhenjiang Jiangsu 212013, China)

Abstract: In order to study the generation and evolution rules of road traffic accident chain and fully reflect the mechanism of road traffic accidents, based on the 100-car natural driving research data from the United States, the variables characterizing pre-accident driver status and behavior are constructed considering the impact of driver's status and driving behavior on road traffic safety, which are utilized along with other variables of influencing factors such as traditional driver, roadway, traffic, and environment characteristics for establishing the Bayesian network model for road traffic accident risk type. By introducing the casual chain theory of accidents, a simple path search algorithm for directed acyclic network is used to generate the accident chain set based on the established Bayesian Network, and the information gain feature selection method is used to identify the key accident chain. The result of critical accident chain derived from the 100-car natural driving research data shows that (1) single vehicle accidents tend to occur on curved and sloped roadways when drivers are not distracted or not engaged in complex secondary tasks; (2) while front, side, and rear-end collision accidents are more usually occurred under horizontal and straight roadway conditions

收稿日期: 2017-04-05

基金项目: 国家自然科学基金项目 (U1564201, 51108209, 50875112, 70972048); 江苏省普通高校研究生科研创新计划项目 (KYLX16_0905)

作者简介: 熊晓夏 (1987-), 女, 江西南昌人, 博士研究生. (x_xiong623@163.com)

with the driver's attention transferring from the roadway ahead to both sides of the window and rearview mirror or the secondary task of the driver becoming more complex. By changing the status probability of each node in the identified accident chain, the evolution path and rule of different types of accidents can be explored, which has overcome the limitation of traditional methods that analyze the causes of road traffic accidents separately, and reflected the interrelationship of the contribution factors and events of accident chain, and provided a new perspective on better control of road traffic risk and realizing the blockage of accident chain.

Key words: traffic engineering; accident chain generation and evolution; Bayesian network model; road traffic accident chain; path search; feature selection

0 引言

面对大量的道路交通事故及安全隐患,国内外在道路交通事故致因分析与预测方面进行了大量研究。其中,使用最广泛的研究方法包括单因素或多因素统计回归模型,如线性回归模型、多项 Logistic 回归模型、泊松回归模型,和负二项回归模型等^[1-3]。然而,这些模型通常仅描述了事故结果与其影响因素之间的定量关系,而未能反映事故发生前的事件和影响因素间的相互联系,即道路交通“事故链”^[4]。根据海因里希事故因果链理论,若能适当干预事故链的发展过程(如消除一个或多个不可能导致事件发生的不利事件或因素)则最终事故可能不会发生^[5]。因此,不同于传统方法中对每个影响事故结果的因素进行独立分析,道路交通事故链的研究将更有利于综合考察道路交通事故链的阻断策略,即通过阻止事故发生前风险事件的发生以达到及时预防道路交通事故的目的。

目前已有多种数据挖掘方法被应用于道路交通事故链的研究,包括分类树法、粗糙集法和神经网络法等^[6-8]。然而,这些方法通常将事故链表示为影响因素或事件的组合,未能进一步揭示这些影响因素/事件之间的相关关系,因此不利于交通事故阻断策略的具体构建。贝叶斯网络(Bayesian Network, BN)由于具有清晰语义的网络结构,可以揭示领域对象间的内在关系^[9],已被广泛应用于道路交通事故致因分析并取得较好的预测效果^[10-11],但尚未被应用于事故链演化路径和内在规律的挖掘研究。此外,现有道路交通事故链研究主要局限于驾驶员特征(年龄、性别等)、道路交通特性(信号类型和交通密度等)以及环境特点(天气和光照条件等)等因素,而较少考虑驾驶员状态和驾驶行为对道路交通安全的影响,因此不能全面反映道路交通事故的产生机理。

本研究充分考虑了交通系统中驾驶员的主体作

用,建立了基于贝叶斯网络的道路交通事故模型,获得了事故类型与相关因素间的关联关系,并以此为基础引入事故因果链理论,探索关键事故链的生成方法和演化规律,为掌控车辆运行风险状态和构建事故预防策略提供了新的思路。

1 相关理论与模型

本研究提出的基于贝叶斯网络的事故链推导方法主要包括两个阶段。第一阶段是建立具有事故结果和影响因素/事件关联关系的贝叶斯网络模型;第二阶段是在建立的贝叶斯网络模型的基础上生成事故链集合并识别关键事故链。以下分别对两阶段相关理论和模型方法进行详细介绍。

1.1 贝叶斯网络模型

贝叶斯网络是由 Pearl 提出的一种基于概率论和图论的不确定性知识表示和推理的模型^[12]。在贝叶斯网络中,不确定知识被表示为一个有向无环图,其中节点集合 $U = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ 代表有某种独立性假设的随机变量,有向边集合代表变量间的条件依赖关系,并用条件概率分布(Conditional Probability Distribution, CPD)进行定量描述:

$$CPD = \{Prob(x_i | pa(x_i)), x_i \in U\}, i = 1, 2, \dots, n, \quad (1)$$

式中 $pa(x_i)$ 表示子节点 x_i 的父节点集(即网络中存在由父节点集 $pa(x_i)$ 指向子节点 x_i 的有向边)。为了提高计算效率,贝叶斯网络设定了条件独立假设:即若给定变量 x_i 的父节点集,则 x_i 独立于其非子孙节点。在 CPD 和条件独立假设的基础上,贝叶斯网络所刻画的节点集合 U 的联合概率分布可表示为:

$$P_{BN}(U) = \prod_{x_i \in U} Prob(x_i | pa(x_i)), i = 1, \dots, n. \quad (2)$$

贝叶斯网络的构建过程可分为3个主要步骤:

(1) 特征选择:通过相关性分析方法确定贝叶斯网络相关的变量节点;

(2) 结构学习:通过专家/先验知识或机器学习算法学习样本数据确定贝叶斯网络的结构,机器学习算法主要包括基于评分搜索 (searching and scoring based) 的方法和基于条件约束 (constraint based) 的方法;

(3) 参数学习:在给定网络拓扑结构条件下,通过最大似然估计法或贝叶斯法进行网络参数 (即 CPD) 学习。

由于基于 MDL (Minimum Description Length) 评分的爬山搜索算法对于事故类型分类问题具有良好的学习能力^[10],故本研究亦采用该算法对 BN 进行结构学习。为了避免模型学习“过拟合”问题,且考虑到可获得的数据样本量有限,本研究采用十折交叉验证法对 BN 结构进行学习:即将所有样本分成 10 份,轮流将其中 9 份作为训练数据,1 份作为测试数据,进行 10 次 MDL 评分计算,并以 10 次结果的平均值作为当前 BN 结构下 MDL 评分的估计值,并以该估计值为基础进行爬山搜索 (包括对网络有向边进行添加、删减或者逆向操作) 直至获取 MDL 评分估计值最高的网络结构作为最终建立的 BN 结构。

类似地,为避免参数数目较多可能造成的过拟合问题,在给定的 BN 拓扑结构下采用贝叶斯参数估计方法学习 BN 各节点变量的条件概率分布。本研究采用基于 Dirichlet 先验分布^[11]的贝叶斯参数估计方法,即设节点变量 x_i (离散变量) 具有先验分布如下:

$$p(\theta) = \text{Dir}(\theta | \alpha_1, \dots, \alpha_r) = \frac{\Gamma(\alpha)}{\prod_{k=1}^r \Gamma(\alpha_k)} \prod_{k=1}^r \theta_k^{\alpha_k-1}, \quad (3)$$

式中, $\alpha = \sum_{k=1}^r \alpha_k$, α_k 为超参数且满足 $\alpha_k > 0$, $r \geq 2$ 表示变量 x_i 取值类别的数目; $\Gamma(\alpha)$ 为伽马函数,则节点变量 x_i 给定数据集 E (即观测样本) 的后验分布即为估计得到的该节点条件概率分布,其亦服从 Dirichlet 分布:

$$p(\theta | E) = \frac{p(\theta)p(E | \theta)}{p(E)} = \frac{p(\theta)p(E | \theta)}{\int p(\theta)p(E | \theta) d\theta} = \frac{\frac{\Gamma(\alpha)}{\prod_{k=1}^r \Gamma(\alpha_k)} \prod_{k=1}^r \theta_k^{\alpha_k-1} \cdot \prod_{k=1}^r \theta_k^{N_k}}{\int \frac{\Gamma(\alpha)}{\prod_{k=1}^r \Gamma(\alpha_k)} \prod_{k=1}^r \theta_k^{\alpha_k-1} \cdot \prod_{k=1}^r \theta_k^{N_k} d\theta} =$$

$$\frac{\Gamma(\alpha + N)}{\prod_{k=1}^r \Gamma(\alpha_k + N_k)} \prod_{k=1}^r \theta_k^{\alpha_k + N_k - 1} = \text{Dir}(\theta | \alpha_1 + N_1, \dots, \alpha_r + N_r), \quad (4)$$

式中, $N = \sum_{k=1}^r N_k$ 为数据集 E 的总样本量, N_k 为数据集 E 中变量 x 取值为第 k 类的样本量。

1.2 事故链生成与识别方法

根据事故因果链理论,一系列具有连锁关系的事件环节/要素 $A_i (i = 1, 2, \dots, n)$ 相继触发最终引起风险事故 A 的发生,若事故链中的任何一个事件环节/要素不发生则事故过程被中止,最终事故将不会发生^[7]。因此,对道路交通风险事故链的研究有利于更好地掌控交通系统的风险状态,可为实现道路交通事故链的及时阻断奠定理论基础。

值得注意的是,事故因果链中各相邻事件环节/要素 $A_i (i = 1, 2, \dots, n)$ 可以具体指系统发生的某个事件/行为,也可以指系统所处的某种状态。因此,事故链中各事件环节/要素的发展路径并不一定代表实际事故的发生过程,而仅代表了不同事件环节/要素和事故结果在逻辑上或时间/空间上的关联关系^[4]。贝叶斯网络可以通过有向链接图对事故链中的各事件要素的因果关系和条件相关关系进行描述,因此适用于风险事故链发展路径的识别。

对于不同的道路交通事故,其对应的事故链发展路径亦不止一条,因此为了更好地掌控风险,需要建立尽可能完整的事故链集合。本研究在构建的贝叶斯网络基础上,利用有向无环网络中简单路径搜索算法^[13]对事故链路径进行搜索,得到完整的风险事故链集合。当某一根节点发展出多条事故链,即其风险传播路径具有多条支路时,需要辨识出其中最有可能的传播路径即关键事故链,以便进一步明确风险控制要点。信息增益 (Information Gain, IG) 作为机器学习中一种常用的特征选择方法^[14],可以衡量某个特征使分类系统信息熵减少的量,即减少分类不确定性的能力,可用于关键事故链的识别。整个识别过程具体如下:

(1) 生成事故链集合:根据有向无环网络中简单路径搜索算法,将有向路径终点设为风险类型 A ,并由 A 开始向上游方向依次将其所有祖先节点以结构体 L 形式存入链表,其中结构体 L 包括 Data, No. 和 Child No. 这 3 个变量,分别代表存入节点的名称、该节点进入链表的序号和该节点的子节点进入链表的序号,最后在链表的基础上搜索网络模型中各根节点

至 A 的所有简单路径 (即无重复出现节点的序列路径) (详见文献 [13]), 得到风险事故链集合。

(2) 识别关键事故链: 对风险传播路径具有多条支路的事事故链, 计算支路上游节点对各支路节点的信息增益^[14], 选取信息增益比最大的支路节点作为关键事故链的下级支路, 即给定上游节点下不确定性减少比例最大的支路节点, 循环该过程直至到达路径终点 A 完成该根节点的关键事故链识别。上游节点变量 T 给支路节点变量 C 带来的信息增益比 $IGP(C|T)$ 定义为: C 原本的熵与给定 T 后的条件熵之差 (即信息增益 $IG(C|T)$) 占 C 原本的熵的比例, 即:

$$IGP(C|T) = IG(C|T)/H(C) = (H(C) - H(C|T))/H(C), \quad (5)$$

式中 $H(C)$ 表示节点变量 C 的信息熵, 假设 C 包含 $C_1, C_2, \dots, C_n$ 共 n 个类别 (即 n 个离散取值), 每一个类别出现的概率依次为 $P(C_1), P(C_2), \dots, P(C_n)$, 则其熵定义为:

$$H(C) = - \sum_{i=1}^n P(C_i) \cdot \log_2 P(C_i). \quad (6)$$

$H(C|T)$ 表示当给定上游节点 T (可取值 t 和 $\bar{t}$) 时节点变量 C 的条件信息熵, 其在式 (6) 的基础上按全概率分解法可分解为:

$$\begin{aligned} H(C|T) &= P(t) \cdot H(C|t) + P(\bar{t}) \cdot H(C|\bar{t}) = \\ &- P(t) \cdot \sum_{i=1}^n P(C_i|t) \cdot \log_2 P(C_i|t) - P(\bar{t}) \cdot \\ &\sum_{i=1}^n P(C_i|\bar{t}) \cdot \log_2 P(C_i|\bar{t}). \end{aligned} \quad (7)$$

式 (5)~(7) 中各概率值可由构建的贝叶斯网络的条件概率分布 CPD 获得。

2 数据来源

本研究用于建立交通事故贝叶斯网络的数据来源于美国弗吉尼亚理工于 2004—2005 年发起的“100-car 自然驾驶研究”所获得数据^[15], 其中包括 19 600 段正常驾驶数据、68 起事故数据和 760 起临近事故数据 (即驾驶员采取了紧急刹车或紧急避让行为的情形)。“自然驾驶”研究即在车辆上安装包括多普勒雷达、摄影头、数据集成记录器等在内的一套数据采集系统, 全时监测和记录驾驶员实际驾驶过程中车辆的运行状态、驾驶员驾驶状态, 以及交通和环境状态等特征信息。在数据采集过程中, 驾驶员不会得到任何特别指示, 而是按照各自日常的驾驶行为和习惯进行驾驶, 以便获取驾驶员在受

干扰最小的自然状况下最真实的驾驶行为特征。与传统的交通事故数据采集方式 (通过事故现场调查及驾驶员和目击证人的证词) 相比, 自然驾驶研究方法可以获取更丰富的事故相关数据, 如事故发生前驾驶员操作行为、驾驶员注意力分散现象、周围交通流量环境等人-车-路-环境动态交互信息, 故可更全面、系统地分析驾驶员的行为和事故原因, 因此更适用于交通事故内在机理的研究。

考虑到 100-car 数据库中发生事故的样本量较少 (68 起), 且研究表明临近事故数据与实际事故数据具有内在一致性可共同进行事故机理分析^[16-18], 本研究将数据库中事故数据和临近事故数据合并得到共 828 个风险事故观测样本。数据库中原变量达 30 多种, 结合其他相关建模经验初步筛选出 22 个变量为模型相关变量。由于其中多个变量 (如驾驶员注意力情况、驾驶员方向盘使用情况、道路照明情况等) 均是基于视频资料通过人工获取, 样本中存在由于视频片段缺失导致变量观测值不完整的情况。通过剔除观测值缺失和错误的无效样本以及特殊场景下 (如进出停车场或 U 形转弯) 的观测样本, 最后获得共 724 个有效风险事故样本进行贝叶斯网络构建和风险事故链辨识。

为满足建模要求, 将所选部分连续变量编码处理为离散变量, 各变量的设置如表 1 所示。由表 1 可以看出, 100-car 数据集中提供了多个关于事故发生前驾驶员行为和注意力状态的变量, 其中驾驶员“非驾驶任务”指驾驶员进行的与驾驶任务无关的任意其他行为 (如打电话或饮食等), 其可按实现该行为的复杂程度进行等级划分, 本研究按文献 [19] 的分类标准将驾驶员的非驾驶任务行为分为简单、中等和复杂 3 大类。

3 贝叶斯网络建立与验证

运用 SPSS 相关性分析选取 AGE (驾驶员年龄)、MNV (事故发生前驾驶员驾驶行为)、ICM (注意力转移至内后视镜)、ILM (注意力转移至左/右后视镜)、ILW (注意力转移至左/右车窗)、NST (非驾驶任务数目)、HST (非驾驶任务最高等级)、SBU (安全带使用情况)、NTL (车道数)、TRD (交通流密度)、TRC (交通控制方式)、RTJ (与交叉口位置关系)、CUR (道路线形)、GRD (道路纵坡)、LDU (土地使用类型)、WTH (天气状况)、SUR (路面状况)、ACT (事故类型) 共 18 个变量建立贝叶斯网络。注意到, 运用相关性分析筛选出的变量包括

表 1 变量设置表
Tab. 1 List of variables

变量类别	变量名称	符号	变量取值和分类描述	均值	标准差
人	驾驶员年龄	AGE	1. 青年 (18 – 35); 2. 中年 (36 – 55); 3. 老年 (大于 55)	1.45	0.66
	驾驶员性别	GEN	1. 男; 2. 女	1.46	0.50
	事故发生前驾驶员操纵行为	MNV	1. 匀速直行; 2. 加速直行; 3. 减速直行; 4. 变换车道; 5. 启动/停止; 6. 左/右转弯	2.38	1.44
	驾驶员注意力是否由前方道路转移至内后视镜	ICM	1. 否; 2. 是	1.07	0.26
	驾驶员注意力是否由前方道路转移至左/右后视镜	ILM	1. 否; 2. 是	1.06	0.24
	驾驶员注意力是否由前方道路转移至左/右车窗	ILW	1. 否; 2. 是	1.13	0.33
	驾驶员非驾驶任务数目	NST	1. 0 个; 2. 1 个; 3. 2 个; 4. 3 个	1.72	0.77
	驾驶员非驾驶任务等级	HST	1. 无; 2. 简单; 3. 中等; 4. 复杂	2.07	1.05
	驾驶员方向盘使用情况	HOW	1. 双手脱离方向盘; 2. 仅一只手握方向盘; 3. 双手握方向盘	2.64	0.78
	驾驶员是否使用安全带	SBU	1. 否; 2. 是	1.94	0.47
道路	交通流向	TRF	1. 物理隔离 (中分带或护栏); 2. 未物理隔离; 3. 单向交通	1.39	0.54
	车道数	NTL	1. 单车道; 2. 双车道; 3. 3 车道; 4. 4 车道; 5. > = 5 车道	2.95	0.98
	交通密度	TRD	1. 服务水平 – A; 2. 服务水平 – B; 3. 服务水平 – C; 4. 服务水平 – D; 5. 服务水平 – E & F	2.30	1.04
	交通控制方式	TRC	1. 无信号或标志标线控制; 2. 信号控制; 3. 标志标线控制	1.44	0.82
	路段类型	RTJ	1. 一般路段; 2. 平面交叉口内部; 3. 平面交叉口外部; 4. 立交内部; 5. 立交进出匝道; 6. 其他	1.88	1.34
	道路线形	CUR	1. 直线; 2. 弯道	1.15	0.36
	道路纵坡	GRD	1. 水平; 2. 坡道	1.03	0.18
环境	土地使用类型	LDU	1. 商业; 2. 居民; 3. 户外; 4. 洲际公路; 5. 其他 (施工区域、学校等)	2.38	1.33
	照明情况	LIG	1. 白天; 2. 黄昏; 3. 夜晚, 有照明; 4. 夜晚, 无照明	1.62	1.02
	天气状况	WTH	1. 晴朗; 2. 多云; 3. 其他 (雨、雪、雾等)	1.33	0.64
	路面状况	SUR	1. 干燥; 2. 湿/滑	1.14	0.34
事故	事故类型	ACT	1. 单车事故; 2. 车辆正面碰撞; 3. 车辆侧面碰撞; 4. 车辆尾部碰撞	2.24	0.78

驾驶员驾驶行为变量 (*MNV*) 和驾驶员注意力不集中变量 (*ICM*, *ILM*, *ILW*, *NST*, *HST*), 与其他相关研究显示的驾驶员行为状态与交通事故存在致因关系的结果一致^[20], 表明有必要将这些变量纳入反映道路交通事故不确定关系的贝叶斯网络进行进一步分析。

根据 1.1 讨论所述, 采用基于交叉验证和 *MDL* 评分的爬山搜索算法及基于 Dirichlet 先验分布的贝叶斯参数估计方法, 应用 Matlab 的 Full - BNT 工具箱进行 BN 的结构和参数学习, 最终获得 BN 学习结果如图 1 所示。

为验证建立的贝叶斯网络模型的预测效果,本研究采用混淆矩阵 (Confusion Matrix) 对比分析预测分类结果和实际观测结果 (见表2), 同时以各分类类别的 ROC 曲线 (Receiver Operating Characteristic Curve, 即受试者工作特征曲线) 综合评价各事故类型的预测效果 (见图2)。ROC 曲线是以预测真阳性

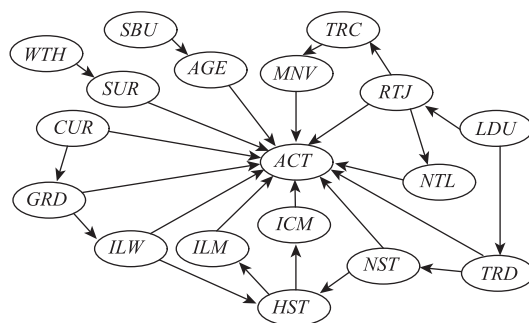


图 1 基于 MDL 评分爬山算法的交通事故 BN 结构

Fig. 1 BN structure of traffic accident using MDL-scoring hill-climbing algorithm

率（即灵敏度）为纵坐标，假阳性率（1 - 特异度）为横坐标绘制的曲线，若曲线下包含面积 AUC （Area Under Curve）越大（ AUC 范围为 0 ~ 1），则表示该类别的综合预测效果越好^[21]。

表2 BN交通事故预测混淆矩阵

Tab.2 Confusion matrix for BN traffic accident prediction

实际事故类型	预测事故类型				命中率/%
	单车	正面	侧面	尾部	
	ACT=1	ACT=2	ACT=3	ACT=4	
单车 ACT=1	67	13	3	0	80.7
正面 ACT=2	25	388	28	16	84.9
侧面 ACT=3	4	13	93	3	82.3
尾部 ACT=4	1	11	2	57	80.3
总体预测正确率/%					82.1

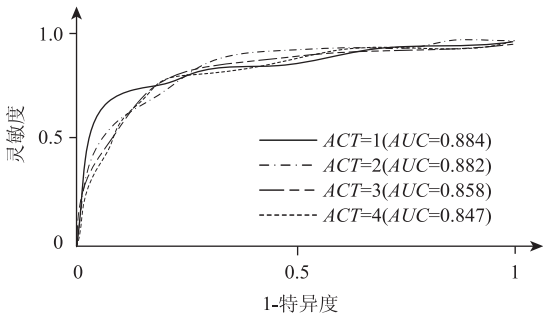


图2 各事故类型BN预测ROC曲线

Fig.2 ROC curves of BN prediction of traffic accident type

由混淆矩阵可以看出,本研究建立的BN预测平均命中率达80%以上,反映了该模型具有较好的预测精度(通常80%以上命中率即可被认为是理想结果^[22])。同时ROC曲线结果显示各事故类型预测

表3 基于BN的交通事故链集合和关键事故链

Tab.3 Set of accident chains and critical accident chain based on BN

编号	风险事故链发展路径						编号	风险事故链发展路径					
1*	SBU	AGE	ACT				8	LDU	TRD	ACT			
2*	WTH	SUR	ACT				9	LDU	RTJ	ACT			
3	CUR	ACT					10	LDU	TRD	NST	ACT		
4	CUR	GRD	ACT				11	LDU	RTJ	NTL	ACT		
5	CUR	GRD	ILW	ACT			12*	LDU	RTJ	TRC	MNV	ACT	
6	CUR	GRD	ILW	HST	ILM	ACT	13	LDU	TRD	NST	HST	ILM	ACT
7*	CUR	GRD	ILW	HST	ICM	ACT	14	LDU	TRD	NST	HST	ICM	ACT

以根节点CUR为例,其关键事故链(即编号7)为相应信息增益比最大的支路节点组成的路径

ROC曲线下面积基本均大于0.85,表明该BN模型对各事故类型的综合预测效果亦较好。为了进一步验证BN模型的预测效果,本研究同时建立了多项Logistic回归(Multinomial Logistic Regression, MNL)模型进行对比,两者预测结果对比如图3所示。结果显示相同样本条件下MNL回归模型的总体预测正确率为81.7%,与BN预测结果相近,说明本研究建立的BN模型可信度较高。因此,可以进一步应用建立的贝叶斯网络得到的交通事故各事件要素间的连锁关系生成风险事故链集合。

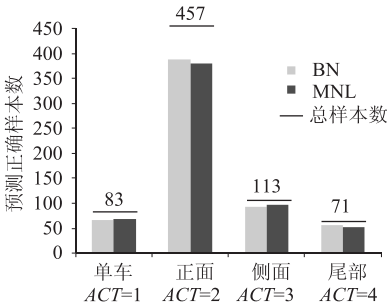


图3 BN和MNL预测结果比较

Fig.3 Comparison of prediction results of BN and MNL

4 交通事故链生成识别与评价

在以上建立的BN模型基础上,由1.2节给出的算法步骤,可获得表3中所列出的14条事故链和4条关键事故链(*标识)。

(如图4所示),计算信息增益比所需的相关节点先验概率和条件概率可由CPD直接获得(如表4所

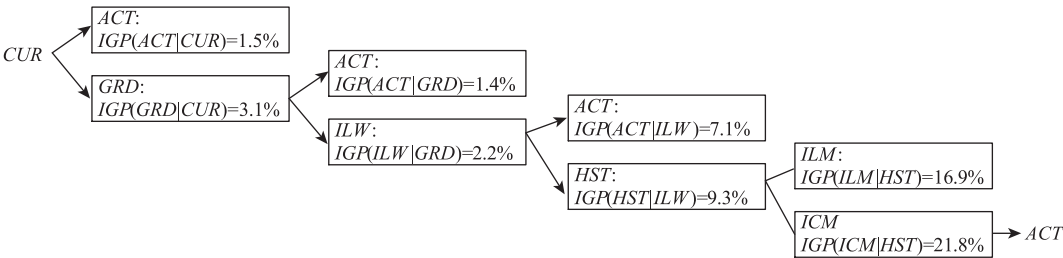


图4 关键事故链(路径7)预测线路

Fig.4 Prediction line of critical accident chain (route No. 7)

示),各节点的状态概率可通过因果前向推理方法计算获得(结果如图5所示)。

表4 关键事故链(路径7)根节点先验概率与其他节点条件概率

Tab.4 Prior probability of root nodes and conditional probability of other nodes of critical accident chain (route No.7)

CUR	GRD	CUR		ILW	GRD	
		直线	弯道		水平	坡道
直线	0.848	水平	0.923	否	0.871	0.913
弯道	0.152	坡道	0.077	是	0.129	0.087

HST	ILW		ICM	HST				ACT	ICM	
	否	是		无	简单	中等	复杂		否	是
无	0.511	0.021	否	0.998	0.991	0.824	0.934	单车	0.122	0.028
简单	0.131	0.022	是	0.002	0.009	0.176	0.066	正面	0.648	0.407
中等	0.275	0.932						侧面	0.138	0.386
复杂	0.082	0.026						尾部	0.092	0.179

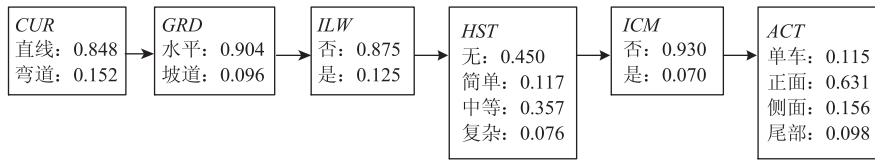


图5 关键事故链(路径7)状态概率表

Fig.5 State probability table of critical accident chain (route No.7)

$$H(GRD|CUR=1) = - \sum_{i=1}^2 P(GRD=i|CUR=1) \cdot \log_2 P(GRD=i|CUR=1) = -0.923 \times \log_2 0.923 - 0.077 \times \log_2 0.077 = 0.392, \quad (10)$$

$$H(GRD|CUR=2) = - \sum_{i=1}^2 P(GRD=i|CUR=2) \cdot \log_2 P(GRD=i|CUR=2) = -0.796 \times \log_2 0.796 - 0.204 \times \log_2 0.204 = 0.730, \quad (11)$$

$$H(GRD|CUR) = P(CUR=1) \cdot H(GRD|CUR=1) + P(CUR=2) \cdot H(GRD|CUR=2) = 0.848 \times 0.392 + 0.152 \times 0.730 = 0.443, \quad (12)$$

$$IGP(GRD|CUR) = (H(GRD) - H(GRD|CUR)) / H(GRD) = (0.457 - 0.443) / 0.457 = 3.1\%. \quad (13)$$

在表4和图5的基础上,可通过改变某类型事故是否发生(将该类型状态概率设为1.0作为输入证据)利用贝叶斯定理进行后向推理,得到证据在整个链中的更新传播结果如图6所示。

由图6可得100-car采集数据显示,单车事故类型更容易在弯道和坡道条件下驾驶员无注意力转移或进行简单非驾驶任务的情况下发生,而正面、侧面和尾部碰撞事故类型的发生往往在水平直线道路条件下,伴随着驾驶员注意力由前方道路转移至

以CUR对GRD的信息增益比IGP(GRD|CUR)为例,其计算过程如下:

由全概率公式查表4得:

$$P(GRD=1) = \sum_{i=1}^2 P(GRD=1|CUR=i) \cdot P(CUR=i) = 0.923 \times 0.848 + 0.796 \times 0.152 = 0.904. \quad (8)$$

同理可计算得 $P(GRD=2) = 0.096$ 。由式(6)和(7)得:

$$H(GRD) = - \sum_{i=1}^2 P(GRD=i) \cdot \log_2 P(GRD=i) = -0.904 \times \log_2 0.904 - 0.096 \times \log_2 0.096 = 0.457, \quad (9)$$

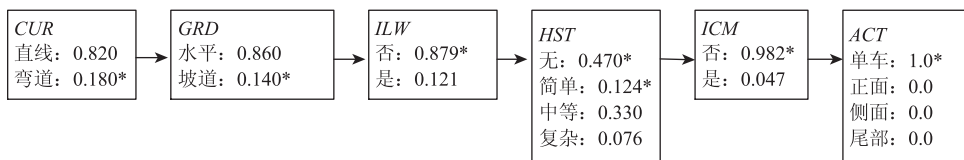
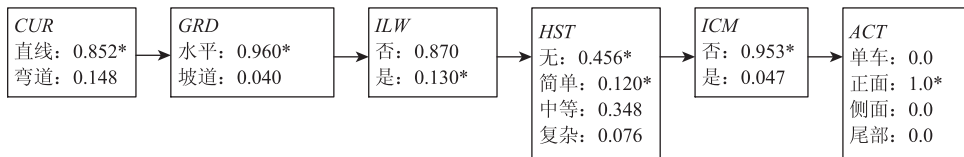
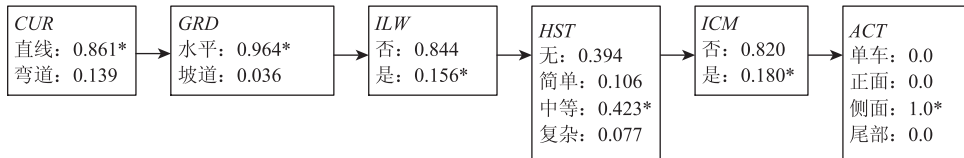
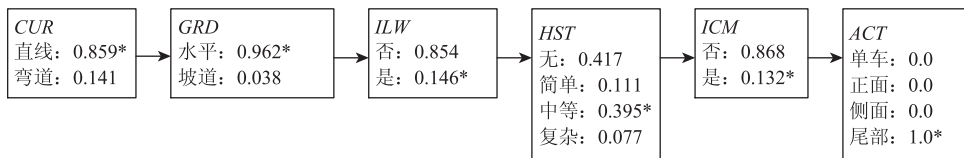
左/右车窗或内后视镜,及驾驶员非驾驶任务变复杂的情况。按照这种方法,通过改变事故链中各节点的状态概率,可探索不同类型事故的形成条件,从而为实现事故链的阻断提供有益信息,例如可对实时道路交通环境和驾驶员状态进行监测,在事故链各事件要素满足上述条件时给驾驶员发送预警信息,从而直接干预ILW/HST/ICM的状态以达到阻断事故发生的目的。

5 结论

本研究在贝叶斯网络模型的基础上挖掘了道路交通事故链的演化路径和内在规律,具体研究结果如下:

(1) 以美国100-car自然驾驶研究数据为数据来源,引入驾驶员状态和行为特征参数等变量,建立了道路交通事故风险类型的贝叶斯网络模型,结果显示该模型对道路交通事故风险类型具有较好的预测效果。

(2) 基于构建的贝叶斯网络模型,引入事故因果链理论,利用有向无环网络中简单路径搜索算法生成事故链集合,并通过信息增益特征选择方法识别关键事故链。

(1) 输入证据: $ACT=1$ (2) 输入证据: $ACT=2$ (3) 输入证据: $ACT=3$ (4) 输入证据: $ACT=4$ 

注: * 表示给定事故类型后概率增大 (与图5中状态概率相比) 的事件。

图6 关键事故链 (路径7) 更新状态概率表

Fig. 6 Updated state probability table of critical accident chain (route No. 7)

(3) 以获取的关键事故链演化路径为基础, 通过改变事故链中各节点的状态概率, 探索不同类型事故的演化规律, 为更好地掌控道路交通风险状态和实现事故链阻断提供了新的思路。

由于现有实际事故样本量不足, 本研究亦将临近事故样本数据纳入了事故机理分析, 未来需获取更多实际事故样本进行对比研究; 且由于目前样本包含的条件变量有限, 辨识得到的风险事故链线路和节点的数目均较少, 未来通过采集更多的条件变量将有利于识别更丰富完整的事故链, 进而设计更有效的事故链阻断策略。另外, 由于100-car自然驾驶数据是在美国采集的, 其反映的道路交通环境和驾驶员行为特性与国内情况可能有所不同, 未来还需针对我国实际道路交通情况进行数据采集, 对本研究提出的模型和识别方法进行实证研究。

参考文献:

References:

[1] 袁伟, 付锐, 郭应时, 等. 考虑坡长因素的纵坡坡度对

交通事故的影响分析 [J]. 公路交通科技, 2008, 25 (5): 130-135.

YUAN Wei, FU Rui, GUO Ying-shi, et al. Influences of Longitudinal Gradient on Traffic Accident Rate Considering Length of Downgrade [J]. Journal of Highway and Transportation Research and Development, 2008, 25 (5): 130-135.

[2] AL-GHAMDI A S. Using Logistic Regression to Estimate the Influence of Accident Factors on Accident Severity [J]. Accident Analysis and Prevention, 2002, 34 (6): 729-741.

[3] LORD D, WASHINGTON S P, IVAN J N. Poisson, Poisson-gamma and Zero-inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory [J]. Accident Analysis and Prevention, 2005, 37 (1): 35-46.

[4] ELVIK R. Assessing the Validity of Road Safety Evaluation Studies by Analyzing Causal Chains [J]. Accident Analysis and Prevention, 2003, 35 (5): 741-748.

[5] 毛敏, 喻翔. 道路交通事故致因分析 [J]. 公路交通

- 科技, 2002, 19 (5): 125 - 127.
- MAO Min, YU Xiang. Analysis of Traffic Accident Causation [J]. Journal of Highway and Transportation Research and Development, 2002, 19 (5): 125 - 127.
- [6] CHANG L Y, WANG H W. Analysis of Traffic Injury Severity: An Application of Non-parametric Classification Tree Techniques [J]. Accident Analysis and Prevention, 2006, 38 (5), 1019 - 1027.
- [7] WONG J T, CHUNG Y S. Rough Set Approach for Accident Chains Exploration [J]. Accident Analysis and Prevention, 2007, 39 (3): 629 - 637.
- [8] DELEN D, SHARDA R, BESSONOV M. Identifying Significant Predictors of Injury Severity in Traffic Accidents Using a Series of Artificial Neural Networks [J]. Accident Analysis and Prevention, 2006, 38 (3): 434 - 444.
- [9] 郑恒, 李长城, 刘煜. 基于贝叶斯网络的路侧安全评价方法 [J]. 公路交通科技, 2008, 25 (2): 139 - 144.
- ZHENG Heng, LI Chang-cheng, LIU Yu. A Roadside Safety Assessment Method Based on Bayesian Network [J]. Journal of Highway and Transportation Research and Development, 2008, 25 (2): 139 - 144.
- [10] DE OÑA J, LÓPEZ G, MUJALLI R, et al. Analysis of Traffic Accidents on Rural Highways Using Latent Class Clustering and Bayesian Networks [J]. Accident analysis and prevention, 2012, 51 (2): 1 - 10.
- [11] 许洪国, 张慧永, 宗芳. 交通事故致因分析的贝叶斯网络建模 [J]. 吉林大学学报: 工学版, 2011, 41 (增1): 89 - 94.
- XU Hong-guo, ZHANG Hui-yong, ZONG Fang. Bayesian Network Modeling for Causation Analysis of Traffic Accident [J]. Journal of Jilin University: Engineering and Technology Edition, 2011, 41 (S1): 89 - 94.
- [12] PEARL J. Probabilistic Reasoning in Intelligent Systems; Networks of Plausible Inference [J]. Computer Science Artificial Intelligence, 1988, 70 (2): 1022 - 1027.
- [13] 刘夫云, 祁国宁, 车宏安. 复杂网络中简单路径搜索算法及其应用研究 [J]. 系统工程理论与实践, 2006, 26 (4): 9 - 13.
- LIU Fu-yun, QI Guo-ning, CHE Hong-an. Research on Algorithm for Detecting Simple Path in Complex Network and Its Application [J]. Systems Engineering - Theory & Practice, 2006, 26 (4): 9 - 13.
- [14] BISHOP C M. Pattern Recognition and Machine Learning [M]. New York: Springer-Verlag New York, Inc. 2006.
- [15] DINGUS T A, KLAUER S G, NEALE V L, et al. The 100-Car Naturalistic Driving Study: Phase II - Results of the 100-Car Field Experiment, DOT HS 810 593 [R]. Washington, D. C.: U. S. Department of Transportation, 2006.
- [16] GUO F. Near-crashes as Crash Surrogate for Naturalistic Driving Studies [J]. Transportation Research Record, 2010, 2147: 66 - 74.
- [17] WANG J, ZHENG Y, LI X, et al. Driving Risk Assessment Using Near-crash Database through Data Mining of Tree-based Model [J]. Accident Analysis and Prevention, 2015, 84: 54 - 64.
- [18] WU K F, JOVANIS P P. Defining and Screening Crash Surrogate Events Using Naturalistic Driving Data [J]. Accident Analysis and Prevention, 2013, 61 (8): 10 - 22.
- [19] DINGUS T A, HULSE M C, ANTIN J F, et al. Attentional Demand Requirements of an Automobile Moving: Map Navigation System [J]. Transportation Research Part A: General, 1989, 23 (4): 301 - 315.
- [20] TALBOT R, FAGERLIND H, MORRIS A. Exploring Inattention and Distraction in the SafetyNet Accident Causation Database [J]. Accident Analysis and Prevention, 2013, 60 (45): 445 - 455.
- [21] MARCOT B G, STEVENTON J D, SUTHERLAND G D, et al. Guidelines for Developing and Updating Bayesian Belief Networks Applied to Ecological Modeling and Conservation [J]. Canadian Journal of Forest Research, 2006, 36 (12): 3063 - 3074.
- [22] COOPER G F, HERSKOVITS E. A Bayesian Method for the Induction of Probabilistic Networks from Data [J]. Machine Learning, 1992, 9 (4): 309 - 347.