

文章编号:1671-8879(2014)05-0109-08

# 交通事故严重程度 C5.0 决策树预测模型

孙轶轩<sup>1</sup>, 邵春福<sup>1</sup>, 赵丹<sup>2</sup>, 欧阳松寿<sup>3</sup>

(1. 北京交通大学 城市交通复杂系统理论与技术教育部重点实验室, 北京 100044; 2. 中国人民公安大学 交通管理系, 北京 102623; 3. 北京市交通委员会 运输管理局, 北京 100053)

**摘要:**根据中国现行交通事故严重程度分类与事故信息数据分布特征, 基于 C5.0 决策树方法, 选取某省会城市城区及周边重点公路 16 009 起交通事故现场数据, 分别将事故严重程度输出变量按照 2 分类和 3 分类, 输入变量按照空间属性、涉事驾驶人及车辆属性和全属性, 建立事故严重程度预测模型, 生成相应规则集并利用测试样本进行检验和模型对比。研究表明: 2 分类和 3 分类事故严重程度预测模型精度分别为 70% 和 61%, 多模型综合优度有所提升; 实证规则集揭示了影响事故严重程度分类的因素主要有, 碰撞类型、道路属性、事故致因和驾驶人类型等。

**关键词:**交通工程; 交通事故严重程度; 预测模型; 数据挖掘; 决策树; C5.0 算法

**中图分类号:**U491.31 **文献标志码:**A

## Traffic accident severity prediction model based on C5.0 decision tree

SUN Yi-xuan<sup>1</sup>, SHAO Chun-fu<sup>1</sup>, ZHAO Dan<sup>2</sup>, OU-YANG Song-shou<sup>3</sup>

(1. Key Laboratory for Urban Transportation Complex Systems Theory and Technology of the Ministry of Education, Beijing Jiaotong University, Beijing 100044, China; 2. Department of Transportation Management, China People's Public Security University, Beijing 102623; 3. Transportation Administration of Beijing Municipal Commission of Transport, Beijing 100053, China)

**Abstract:** Based on the algorithm of C5.0 decision tree, current severity classification of traffic accidents and the distribution characteristics of accident information data, this paper used the field data of 16,009 traffic accidents which occurred in some main highways of the urban area in and around a certain capital city to analyze the accident severity, and established a prediction model of accident severity according to the output variables based on dichotomy and trichotomy as well as the input variables based on the spatial attributes, the driver involved, vehicle attributes and the overall attributes. Through the test, this paper got the appropriate rule set and used the test samples for inspection and the comparison of models. The results show that the accuracy of the prediction model in accident severity is 70% and 61% separately based on dichotomy and trichotomy, and the integrated goodness of multi-model is improved. The empirical rule set reveals that the factors influencing accident severity classification are mainly the type of collision, road attributes, accident causation and the type of driver. 5 tabs, 2 figs, 14 refs.

**Key words:** traffic engineering; traffic accident severity; prediction model; data mining; decision tree; C5.0 algorithm

收稿日期: 2014-03-20

基金项目: 国家重点基础研究发展(973)计划资助项目(2012CB725403); 国家自然科学基金国际合作重大项目(71210001)

作者简介: 孙轶轩(1982-), 男, 山西太原人, 工学博士研究生, E-mail: squallsyx@163.com.

## 0 引言

道路交通事故是当今社会公民生命财产安全的主要威胁之一。事故严重程度预测研究逐渐成为交通安全管理者、研究者及车辆厂家关注的重要内容。交通事故严重程度影响因素主要包括人、车、路和环境等。国外学者针对交通事故严重程度预测和影响因素的研究开展较早,主要研究方法包括经典统计回归模型,如二项或多项 Logit、Probit 模型<sup>[1-2]</sup>;考虑事故严重程度分类有序离散属性(ordered discrete)的广义或贝叶斯有序 Logit 或 Probit 模型<sup>[3-4]</sup>;考虑选择枝间相关性约束条件的嵌套 Logit 模型、混合 Logit 模型以及马尔科夫链多元 Logit 模型<sup>[5-7]</sup>;也有研究者使用数据挖掘技术,如人工神经网络建立事故严重程度预测模型<sup>[8]</sup>。中国有关事故严重程度的相关研究主要有:李世民等使用累积 Logit 模型分析发生于北京市无信号交叉口的事故严重程度与交叉口属性之间的关系<sup>[9]</sup>;马壮林等使用 Logit 模型研究公路隧道交通事故严重程度的影响因素,使用模糊和灰色 Dlephi 法研究基于事故严重程度的道路安全评价<sup>[10-11]</sup>;侯树展等利用主成分分析(PCA)技术,研究交通流因素对事故严重程度的影响<sup>[12]</sup>。

经典统计回归模型由于其严格的模型假设以及输入、输出变量间既定的函数形式,在交通事故严重程度研究的实际应用中对数据集结构要求较高,易出现由于有偏参数估计导致的模型解释能力下降的现象。决策树(decision tree)是通过归纳和提炼现有数据包含的规律,并用于新数据分类预测的一种非参数方法。其特点是不固定特定函数形式,且不需要任何样本数据先验分布假设。考虑到中国现行道路现场信息数据结构中,输入变量繁多且多为分类变量的数据特征,本文以决策树模型为工具,通过对实证数据的决策规则集发掘建立事故严重程度预测模型,对影响事故严重程度的安全因素进行分析,探讨决策树应用于事故严重程度分析的可行性,为相关政策法规的制定提供理论支持。

## 1 决策树模型与算法

### 1.1 决策树的构造机理

决策树是由 1 个根节点和数个叶节点及中间节点组成的树状结构,每个节点均是具有一定样本量的样本集合,根节点的样本量最大,其他节点的样本

量依层递减。决策树学习过程是数据驱动的自顶向下的递归方法,每一层节点依照某一属性向下分子节点,待分类目标在每一节点处与该节点相关属性进行比较,根据结果向响应的子节点扩展,直到到达决策树的叶节点时结束。

决策树建模过程体现了对样本数据不断分组的过程,即考虑训练数据集  $x$ ,根据输出变量的取值和分组规则将  $x$  分为 2 个或更多子集的所有可能分枝,从而生成分类规则集。与其他分类预测模型相比,决策树模型是基于布尔逻辑的分类预测,即利用 If……then……形式,通过输入变量取值的逻辑性预测输出变量的取值。

### 1.2 决策树的生长和修剪算法

决策树学习过程包括生长和修剪 2 个部分。生长算法即分枝准则,是通过最佳分组变量和组内取值最佳分割点的确定,使样本分组差异显著减少,分枝输出变量尽快趋同,当新的分枝不再有意义时,生长过程结束。修剪算法是通过预修剪或后修剪 2 种方法限制决策树的成长深度,避免由于生长算法对样本特征的过度描述引起过拟合(over-fitting)效应。预修剪直接限制决策树的生长深度,而后修剪则是允许决策树充分生长基础上,根据训练样本和测试样本集的过度拟合程度再进行修剪。

### 1.3 基于信息增益与误差估计的 C5.0 算法

本文采用 C5.0 算法构造决策树。C5.0 算法引入了自适应增强 Boosting 技术,即通过对现有加权样本的反复抽样模拟增加样本集以提高模型在测试集上的稳健性。考虑到中国事故数据输入变量多数为分类且非二值变量,决策树形态则选择多叉树进行事故严重程度分析,以避免二叉树结构造成的信息流失。此外,在相同预测精度的约束条件下,多叉树的分裂次数一般少于二叉树,减少判别次数可提高决策树的生长效率。

#### 1.3.1 C5.0 生长算法

与经典 CART 算法采用 Gini 系数减少量作为异质性下降程度的测度不同,C5.0 算法的生长算法是以信息增益率为标准的分枝准则,寻找最佳分组变量和分割点。

信息论定义信息传递是由信源(发送端)、信道、信宿(接收端)组成的系统活动,发送信息记为  $U$ ,接受信息记为  $V$ ,则信道模型由条件概率  $P(U/V)$  矩阵组成,记为

$$\begin{bmatrix} P(u_1|v_1) & P(u_2|v_1) & \cdots & P(u_r|v_1) \\ P(u_1|v_2) & P(u_2|v_2) & \cdots & P(u_r|v_2) \\ \vdots & \vdots & \ddots & \vdots \\ P(u_1|v_q) & P(u_2|v_q) & \cdots & P(u_r|v_q) \end{bmatrix} \quad (1)$$

$P(u_i|v_j)$  表示信宿收到信息  $v_j$  同时信源发出信息  $u_i$  的概率,且满足  $\sum P(u_i|v_j)=1$ 。

信息量描述了通信过程中消除随机不确定性的 大小,包括来源于信源的先验不确定性和系统干扰造成的信宿对信源信息接收的后验不确定性两部分。定义为以 2 为底的对数形式,单位为比特。

$$I(u_i) = \log_2 \left( \frac{1}{P(u_i)} \right) \quad (2)$$

信息量的数学期望  $E(u_i)$  称为信息熵,表示信息发出前的先验不确定性,称为先验熵,即

$$E(U) = \sum_i P(u_i) \log_2 \left( \frac{1}{P(u_i)} \right) = - \sum_i P(u_i) \log_2 (P(u_i)) \quad (3)$$

当已知信号  $U$  的概率分布为  $P(U)$ ,收到信号  $V=v_j$  时,发出信号的概率分布  $P(U|v_j)$ ,则信源的平均期望为  $E(U|v_j)$ ,称为后验熵,即

$$E(U|v_j) = \sum_i P(u_i|v_j) \log_2 \left( \frac{1}{P(u_i|v_j)} \right) = - \sum_i P(u_i) \log_2 (P(u_i|v_j)) \quad (4)$$

收到信号  $V$  也是随机变量,考虑信号  $V$  条件概率的期望称为条件熵,表示信宿收到信息  $V$  之后对信号  $U$  仍存在的平均不确定性,由系统随机干扰引起。通常情况下  $E(U|V) < E(U)$ ,则

$$\text{Gains}(U,V) = E(U) - E(U|V) \quad (5)$$

式中: $\text{Gains}(U,V)$  为信息增益,反映信息传递消除随机不确定性的程度。

C5.0 生长算法将输出变量看作信源发出的信息  $U$ ,输入变量看作信宿收到的信息  $V$ 。生长过程就是选择信息增益最大的输入变量作为最佳分枝变量,快速消除分组过程的平均不确定性,使得输出变量在组内趋同程度最高。另外,C5.0 算法使用信息增益率代替信息增益值,不仅考虑信息增益的大小,还要兼顾信息增益的消费,即

$$\text{Gains}(R(U,V)) = \text{Gains}(R/E(V)) \quad (6)$$

从而克服输入变量分类值较多造成的自身信息熵偏大问题。

### 1.3.2 C5.0 修剪算法

C5.0 修剪算法属于叶节点向上逐层修剪的后修剪算法,利用统计学置信区间的估计方法,直接在

训练样本集上估计误差。

设第  $i$  个节点包含  $N_i$  个观测,其中有  $E_i$  个错误预测,误差率  $f_i = E_i/N_i$ ,则对节点  $i$  的真实误差  $e_i$  在置信度  $1-\alpha$  有

$$P \left[ \frac{f_i - e_i}{\sqrt{f_i(1-f_i)/N_i}} < |z_{\alpha/2}| \right] = 1 - \alpha \quad (7)$$

则节点  $i$  真实误差  $e_i$  的置信区间上限值为

$$e_i = f_i + z_{\alpha/2} \sqrt{\frac{f_i(1-f_i)}{N_i}} \quad (8)$$

在对每个节点进行误差估计的基础上,对叶节点计算加权误差  $\sum_{i=1}^k p_i e_i > e$ ,  $i=1,2,\dots,k$ ,  $k$  为父节点包含的全部叶节点个数; $P_i$  为  $i$  节点的样本量占整个分枝样本量的比例; $e$  为父节点的误差估计。若待剪子树中叶节点的加权误差大于父节点误差,则可以修剪。

## 2 模型建立与实证研究

### 2.1 事故严重程度与事故信息数据结构特征

各国对道路交通事故严重程度分类标准不一。日本和香港将交通事故严重程度分为 3 级,即死亡、重伤和轻伤事故;德国将交通事故严重程度分为 4 级,即死亡、重伤、轻伤和物损事故;美国将交通事故严重程度分为 5 级,即死亡、致残受伤、非致残受伤、轻微伤和无伤事故;中国的分类方法按照人身伤亡或者财产损失的程度和数额,将交通事故分为轻微事故、一般事故、重大事故和特大事故;另外亦直接根据事故后果种类分为死亡、重伤、轻伤和物损事故 4 级。由此可见尽管各国和地区对事故严重程度的分类标准不一,但基本上都包括了死亡事故、致伤事故和物损事故 3 个等级。

中国公安部“道路交通事故信息系统”事故处理相关业务信息包括案件受理信息、事故基本信息、调查取证信息、事故认定信息、损害赔偿调解信息等,以对事故发生的客观条件进行全面描述<sup>[13]</sup>。其中事故基本信息是整个事故处理信息结构的最主要部分,涉及到事故严重程度评定的指标为死亡人数、受伤人数和损失折款,其他指标包括事故自身属性、致因属性、时空属性和行政属性等。

### 2.2 建模数据背景及分析

选取某省会城市城区及周边重点公路单年事故数据作为实证研究对象,调取事故数据信息 16 009 起。其中,财产损失事故 10 342 起,占 64.61%;受伤事故 4

685起,占29.27%;死亡事故981起,占6.13%。将分类变量标签分别定义为2分类严重程度和3分类严重程度两类,如表1所示。

表1 输出变量定义

Tab.1 Output variables definition

变量	变量名	变量说明
2分类事故严重程度	2 severity	1(低风险,财产损失事故); 2(高风险,伤亡事故)
3分类事故严重程度	3 severity	1(财产损失事故);2(致人受伤事故);3(致死事故)

表2 部分输入变量定义

Tab.2 Input variables definition

变量类别	变量	变量名	变量说明
时间属性	月分布	month	1~12(月份)
	季分布	season	1(春季);2(夏季);3(秋季);4(冬季)
	日分布	date	1~31(日期)
	小时分布	time	1~24(小时)
	时段分布	T-interval	1(夜间);2(上午);3(下午);4(晚间)
	周日分布	weekday	1~7(周日)
空间属性	控制方式	Ctrl-type	1(主动控制);2(被动控制);3(无控制)
	光照条件	lighting	1(白天);2(夜间有路灯);3(夜间无路灯)
	天气条件	weather	1(晴阴);2(雨雪);3(大风)
	路面类型	pavement	1(沥青);2(水泥);3(沙土)
	路口路段类型	S/J	1(3枝路口);2(4枝路口);3(其他路口);4(正常路段);5(人行横道);6(非正常路段)
	道路线型	alignment	1(弯道);2(坡道);3(弯坡组合);4(平直)
	道路属性	road	1(公路);2(城市道路)
	道路类型	R-type	1(高速、1级公路);2(2级及以下);3(快速路);4(主干路);5(次干路);6(支路及以下)
	道路断面	R-cross	1(分向式);2(分车式);3(分车分向式);4(混合式)
地形条件	land	1(平原);2(丘陵、山区)	
涉事驾驶人及车辆属性	年龄	age	1(25岁以下);2(25~35岁);3(35~45岁);4(45岁以上)
	驾龄	D-age	1(3年以下);2(3~8年);3(8~15年);4(15年以上)
	行驶状态	D-behavior	1(直行);2(转向及并线);3(倒车及掉头);4(驶离路面等其他行为)
	驾驶人类型	D-type	1(职业驾驶人);2(非职业驾驶人);3(非驾驶人)
	投保类型	insurance	1(投保);2(未投保)
	车损程度	V-loss	1(报废);2(严重);3(一般);4(轻微);5(未损坏)
	事故原因	A-cause	1(车辆原因);2(主动违章);3(驾驶失误);4(其他驾驶人原因);5(非机动车原因)
	事故形态	A-form	1(正面相撞);2(侧面相撞);3(尾随相撞);4(对面刮擦);5(同向刮擦);6(其他)
	碰撞类型	vehicle	1(机非事故);2(单车事故);3(多车事故)
伤残等级	injured	1(死亡);2(重伤);3(轻伤);4(无伤)	

### 2.3 决策树结构及结果分析

本文依次输入空间属性、涉事驾驶人及车辆属性和全属性,分别针对2种事故严重程度分类构建决策树,从而进行输入属性集和模型优度的比较。以全输入变量为例,生成决策树预测模型如下页图1所示。由图1可知,2 severity和3 severity均为深度为5的多叉树型结构,26个输入的影响因素

输入变量集使用事故现场信息代码,主要包括时间属性、空间属性和涉事驾驶人及车辆属性3部分。考虑到事故信息代码本身较为繁复,如事故原因分类代码中,仅代指由机动车驾驶人因素引起的事故原因代码就包含编码11~49,共计23个分类项,导致样本观测过离散,且作为模型变量解释能力变差<sup>[14]</sup>。为尽可能包含事故完全信息和实证观测数据的近似正态分布,根据其描述特征将事故信息代码进行合并分类处理,使用合并后分类水平减少的“合成”变量,如表2所示。

中,有5个因素对本实证数据集事故严重程度影响显著。其中具有最高信息增益率的碰撞类型和道路属性为决策树第1层和第2层属性,从第3层开始事故原因、地形条件、驾驶人类型等因素的分类概率值逐渐趋同。

根据树状结构生成推理规则集(rule set)分别对应输出变量的各个分类,截取部分如下页表3所示。

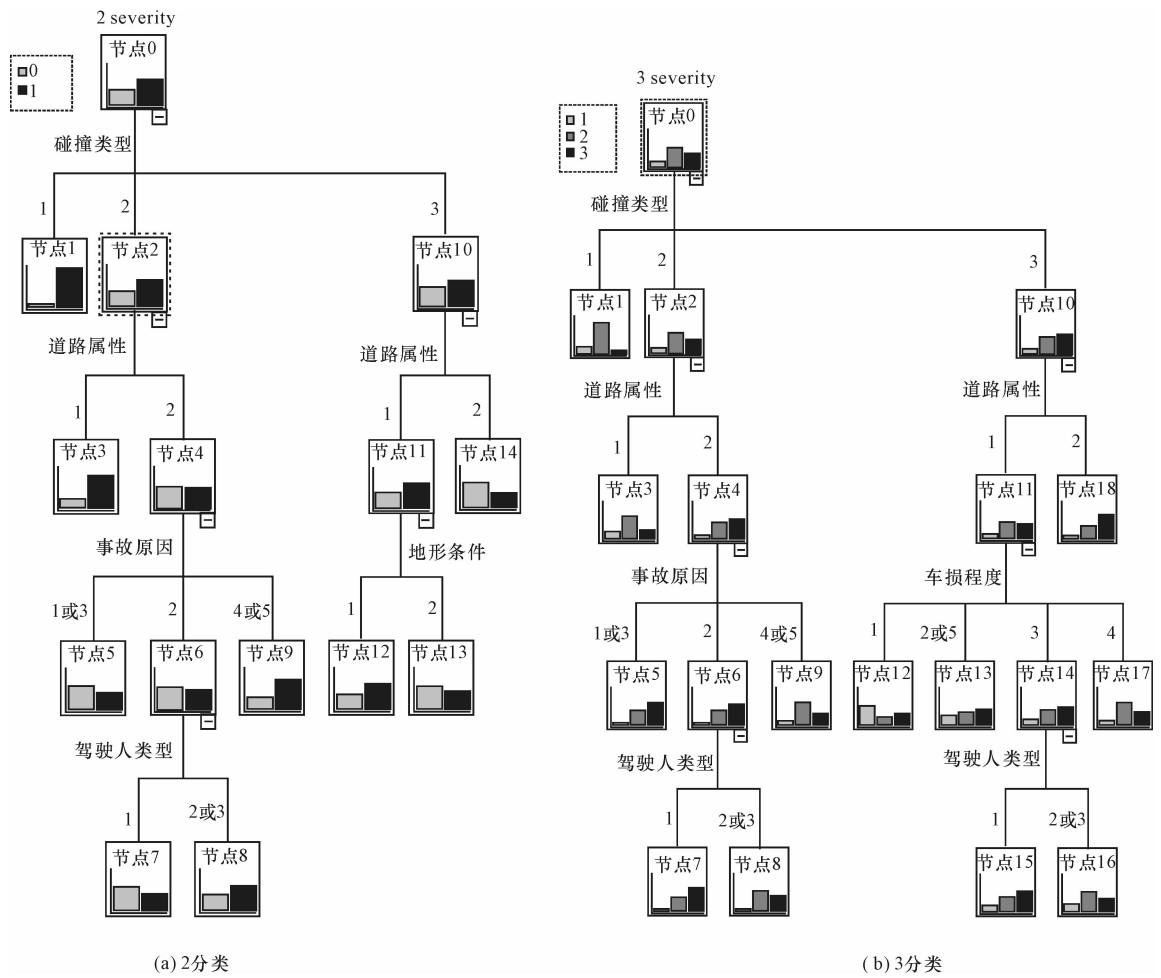


图 1 预测模型树形结构

Fig. 1 Prediction model tree structure

表 3 推理规则集

Tab. 3 Rule sets

分类	输出变量	规则集	观测(Obs)	精度/%
2 分类	高风险	机非事故	785	94
		单车事故;公路	1 293	78
		单车事故;城市道路;非职业驾驶人或非驾驶人;驾驶人主动违章	140	65
		单车事故;城市道路;非机动车或其他驾驶人原因	186	75
		多车事故;公路	1 244	66
3 分类	致伤事故	机非事故	785	78
		单车事故;公路	1 293	69
		单车事故;城市道路;非职业驾驶人或非驾驶人;驾驶人主动违章	140	55
		单车事故;城市道路;非机动车或其他驾驶人原因	186	61
		多车事故;公路;非职业驾驶人或非驾驶人;车辆损坏程度为一般	204	55
	致死事故	多车事故;公路;车辆损坏程度为轻微	571	58
		单车事故;城市道路;车辆故障或主动违章	447	57
		单车事故;城市道路;职业驾驶人;主动违章	317	60
		多车事故;公路;车辆损坏程度为严重	86	40
		多车事故;公路;职业驾驶人;车辆损坏程度为一般	558	50
		多车事故;城市道路	820	62

注:机非事故表示机动车与非机动车事故;多车事故表示机动车与机动车事故。

以上针对 2 种事故严重程度分类,分别建立了空间属性(\$C)、涉事驾驶人及车辆属性(\$C1)和全属性(\$C2)3 个决策树预测模型,模型在测试集上的总体预测精度如表 4 所示。

为了保证模型稳健性,建模中进行数据训练集和测试集的分组及交叉抽样;同时,为了避免过拟合问题,对部分“合成”后输入变量通过 ChiMerge 分箱法减少过度分枝。最后进行模型的横向比较,以评价同类决策树预测模型的误差和综合优度。表 5 为模型间评价分析结果。

表 4 预测模型结果

Tab. 4 Results of prediction models

模型种类	平均预测精度/%	置信度分布	平均正确性	平均不正确性
\$C-2 severity	65	0.501~1.000	0.846	0.770
\$C1-2 severity	70	0.500~1.000	0.811	0.704
\$C2-2 severity	69	0.512~0.940	0.721	0.649
\$C-3 severity	55	0.444~0.714	0.563	0.560
\$C1-3 severity	60	0.500~1.000	0.858	0.808
\$C2-3 severity	61	0.397~0.782	0.607	0.577

表 5 模型评价分析结果

Tab. 5 Analysis results of prediction models

2 severity		训练集		测试集	
一致性检验	一致	3 036	55.03%	1 299	54.44%
	不一致	2 481	44.97%	1 087	45.56%
一致性分布	正确	2 490	82.02%	1 051	80.91%
	错误	546	17.98%	248	19.09%
一致性置信度指标	置信度分布	0.516~0.980		0.517~0.980	
	平均正确性	0.839		0.841	
	平均不正确性	0.775		0.784	
	正确性始终高于	0.980		0.908	
	不正确性始终低于	0.528		0.545	
	90.01% 以上精度	0.815		0.864	
	2.0 以上折叠精度	0.910		0.906	
3 severity		训练集		测试集	
一致性检验	一致	2 895	52.47%	1 236	51.80%
	不一致	2 622	47.53%	1 150	48.20%
一致性分布	正确	1 981	68.43%	824	66.67%
	错误	914	31.57%	412	33.33%
一致性置信度指标	置信度分布	0.487~0.832		0.491~0.832	
	平均正确性	0.693		0.690	
	平均不正确性	0.674		0.674	
	正确性始终高于	0.810		0.791	
	不正确性始终低于	0.527		0.527	
	90.01% 以上精度	0.785		0.791	
	2.0 以上折叠精度	0.821		0.857	

使用提升度(Lift)指标进行模型整体性能而非单个节点精度的评价,即能满足评价各模型在数据学习过程中特征提炼和规律归纳能力的综合优度。其定义为

$$P_{\text{Lift}} = P(\text{Se}_i | \text{node}_j) / P(\text{Se}_i | \text{pop}) \quad (9)$$

式中: $P_{\text{Lift}}$ 为 Lift 的计算参数; $P(\text{Se}_i | \text{node}_j)$ 为节点  $j$  所属样本空间中输出变量值为事故严重程度分类  $i$  的概率; $P(\text{Se}_i | \text{pop})$ 为整个样本空间内输出变量为  $i$  的概率。

提升度指标是节点规则集概括能力及置信度的

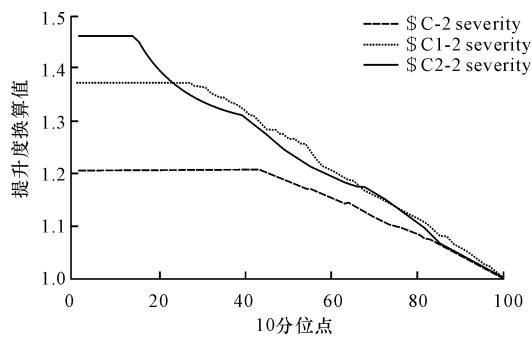
综合反映,下页图 2 为 2 种事故严重程度分类决策树模型的提升度收益对比图。

分析表 4、表 5 可知,单独 C5.0 决策树模型在本实证研究测试样本集整体预测精度一般,2 分类和 3 分类事故严重程度预测模型在测试集上的最优精度为 70%和 61%。多模型联合评价方面,总体模型优度有所提升。以 2 severity 测试样本集为例,3 个模型有 54%的一致性并且呈 81%的正确率,置信度取值范围在 0.516~0.980 之间,预测置信度在 0.864 以上的样本有 90%预测正确。从模型提

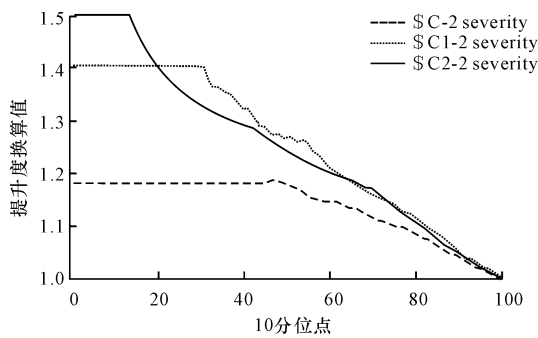
提升度收益对比来看,3 个模型总体差异不大,全属性输入模型略优于涉事驾驶人及车辆属性输入模型,显著优于空间属性输入模型。

另外,观察提升度累积曲线和决策树规则集可

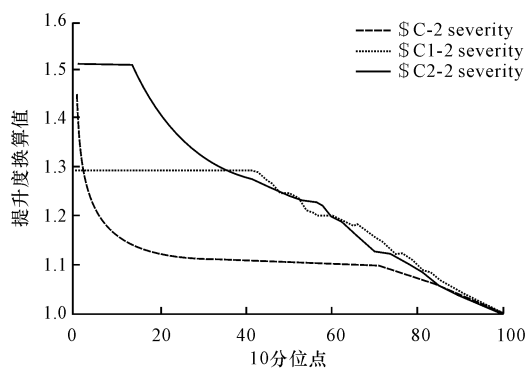
知,浅层节点提升度较高且对应规则集逻辑条件清晰。规则集显示碰撞类型、道路属性、事故原因是影响本实证对象事故严重程度分类的主要影响因素,能够反映道路的实际真实情况。



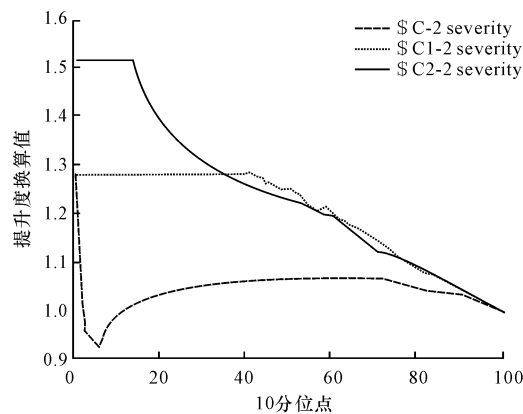
(a) 2 severity=1训练集



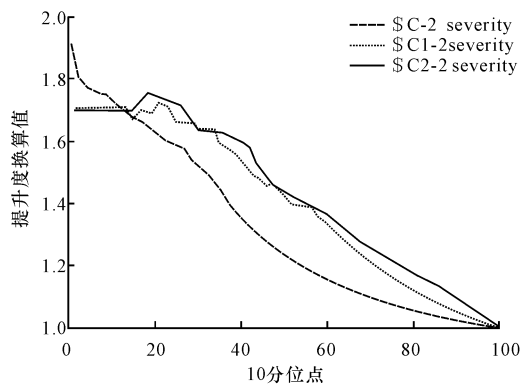
(b) 2 severity=1测试集



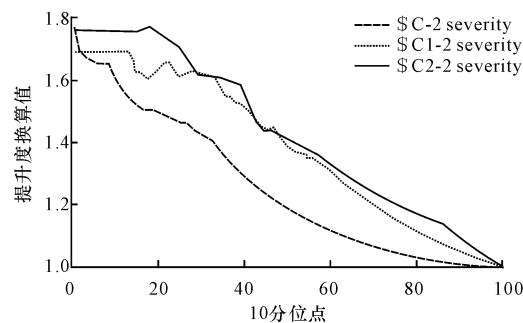
(c) 3 severity=2训练集



(d) 3 severity=2测试集



(e) 3 severity=3训练集



(f) 3 severity=3测试集

图 2 模型提升度收益对比

Fig. 2 Comparison of gains by improvement of model

### 3 结 语

(1)中国现行事故数据包括了事故本体、人员、车辆、道路及环境等多角度信息记录。然而人工采集、后期录入的采集模式易导致数据缺失,减弱了数据的潜在价值。数据完整性是影响事故成因分析、

交通安全评价、事故预测等道路交通安全研究的重要因素。

(2)交通事故的发生受到人、车、路、环境等众多因素的影响,现行事故数据采集侧重于调查人的因素,强调为事故定责提供依据,直接导致数据集虽然结构复杂,但影响因素间关联性不强,难以深层次挖

掘事故成因和机理。

(3) C5.0 决策树方法与其他分类模型相比,在面对数据缺失和高维输入变量的问题时更为稳健,也能保证一定的分类精度。且由于其基于逻辑的树形结构生成方法,C5.0 模型易于理解,规则集的生成与退出也有非常直观的解释。

(4) 考虑到实际事故数据常出现的各类问题,即数据完整性与客观性问题,本文选用 C5.0 决策树模型对事故严重程度进行分析,实证研究表明 C5.0 决策树模型在特征信息未获得或无法还原的情况下,可以较好地处理各种数据噪声干扰,在一定程度上对分类目标做出较高精度的预测。然而,基于逻辑决策的分类规则在处理受伤事故和致死事故此类界限较为模糊的分类问题时,依然存在精度有限的问题,存在一定程度的误判,影响了模型的整体表现,而受伤事故和致死事故是事故严重程度研究的重要内容。因此有必要在今后的研究中,从数据获取、抽样统计、分类器选择和算法改进等多角度展开对事故严重程度的进一步分析研究。

#### 参考文献:

#### References:

- [1] Chang H, Yeh T. Risk factors to driver fatalities in single-vehicle crashes: comparisons between non-motorcycle drivers and motorcyclists [J]. *Journal of Transportation Engineering*, 2006, 132(3): 227-236.
- [2] Malyshkina N, Mannering F. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents [J]. *Accident Analysis and Prevention*, 2010, 42(1): 131-139.
- [3] Yamamoto T, Shankar V. Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects [J]. *Accident Analysis and Prevention*, 2004, 36(5): 869-876.
- [4] Helai H, Chor C, Haque M. Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis [J]. *Accident Analysis and Prevention*, 2008, 40(1): 45-54.
- [5] Lee J, Mannering, F. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis [J]. *Accident Analysis and Prevention*, 2002, 34(2): 149-161.
- [6] Eluru N, Bhat C, Hensher D. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes [J]. *Accident Analysis and Prevention*, 2008, 40(3): 1033-1054.
- [7] Malyshkina N, Mannering F. Markov switching multinomial logit model: an application to accident-injury severities [J]. *Accident Analysis and Prevention*, 2009, 41(4): 829-838.
- [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] 李世民, 孙明玲, 关宏志. 基于累积 Logistic 模型的道路交通事故严重程度预测模型 [J]. *交通标准化*, 2009(3): 168-171.
- LI Shi-ming, SUN Ming-ling, GUAN Hong-zhi. Prediction model cumulative logistic for severity of road traffic accident [J]. *Transport Standardization*, 2009(3): 168-171. (in Chinese)
- [10] 马壮林, 邵春福, 李霞. 基于 Logistic 模型的公路隧道严重事故严重程度的影响因素 [J]. *吉林大学学报: 工学版*, 2010, 40(2): 423-426.
- MA Zhuang-lin, SHAO Chun-fu, LI Xia. Analysis of factors affecting accident severity in highway tunnels based on logistic model [J]. *Journal of Jilin University: Engineering and Technology Edition*, 2010, 40(2): 423-426. (in Chinese)
- [11] Pei Y L, Fu C Y. Investigating crash Injury severity at unsignalized intersections in hei longjiang province, china [J]. *Journal of Traffic and Transportation Engineering: English Edition*, 2014, 1(4): 272-279.
- [12] 侯树展, 孙小瑞, 贺玉龙, 等. 高速公路交通事故严重程度与交通流特征的关系研究 [J]. *中国安全科学学报*, 2011(9): 106-111.
- HOU Shu-zhan, SUN Xiao-rui, HE Yu-long, et al. Relationships between crash severity and traffic flow characteristics on freeways [J]. *China Safety Science Journal*, 2011(9): 106-111. (in Chinese)
- [13] GA/T 859-2010, 道路交通事故处理信息数据结构 [S].
- GA/T 859-2010, Data structure for accident information [S].
- [14] GA/T 16.1-11, 道路交通事故信息代码 [S].
- GA/T 16.1-11, Codes for road traffic accident scene [S].



