

Traffic accident risk identification of two-wheeled vehicles based on improved association rules

MI Jun-song¹, LI Ping-fei^{1,2}, WANG Mi-yang¹, YANG Hong¹, XU Chang¹, REN Yao¹, HE Chao-ming³

¹College of Automotive and Transportation, Xihua University, Chengdu, China, 610039

²Sichuan Xihua Jiaotong Forensics Center, Chengdu, China, 610039

³Sichuan Dingcheng Forensic Center, Chengdu, China, 610039

Abstract: In order to obtain the key cause factors of death of two-tricycle drivers in vehicle-two-tricycle traffic accidents, this paper selected 1146 cases of vehicle-two-tricycle accident data from NAIS as the research object, and applied machine learning algorithm random forest to select the characteristic variables. Then, binary logistic regression algorithm and BP neural network algorithm are used to assign weights to the characteristic variables and the dimensions in which the variables are located. By substituting different weight characteristic variables with Apriori for analysis, the output results show that vehicle characteristics, vehicle motion state, driver's own characteristics and road characteristics have an important impact on the severity of injury of two-tricycle drivers. The results show that the Apriori algorithm proposed in this paper can be used as an effective analysis method in the analysis of personnel injury factors in traffic accidents, and the results can also be used to accelerate the development of autonomous driving technology and theoretical support for road safety management.

Keywords: traffic safety; Accident injury analysis; Data mining; Association rule algorithm; Machine learning algorithm

基于改进关联规则的二轮车交通事故风险辨识

米俊松¹, 李平飞^{1,2}, 王咪杨¹, 杨鸿¹, 徐畅¹, 任耀¹, 何朝明³

¹ 西华大学 汽车与交通学院, 成都, 中国, 610039

² 四川西华交通司法鉴定中心, 成都, 中国, 610039

³ 四川鼎诚司法鉴定中心, 成都, 中国, 610023

摘要: 为得到车-二轮车交通事故中, 二轮车驾驶员死亡事故中关键致因因素, 本文从 NAIS 数据库中筛选出 1146 例车-二轮车事故数据作为研究对象, 应用机器学习算法随机森林实现对特征变量的选择。进而通过二元逻辑回归算法以及 BP 神经网络算法分别对特征变量以及变量维度进行权重赋值。将不同权重的特征变量代入 Apriori 进行分析, 结果显示, SUV 相较于轿车发生死亡事故概率更高, 事故发生时乘用车未采取采取转向和紧急制动更容易导致死亡事故、中老年二轮车驾驶员事故受到的危害会更大, 夜间的普通路段也是致命事故的高发段。本文提出的基于权重改进的 Apriori 算法可以作为交通事故中人员损伤因素分析中一种有效分析方法。

关键词: 交通安全; 事故人员损伤分析; 数据挖掘; 关联规则算法; 机器学习算法

1 引言

据《2020 年中国交通统计年鉴》非机动车交通事故发生数量为 29969 起, 同比增长 3.2%; 摩托车交通事故发生数量为 45789 起, 同比增长 0.3%。^[1]二轮车作为交通弱势参与者, 事故发生时容易出现致命损伤。二轮是国内重要的交通工具, 二轮车的销量还在连年增长, 所以研究二轮车人员致伤的主要因素, 改善人们的出行方式以及事故发生时的补救措施以减少人员的伤亡非常重要。

传统的事故分析方法 FTA 以及 FP-Tree 等可以用于分析交通事故整体特征, 但无法实现对交通事故致因的

多个变量之间的关联性分析，关联规则分析可以很好的解决这个问题，其也是用于交通事故数据挖掘的常用方法，Apriori 算法是关联规则分析的主要方法之一。袁振洲应用极限梯度提升关联规则挖掘算法识别城市道路老年行人交通事故风险因子^[2]；柳本民采用层次分析法加权的改进 Apriori 算法，杨洋等人提出了一种考虑定向约束和指标赋权的多维度交互改进 Apriori 关联规则挖掘算法^[3]。现有的研究虽然都基于机器学习算法以及赋权对传统关联规则算法进行了改进，但是存在以下问题：未在运用关联规则算法分析之前考虑自变量是否对结果变量有显著影响，运用主观赋权的方法不符合数据分析的客观性。本文利用车-二轮车事故数据，筛选对二轮车驾驶员受伤有显著影响的自变量，利用机器学习算法从变量不同维度赋予各变量权重，再结合 Apriori 算法找到驾驶员死亡关键影响因素，为找到交通事故中关键致损因素提供了一种新的分析方法。

2 数据来源与统计

本研究的数据来自于国家车辆事故深度调查体系 (National Automobile Accident in-depth Investigation System, 简称 NAIS)，NAIS 充分吸纳综合了美国的 NHTSA 以及德国 GIDAS。NAIS 车辆事故数据由分布于全国 7 个地区的 8 个合作单位完成采集、分析^[4]。2018-2021 年 NAIS 共采集 5000 余例事故数据，其中车-二轮车事故共有 2331 例，剔除缺少本论文研究所需关键信息的部分数据，最终得到 1146 例车-二轮车有效事故数据作为本文的研究对象。由图 1 知涉及轿车的事故最多，其次是 SUV；在二轮车驾驶员损伤方面，死亡和重伤(最大 AIS<2 为轻伤，最大 AIS>2 为重伤)率高达了 95.3%，这与 NAIS 事故采集条件为交通参与者中人员重伤以上有关。

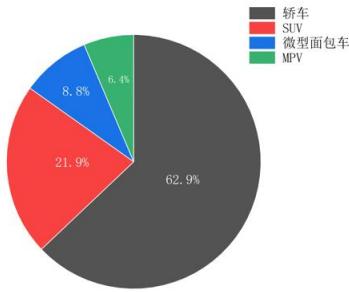


Figure 1. Proportion of vehicle types

图 1. 车辆类型占比

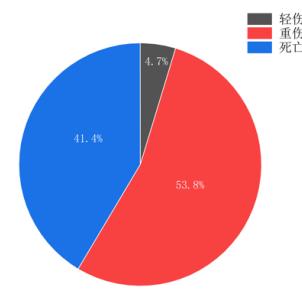


Figure 2. Injury distribution of two tricycle drivers

图 2. 二轮车驾驶员损伤分布

3 关键特征变量选取

本文初始选取 11 个特征变量作为评价指标，初次提取变量及各变量对应的水平如表 1 所示，结果变量为二轮车驾驶员死亡。

Table 1. Each variable and its assignment

表 1. 各变量及赋值情况

| 变量名称 | 水平变量 |
|-------------|---------------------------------------|
| 路段信息 | 普通路段，四枝分岔口，三枝分岔口，路段进出处，特殊路段 |
| 事故发生时段 | 日间，夜间 |
| 主车类型 | 轿车，SUV，微面，MPV |
| 是否有救护车到现场 | 是，否 |
| 碰撞方式 | 正面碰撞，侧面碰撞，追尾碰撞 |
| 二轮车碰后运动状况 | 右侧倒地，左侧倒地，保持原状 |
| 主车避让反应 | 未转向，左转向，右转向 |
| 主车刹车反应 | 刹车，未刹车 |
| 二轮车驾驶员年龄(岁) | 青少年(7-17)，青年(18-45)，中年(46-69)，老年(>69) |
| 是否戴头盔 | 是，否 |
| 衣着厚度 | 薄，中等，厚 |
| 主要致伤方式 | 碰撞伤，抛掷或摔跌伤，刮擦与拖擦伤，碾压伤，其它 |

3.1 随机森林-变量信息影响量计算

在 NAIS 收集到的事故数据中导致二轮驾驶员受伤的变量有很多，但是有些变量的影响甚微不能作为关键变量进行分析，人为的挑选变量比较主观，不能得到与结果高度相关的特征变量，机器学习算法可以很好的解决这个问题。

为了从中挑选出合理数量的特征变量，不遗漏关键变量的信息，本文利用随机森林算法的功能特性，再结合递归特征消除算法对因素进行筛选，得到与驾驶员受伤高度相关的特征因素。随机森林是一种监督式算法，使用由众多决策树组成的一种集成学习方法，得到的结果是对问题最佳答案的一个共识，常用来分类或回归，也可以用来计算数据集中不同特征变量的重要程度。据此条件随机森林的算法步骤为：

- (1) 在 N ($N=1146$) 个原始数据样本中，每轮从原始数据样本集中利用 Bootstrapping(有放回抽样)的方式抽取 N 个样例，得到一个样本数为 N 的训练集，由这 N 个样本构成一颗决策树 H_n ，在这次抽取过程中可能有重复数据，也有未被抽取数据，没有被抽到的作为袋外数据 N_n ，即测试集；
- (2) 无放回的从训练集中随机取 m ($m < 12$) 个指标，计算每一个指标对结果指标信息影响量的大小，并从所有指标选择一个分类能力最佳的作为根节点进行分裂。
- (3) 决策树的形成过程中都按步骤 2 进行分裂，直到不能分裂为止，形成决策树的整个过程未进行剪枝处理。
- (4) 对步骤 1 至步骤 3 重复 p 次，形成由 P 棵决策树组成的决策森林 $f = \{s_1, s_2, \dots, s_p\}$ ；
- (5) 分类 N_n 训练集，根据生成的 p 棵决策树投票确定最终分类结果并计算训练集数据的正确率
- (6) 将原始特征变量设定为 λ_l ($l = 1, 2, \dots, 11$)，把初始指标作为噪声依次对测试集数据进行随机添加，由此获新的测试集数据 N_{nl} ，重复步骤 (5)，得到新的正确率 R_n' ；
- (7) 根据两次得到的正确率的不同计算所有变量指标对结果指标信息影响量的大小，公式如下：

$$V_1 = \frac{1}{p} \sum_{i=1}^p (P_n - P_n') \quad (1)$$

通过公式 (1) 就可以得到每个特征变量的信息影响量的大小，其中， V_1 对应的是 λ_1 信息影响量； p 为决策树的数量，此参数可根据自身数据调整，本文设置为 1000。

3.2 随机森林-递归特征消除算法

递归特征消除是一种基于特征信息影响量的大小对特征进行筛选的方法。在前面根据随机森林计算求出了每个特征对结果变量信息影响量大小，接着用随机森林计算指标数据的正确率，因为初选的特征变量为 12 个，所以迭代 12 次就可以得到特征个数与正确率之间的关系，以此对特征变量进行筛选。

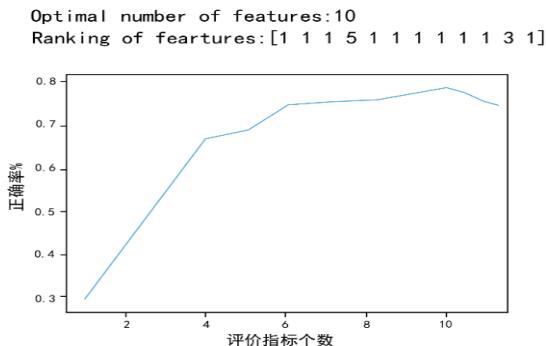


Figure 3. Relation between the number of indicators and the accuracy rate

图 3. 指标个数与正确率关系

由图 3 可以看出，随着指标个数的增加，正确率也逐渐增加，在指标个数为 10 时正确率达到最大，此后正确率开始下降。说明指标个数大于 10 的话会对模型产生干扰，冗余信息也随之增加。根据图 3 输出的结果，这 12 个变量的得分分别为 1,1,1,5,1,1,1,1,1,3,1，对应的变量分别为路段信息、事故发生时段、主车类型、是否有救护车到达现场、碰撞方式、二轮车碰后运动状况、主车避让反应、主车刹车反应、二轮车驾驶员年龄、是否

戴头盔、衣着厚度、主要致伤方式。变量分数为 1，说明该变量对结果变量有显著影响，去除得分为 5、3 的两个变量。根据模型的输出，本文选择正确率最高的 10 个评价指标，分别为：路段信息、事故发生时段、主车类型、碰撞方式、二轮车碰撞后运动状况、主车事故时转向避让反应、主车刹车反应、二轮车驾驶员年龄、是否戴头盔、主要致伤方式。

4 二轮车驾驶员损伤识别流程

为找到这些变量的耦合与二轮车驾驶员损伤之间的关联性，利用 Apriori 算法进行数据挖掘。传统关联规则算法认为数据库每个变量的权重相同，但实际对于交通事故中造成乘员损伤的各变量应有不同的重要性，比如车辆在不同速度下对乘员造成的损伤是不同的，车辆的因素和环境的因素对于乘员的损伤也具有不同的重要程度。因此，考虑权重进行关联规则的挖掘对于结果的合理性是十分必要的。本文基于逻辑回归确定各变量的水平权重，再利用 BP 神经网络确定各维度及其变量的权重。

4.1 关联规则基本定义

(1) 支持度 (Support)：是指相关联的事故特征在数据集中同时出现的次数占总数据集的比，数据集 D 中包含的项目集 X 的事务数为项目集 X 的支持数，表示为 d，项目集 X 的支持度记为 support (X)，即概率 P (X)：

$$support(d) = \frac{d}{|D|} \quad (2)$$

其中， $|D|$ 为数据集 D 的事务数。

项目集 $(X \cup Y)$ 的支持度为关联规则 $X \Rightarrow Y$ 的支持度，即为 D 中事务包含 $(X \cup Y)$ 的比例，为概率 $P(X \cup Y)$ ，记为 Support ($X \Rightarrow Y$)：

$$Support(X \Rightarrow Y) = P(X \cup Y) = \frac{|X \cup Y|}{|D|} = support(X \cup Y) \quad (3)$$

(2) 置信度：2 个定义在 I 和 D 上的关联规则 $X \Rightarrow Y$ ，它的置信度是指满足 X 特征后，Y 特征也出现的概率，或是在特征 X 的条件下出现特征 Y 的条件概率 $P(Y|X)$ ，记为 Confidence($X \Rightarrow Y$)。

$$Confidence(X \Rightarrow Y) = P(Y|X) = \frac{P(X,Y)}{P(X)} = \frac{Support(X \Rightarrow Y)}{Support(X)} = \frac{Support(X \cap Y)}{Support(X)} \quad (4)$$

其中， $X \subseteq I$, $Y \subseteq I$, $X \cap Y = \emptyset$

(3) 提升度：表示含有 Y 的条件下同时含有 X 的概率，与 X 总体发生的概率之比，用以表征先导项和后继项的关联程度。当提升度大于 1 则表示是强关联规则，此时 X 和 Y 正相关；当提升度小于或者等于 1，则表示是无效强关联规则，说明 X 和 Y 没有关联度。因此，为了避免关联结果受到伪强关联规则的影响，特此引入提升度作为筛选有效关联规则的指标。

$$Lift(X \Rightarrow Y) = \frac{P(Y|X)}{P(Y)} = \frac{Confidence(X \Rightarrow Y)}{Support(Y)} = \frac{Support(X \cap Y)}{Support(X)Support(Y)} \quad (5)$$

4.2 逻辑回归确定变量水平权重

(1) 逻辑回归模型及验证

本文选取的 1146 例事故案例中，把它分为二轮车驾驶员是否死亡两类，由信息熵理论，变量的水平权重与该变量下导致驾驶员死亡的概率有很大关系。利用逻辑回归算法计算数据库中造成二轮车驾驶员死亡的概率，将概率转化为变量的水平权重。首先建立二元逻辑回归模型，被解释变量为该事故是否造成驾驶员死亡：1 表示驾驶员死亡；2 表示驾驶员未死亡。解释变量为前文由随机森林-递归特征消除算法选取的 10 个评价指标。在二元逻辑回归模型中，造成人员死亡的概率可表示为

$$P = \frac{\exp[\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m]}{1 + \exp[\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m]} \quad (6)$$

式中： x_1 、 x_2 、 \dots 、 x_m 为选择的影响因素； β_0 为常数项； β_1 、 β_2 、 \dots 、 β_m 为影响因素 x_1 、 x_2 、 \dots 、 x_m 的回归系数，表示解释变量与被解释变量之间的相关性。因此，事故中二轮车驾驶员死亡的概率与未死亡概率的比值，即优势比 (odds ratio, OR) 为

$$S_{odds} = \frac{P}{1-P} = \exp[\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m] \quad (7)$$

为检验回归模型的正确率，根据模型测试集预测概率绘制受试者工作特征 (ROC) 曲线，可以评价 Logistic

模型的预测效应，是反映敏感性与特异性之间的关系曲线。

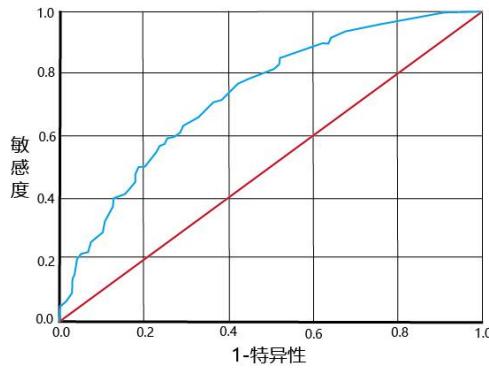


Figure 4. ROC curve of the predictive model

图 4. 预测模型 ROC 曲线

测试集 ROC 曲线下的面积为 0.784，预测正确率是 78.4%，说明预测模型的效果较好。Omnibus 检验是模型系数的综合检验，显著性小于 0.05，表示逻辑回归模型总体有意义，Hosmer 和 Lemeshow 是模型拟合度检验，目的是探究构建的模型能否较好的预测结果，该检验结果的显著性为 0.724 其值不显著，说明模型的拟合度较好，表明自变量可以有效的预测结果。

(2) 变量水平权重的确定

Table 2. Horizontal weights of variables

表 2. 变量水平权重

| 变量, V | 水平变量, p | 优势比, OR | 水平权重, w |
|-----------|---------|---------|---------|
| 路段信息 | 普通路段 | 0.547 | 0.139 |
| | 四枝分叉口 | 0.633 | 0.161 |
| | 三枝分叉口 | 0.930 | 0.237 |
| | 路段进出处 | 0.122 | 0.031 |
| | 特殊路段 | 1.691 | 0.432 |
| 事故发生时段 | 日间 | 0.823 | 0.400 |
| | 夜间 | 1.236 | 0.600 |
| 主车类型 | 轿车 | 1.624 | 0.268 |
| | SUV | 1.678 | 0.277 |
| | 微面 | 1.230 | 0.204 |
| | MPV | 1.520 | 0.251 |
| 碰撞方式 | 正面碰撞 | 1 | 0.403 |
| | 侧面碰撞 | 0.81 | 0.326 |
| | 追尾碰撞 | 0.673 | 0.271 |
| 二轮车碰后运动情况 | 右侧倒地 | 1 | 0.359 |
| | 左侧倒地 | 0.945 | 0.339 |
| | 保持原状 | 0.839 | 0.301 |
| 主车避让反应 | 未转向 | 1.6 | 0.474 |
| | 左转向 | 0.877 | 0.260 |
| | 右转向 | 0.897 | 0.266 |
| 主车刹车反应 | 刹车 | 0.635 | 0.320 |
| | 未刹车 | 1.362 | 0.680 |
| 二轮车驾驶员年龄 | 青少年 | 1.264 | 0.294 |
| | 青年 | 0.768 | 0.179 |
| | 中年 | 0.832 | 0.193 |
| | 老年 | 1.438 | 0.334 |
| 是否戴头盔 | 是 | 1 | 0.523 |
| | 否 | 0.913 | 0.477 |
| 主要致伤方式 | 碰撞伤 | 1.4 | 0.151 |
| | 抛掷或摔跌伤 | 2.804 | 0.302 |
| | 刮擦与拖擦伤 | 0.28 | 0.030 |
| | 碾压伤 | 3.527 | 0.380 |
| | 其它 | 1.28 | 0.137 |

由逻辑回归模型得到每一个子变量的水平优势比，即以第一个子变量为参考，其余子变量导致二轮车驾驶死亡事故的概率，将每个变量下所有子变量的权重值总和设为 1，即归一化，由此衡量每个子变量在该变量下对乘员损伤的影响的重要程度^[5]。

4.3 BP 神经网络确定维度权重

在场景中，每个变量对驾驶员损伤影响的权重不一致，不同变量所属的维度对驾驶员损伤的程度也存在区别，比如车辆因素和环境因素它们的贡献度是不同的。在利用二元逻辑回归模型确定了各变量的水平权重后，为了进一步确定每个维度因素对于驾驶员影响程度的大小，本文利用 BP 神经网络确定各变量以及维度的权重。

首先对 1146 例车-二轮车事故数据进行处理，设置其初始参数，据数据需要设置隐藏层层数为 1，隐藏层中神经单元为 7 个，输入层为随机森林筛选选的 10 个变量的所有数据，输出层即二轮车驾驶员是否死亡数据，神经单元数为 1。把数据分为训练集和测试集，训练集应占到样本量的 70% 以上较好。经过 BP 神经网络学习后得到均方误差 MSE 为 0.100，并得到输入层与隐藏层之间的权重系数 ω_{ki} ，以及隐藏层与输出层之间的权值系数 ω_{jk} ，将各神经元之间的权值系数转化为各指标权重还需要对数据进行以下处理：

(1) 相关显著性系数

$$r_{ij} = \sum_{k=1}^p \omega_{ki} (1 - e^{-\omega_{jk}}) / (1 + e^{-\omega_{jk}}) \quad (8)$$

(2) 相关指数

$$R_{ij} = |(1 - e^{-r_{ij}}) / (1 + e^{-r_{ij}})| \quad (9)$$

(3) 绝对影响系数

$$S_{ij} = R_{ij} / \sum_{i=1}^m R_{ij} \quad (10)$$

上述三个公式中₁ 为输入神经单元， $i = 1, \dots, m$ ，本文中 $m = 33$ ， j 为输出单元， $j = 1, \dots, n$ ， k 为隐藏层单元， $k = 1, \dots, P$ ，这里 P 为 7， ω_{ki} 为隐藏层与输出层之间的权值系数，为 11 行一列数据， ω_{jk} 为输入层与隐藏层之间的权值系数，为 11 行 33 列数据。将 BP 神经网络算法结果经过以上权重处理得到不同变量权重和维度权重。

Table 3. Calculation results of variable weights and dimension weights

表 3. 变量权重和维度权重计算结果

| 维度信息 | 维度权重 | 变量信息 | 变量权重 |
|------|-------|-----------|-------|
| 主车 | 0.312 | 主车类型 | 0.326 |
| | | 主车避让反应 | 0.268 |
| | | 主车刹车反应 | 0.406 |
| 目标车 | 0.245 | 二轮车碰后运动状况 | 0.212 |
| | | 二轮车驾驶员年龄 | 0.342 |
| | | 是否戴头盔 | 0.446 |
| 碰撞形态 | 0.263 | 碰撞方式 | 0.318 |
| | | 主要致伤方式 | 0.682 |
| 环境信息 | 0.177 | 路段信息 | 0.425 |
| | | 事故发生时段 | 0.575 |

4.4 结果分析

将得到的 4 个维度权重和变量的权重带入 Apriori 算法，支持度设置为 0.1，置信度设置为 0.3，最终输出 149 条有价值强关联规则结果，由图 5 可以看出，大部分频繁项集的支持度和置信度分别集中在 0.1-0.15 和 0.35-0.5 这个区间，高支持度和置信度关联规则表示此类频繁项集发生致死的概率较高，在车-二轮车事故中要尽可能避免这种情况出现。表 4 是从图 5 中 149 个频繁项集中分别根据支持度或置信度大小排序选出的 10 条频繁项集（规则 1-5 是支持度排序结果，6-10 为置信度大小排序结果）。表中 S、C、L 分别代表支持度、置信度和提升度，

支持度和置信度的数值越高说明该项集对结果变量即驾驶员死亡的关联性就越大，提升度大于 1，则关联规则是一个强关联规则；如果提升度小于等于 1，则为无效的关联规则。

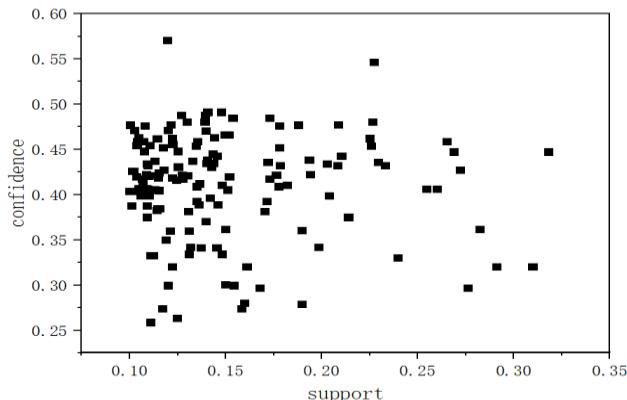


Figure 5. Mining results of improved Apriori algorithm

图 5. 改进 Apriori 算法挖掘结果

Table 4. Relative support and confidence ranking outputs valuable association rules

表 4. 相对支持度以及置信度排序输出有价值关联规则

| Rule No | Antecedent | S(%) | C(%) | L |
|---------|--|------|------|-----|
| 1 | 主车避让反应=未转向, 夜间, 老年, 主车类型=SUV, | 31 | 36 | 1.3 |
| 2 | 主要致伤方式=抛掷或摔跌伤, 主车避让反应= 未转向, 是否戴头盔=否, 道路类型=普通路段 | 24 | 33 | 1.1 |
| 3 | 主车类型=SUV, 是否戴头盔=否, 主要致伤 方式=碰撞伤, 主车刹车反应=未刹车 | 19 | 36 | 1.1 |
| 4 | 是否戴头盔=否, 中年, 二轮车碰后状况= 右侧倒地, 道路类型=普通路段, 夜间 | 16 | 28 | 1.4 |
| 5 | 主要致伤方式=抛掷或摔跌伤, 是否戴头盔= 否, 主车刹车反应=未刹车, 主车类型= 轿车 | 15 | 30 | 1.2 |
| 6 | 主车避让反应=未转向, 主要致伤方式=碰撞 伤, 老年 | 12 | 57 | 1.1 |
| 7 | 碰撞方式=正面碰撞, 主车避让反应=未转 向, 二轮车碰后状况=右侧倒地 | 13 | 48 | 1.2 |
| 8 | 道路类型=普通路段, 主车避让反应=未转 向, 是否戴头盔=否, 二轮车碰后状况=右侧倒地 | 14 | 47 | 1.1 |
| 9 | 道路类型=普通路段, 主车避让反应=未转 向, 是否戴头盔=否, 夜间 | 11 | 42 | 1.3 |
| 10 | 碰撞方式=正面碰撞, 是否戴头盔=否, 二轮车碰 后状况=右侧倒地, 主要致伤方式=碰撞伤, 中年, 主车刹车反应=未刹车, | 14 | 37 | 1.1 |

4.5 结论

(1) 主车方面，发生致命事故以轿车和 SUV 为主，其次 SUV 拥有更大的体积和重量，发生致死事故的概率更多。在事故发生前，乘用车驾驶员未采取及时的转向避让措施和紧急制动也更容易导致二轮车驾驶员死亡。

(2) 在道路类型方面，数据反映出普通路段更容易发生致死事故，这可能是因为普通路段交通设施差，交通参与者更容易违反交通法规；规则 4、9 中普通路段与夜间相关联，说明在夜间光线条件不足的情况下更容易导致重大事故。

(3) 在二轮车方面，中老年人发生交通事故致死的概率相较于青少年更高，二轮车驾驶员未佩戴头盔也易造成严重的致死事故。另一个有趣的现象是事故发生后二轮车向右侧倒地，驾驶员受到致命伤害的可能性更高。

(4) 碰撞形态方面，输出的关联规则结果表明正面碰撞较容易引起死亡事故，这是因为正面碰撞是两个车速

度的叠加，碰撞能量更高。二轮车驾驶员受到抛掷或摔跌伤也更容易导致死亡，因为在这种致伤方式下，二轮车驾驶员会受到二次伤害。

5 讨论

道路交通事故中发生致命事故不管是对个人还是社会的影响极大，研究交通事故的特性以减少此类事故的发生是极为重要的，希望本文的研究结果能对发展自动驾驶的必要性以及改善交通治理有一定支撑意义。通过以上分析从各个层面给出如下建议：

(1)对于二轮车驾驶员，骑车时佩戴头盔是很重要的，可以有效的避免致命伤害；对于中老年来说，发生事故致死率相对较高，尤其是在夜间，应加大中老年人出行的监管。

(2)在乘用车方面，致命事故的发生多是因为乘用车未及时采取转向避让以及紧急制动，所以加大驾驶辅助系统的投入 AEB 等是很有必要的；

(3)在交通管理者方面，需加强普通道路的管理，例如增加普通道路的摄像头的投放、保证夜间的照明条件等，因为不同质量的制动距离是不一样的，所以对于不同车辆进行不同的限速也是有必要的。

本文的研究主要存在以下的不足：一是数据样本量相对较少，对于数据挖掘来说，越多的样本量越能得到更客观和真实的结果；二是未对关联规则算法进行定向约束，导致产生了很多无效关联规则，希望可以在今后的研究中加以改进。

参考文献 (References)

- [1] Song Dongdong et al." Study on driver injury severity of urban road accidents based on mean heterogeneity random parameter Logit Model. *Transportation Systems Engineering and Information* 21.3(2021):7."
- [2] 宋栋栋等."基于均值异质性随机参数 Logit 模型的城市道路事故驾驶员受伤严重程度研究." *交通运输系统工程与信息* 21.3(2021):7.
- [3] Yuan Zhenzhou, Guo Manze, Peng Yongxin,& Yang Yang.(2022). *Traffic accident risk identification for elderly pedestrians based on gradient association rules*.*Transportation Systems Engineering and Information*,22(1),195.
袁振洲、郭曼泽、彭泳鑫、& 杨洋。(2022). 基于梯度关联规则的老年行人交通事故风险识别. *交通运输系统工程与信息*,22(1), 195.
- [4] Yang Yang, Yuan Zhenzhou, Wang Yinhai, Wang Wencheng, & Sun Dongye. (2021). *Highway traffic accident risk identification based on womdi-apriori algorithm*.*Road Traffic and Safety* (006), 021.
杨洋、袁振洲、王印海、王文成、& 孙东治。(2021). 基于 womdi-apriori 算法的高速公路交通事故风险识别. *道路交通与安全*(006), 021.
- [5] Xiao Le." Analysis of influencing factors of Vehicle-vehicle accident severity at intersection based on NAIS data. *Highway and Automobile Transportation* 6(2022):37-40."
肖乐."基于 NAIS 数据的交叉口车-车事故严重程度影响因素分析." *公路与汽运* 6(2022):37-40.
- [6] Li Pingfei, Jin Siyu, Hu Wenhao, Gao Li, Che Yao-yue,& Tan Zheng-Ping et al.(2022). *Vehicle-vehicle accident scenario complexity evaluation for autonomous driving simulation test*.*Journal of Automotive Safety and Energy Efficiency*,13(4),697-704.
李平飞、金思雨、胡文浩、高立、车瑶栎、& 谭正平等。(2022). 用于自动驾驶仿真测试的车——车事故场景复杂度评价. *汽车安全与节能学报*, 13(4), 697-704.
- [7] A, S. M. K. , A, P. P. , & B, B. B. . *Classification and association rule mining of road collisions for analyzing the fatal severity*, a case study.
- [8] Wahab, L. , & Jiang, H. . (2019). *Severity prediction of motorcycle crashes with machine learning methods*. *International Journal of Crashworthiness*, 1-8.
- [9] Suganya, E. , & Vijayarani, S. . (2018). *Analysis of road accidents in India using data mining classification algorithms*. *International Conference on Inventive Computing & Informatics* (pp.1122-1126).
- [10] Lukuman, Wahab, Haobin, & Jiang. (2019). *A comparative study on machine learning based algorithms for prediction of motorcycle crash severity*. *Plos One*.
- [11] Ye Yingjie. *Traffic accident risk factor mining and prediction model construction based on association rules*.Beijing University of Technology.
叶颖婕. 基于关联规则的交通事故风险因素挖掘及预测模型构建. 毕业论文. 北京工业大学.
- [12] Liu Benmin, & Deng Runzhu. (2022). *Risk identification of mountain road operation based on improved ahp-apriori algorithm*.*Traffic and Transportation* (038-003).
柳本民、& 邓润竹。(2022). 基于 ahp-apriori 改进算法的山区公路运营风险辨识. *交通与运输*(038-003).
- [13] Xu Yan, Su Nuol, & Feng Zhongshu. (2022). *Study on secondary choice behavior of evacuation exit based on logistic regression analysis*.*Science Technology and Engineering*, 22(36), 16260-16268.
许岩、苏诺尔、& 冯中姝。(2022). 基于 logistic 回归分析的人员疏散出口二次选择行为研究. *科学技术与工程*, 22(36), 16260-16268.
- [14] He Y Q, Duan Y Y & Wang C. (2021). *Analysis and countermeasures of pedestrian traffic accident severity based on Cumulative logistic model*.*Journal of Safety and Environment*, 21(3), 8.
何雅琴、段雨阳、& 王晨。(2021). 基于累积 logistic 模型的行人交通事故严重程度分析及对策研究. *安全与环境学报*, 21(3), 8.
- [15] Tang Yu, Dai Qi, Yang Zhiwei, Yang Aimin, & Chen Lifang. (2023). *Research on software defect prediction algorithm based on optimal random forest*.*Computer Engineering and Science*, 45(5), 830-839.
唐宇、代琪、杨志伟、杨爱民、& 陈丽芳。(2023). 基于优化随机森林的软件缺陷预测算法研究. *计算机工程与科学*, 45(5), 830-839.
- [16] Liu Xin, Xiao Le, and Jin Siyu." *Characteristics and causation analysis of serious traffic accidents on Expressways based on NAIS*." *Highway and Automobile* 005(2022):000.
刘鑫、肖乐、& 金思雨."基于 NAIS 的高速公路严重交通事故特征及致因分析." *公路与汽运* 005(2022):000.
- [17] Luo Junren, Zhang Dao-Wen, Zhang You,& Che Yao-Li.(2021). *Study on accident scene at intersection of automobile vs two-wheeler*.
罗俊任、张道文、张友、& 车瑶栎。(2021). 汽车 vs 二轮车十字路口事故场景研究.