

基于改进 Apriori 算法的高速公路交通事故关联分析

邱文利¹, 杨海峰¹, 张少波¹, 邱宇², 赵姣^{3*}

(1.河北雄安京德高速公路有限公司,河北保定 071000;2.河北交投智能交通技术有限责任公司,河北石家庄 050000;
3.长安大学 运输工程学院,陕西西安 710064)

摘要:针对现有交通数据中事故影响因素间的关联特性,为高速公路运营及管理部门提供精准化、细粒度的决策支持信息,该文考虑驾驶人、环境、道路和车辆4个维度,建立带约束的改进 Apriori 算法,挖掘影响高速公路交通事故的关联规则。在传统 Apriori 算法的基础上,规则前项和后项约束的增加,可以提高关联规则的准确性和挖掘效率。结合3178条高速公路交通事故数据的分析结果表明:改进 Apriori 算法通过精准挖掘潜在因素和事故等级间的关联程度,降低无效关联规则数量,关联规则准确性、挖掘效率均大大提升。驾驶人性别、年龄、照明强度、车辆类型均与高速公路事故严重程度有强关联性;路面湿滑会使交通事故升级为一般事故;黑夜下的照明状况,是使轻微事故升级为一般和严重事故的主要因素。

关键词:交通安全;高速公路;关联规则;Apriori 算法;交通事故

中图分类号:U495

文献标志码:A

Correlation Analysis of Highway Traffic Accidents Based on Improved Apriori Algorithm

QIU Wenli¹, YANG Haifeng¹, ZHANG Shaobo¹, QIU Yu², ZHAO Jiao^{3*}

(1.Hebei Xiong'an Jingde Expressway Limited Company, Baoding, Hebei 071000, China;2.Hebei Jiaotou Intelligent Transportation Technology Co., Ltd., Shijiazhuang, Hebei 050000, China;3. School of Transportation Engineering, Chang'an University, Xi'an, Shaanxi 710064, China)

Abstract: In order to explore the association characteristics between influencing factors of accidents from existing traffic data and provide precise and fine-grained decision support information for highway operation and management departments, a constrained improved Apriori algorithm was established to mine the association rules affecting highway traffic accidents by considering four dimensions: driver, environment, road, and vehicle. On the basis of the traditional Apriori algorithm, the improvement of antecedent and consequent constraints of the rule may increase the accuracy and efficiency of mining association rules. Based on the analysis of 3178 highway traffic accident data, it is shown that the improved Apriori algorithm reduces the number of invalid association rules by accurately mining the correlation between potential factors and accident levels and improves the accuracy of association rules and mining efficiency. There is a strong correlation between driver gender, age, lighting intensity, vehicle type and the severity of highway traffic accidents. Wet and slippery road surfaces can escalate traffic accidents into general accidents. The lighting conditions under the dark night are the main factors that escalate minor accidents into general and serious accidents.

Keywords: traffic safety; highway; association rules; Apriori algorithm; traffic accidents

0 引言

高速公路在国家社会经济发展中具有不可替代

的地位和作用,是衡量一个国家公路交通运输和汽车工业现代化水平的标准之一^[1-3]。近年来,国内外学者对道路交通安全问题进行了研究。Bucsuházy

收稿日期:2023-06-08

基金项目:国家自然科学基金资助项目(编号:71971030)

作者简介:邱文利,男,大学本科,正高级工程师.E-mail:qiuwenli@sohu.com

*通信作者:赵姣,女,博士,讲师.E-mail:jiaozhao@chd.edu.cn

等^[4]根据捷克的道路交通事故数据,探讨人为因素对事故的影响,发现发生事故的主要原因是驾驶员疏忽;邓国忠等^[5]使用有序Logit模型研究发现,事故发生天气和事故发生时段明显影响事故严重程度。此外,一些学者研究了高速公路交通事故影响因素间的关联规则问题,以揭示交通事故发生的本质规律。Cai^[6]考虑事故数据的多层次和维度,基于改进Apriori算法,提出坡道入口/出口处,路面干燥和道路平直容易引发交通事故;牛毅等^[7]根据交通事故发生的规律,使用Apriori算法对货车交通事故发生时间、事故形态等进行关联分析,提出不同路段在安全状况和事故发生时间上差异性较大;冯伟等^[8]采用灰关联分析方法对沥青路面车辙深度的影响因素进行比较,发现适度降低路面温度,并加强对重载、超载车辆的运输管理,可以有效减少沥青路面的车辙病害;周小勇等^[9]采用模糊聚类法和系统聚类法,并从人、车、路和环境方面考虑,分析机动车司机的不同违章驾驶行为和交通事故的关联性,发现违章行为易引发交通事故;徐金华等^[10]对交通事故不同影响因素在平面交叉口的关联特征进行研究,发现路面湿滑是使道路交叉口事故升级为重大事故的因素;李珏等^[11]使用Apriori算法挖掘事故致因间的关联规则,为高处坠落事故的预控提供参考。

上述研究在高速公路交通事故影响因素关联规则分析与挖掘方面取得了较好的进展。然而,仍存在以下不足:① 现有算法在挖掘高速公路交通事故关联规则时,未从事故维度对规则后项进行约束,使关联规则出现因果颠倒的现象,导致算法准确性较低;② 在挖掘交通事故数据间关联性时,未针对不同类型的交通事故设置不同的参数,导致部分关联规则的缺失。

因此,该文提出一种改进Apriori算法的高速公路交通事故关联分析方法,弥补上述研究的不足。通过加入定向约束,对候选集中未包含事故类型的数据进行筛选,提高算法的准确性和求解效率;对不同等级的交通事故设置不同阈值,实现不同影响因素与交通事故等级间更细粒度的关联分析,并将事故维度下的关联规则结果以图论的形式进行可视化,为交通事故预防提供理论参考。

1 相关定义

基于高速公路交通事故的关联规则问题可表示

为:设 $X=\{X_1, X_2, \dots, X_m\}$ 为高速公路交通事故的指标数据集合, m 为指标的数量; $S=\{S_1, S_2, \dots, S_n\}$ 为高速公路交通事故数据事务集合, n 为高速公路交通事故数据总数,其中每个事务 $S_i(1 \leq i \leq n)$ 均为项目集合 X 的子集,即 $S_i \in X^{[12]}$ 。

1.1 支持度(Support)

高速公路交通事故指标项 A 、 B 同时出现的概率称为高速公路交通事故致因关联规则支持度,表示为:

$$\text{Support}(A \rightarrow B) = P(A \cap B) = \frac{|A \cup B|}{|N|} \quad (1)$$

式中: $|A \cup B|$ 表示 A 和 B 同时出现的支持计数; $|N|$ 表示所有事务出现的个数。

1.2 置信度(Confidence)

若高速公路交通事故指标项 A 出现,则高速公路交通事故指标项 B 也出现的概率称为高速公路交通事故致因关联规则置信度,表示为:

$$\text{Confidence}(A \rightarrow B) = P(A|B) = \frac{|A \cup B|}{|A|} \quad (2)$$

式中: $|A|$ 表示包含影响因素集合 A 的事故数量。

1.3 提升度(Lift)

高速公路交通事故指标项 A 出现对高速公路交通事故指标项 B 出现概率的提升,表示为:

$$\text{Lift}(A \rightarrow B) = \frac{\text{Confidence}(A \rightarrow B)}{\text{Support}(B)} \quad (3)$$

高速公路交通事故关联规则表达式为: $A \rightarrow B$,其中 A 和 B 分别表示规则的前项(Left-Hand-Side, LHS)和后项(Right-Hand-Side, RHS),有 $A \in X, B \in X$ 且 $A \cap B = \emptyset$,关联规则不低于最小支持度和最小置信度,同时提升度大于1。

2 高速公路交通事故关联规则挖掘原理

为了准确、高效地分析高速公路交通事故指标间的关联性,按照图1所示的关联分析流程挖掘高速公路交通事故关联规则。

2.1 交通事故数据预处理

2.1.1 样本量确定

研究选取的数据为英国2019年高速公路交通事故的相关数据^[13],数据量3 178条,采用统计学中样本量确定方法验证数据量的充足性。样本量的数值需根据使用场合确定方法,在区间估计场合,限制置

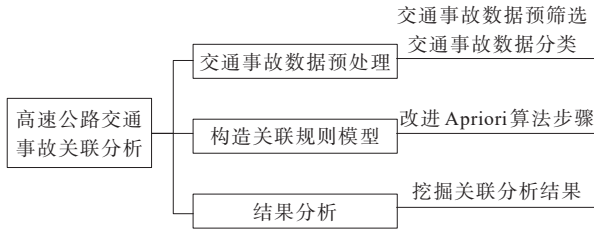


图1 高速公路交通事故关联规则分析流程
Figure 1 Association rule analysis process of highway traffic accident

信区间长度不超过 $2d$ 的条件下确定样本量 n ,其中 d 是事先给定的置信区间半径。计算可知数据标准差 $\sigma=5.12$,在标准差 σ 已知时,通过式(4)确定样本量。

$$n \geq \left(\frac{u_{1-c/2}\sigma}{d} \right)^2 \quad (4)$$

式中:显著性水平 c 一般取0.05,置信区间 $u_{1-c/2}=1.96$ (置信度取95%)。

经计算,在95%置信水平下,总体均值的置信区间为(7.709 5,8.009 5),置信区间半径为0.3。置信区间长度越短,精度越高,为提高数据的精度,取置信区间半径 $d=0.2$,得出 $n \geq 2\ 518$ 。因此,选取样本量为3 178条数据,满足最小样本量的要求。

2.1.2 指标数据筛选

为便于交通事故的关联规则挖掘,需要将相关连续性特征指标进行离散化处理^[14],利用K-means无监督聚类算法对连续性变量进行聚类分析。使用K-means算法对限速值和驾驶人年龄这两类连续变量进行聚类分析,提高分类的科学性和有效性,通过计算数据集中所有样本点到簇中心距离之和的平方,使用肘部法确定误差平方和突然变小时对应的 k 值,即分类值。限速值和驾驶人年龄对应的分类值分别为2和3,具体分类比例如表1所示。

表1 连续性指标离散化结果

Table 1 Discretization results of continuity indicators

指标描述	分裂点	分类比例/%	
限速值	96.56 km/h	12	
		88	
驾驶人年龄	16~30岁	29	
		>30~55岁	54
		>55岁	17

选取英国2019年高速公路交通事故数据,查阅

相关文献,选择以下数据类型作为关联规则挖掘的集合。从可能造成交通事故的驾驶人、车辆、道路和环境4个维度对高速公路交通事故数据进行统计。具体数据类型和统计特性如表2所示^[15]。为便于关联规则挖掘,将不同指标进行离散化处理,根据Apriori算法特性,将输入模型前的数据进行二进制转换,且将包含该变量的数据记为1,否则,记为0^[10]。

表2 变量及数据统计

Table 2 Variables and data statistics

维度	属性	描述	频数	频率
驾驶人	性别	男性 I_1	2 355	0.74
		女性 I_2	823	0.26
	年龄	16~30岁 I_3	906	0.29
		>30~55岁 I_4	1 716	0.54
		>55岁 I_5	556	0.17
是否左侧行驶	左侧行驶 I_6	77	0.02	
	非左侧行驶 I_7	3 101	0.98	
车辆类型	汽车 I_8	2 457	0.77	
	货车 I_9	721	0.23	
车辆位置	交叉口附近 I_{10}	404	0.13	
	非交叉口附近 I_{11}	2 774	0.87	
车辆	打滑翻车	无 I_{12}	2 411	0.76
		打滑 I_{13}	495	0.15
	碰撞情况	翻车 I_{14}	272	0.09
		未碰撞 I_{15}	87	0.03
		前端碰撞 I_{16}	1 957	0.61
	路面状况	后方碰撞 I_{17}	461	0.15
		侧向碰撞 I_{18}	673	0.21
		干燥 I_{19}	2 087	0.66
道路	潮湿 I_{20}	1 091	0.34	
	限速值	<96.56(km/h) I_{21}	367	0.12
≥ 96.56 (km/h) I_{22}		2 811	0.88	
照明情况	白天 I_{23}	2 134	0.67	
	黑夜 I_{24}	1 044	0.33	
环境	天气状况	晴朗 I_{25}	2 516	0.79
		雨/雪/雾 I_{26}	662	0.21
日期	日期	工作日 I_{27}	2 238	0.70
		周末 I_{28}	940	0.30
		轻微事故 I_{29}	2 610	0.82
事故属性	事故属性	一般事故 I_{30}	488	0.15
		严重事故 I_{31}	80	0.03

2.2 基于改进 Apriori 算法的关联规则模型

结合传统 Apriori 算法匹配交通事故致因挖掘的特征,提出包含规则后项事故约束的改进 Apriori 算法,对高速公路不同因素、多维属性间进行关联规则挖掘。通过频繁项集挖掘到的关联规则形如“ $A \rightarrow B$ ”的形式,其中, A 表示为前项; B 表示为后项。针对高速公路交通事故数据特点,在得到的关联规则中,前项仅为致因维度的属性(如驾驶员维度、车辆维度、道路维度和环境维度),后项仅为事故属性,对诸如{晴朗,干燥} \rightarrow {白天}这类无效关联规则进行删除,提高了算法在高速公路事故关联规则挖掘时的准确性^[1]和运行效率。改进后的 Apriori 算法步骤如下:

步骤 1: 输入高速公路交通事故数据集 $X = \{X_1, X_2, \dots, X_m\}$, 设定 n 个交通事故数量(项集), m 个交通事故数据维度, 得到式(5):

$$X = \begin{bmatrix} X_{11} & \dots & X_{1m} \\ \vdots & \dots & \vdots \\ X_{n1} & \dots & X_{nm} \end{bmatrix} \quad (5)$$

式中: X_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$) 表示第 i 个事故关于第 j 个指标的值。

步骤 2: 频繁 k 项集的产生。扫描相应交通事故数据, 设定最小支持度阈值 \min_sup , 并对候选集 C 、频繁项集 L 进行初始化, 其中, \emptyset 为空集, 根据 Apriori 算法生成候选集 C_k , 对候选集 C_k 剪枝, 确定满足最小支持度阈值的频繁 k 项集 L_k 。若 k 项频繁集中出现无项集或只有 1 个项集时, 迭代停止。

步骤 3: 对频繁项集后项进行事故属性筛选, 并根据最小支持度阈值、最小置信度阈值以及提升度 > 1 进行筛选, 得到约束后的频繁 k 项集。

步骤 4: 得到满足定向约束下的关联规则结果。

改进 Apriori 算法的流程见图 2。

3 关联规则生成及分析

3.1 关联规则生成

将改进 Apriori 算法和传统 Apriori 算法挖掘结果比较(表 3), 分析支持度和置信度阈值存在差异时, 不同算法间生成的有效关联规则数和所耗时间。由表 3 可知: 相较于传统 Apriori 算法, 改进 Apriori 算法准确性、挖掘效率均大大提升。

根据改进 Apriori 算法挖掘的关联规则, 设定最小支持度和置信度阈值均为 0.01 进行规则挖掘。初

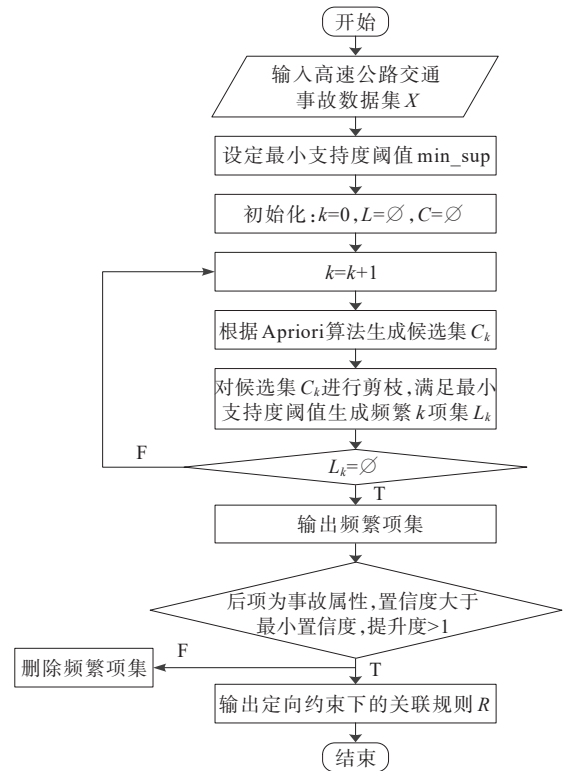


图 2 改进 Apriori 算法流程框图

Figure 2 Flow chart of improved Apriori algorithm

表 3 传统 Apriori 算法与改进 Apriori 算法挖掘结果比较
Table 3 Comparison of mining results between traditional Apriori algorithm and improved Apriori algorithm

算法	支持度	置信度	规则数	耗时/s
传统 Apriori 算法	0.2	0.8	4 610	2 898.20
	0.3	0.8	1 260	514.88
改进 Apriori 算法	0.2	0.8	775	259.30
	0.3	0.8	259	93.06

始数据中, 一般事故和严重事故两种数据类型所占比例较小, 导致不同事故等级下关联规则的支持度和置信度存在差异, 结合关联规则对应支持度、置信度散点图(图 3), 即置信度区间在 0.82~1.00 时, 后项主要为轻微事故类型的关联规则 37 883 条; 置信度区间在 0.15~0.30 时, 后项以一般事故为主的关联规则 8 020 条; 置信度区间在 0.02~0.05 时, 后项主要为严重事故类型的关联规则 197 条。结合关联规则结果, 随着最小支持度和最小置信度阈值的增加, 关联规则数量减少, 最终趋于平稳, 选取关联规则条数变化不明显时的数值为不同事故等级下最小支持度和置信度阈值^[16]。

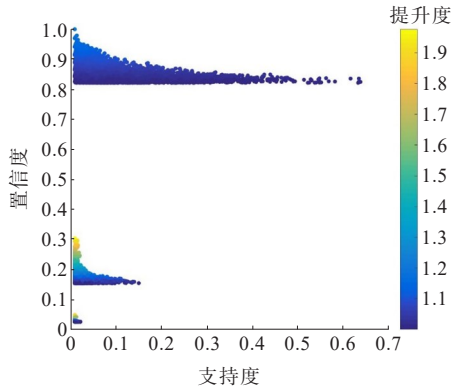


图3 支持度、置信度散点图

Figure 3 Scatter diagram of support degree and confidence

3.2 关联规则结果分析

3.2.1 特定致因维度与事故维度的定向挖掘

为深层次分析不同致因维度对高速公路交通

事故的影响,结合表2变量及数据统计信息,分别从规则前项为致因维度(驾驶员维度、车辆维度、道路维度和环境维度)与规则后项为不同事故维度类型(轻微事故、一般事故和严重事故)的关联规则进行分析,直观地比较致因维度与事故维度的关联状况。

(1) 驾驶人维度与事故维度关联规则

筛选后关联规则前项为驾驶人维度,后项为轻微事故的规则17条、一般事故的规则17条、严重事故的规则10条。表4为驾驶员维度与事故维度关联规则的部分结果,从性别视角分析,可知女性驾驶员与轻微事故的关联性较大,而男性驾驶员较容易发生一般或者严重事故。年龄在55岁及以下时,易引发轻微事故;大于55岁时,容易引发一般事故;此外,相较于左侧行驶而言,驾驶员非左侧行驶易导致事故发生。

表4 前项为驾驶人维度的关联规则(部分)

Table 4 Association rules with antecedent of driver (part)

前项	后项	支持度	置信度	提升度
{女性,年龄(30~55岁),非左侧行驶}	{轻微事故}	0.116 7	0.866 8	1.055 5
{女性,年龄(16~30岁),非左侧行驶}	{轻微事故}	0.072 4	0.836 4	1.018 4
{男性,非左侧行驶}	{一般事故}	0.117 4	0.163 3	1.063 5
{年龄(30~55岁),非左侧行驶}	{轻微事故}	0.433 3	0.829 0	1.009 4
{男性,年龄大于55岁,非左侧行驶}	{一般事故}	0.024 2	0.179 1	1.166 2
{年龄大于55岁,非左侧行驶}	{一般事故}	0.030 2	0.176 5	1.149 2
{男性,年龄(30~55岁),非左侧行驶}	{严重事故}	0.010 4	0.026 8	1.063 2
{男性,非左侧行驶}	{严重事故}	0.018 2	0.025 4	1.008 8

(2) 车辆维度与事故维度关联规则

筛选后关联规则前项为车辆维度,后项为轻微事故的规则48条、一般事故的规则35条、严重事故的规则16条。表5为车辆维度与事故维度关联规则的部分结果,前端碰撞易导致事故发生,以车辆类型而言,汽车发生轻微和一般事故的可能性更大;当车辆打滑或翻车时,较容易引发一般事故。

(3) 道路维度与事故维度关联规则

筛选后,关联规则前项为道路维度,后项为轻微事故的规则6条、一般事故的规则6条、严重事故的规则5条。表6为道路维度与事故维度关联规则的部分结果。以路面状况而言,路面状况干燥容易引发轻微事故;当车辆限速值在96.56 km/h以下时,较容易引发轻微事故;路面状况潮湿且限速值为96.56 km/h及以上时,易引发一般事故,由于路面潮

湿降低路面抗滑性能,导致事故发生。

(4) 环境维度与事故维度关联规则

筛选后,关联规则前项为环境维度,后项为轻微事故的规则17条、一般事故的规则15条、严重事故的规则2条。表7为环境维度与事故维度关联规则的部分结果,雨、雪等不良天气的出现,更易引发一般事故;结合照明情况,白天光照较强时,易发生轻微事故;光照较暗的黑夜时,驾驶员容易神经系统松懈,引发一般事故或严重事故。

3.2.2 全维度与事故维度的定向挖掘

为深层次挖掘不同事故等级下的关联规则结果,针对事故类型,分别设定不同支持度和置信度阈值,并对结果可视化,其中方框内表示数据类型,连线表示不同类型连接关系,数值为不同类型数据在关联规则中的计数情况。

表 5 前项为车辆维度的关联规则(部分)

Table 5 Association rules with antecedent of vehicle (part)

前项	后项	支持度	置信度	提升度
{货车,非交叉口附近,无打滑翻车,侧向碰撞}	{轻微事故}	0.040 3	0.847 7	1.032 2
{汽车,无打滑翻车,前端碰撞}	{轻微事故}	0.283 8	0.823 7	1.003 0
{汽车,非交叉口附近,无打滑翻车,前端碰撞}	{轻微事故}	0.249 8	0.822 8	1.001 9
{汽车,非交叉口附近,翻车}	{一般事故}	0.011 0	0.251 8	1.639 8
{汽车,打滑,前端碰撞}	{一般事故}	0.015 1	0.170 2	1.108 5
{汽车,非交叉口附近,打滑,前端碰撞}	{一般事故}	0.013 2	0.162 8	1.060 1
{货车}	{严重事故}	0.010 1	0.044 4	1.763 1
{非交叉口位置,无打滑翻车,前端碰撞}	{严重事故}	0.011 6	0.028 5	1.134 1
{男性,右侧行驶}	{严重事故}	0.018 2	0.025 4	1.008 8

表 6 前项为道路维度的关联规则(部分)

Table 6 Association rules with antecedent of road (part)

前项	后项	支持度	置信度	提升度
{干燥,限速值<96.56 km/h}	{轻微事故}	0.073 3	0.866 2	1.054 7
{潮湿,限速值<96.56 km/h}	{轻微事故}	0.026 4	0.857 1	1.043 7
{干燥}	{轻微事故}	0.540 0	0.822 2	1.001 2
{潮湿,限速值≥96.56 km/h}	{一般事故}	0.050 0	0.160 1	1.042 8
{干燥,限速值≥96.56 km/h}	{一般事故}	0.090 6	0.158 4	1.031 7
{干燥,限速值≥96.56 km/h}	{严重事故}	0.014 8	0.025 9	1.027 0
{干燥}	{严重事故}	0.016 7	0.025 4	1.008 8
{限速值≥96.56 km/h}	{严重事故}	0.022 3	0.025 3	1.003 4

表 7 前项为环境维度的关联规则(部分)

Table 7 Association rules with antecedent of environment (part)

前项	后项	支持度	置信度	提升度
{工作日,白天,晴朗}	{轻微事故}	0.342 7	0.840 9	1.023 9
{白天,晴朗}	{轻微事故}	0.477 0	0.838 5	1.021 0
{周末,白天,雨、雪等不良天气}	{轻微事故}	0.029 9	0.826 1	1.005 9
{周末,黑夜}	{一般事故}	0.018 3	0.185 9	1.210 6
{周末,黑夜,晴朗}	{一般事故}	0.012 0	0.178 4	1.161 8
{周末,雨、雪等不良天气}	{一般事故}	0.011 6	0.172 9	1.126 0
{工作日}	{严重事故}	0.018 3	0.025 9	1.029 5
{黑夜}	{严重事故}	0.013 8	0.042 1	1.674 2

(1) 当最小支持度和置信度阈值分别为 0.2 和 0.8 时,得到后项为轻微事故的关联规则 774 条(图 4)。

由图 4 可知:轻微事故与方框内数据类型相关,其中男性、非左侧行驶、无打滑翻车、白天、天气晴朗、限速值 ≥ 96.56 km/h 与之关联性较大。男性驾驶员在限速值 ≥ 96.56 km/h 下非左侧行驶时,因汽车前

端碰撞发生轻微事故的可能性较大。车辆行驶速度较大,驾驶员反应时间短,易引发轻微事故。其中不同性别驾驶员均会引发轻微交通事故,且主要表现为男性驾驶员,可能与男性驾驶员基数较大有关。

(2) 当最小支持度和置信度阈值分别为 0.05 和 0.15 时,得到后项为一般事故的关联规则 334 条(图 5)。

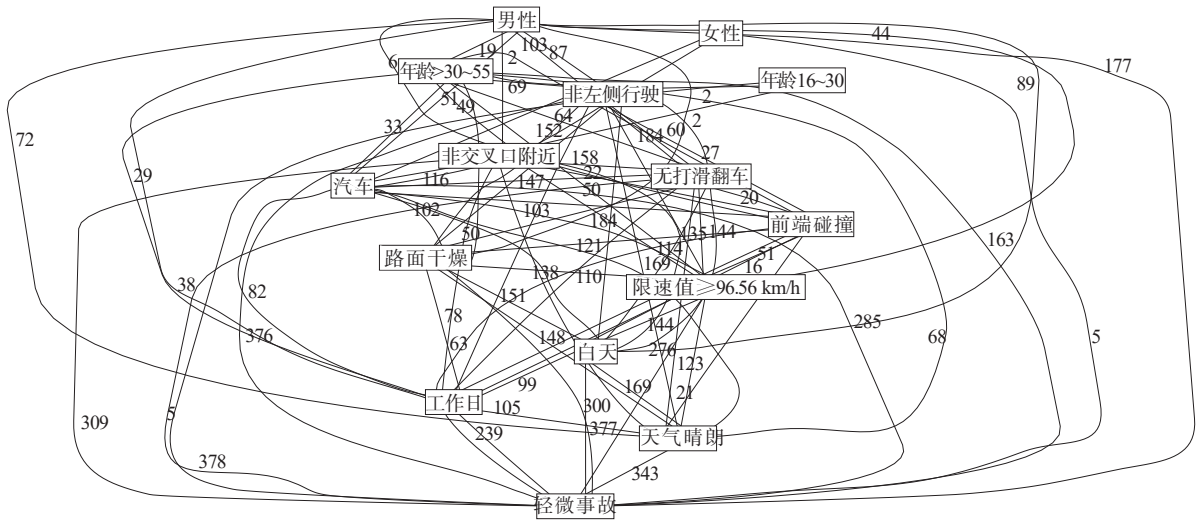


图4 轻微事故的关联规则网络

Figure 4 Association rule network of minor accidents

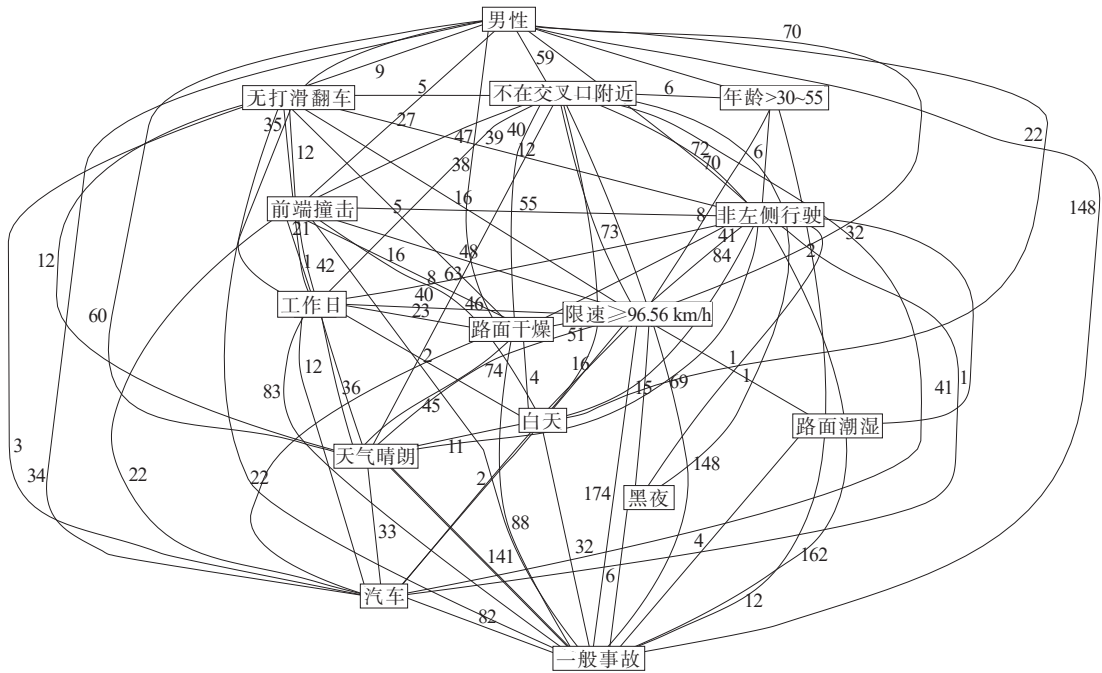


图5 一般事故的关联规则网络

Figure 5 Association rule network of general accidents

由图5可知:一般事故与方框内数据类型相关,其中男性、非左侧行驶、限速 ≥ 96.56 km/h、汽车与之关联性较大。男性驾驶员在非左侧行驶时,因路面湿滑会引发一般事故;男性驾驶员在黑夜状况下,且不在交叉口附近时会引发一般事故。路面湿滑会使汽车制动距离变长,引发一般事故。限速 ≥ 96.56 km/h,车辆运行速度较快,使其在相同制动时间内制动距离较长,增加事故发生概率。

(3) 当最小支持度和置信度阈值均为0.01时,得到后项为严重事故的关联规则197条(图6)。

由图6可知:严重事故与方框内数据类型相关,其中男性、非左侧行驶、前端碰撞、货车与之关联性较大。男性驾驶员年龄为30~55岁,黑夜光线不佳时,驾驶员不易注意到前方障碍物,且一旦车辆前端发生碰撞,不能及时操纵车辆进行避险操作,或因判断出现偏差引发严重事故。货车在速度较大时发生

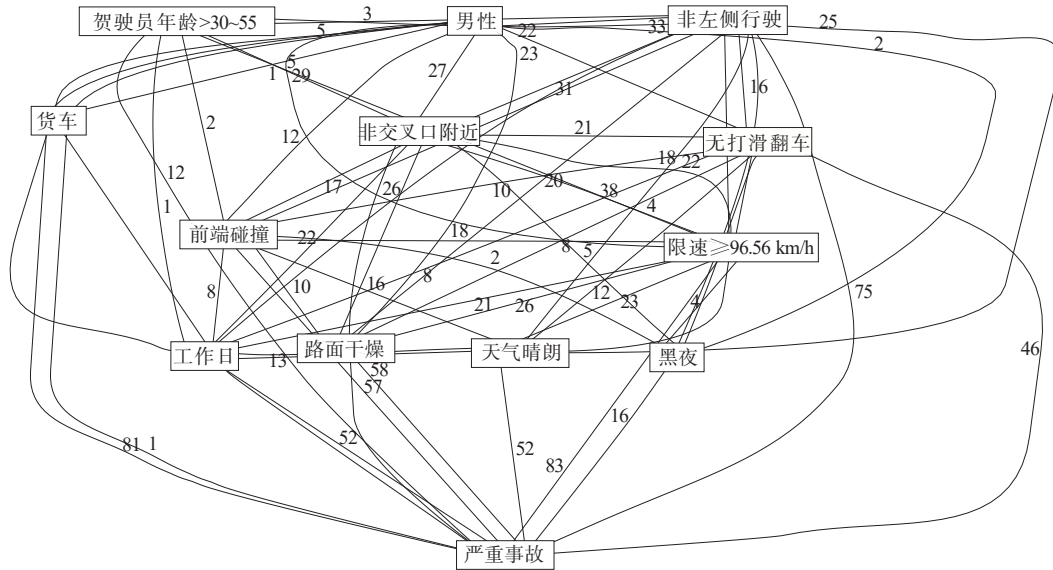


图6 严重事故的关联规则网络

Figure 6 Association rule network of serious accidents

严重事故的可能性较高,且货车体积大、载货重,事故发生后,易引发连环事故,加剧事故等级。

结合上述分析,为降低高速公路事故发生率及严重程度,建议:加强恶劣气象条件下高速公路的交通管控水平;设置LED可变信息情报板,当路面潮湿时,提醒驾驶员注意行车速度和行车环境;在光照强度不足时,提醒货车驾驶员避免疲劳驾驶,谨防严重事故的发生。此外,加强对中年驾驶员的安全教育;建议驾驶能力较低的驾驶人采用辅助驾驶系统提高行车安全。

4 结论

为深入挖掘高速公路不同事故等级的影响因素,本文提出一种考虑事故等级约束的改进 Apriori 算法,应用英国高速公路的交通事故数据集进行验证,分析结果并提出建议,得出以下结论:

(1) 改进 Apriori 算法对频繁项集中事故类型进行筛选,通过关联规则后项约束为事故维度减少冗余规则的产生,提高算法的准确性以及规则挖掘效率。

(2) 年龄 30~55 岁的驾驶人在高速公路易引发交通事故,可能是由于该年龄段的驾驶员因驾驶经验丰富而过分自信,对危险工况判断出现偏差,导致事故发生。

(3) 路面湿滑易引发一般事故,由于湿滑路面的道路摩擦系数低于干燥路面,车辆刹车距离变长,致

使事故发生。黑夜易引发一般和严重事故,增设激光示警灯缓解因单一枯燥的行车环境引起的疲劳,提高高速公路的安全通行水平。货车发生事故时,多为严重事故,应加强对货车司机的安全教育。

(4) 在后续研究中,可采用带约束的算法对最小支持度阈值进行求解,以针对不同需求确定最佳的最小支持度。

参考文献:

References:

[1] 袁振洲,娄晨,杨洋. 时间差异条件下的高速公路交通事故致因分析[J]. 北京交通大学学报,2021,45(3):1-7.
YUAN Zhenzhou, LOU Chen, YANG Yang. Analysis of highway traffic accidents causes under time differences[J]. Journal of Beijing Jiaotong University,2021,45(3):1-7.

[2] 赵佳军,吴昊,张文浩. 大流量集中养护下高速公路交通可行性研究[J]. 中外公路,2021,41(6):281-287.
ZHAO Jiajun, WU Hao, ZHANG Wenhao. Research on feasibility of expressway traffic under large flow centralized maintenance[J]. Journal of China & Foreign Highway,2021,41(6):281-287.

[3] 徐春明,汪春桃,吴家铭. 高速公路改扩建工程勘察技术要点概述[J]. 中外公路,2020,40(增刊 2):1-4.
XU Chunming, WANG Chuntao, WU Jiaming. Summary of survey technical points of expressway reconstruction and expansion project[J]. Journal of China & Foreign Highway,

- 2020,40(sup 2):1-4.
- [4] BUCSUHÁZY K, MATUCHOVÁ E, ZŮVALA R, et al.. Human factors contributing to the road traffic accident occurrence[J]. *Transportation Research Procedia*, 2020, 45: 555-561.
- [5] 邓国忠,曹帆,吴勇,等.互通式立交与隧道出口小间距路段事故影响因素分析[J]. *中外公路*, 2019, 39(4): 283-287.
DENG Guozhong, CAO Fan, WU Yong, et al. Analysis on accidents influence factors of the small spacing section between intersection and tunnel exit[J]. *Journal of China & Foreign Highway*, 2019, 39(4): 283-287.
- [6] CAI Q R. Cause analysis of traffic accidents on urban roads based on an improved association rule mining algorithm[J]. *IEEE Access*, 2020, 8: 75607-75615.
- [7] 牛毅,李振明,樊运晓.基于数据挖掘的高速公路货车交通事故影响因素关联分析研究[J]. *安全与环境工程*, 2020, 27(4): 180-188.
NIU Yi, LI Zhenming, FAN Yunxiao. Correlation analysis of influencing factors of truck traffic accidents on expressways[J]. *Safety and Environmental Engineering*, 2020, 27(4): 180-188.
- [8] 冯伟,刘朝晖,柳力,等.半刚性基层沥青路面车辙有限元模拟及灰关联分析[J]. *中外公路*, 2021, 41(6): 60-65.
FENG Wei, LIU Zhaohui, LIU Li, et al. Finite element simulation and grey correlation analysis of rutting on asphalt pavement with semi-rigid base[J]. *Journal of China & Foreign Highway*, 2021, 41(6): 60-65.
- [9] 周小勇,宁建标,张兵,等.基于外部环境因素影响的高速公路交通事故聚类分析[J]. *公路与汽运*, 2017(4): 36-39.
ZHOU Xiaoyong, NING Jianbiao, ZHANG Bing, et al. Cluster analysis of expressway traffic accidents based on the influence of external environmental factors[J]. *Highways & Automotive Applications*, 2017(4): 36-39.
- [10] 徐金华,李岩,张玉婷.一种挖掘交叉口事故严重程度影响因素关联特性的算法[J]. *安全与环境学报*, 2022, 22(3): 1412-1420.
XU Jinhua, LI Yan, ZHANG Yuting. An algorithm to mine the association rules among causes of accident severity at intersections[J]. *Journal of Safety and Environment*, 2022, 22(3): 1412-1420.
- [11] 李珏,李世杰.基于文本挖掘的高处坠落事故致因及关联规则分析[J]. *长沙理工大学学报(自然科学版)*, 2020, 17(2): 61-67, 74.
LI Jue, LI Shijie. Analysis of causes and association rules for falling accidents based on text mining method[J]. *Journal of Changsha University of Science & Technology (Natural Science)*, 2020, 17(2): 61-67, 74.
- [12] YÜRÜŞEN N Y, UZUNOĞLU B, TALAYERO A P, et al. Apriori and K-means algorithms of machine learning for spatio-temporal solar generation balancing[J]. *Renewable Energy*, 2021, 175: 702-717.
- [13] PASIDIS I. Congestion by accident? A two-way relationship for highways in England[J]. *Journal of Transport Geography*, 2019, 76: 301-314.
- [14] 胡松,林鹏飞,翁剑成,等.基于改进 Apriori 算法的乘客公共交通依赖性层级转移分析[J]. *东南大学学报(自然科学版)*, 2022, 52(2): 344-351.
HU Song, LIN Pengfei, WENG Jiancheng, et al. Hierarchy transfer analysis of passengers' dependence on public transport based on improved Apriori algorithm[J]. *Journal of Southeast University (Natural Science Edition)*, 2022, 52(2): 344-351.
- [15] 刘志强,王玲,张爱红,等.基于贝叶斯模型的雾霾天高速公路交通事故发生机理研究[J]. *重庆理工大学学报(自然科学)*, 2018, 32(1): 43-49.
LIU Zhiqiang, WANG Ling, ZHANG Aihong, et al. Study on traffic accidents occurrence mechanism in haze weather on the highway[J]. *Journal of Chongqing University of Technology (Natural Science)*, 2018, 32(1): 43-49.
- [16] 程明,梅傲然.基于关联规则的跑道着陆事件演化网络构建研究[J]. *中国安全生产科学技术*, 2022, 18(12): 203-208.
CHENG Ming, MEI Aoran. Research on construction of evolution network for runway landing event based on association rules[J]. *Journal of Safety Science and Technology*, 2022, 18(12): 203-208.