

安全与环境学报
Journal of Safety and Environment
ISSN 1009-6094, CN 11-4537/X

《安全与环境学报》网络首发论文

题目：一种挖掘交叉口事故严重程度影响因素关联特性的算法
作者：徐金华，李岩，张玉婷
DOI：10.13637/j.issn.1009-6094.2021.1553
收稿日期：2021-09-03
网络首发日期：2021-12-02
引用格式：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/OL]. 安全与环境学报.
<https://doi.org/10.13637/j.issn.1009-6094.2021.1553>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

一种挖掘交叉口事故严重程度影响因素关联特性的算法

徐金华, 李岩, 张玉婷*

(长安大学运输工程学院, 西安 710064)

摘要: 为探究平面交叉口不同严重程度交通事故影响因素的关联特性, 从驾驶人、环境、道路、车辆四个维度出发, 建立了带约束的二进制粒子群-蚁群算法以挖掘交叉口事故严重程度的关联规则。首先通过二进制粒子群算法确定最优提升度和支持度阈值, 再利用蚁群算法挖掘最大频繁项集, 并增加规则前项与后项的约束, 以提高关联规则挖掘效率。对 2022 条平面交叉口事故数据的分析结果表明, 论文提出的 BPSO-ACA 算法可在精准识别潜在因素与事故等级关联结果的基础上减少冗余无效关联规则, 并将规则挖掘效率提升 17%~29%。驾驶人年龄、分心状态、交叉口形态、交叉口车道数和天气均与交叉口事故严重程度有强关联性; 车道数少是导致轻微事故升级为一般事故的重要因素; 路面湿滑是导致交叉口事故升级为重大事故的关键因素, 尤其在雨雪雾等不良天气条件下, 机非事故为重大等级的可能性最高。本研究成果可为交通管理部门的主动防控措施提供理论指导, 减少交叉口事故数量及降低事故严重程度。

关键词: 安全工程; 事故严重程度; 关联规则; 致因分析; 二进制粒子群; 蚁群算法; 交叉口

中图分类号: X951 **文献标志码:** A

DOI: 10.13637/j.issn.1009-6094.2021.1553

0 引言

平面交叉口作为道路网络的重要节点, 交通运行环境复杂, 各类车流在交叉口冲突区域反复交织冲突, 是交通事故高发的区域^[1]。据统计, 我国城市交通事故约有 30% 发生在交叉口^[2]。根据交通事故中受伤人数以及财产损失的不同, 事故被划分为轻微、一般、重大和特大四种等级。降低事故的严重程度是提升道路交叉口交通安全水平的重要途径。因各严重程度等级交通事故的致因有所差异, 所以识别各严重程度事故的影响因素可有针对性的制定安全防控策略, 从而有效降低交叉口交通事故的严重程度, 提高其安全水平。研究成果可有针对性、分层级地为交通设施的规划、设计和管理等方面提供理论支撑。

事故严重程度致因多从人、车、道路和环境等因素分析^[3-5], 常用方法主要为参数模型, 2002 年, Al Ghamdi AS^[6]首次应用 logistic 回归方法分析发现事故地点等与事故严重程度显著相关; 何雅琴等^[7]则采用累积 Logistic 回归分析了行人交通事故严重程度的影响因素, 发现照明条件和行人饮酒情况等 6 个因素与事故严重程度显著相关; 马明等^[8]构建广义估计方程负二项回归模型分析了影响信号交叉口安全的因素; 秦雅琴等^[9]发现多项 Logit 模型更适用于分析山区双车道公路摩托车事故严重程度的影响因素。然而, 参数模型需要预先设定参数服从的概率分布, 当样本分布不均衡时易出现欠拟合问题, 结果精度不高。为克服此缺陷, 如贝叶斯网络^[10]、随机森林^[11]、分类树^[12]等非参数模型被用于识别事故严重程度的致因。非参数模型在自变量因素较多时需要的样本量较大, 且易出现过拟合问题。上述方法均考虑了某些因素对事故的影响, 但忽略了因素间的潜在关联性, 难以挖掘多维因素相互作用对事故严重程度的影响。

关联规则是分析数据集各变量间关联特性的有效方法, 可用于挖掘历史交通事故数据中潜藏的事故因素间内在关联性。与参数模型不同, 关联规则无需预先确定假设及函数形式; 相比于其他非参数模型, 样本量较少时关联规则也可进行有效分析, 并获取各因素与事故间的关联特征。2016 年, LIU 等^[13]最先应用关联规则算法分析了道路作业区事故伤亡程度的影响因素。牛毅等^[14]将传统关联规则算法与社会网络分析法相结合来分析高速公路货车交通事故的影响因素, 并提出了具有针对性的事故预防对策与建议。为减少冗余关联规则的产生, 袁振洲等^[15]采用定向约束的 Apriori 算法分别分析了工作日和休息日下高速公路交通事故致因, 发现工作日和休息日发生的不同等级交通事故形态存在差异。然而传统关联规则算法需要人为设置支持度和置信度阈值, 易产生冗余规则或忽略重要规则。

综上, 为挖掘平面交叉口不同严重程度交通事故各因素变量间的相互耦合关联性, 并克服传统关联规则方法需要人为主观地设置阈值的缺陷, 减少冗余无效规则的生成, 构建了带约束的二进制粒子群-蚁群(Binary particle swarm optimization - Ant colony algorithm, BPSO-ACA)关联规则挖掘算法。在分析了驾驶人、车辆、道路和环境四种单因素与事故严重程度相关性的基础上, 深入挖掘这四种维度因素的耦合作用与平面交叉口事故严重程度的关联性。

收稿日期: 2021-09-03

作者简介: 徐金华, 博士研究生, 从事交通安全、交通规划和大数据方面研究; 李岩, 教授, 博导, 从事交通安全、交通行为和交通信控方面研究; 张玉婷, 博士, 讲师, 从事交通安全、驾驶行为和交通规划方面研究, ytzhang@chd.edu.cn

基金项目: 国家自然科学基金资助(基于驾驶仿真实验的非故意闯红灯行为规律及其事故智能预警方法研究, 71901036), 国家重点研发计划资助(道路交通安全主动防控技术及系统集成, 2017YFC0803906), 陕西省自然科学基金资助(手动-自动车辆混合异质交通流下的智慧高速公路加速车道长度研究, 2020JM-222; 空气污染环境下考虑出行者健康的交通管控方法, 2020JM-237)

1 关联规则基本概念

关联规则算法可挖掘数据库中各项要素间的内在联系。关联规则可用 $X \Rightarrow Y$ 的形式表述，其中 $X \cap Y = \emptyset$ ， X 为规则前项(Left-Hand-Side, LHS)， Y 为规则后项(Right-Hand-Side, RHS)。

如式(1)所示，关联规则 $X \Rightarrow Y$ 的支持度(support)为规则前项 X 和规则前项 Y 在整个数据集同时发生的概率。最低支持度 S_{\min} 可表示项目集统计意义上的最低重要性。

$$S(X \Rightarrow Y) = P(X \cap Y) \quad (1)$$

式中， $S(X \Rightarrow Y)$ 为规则前项 X 指向规则前项 Y 的支持度； $P(X \cap Y)$ 为事务 X 、 Y 同时发生的概率。

关联规则 $X \Rightarrow Y$ 的置信度(confidence)是指在前项 X 出现的情况下，后项 Y 出现的概率，可应用式(2)计算。最小置信度 C_{\min} 表示规则的最低可靠性。

$$C(X \Rightarrow Y) = P(Y|X) = \frac{S(X \cup Y)}{S(X)} \quad (2)$$

关联规则 $X \Rightarrow Y$ 的提升度(lift)被定义为由式(3)计算的在含有事务 X 的条件下同时含有事务 Y 的概率与事务 Y 在整个数据集中出现的概率之比。提升度表明了事务 X 对事务 Y 影响力的大小，有效强关联规则的最小提升度 L_{\min} 需大于 1^[13]。

$$L(X \Rightarrow Y) = \frac{P(Y|X)}{P(Y)} = \frac{C(X \Rightarrow Y)}{P(Y)} \quad (3)$$

2. BPSO-ACA 关联规则挖掘算法

2.1 算法框架

经典的 Apriori 关联规则挖掘算法采用的是逐层搜索迭代的思路，从单元素项集开始逐层搜索满足条件的项集，在此过程中会形成大量候选项集。当数据集较大时，算法的复杂度也随之增加。在筛选关联规则时，通常以 S_{\min} 和 C_{\min} 为指标。然而，该两项指标通常由人为主观确定，易导致生成大量冗余规则，并无法有效识别重要规则。为克服上述问题，本文建立了带约束的 BPSO-ACA 关联规则挖掘算法用以挖掘高质量的规则。由于粒子群算法具有实数优化性能好、搜索能力强、收敛速度快等诸多优点^[16]，因此选择粒子群算法来搜索关联规则算法的 S_{\min} 和 C_{\min} 阈值。蚁群算法具有良好的鲁棒性和全局搜索能力，其使用信息素决定移动方向，只需扫描一次数据集^[17]即可有效、快速地获取各因素与事故严重程度关系，提高算法性能，因此选取蚁群算法生成频繁项。最后，增加对规则前项与后项要素的约束，即规则前项不包含事故等级要素，规则后项仅为事故等级要素。

2.2 规则生成阈值确定

关联规则算法中确定最优支持度和置信度阈值是离散问题，可采用二进制粒子群优化算法求解。在粒子群算法中粒子的适应度越高，其位置越好。通过构造适应度函数，搜索粒子在全局的最优位置，并将该位置的支持度和置信度作为关联规则生成的最优最小支持度和最小置信度。适应度函数 f 的计算公式如下：

$$f = \gamma \cdot S + \theta \cdot C \quad (4)$$

式中， S 为支持度； C 为置信度； γ 与 θ 为常数，用于调整关联规则评价的偏重，取值范围为 $[0,1]$ ，为获取更多关联规则，需要提高规则支持度的权重。为了获取高置信度的规则，参照文献^[18]的结果，本文取 $\gamma = 0.3$ ， $\theta = 0.7$ 。

在 BPSO 算法中每个粒子都有自身的位置和速度信息及其适应值，每个粒子通过追踪个体极值和全局极值来不断地更新自身。

设 D 表示搜索空间的维度， $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})^T$ 表示粒子 i 当前位置， $p_i = (p_{i1}, p_{i2}, \dots, p_{iD})^T$ 表示粒子 i 曾经到达的最优位置，最优粒子序号为 g ， $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})^T$ 是粒子 i 的速度，每个粒子依据式(5)和式(6)更新自身的速度与位置：

$$v_{id}(t+1) = \omega v_{id}(t) + c_1 r_1 (p_{id}(t) - x_{id}(t)) + c_2 r_2 (p_{gd}(t) - x_{id}(t)) \quad (5)$$

$$x_{id}(t+1) = \begin{cases} 1, & r_3 < \text{sig}(v_{id}(t+1)) \\ 0, & r_3 \geq \text{sig}(v_{id}(t+1)) \end{cases} \quad (6)$$

式中， t 为迭代次数； c_1, c_2 为学习因子，取 $c_1 = 1$ ， $c_2 = 1$ ； r_1, r_2 为 $[0,1]$ 间随机数； ω 为惯性权重系数； r_3 为 $(0,1)$ 间的随机数； $\text{sig}()$ 为 Sigmoid 函数。

重复上述过程，直到满足迭代要求，输出全局极值 g_{best} ，最终得到最优的 S_{\min} 和 C_{\min} 。

2.3 频繁项集生成

蚁群算法在挖掘关联规则时需将问题转换为旅行商问题。根据数据集中各离散属性间关系建立无向图。将数据中的各项要素均作为一个节点，将无向图节点 i 到其他节点 j 的支持度 S_{ij} 作为边 (i, j) 的权值， S_{ij} 的倒数为节点 i 与 j 间距离。

选取 m 只蚂蚁放入无向图进行遍历。蚂蚁在遍历过程中根据转移概率来选择下一节点，转移概率函数可表示为：

$$P_{ij}^k(t) = \begin{cases} \frac{\tau_{ij}(t)^\alpha \eta_{ij}(t)^\beta}{\sum_{s \in \text{allowed}_k} \tau_{is}(t)^\alpha \eta_{is}(t)^\beta}, & j \in A_k \\ 0, & \text{其他} \end{cases} \quad (7)$$

式中, $\tau_{ij}(t)$ 表示 t 时刻的信息素浓度, $\tau_{ij} \in (\tau_{\min}, \tau_{\max})$, 设 $\tau_{ij}(0) = \tau_{\max}$; $\eta_j(t)$ 为启发函数, 令 $\eta_j(t) = S_{ij}$; α 与 β 分别表示信息素和启发因子的相对重要程度; A_k 表示蚂蚁 k 可选择的节点集合。

当蚂蚁选择了下一节点时, 需要更新当前节点的信息素, 即局部信息素的更新, 规则为:

$$\tau_{ij}(t+1) = \varepsilon \cdot \tau_{ij}(t) + (1 - \varepsilon) \cdot S_{ij} \quad (8)$$

式中, ε 为信息素挥发系数, $\varepsilon \in (0,1)$ 。

蚂蚁 k 选择了节点 i 加入路径后, 判断此路径上的节点是否满足 S_{\min} 要求, 若满足要求可保留该路径, 否则从该路径中删除该节点 i 。当所有节点都遍历完成时, 可获取本次循环中具有最大支持度的频繁项集 L_{best} 。

当所有节点都遍历完时需要更新全局信息素, 更新规则为:

$$\tau_{ij}(t+1) = \rho \cdot \tau_{ij}(t) + (1 - \rho) \cdot \Delta\tau_{ij}(t) \quad (9)$$

式中, ρ 为信息残留系数, $\rho \in (0,1)$; $\Delta\tau_{ij}(t)$ 为信息素的增量, 若 ij 在频繁项集中, 则 $\Delta\tau_{ij}(t) = 1/L_{\text{best}}$, 若 ij 不在频繁项集中, 则 $\Delta\tau_{ij}(t) = 0$ 。

循环上述过程直到满足迭代要求, 最终可输出所有的频繁项集。

2.4 规则要素约束

对每个频繁项集 L_k 的任一非空子集 A , 若 $\frac{S(L_k)}{S(A)} > C_{\min}$, 则生成规则 $A \Rightarrow (L_k - A)$ 。但在此过程中容易出现无效的规则。例如频繁项集{白天, 轻微, 干燥}, 假设其生成的规则均满足阈值要求, 则共可生成 6 条规则, 但其中有效规则为“{白天, 干燥} \Rightarrow {轻微}”, 而诸如“{轻微, 干燥} \Rightarrow {白天}”的规则为无效规则。为了筛除此类无效规则, 对规则增加维度约束, 即判断规则 $A \Rightarrow (L_k - A)$ 是否为前项 A 中不包含事故等级要素且后项 $L_k - A$ 中仅包含事故等级要素的规则, 若是则保留规则, 若不是则筛除。

2.5. BPSO-ACA 关联规则生成算法

因此, 带约束的 BPSO-ACA 关联规则生成算法包含规则生成阈值确定、频繁项集生成、规则要素约束等步骤, 算法流程图如图 1 所示。

步骤 1: 初始化算法各参数, 生成原始种群;

步骤 2: 计算粒子群的速度和位置, 计算粒子的适应度函数值, 更新局部最优位置和适应度;

步骤 3: 更新全局最优位置和适应度, 更新粒子的速度和位置;

步骤 4: 若满足迭代条件, 则输出最优 S_{\min} 和 C_{\min} , 进入步骤 5, 否则, 返回步骤 2;

步骤 5: 将 m 只蚂蚁放入无向图中开始遍历;

步骤 6: 蚂蚁 k 随机选择节点 i 开始, 并将该节点 i 放入禁忌表 $tabu_k$ 中, 计算当前节点到相邻节点的转移概率, 选择转移概率最大值作为下一节点, 根据局部信息素更新规则更新当前节点信息素;

步骤 7: 若所有节点都遍历完, 则停止遍历, 并将此过程中满足 S_{\min} 要求的路径节点记录到频繁项元素表中, 否则转入步骤 6;

步骤 8: 更新全局信息素, 更新所有被蚂蚁遍历过的节点信息素;

步骤 9: 若达到最大迭代次数, 则生成最大频繁项集, 否则转入步骤 5;

步骤 10: 对每个频繁项集 L_k 的任一非空子集 A , 若 $\frac{S(L_k)}{S(A)} > C_{\min}$, 则生成规则 $A \Rightarrow (L_k - A)$, 并判断规则 $A \Rightarrow (L_k - A)$ 是否满足维度约束, 若是则保留输出规则, 若不是则筛除。

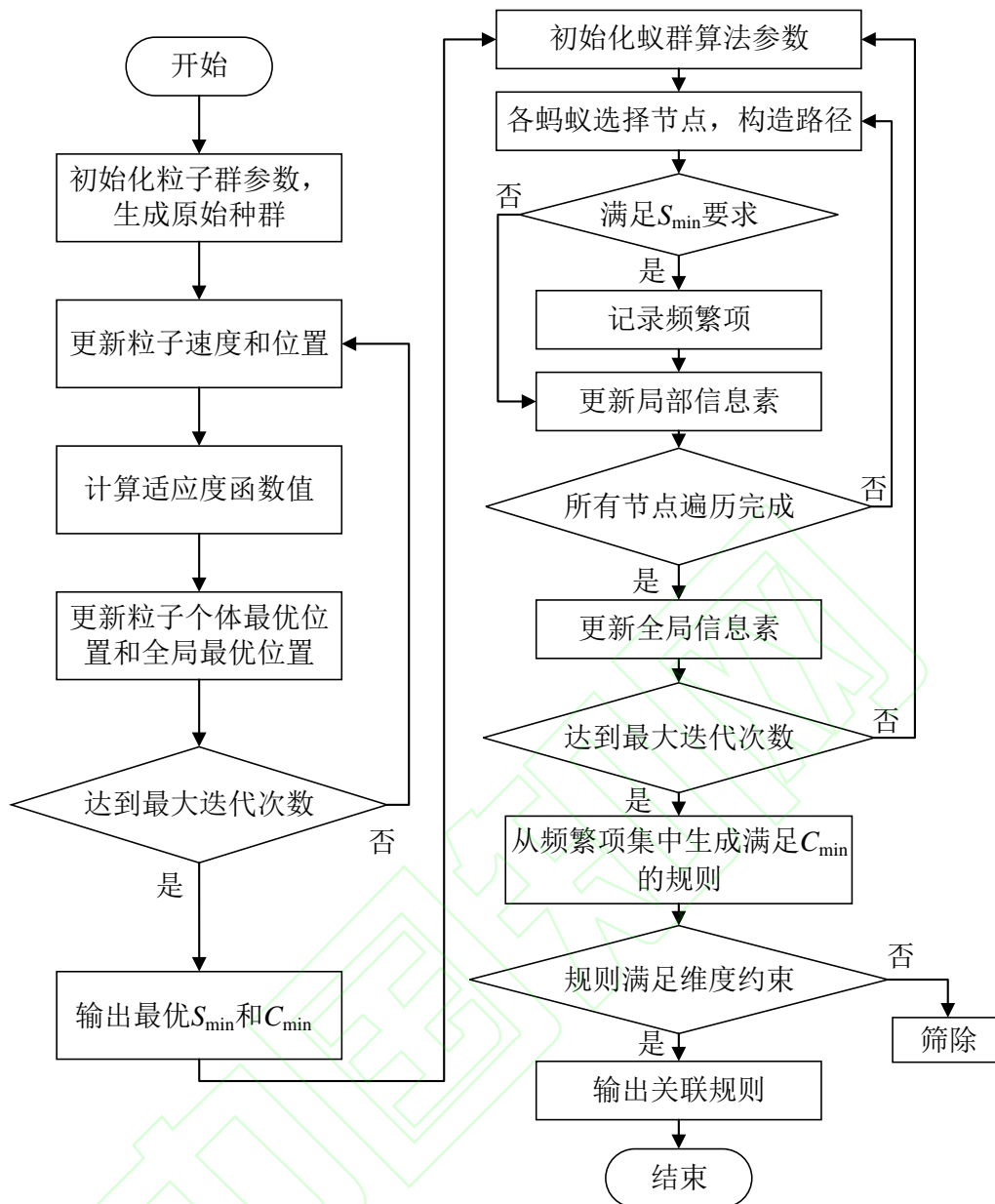


图1 BPSO-ACA 算法流程图

Fig. 1 BPSO-ACA algorithm flow chart

3 数据及结果

3.1 事故数据

选取国内某地交通管理部门统计的平面交叉口事故数据, 共包含有效数据 2022 条。从可能导致事故发生的环境、道路、车辆和驾驶人四个维度统计平面交叉口的事故严重等级数据。数据集的统计特性如表 1 所示, 由于特大交通事故占比不到 1%, 数量太少无法生成规则, 故将其与重大事故进行合并。为加快算法的求解效率, 在数据输入模型前, 对原始数据进行二进制转换, 即将每条数据中包含的变量记为 1, 不包含的变量记为 0, 例如, 某条数据包含的变量有白天、晴朗、干燥、2 车道、十字交叉、机动车、男性、20 至 39 岁、分心和一般, 则其经过转换后的数据如图 2 所示。并采用密西根方法 (Michigan approach) [19] 对关联规则编码。

表 1 变量及数据统计

Table 1 Variables and data statistics

维度	属性	描述	频数	频率	维度	属性	描述	频数	频率
环境	光照	白天 I_1	1574	0.78	车辆	涉及车辆	机动车 I_{17}	1241	0.61
		黑夜 I_2	448	0.22		类型	机动车/非机动车 I_{18}	781	0.39
	天气	晴朗 I_3	1400	0.69	驾驶人	性别	男性 I_{19}	1091	0.54
		阴 I_4	387	0.19		女性 I_{20}	931	0.46	
		雨/雪/雾 I_5	235	0.12	驾驶人	年龄	小于 20 岁 I_{21}	194	0.10
路面	干燥 I_6	1627	0.80	驾驶人		20 至 39 岁 I_{22}	855	0.42	
条件	湿滑 I_7	395	0.20	年龄		40 至 59 岁 I_{23}	520	0.26	
道路	车道数	1 车道 I_8	20	0.01			≥ 60 岁 I_{24}	453	0.22
		2 车道 I_9	776	0.38		驾驶人是	未分心 I_{25}	1416	0.70
		3 车道 I_{10}	404	0.20			否分心	分心 I_{26}	606
		4 车道 I_{11}	531	0.26		事故等级	轻微 I_{27}	1052	0.52
		5 车道 I_{12}	192	0.09			一般 I_{28}	772	0.38
		≥ 6 车道 I_{13}	99	0.06			重大 I_{29}	198	0.1
交叉口形态	三路交叉 I_{14}	212	0.10						
	十字交叉 I_{15}	1607	0.80						
	五路交叉 I_{16}	203	0.10						

I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}	I_{16}	I_{17}	I_{18}	I_{19}	I_{20}	I_{21}	I_{22}	I_{23}	I_{24}	I_{25}	I_{26}	I_{27}	I_{28}	I_{29}
1	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	1	0	1	0	0	1	0	0	0	1	0	1	0

图 2 二进制转换示例

Fig. 2 Example of binary transformation

3.2 关联规则生成

利用本文提出的 BPSO-ACA 算法求得最优支持度和置信度分别为 0.02 和 0.35。将 BPSO-ACA 算法与经典 Apriori 算法和改进的 Apriori 算法^[15]进行对比, 结果如表 2 所示。通过分析不同支持度和置信度阈值条件下不同算法生成的有效关联规则数和耗时, 可以看出本文提出的 BPSO-ACA 算法挖掘效率最高。相较于经典 Apriori 算法, BPSO-ACA 算法与改进 Apriori 算法均增加了维度约束, 可以避免生成大量冗余无效规则。与改进 Apriori 算法相比, BPSO-ACA 算法中的智能体(蚂蚁)会优先选择信息素浓度较高的路径, 更有效地降低了低支持度规则的生成, 从而提高关联规则的挖掘效率。

表 2 BPSO-ACA 算法与改进 Apriori 算法挖掘结果比较

Table 2 Comparison of result between BPSO-ACA algorithm and the improved Apriori algorithm

算法	支持度	置信度	规则数	耗时/s
BPSO-ACA 算法	0.01	0.20	19183	76.22
	0.02	0.35	7948	41.94
	0.10	0.50	765	3.71
改进 Apriori 算法 ^[15]	0.01	0.20	21153	107.35
	0.02	0.35	8606	53.06
	0.10	0.50	778	4.49
经典 Apriori 算法	0.01	0.20	2594144	106.23
	0.02	0.35	726925	52.78
	0.10	0.50	39223	4.31

3.3 关联规则分析

1)环境维度的影响

筛选前项仅为环境因素的规则，得到后项为轻微的规则 3 条，后项为一般的规则 0 条，后项为重大的规则 2 条。如表 3 所示，在晴朗的天气条件下，交叉口交通事故严重程度较低，主要为轻微事故；而在雨/雪/雾恶劣天气条件下，特别是在黑夜，交叉口交通事故的严重程度有提升倾向。

表 3 前项为环境维度的关联规则

Table 3 Association rules with environmental dimensions LHS					
序号	前项	后项	支持度	置信度	提升度
1	{晴朗, 黑夜}	{轻微}	0.082	0.576	1.108
2	{晴朗}	{轻微}	0.399	0.576	1.107
3	{晴朗, 白天}	{轻微}	0.317	0.576	1.106
4	{雨/雪/雾}	{重大}	0.062	0.532	2.662
5	{雨/雪/雾, 黑夜}	{重大}	0.045	0.529	2.646

2)道路维度的影响

筛选前项仅为道路因素的规则，得到后项为轻微的规则 10 条，后项为一般的规则 2 条，后项为重大的规则 2 条。表 4 列出各等级事故规则提升度最大的前 5 条规则，可以看出十字路口发生的交通事故等级多为轻微，而在五路交叉口发生的事故等级多为一般；在交叉口的路面为干燥的条件下，发生交通事故的严重程度多为轻微等级，而当路面湿滑时发生交通事故的严重程度等级会有提升的倾向；当交叉口车道数较多时，发生的交通事故等级也较低，这应该是由于车道数较多，车流冲突较少，故而所发生事故的严重程度也较低；而当车道数较少为 2 车道时，交叉口发生的交通事故多为一般或重大等级，尤其是湿滑的 2 车道交叉口发生重大等级交通事故的可能性更高。

表 4 前项为道路维度的关联规则 (部分)

Table 4 Association rules with road LHS (part)					
序号	前项	后项	支持度	置信度	提升度
1	{十字交叉, 干燥}	{轻微}	0.032	0.727	1.398
2	{十字交叉, ≥6 车道}	{轻微}	0.035	0.696	1.338
3	{≥6 车道}	{轻微}	0.034	0.683	1.313
4	{4 车道, 十字交叉, 干燥}	{轻微}	0.109	0.623	1.198
5	{5 车道, 十字交叉, 干燥}	{轻微}	0.038	0.606	1.165
6	{五路交叉}	{一般}	0.037	0.468	1.673
7	{2 车道, 五路交叉}	{一般}	0.016	0.371	1.325
8	{4 车道, 湿滑}	{重大}	0.021	0.5	2.502
9	{2 车道, 湿滑}	{重大}	0.036	0.48	2.404

3)车辆维度的影响

筛选前项仅为车辆因素的规则为 0 条，可能由于事故库中车辆维度仅 2 个因素，数量太少，未进入规则。

4)驾驶人维度的影响

筛选前项仅为环境因素的规则，得到后项为轻微的规则 3 条，后项为一般的规则 0 条，后项为重大的规则 0 条。如表 5 所示，年龄小于 20 岁的驾驶人在交叉口发生轻微等级交通事故的可能性较高。可能是因为年龄较小的驾驶人驾驶经验不足，易在交叉口发生剐蹭等轻微交通事故。

表 5 前项为驾驶人因素的关联规则

Table 5 Association rules with driver LHS					
序号	前项	后项	支持度	置信度	提升度
1	{小于 20 岁, 未分心, 男性}	{轻微}	0.021	0.656	1.261
2	{小于 20 岁, 男性}	{轻微}	0.031	0.618	1.187
3	{小于 20 岁, 未分心}	{轻微}	0.04	0.597	1.147

5)多维度的综合影响

为探究环境、道路、车辆和驾驶人四种维度对事故严重程度的综合影响，筛选提升度不低于 1 的关联规则，得到后项为

轻微的规则 3651 条，后项为一般的规则 634 条，后项为重大的规则 179 条，表 6 列出各等级事故规则提升度最大的前 8 条规则。规则①、⑥和⑧表明在天气、光照良好时，小于 20 岁的驾驶人在十字路口与机动车发生冲突后，造成事故等级为轻微的可能性较高。规则④和⑤表明在天气、光照条件良好时，男性驾驶机动车在三路交叉口因与机动车冲突造成事故等级为轻微的可能性较高。规则②~③和⑦表明在天气、光照良好的条件下，十字路口机非碰撞事故等级为轻微的可能性较高。规则⑨和⑩表明在天气、光照良好时，年龄超过 60 岁的驾驶人驾驶机动车在十字路口与非机动车发生冲突后，造成事故等级为一般的概率较高，约 80.0%。规则⑪~⑬表明天气、光照良好时，五路交叉路口的机非事故等级为一般的概率较大。规则⑭~⑯表明天气、光照良好时，在 2 车道或 3 车道的十字路口发生的仅涉及机动车的事故等级为一般的概率较大。规则⑰、⑱和⑳表明在雨/雪/雾天气或路面湿滑条件下，交叉口机非事故为重大等级的概率较大。规则㉑表明在雨/雪/雾天气下，当机动车驾驶人处于分心状态时发生重大事故的概率为 70.1%。规则㉒表明发生在黑夜五路交叉路口的机非事故等级为重大的概率为 66.3%。规则㉓表明在雨/雪/雾天气下，发生于 2 车道交叉路口的机非事故为重大等级的概率为 53.5%。

表 6 前项为多维度的关联规则(部分)

Table 6 Association rules with multiple dimensions LHS (part)

序号	前项	后项	支持度	置信度	提升度
1	{十字路口, 小于 20 岁, 干燥, 晴朗, 机动车, 男性}	{轻微}	0.042	0.903	1.736
2	{20 至 39 岁, 十字路口, 干燥, 晴朗, 机动车/非机动车, 白天}	{轻微}	0.049	0.884	1.699
3	{4 车道, 十字路口, 干燥, 晴朗, 机动车/非机动车, 男性, 白天}	{轻微}	0.034	0.872	1.676
4	{三路交叉, 机动车, 男性, 白天}	{轻微}	0.03	0.871	1.675
5	{三路交叉, 晴朗, 机动车, 男性, 白天}	{轻微}	0.023	0.868	1.668
6	{十字路口, 小于 20 岁, 机动车, 男性}	{轻微}	0.021	0.857	1.647
7	{5 车道, 十字路口, 干燥, 机动车/非机动车}	{轻微}	0.021	0.857	1.647
8	{十字路口, 小于 20 岁, 晴朗, 机动车, 白天}	{轻微}	0.037	0.851	1.635
9	{十字路口, ≥60 岁, 干燥, 晴天, 机动车/非机动车, 男性}	{一般}	0.02	0.8	2.858
10	{十字路口, ≥60 岁, 机动车/非机动车, 男性, 白天}	{一般}	0.023	0.797	2.846
11	{五路交叉, 干燥, 晴天, 机动车/非机动车, 白天}	{一般}	0.025	0.794	2.835
12	{五路交叉, 未分心, 机动车/非机动车, 白天}	{一般}	0.022	0.789	2.82
13	{五路交叉, 晴天, 机动车/非机动车, 白天}	{一般}	0.025	0.785	2.803
14	{3 车道, 十字路口, 干燥, 机动车, 女性}	{一般}	0.021	0.741	2.649
15	{2 车道, 十字路口, 干燥, 晴天, 机动车, 女性, 白天}	{一般}	0.024	0.727	2.598
16	{2 车道, 十字路口, 晴天, 机动车, 女性, 白天}	{一般}	0.024	0.696	2.485
17	{机动车/非机动车, 湿滑, 雨/雪/雾}	{重大}	0.023	0.73	3.654
18	{分心, 雨/雪/雾}	{重大}	0.023	0.701	3.511
19	{五路交叉, 黑夜, 机动车/非机动车}	{重大}	0.029	0.663	3.318
20	{十字路口, 机动车/非机动车, 湿滑, 白天}	{重大}	0.023	0.657	3.289
21	{十字路口, 湿滑, 男性, 雨/雪/雾}	{重大}	0.021	0.642	3.212
22	{未分心, 机动车/非机动车, 湿滑}	{重大}	0.026	0.589	2.947
23	{20 至 39 岁, 十字路口, 湿滑, 白天}	{重大}	0.026	0.57	2.852
24	{2 车道, 机动车/非机动车, 雨/雪/雾}	{重大}	0.023	0.535	2.677

将提升度>1 的关联规则绘制气泡图^[13]，图中横坐标为后项，纵坐标为前项，如图 4 所示。选取前项中最有代表性的项目分组，括号内的项即为代表项，“+n Items”表示还有 n 个项目在该组中，气泡图中图标大小表示该组累积支持度，图标颜色表示该组累积提升度。

综合图 3 和表 3-6 结果可发现，轻微事故多发于晴朗、路面干燥、白天条件下。年龄小于 20 岁的驾驶人可能因为驾驶经验不足，在交叉口易发生剐蹭等轻微交通事故。年龄为 20 至 39 岁的驾驶人也易发生轻微事故，此类驾驶人驾驶经验较丰富，更易采取冒进的风险行为^[20]。轻微事故也多发于车道数较多的交叉口，表明车道数较多的情况下，避险空间较大，发生

事故的等级较低。

一般事故多发生于晴天、干燥环境下进出口车道数较少的交叉口。其原因可能是因为车道数量少造成的车辆换道空间不足，冲突增加。当事故发生在五路交叉口且涉及到非机动车时，事故等级也会呈现加重可能。在雨/雪/雾环境下，女性驾驶人发生一般等级事故的潜在可能性有增加倾向，其原因可能为女性驾驶人在面对紧急情况时控制车辆的经验欠佳^[21]。此外，在同样晴天、干燥的十字交叉口，年龄大于 60 岁的驾驶人可能因为身体机能下降，反应不够灵敏，避险能力减弱^[22]，发生一般事故的可能性较高。

重大事故易发生于湿滑、黑夜、雨/雪/雾的环境下，且多为机非事故。其原因可能为在遭遇雨/雪/雾等不良天气时，路面容易湿滑，车辆的制动性能降低；黑夜驾驶时，驾驶人受光线的影响，视线不佳，不易察觉非机动车，特别是在五路交叉口，路况较为复杂，一旦与非机动车发生碰撞，驾驶人易出现因危险状况不能及时察觉，或者判断决策失误而发生重大事故。发生重大交通事故的交叉口进出口车道数常较少，车辆避险可选择空间较少，当遇到紧急情况时，机动车驾驶人来不及避让，从而造成的交通事故等级更严重。此外，交叉口的重大交通事故还与机动车驾驶人的分心状态有较大关联性。

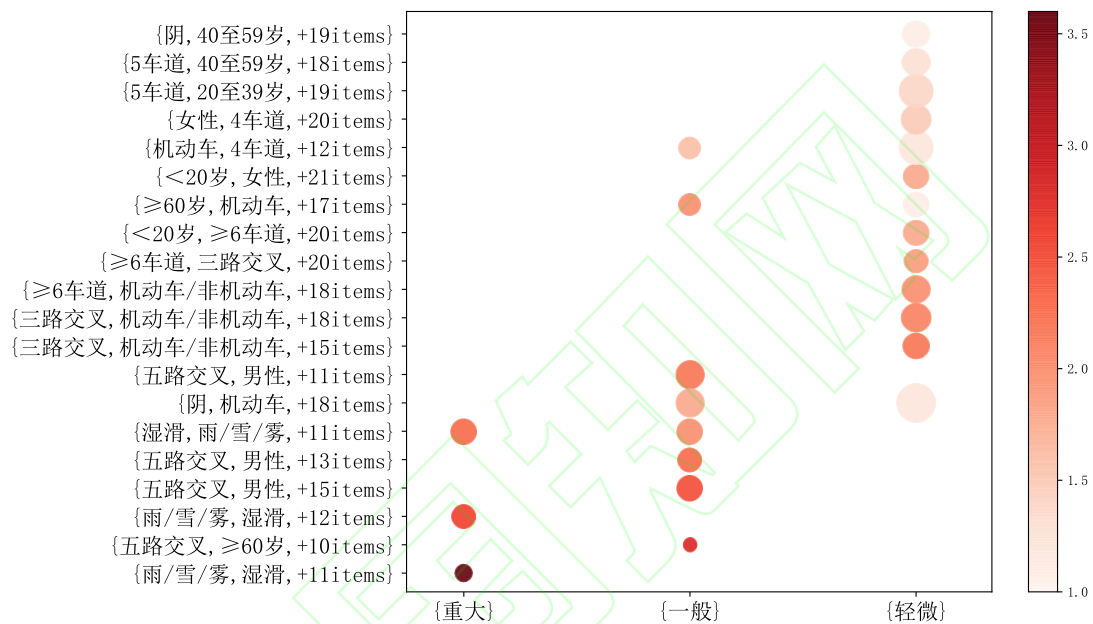


图3 关联规则气泡图

Fig. 3 Grouped balloon plot of rules

4 结论

为深度挖掘平面交叉口不同严重程度事故的影响因素，本文提出了一种带约束的 BPSO-ACA 关联规则算法，并应用我国某地的事故数据集进行验证，得出如下结论：

- 1) 带约束 BPSO-ACA 算法可有效识别事故严重程度影响因素，并显著提高关联规则生成效率，减少冗余无效规则，同时解决了经典关联规则算法需要人为确定最小支持度和置信度阈值的问题。
- 2) 高龄驾驶人（ ≥ 60 岁）和低龄驾驶人（ < 20 岁）在交叉口发生事故等级高于其他年龄段驾驶人。
- 3) 路面湿滑是重大等级交通事故的最核心致因；其次黑夜、畸形交叉口形态、机动车驾驶人分心驾驶和非机动车的影响也会使事故严重等级有加重倾向；车道数与交叉口事故等级也具有较大关联性。
- 4) 交叉口安全防控的目标应在避免特大和重大事故的基础上尽可能降低一般与轻微事故的数量，根据此原则并结合上述研究成果，建议：加强在雨/雪/雾等恶劣气象条件下畸形交叉口的交通管控，可增设视线诱导或夜间照明等设施；结合交叉口几何设计合理设置机非分隔设施，严格规范非机动车行为。此外，还建议加大分心驾驶行为的惩罚力度，加强对低龄机动车驾驶人的安全教育；要求高龄驾驶人须定期体检，建议驾驶能力较低的驾驶人使用辅助驾驶系统以提高行车安全，否则不建议其继续单独驾驶车辆。
- 5) 在后续研究中可深入探究不同规则评价指标对结果的影响，并研究分布式计算方法以应对规模更庞大的数据集。

参考文献

- [1] YAN X, ZHANG Y, LU M. The influence of in-vehicle speech warning timing on drivers' collision avoidance performance at signalized intersections [J]. Transportation Research Part C: Emerging Technologies, 2015, 51: 231-242.
- [2] 潘福全, 陆键, 项乔君, 张国强. 公路信号平面交叉口安全服务水平研究[J]. 东南大学学报(自然科学版), 2008(02): 298-303.

- PAN F Q, LU J, XIANG Q J, et al. Level of safety service for highway signalized intersections[J]. Journal of Southeast University: Natural Science Edition, 2008, 38(2): 298-303.
- [3]张圆,邓院昌,林庆丰,史晨军.不同性别驾驶员小轿车交通事故的影响因素分析[J].安全与环境工程,2019,26(3):166-170.
- ZHANG Y, DENG Y C, LIN Q F, et al. Analysis of influencing factors of car accidents based on drivers' gender [J]. Safety and Environmental Engineering, 2019, 26(3): 166-170.
- [4]汪帆,程慧婷,刘建蓓,等. 超高海拔区域驾驶人超车注视特性[J]. 中国公路学报,2018,31(4):10-17.
- WANG F, CHENG H T, LIU J B, et al. Eye gaze characteristics of overtaking process in very high altitude area [J]. China Journal of Highway and Transport, 2018, 31(4): 10-17.
- [5] LI Y, WANG F, KE H, et al. A driver's physiology sensor-based driving risk prediction method for lane-changing process using hidden markov model [J]. Sensors, 2019, 19(12): 2670.
- [6] AS A G. Using logistic regression to estimate the influence of accident factors on accident severity [J]. Accident Analysis and Prevention, 2002, 34(6): 729-741.
- [7] 何雅琴,段雨阳,王晨.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[J].安全与环境学报,2021,21(3):1165-1172.
- HE Y Q, DUAN Y Y, WANG C. Analysis of the pedestrian traffic accidents and countermeasures for heightening such safety based on the cumulative Logistic model [J]. Journal of Safety and Environment, 2021,21(3):1165-1172.
- [8] 马明,严新平,吴超仲,等. 信号交叉口交通事故频次显著影响因素的作用[J]. 吉林大学学报(工学版),2010,40(2):417-422.
- MA M, YAN X P, WU C Z, et al. Effects of significant factors on signalized intersection crash frequency [J]. Journal of Jilin University: Engineering and Technology Edition, 2010, 40(2): 417-422.
- [9] 秦雅琴,谢碧珊,杨文臣,等. 山区双车道公路摩托车事故严重度分析模型比较[J]. 安全与环境学报,2021,21(4):1397-1404.
- QIN Y Q, XIE B S, YANG W C, et al. Comparative study between an or-dered Logit model and the multinomial Logit model for analysis of motorcycle accidents on the mountainous two-lane roads [J]. Journal of Safety and Environment, 2021,21(4):1397-1404.
- [10]蔡明,程国柱,张素禄,等. 基于贝叶斯网的公路路侧事故风险评价[J]. 交通运输系统工程与信息,2019,19(2):202-208.
- CAI M, CHENG G Z, ZHANG S L, et al. Risk assessment of highway roadside accidents based on bayesian network [J]. Journal of Transportation Systems Engineering and Information Technology, 2019, 19(2): 202-208.
- [11] HARB R, YAN X, RADWAN E, et al. Exploring precrash maneuvers using classification trees and random forests [J]. Accident Analysis & Prevention, 2009, 41(1): 98-107.
- [12] KASHANI A T, MOHAYMANY A S. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models [J]. Safety science, 2011, 49(10): 1314-1320.
- [13] WENG J, ZHU J, YAN X, et al. Investigation of work zone crash casualty patterns using association rules [J]. Accident Analysis & Prevention, 2016, 92: 43-52.
- [14] 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 Y, LI Z M, FAN Y X. Correlation analysis of influencing factors of truck traffic accidents on expressways [J]. Safety and Environmental Engineering, 2020, 27(4): 180-188.
- [15]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 Z Z, LOU C, YANG Y. Analysis of freeway traffic crash causes under time differences [J]. Journal of Beijing Jiaotong University, 2021, 45(3): 1-7.
- [16]钟倩漪,钱谦,伏云发,等. 粒子群优化算法在关联规则挖掘中的研究综述[J]. 计算机科学与探索,2021,15(5):777-793.
- ZHONG Q Y, QIAN Q, FU Y F, et al. Survey of particle swarm optimization algorithm for association rule mining [J]. Journal of Frontiers of Computer Science & Technology, 2021, 15(5): 777-793.
- [17] TELI KA NI A, GANDOMI A H, SHAHBAHRAMI A. A survey of evolutionary computation for association rule mining [J]. Information Sciences: An International Journal, 2020, 524:318-352.
- [18]曾本冲,万旺根. 基于改进 PSO-Aprori 算法的恐怖组织关联特征分析[J]. 电子测量技术,2020,43(1):46-51.
- ZENG B C, WAN W G. Analysis of correlation features of terrorist organizations based on improved PSO-Apriori algorithm [J]. Electronic Measurement Technology, 2020, 43(1): 51-56.

[19]王琳,张永健,钟诗胜,等. 基于二进制粒子群优化的卫星典型件工艺知识挖掘[J]. 东北大学学报(自然科学版),2015,36(1):119-123.

WANG L, ZHANG Y J, ZHONG S S, et al. Binary particle swarm optimization based process knowledge mining for typical parts of satellite [J]. Journal of Northeastern University:Natural Science, 2015, 36(1): 119-123.

[20] YASAK Y, ESIYOK B. Anger amongst Turkish drivers: Driving Anger Scale and its adapted, long and short version [J]. Safety Science, 2009, 47(1): 138-144.

[21] 冯忠祥,刘强,张卫华,等. 女性驾驶人特性与行为机制研究现状及展望[J]. 中国安全生产科学技术,2016,12(4):124-130.

FENG Z X, LIU Q, ZHANG W H, et al. The current situations and futures of female drivers driving characteristics and behavior mechanisms [J]. Journal of Safety Science and Technology, 2016, 12(4): 124-130.

[22]李伟楠,王雪松.基于事故深度调查的城市交通事故特征分析[J].中国安全科学学报,2012,22(12):89-95.

LI W N, WANG X S. Urban crash features analysis based on in-depth investigation [J]. China Safety Science Journal, 2012, 22(12): 89-95.

An algorithm to mine the association rules among causes of accident severity at intersection

XU Jin-hua, LI Yan, ZHANG Yu-ting*

(College of Transportation Engineering, Chang'an University, Xi'an 710064, China)

Abstract: In order to access the causing factors of accidents severity at intersections, a constraint Binary Particle Swarm Optimization - Ant Colony Algorithm (BPSO-ACA) is established to acquire the association rules among the causing factors in the prospects of driver, environment, road, and vehicle. Firstly, the BPSO algorithm is used to determine two association rule generation thresholds, minimum support and minimum confidence. Secondly the ACA algorithm is applied to mine the frequent itemsets. Finally, dimensional constraints on the antecedent and consequent terms of the rule are added to remove invalid rules. The proposed algorithm makes use of heuristic information and the parallel computing features of the ant system to overcome the disadvantages of traditional algorithms, such as low mining efficiency and a large number of redundant rules caused by repeatedly scanning the database when searching frequent itemsets. The results using 2022 intersection accident data from a province in China entries indicate that the proposed BPSO-ACA algorithm can identity correlating potential factors in accident severity accurately. Compared with the improved Apriori algorithm, the proposed BPSO-ACA algorithm can avoid the generation of invalid rules and improve the efficiency of mining rules by 17%-29%, and the efficiency improvement is even better when the support level is low. Driver age, distraction status, intersection geometry, lane number and weather condition are strongly related with intersection accident severity. Limitation of lane number was a critical factor in the transformations of severity that from slight to ordinary. Wet pavement condition is a key factor in escalating accidents at intersections into severe accidents. Especially in the adverse weather conditions such as rain, snow and fog, the possibility of severe accidents is the highest for motor or non-motor vehicle accident. The results can provide a theoretical guidance to traffic management departments for active prevention, which can reduce the accident number and severity at intersections.

Keywords: safety engineering; accident severity; association rules; cause analysis; BPSO; ACA; intersection

CLC Number: X951 **Document code:** A