

基于关联规则挖掘的路面损坏状况影响因素分析^①



曹磊, 徐磊, 杨菲, 贾彭斐

(长安大学 信息工程学院, 西安 710064)

通讯作者: 曹磊, E-mail: 973674829@qq.com

摘要: 本文基于路面评价指标中车辙深度指数和行驶质量指数来评价路面的损坏情况, 使用关联规则挖掘环境、交通、路面等影响因素与路面状况之间的关联程度. 针对关联规则 Apriori 算法复杂度和耗时的缺点, 提出一种不生成候选集的方法来产生频繁集的改进 Apriori 算法, 并通过实验对比证明改进的 Apriori 算法能够有效提升速度和性能. 使用改进的 Apriori 算法分析路面评价指标及其影响因素之间的强关联规则, 得到不同环境路面损坏的主要成因. 本文结论能够对路面养护提供科学可靠的支持, 可为路面养护部门提供合理的养护建议与数据支撑.

关键词: 数据挖掘; 关联规则; 改进 Apriori 算法; 路面损坏分析

引用格式: 曹磊, 徐磊, 杨菲, 贾彭斐. 基于关联规则挖掘的路面损坏状况影响因素分析. 计算机系统应用, 2021, 30(1): 186-193. <http://www.c-s-a.org.cn/1003-3254/7754.html>

Influencing Factors Analysis of Pavement Damage Based on Mining Association Rules

CAO Lei, XU Lei, YANG Fei, JIA Peng-Fei

(School of Information Engineering, Chang'an University, Xi'an 710064, China)

Abstract: Based on the rutting depth index and driving quality index in the pavement evaluation index, the pavement damage was evaluated in this study. The association rules were used to mine the degree of association between influencing factors such as environment, traffic, and road surface and road surface conditions. Aiming at the shortcomings of the complexity and time-consuming of the association rule Apriori algorithm, an improved Apriori algorithm that does not generate candidate sets to generate frequent sets was proposed. The experiments show that the improved Apriori algorithm can effectively improve the speed and performance. The improved Apriori algorithm was used to analyze the strong association rules between evaluation indexes and influencing factors, and the main causes of pavement damage in different environments were obtained. The conclusion of this paper can provide scientific and reliable support for the pavement maintenance, reasonable maintenance suggestions, and data support for the pavement maintenance department.

Key words: data mining; association rules; improved Apriori algorithm; pavement damage analysis

随着我国交通行业的发展, 公路里程数和线网复杂度也在不断提升, 越来越多的道路出现裂缝、坑槽、车辙等各种损坏情况, 同时高温、降雨等各种恶劣天气对路面的性能的影响也极其严重, 路面的养护管

理工作难度也在不断提升, 在这种情况下路面的科学养护和预防养护就显得尤为重要.

目前关于路面损坏及其影响因素的研究主要集中于公路领域^[1], 而针对两者之间综合关系的研究则较

① 基金项目: 陕西省交通运输厅 2018 年度交通科研项目 (18-31X); 长安大学中央高校基本科研业务费专项资金 (300102249102)

Foundation item: Transportation Scientific Research Project of Transportation Department of Shaanxi Province in Year 2018 (18-31X); The Fundamental Research Funds for the Central Universities of China of Chang'an University (300102249102)

收稿时间: 2020-06-05; 修改时间: 2020-06-30; 采用时间: 2020-07-10; csa 在线出版时间: 2020-12-31

少^[2]. 传统的路面损坏情况影响因素的研究的主要方法有经验分析、实地调研、数据统计等. 这些分析方法主要基于当前的路面状况, 分析过程简单、直接, 因此所得到的路面损坏及其影响因素之间的关系往往单一、独立, 所涉及的影响因素较少, 分析维度较低, 无法客观全面地认识路面综合损坏情况.

部分学者尝试借鉴其他学科领域的方法, 对路面损坏影响因素关系进行研究. 数据挖掘关联分析是数据挖掘的一个重要分支, 关联规则分析能够找出数据之间的潜在联系, 在不同的维度对数据进行全方位的分析, 从而能够更好地做出更加理想的决策, 能够更加准确的预测事物未来的发展趋势等^[3]. 文献[4]采用关联规则分析通过大量数据集中客观地发现了路面病害成因. 文献[5]采用关联分析技术对水泥混凝土路面病害关系进行了研究, 得到了填缝料损坏、板底脱空、错台、断板之间的关联规则, 并得出了它们之间的转化关系.

在此基础上, 本文对关联规则 Apriori 算法进行了改进以提升速度和性能. 基于大量的路面数据, 通过挖掘环境因素、交通因素、路面因素和路面破损情况之间存在的内在联系, 分析不同环境下路面破损的主要影响因素, 进而对路面养护提供科学可靠的支持, 为路面养护部门提供合理的养护建议与数据支撑.

1 路面状况影响因素及其评价指标

1.1 路面状况的影响因素分析

路面状况错综复杂, 影响路面状况的因素也是多方面的. 路面在结构上可分为面层、基层和垫层. 通常情况下, 路面受损坏的情况会随着路面深度的增加而逐渐减弱. 面层直接受到降水、风雪、强光、高温、行车荷载等的直接或间接反复作用, 因此本文主要研究面层损坏情况与影响因素之间的关联性. 在诸多影响因素中, 环境因素、路面因素和交通因素为主要影响因素.

环境因素主要包含温度、湿度、降水量、蒸发量等气候条件, 当环境因素变化较大时, 路表面的物理特性也会发生变化, 如温度变化会引发热膨胀系数的变化而导致路面发生开裂或老化现象.

路面因素主要由路面的类型和材料来决定. 由于路面的材料及特性差异, 即使是相同环境因素的影响下, 不同类型路面发生病害的类型和严重程度也不相同.

交通因素主要指路面的交通流量, 城市和乡村的交通流量差距很大, 损坏程度也有很大差别, 因此交通因素也是本文考虑的一个重要的影响因素.

1.2 路面状况的评价指标

在诸多影响因素影响下, 路面破损情况的衡量指标对这些影响因素关键与否起着至关重要的作用, 依据交通运输部发布的《公路技术状况的评定标准》, 路面质量指数 (Pavement Quality Index, *PQI*) 主要由路面损坏指数 (*PCI*)、行驶质量指数 (*RQI*)、路面车辙深度指数 (*RDI*)、路面抗滑性能指数 (*SRI*) 来评价:

$$PQI = w_{PCI}PCI + w_{RQI}RQI + w_{RDI}RDI + w_{SRI}SRI \quad (1)$$

其中, w_{PCI} 、 w_{RQI} 、 w_{RDI} 和 w_{SRI} 为各项指标的参数, 对于沥青路面, 这4个参数值分别为0.35, 0.4, 0.15, 0.1.

式(1)的4个参数中, 路面损坏指数 (*PCI*) 和路面行驶质量指数 (*RQI*) 主要是通过路面平整度 (*IRI*) 来衡量, 且 *IRI* 与 *PCI* 具有正相关的关系. 而由于 *SRI* 的测量方式复杂且权重较小, 本文将不作考虑. 因此通过建立环境、路面、交通等影响因素与路面车辙 *RDI* 和平整度 *IRI* 之间的关联规则即可得到这些影响因素与路面损坏情况之间的关联性.

2 关联规则及 Apriori 算法改进

2.1 关联规则相关概念

关联规则的形式: 两个不相交的非空集合 X 、 Y , 如果有 $X \Rightarrow Y$, 那么 $X \Rightarrow Y$ 就是一条关联规则. 以下是一些比较重要的概念:

项目: 数据库中的一个字段, 以一次事务来看如 $\{A, B, C\}$, 其中每一个事物 A 、 B 、 C 都分别是一个项目.

事务: 事务本身也是一个项集, 但是该项集包含了一次交易记录中发生的所有项目的集合.

项集: 指在一次事务中, 由零个或多个项目组成的集合, 一般会大于零个, 通常我们将由 m 个项目组成的集合称为 m -项集, 如 $\{A, B\}$ 就是一个二项集.

支持度: 项集 $\{X, Y\}$ 在总项集中所占的百分比, 即在总项集中项集 $\{X, Y\}$ 出现的次数 $\text{count}(X \cup Y)$ 与整个数据集中的项集总数 $|D|$ 之比 $\text{support}(X \Rightarrow Y) = \text{count}(X \cup Y) / |D|$, 支持度能够反映一个项集出现的概率.

频繁项集: 支持度大于或等于设定的最小支持度的项集.

置信度: 在已知项集 X 存在并发生的条件下, 项集 Y 发生的概率, 即项集 $\{X, Y\}$ 数量与项集 $\{X\}$ 数量的

比值 $\text{confidence}(X \Rightarrow Y) = \text{support}(X \Rightarrow Y) / \text{support}(X)$

强关联规则: 支持度与置信度均大于或者等于设定的最小支持度和最小置信度的关联规则就是强关联规则。

关联规则挖掘过程主要包含两个阶段:

第1阶段: 在数据集中找出所有的频繁项集;

第2阶段: 由频繁项集产生强关联规则。

关联规则挖掘的第1阶段要求产生项集的支持度必须大于设定的最小支持度, 其中项集的个数可以从1到K递推, 得到频繁K项集; 第2阶段中这些频繁项集对应的关联规则的置信度必须大于设定的最小置信度, 才能形成强关联规则。

2.2 Apriori 算法

Apriori 算法主要是通过扫描数据库产生候选项集, 然后通过对候选项集的剪枝、筛选等操作得到频繁数据集最终得到强关联规则的一种算法。该算法的主要关键点是产生频繁项集即由 $k-1$ 项集到生成 k 项集。该算法的流程如图1所示。

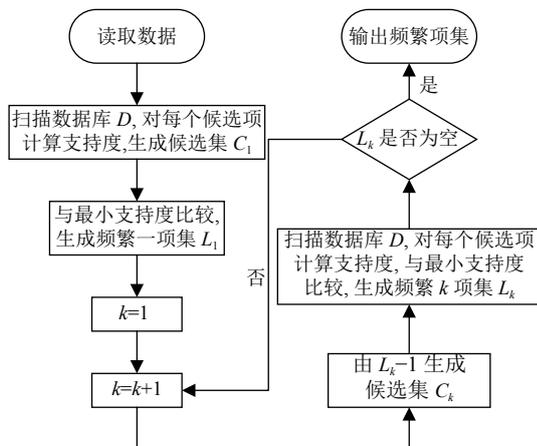


图1 Apriori 算法流程图

Apriori 算法根据自身定理性质, 在产生候选项集时, 减少了频繁项集的个数, 节约了一定的时间, 取得了很好的性能。但还存在两大缺点:

(1) 会产生大量的候选项集;

(2) 在产生频繁项集时, 都需要不断的扫描事务数据库, 因此比较耗时。

如当频繁一项集 L_1 有 1000 个时, 候选二项集 C_2 个数将会达到几十万。而在执行每一步时都需要不断的扫描数据集, 这样会使得 Apriori 算法的效率大大降低, 针对这个问题, 本文对该算法进行了改进。

2.3 改进的 Apriori 算法

对于关联规则挖掘来说, 其目的就是在数据项目中挖掘出频繁多项集, 原 Apriori 算法中通过对候选项集的操作得到频繁项集。而当项目较多时, Apriori 算法在产生候选项集时需要耗费大量的时间, 从而影响算法的效率, 在产生候选二项集时这种情况尤为明显, 因此本文采用一种不生成候选二项集的方法来产生二项集, 减少大量的候选集扫描时间^[6], 因此可以极大地提高算法的效率。

设初始数据集为 D , 包含 $T_k (k=1, 2, \dots, n)$, $T_k = (t_1, t_2, t_3, \dots)$, 即每条记录 T_k 由很多项目 t_i 构成。每个流程中事务集为 $TID (1, \dots, t)$, 最小支持度设为 $\text{min_sup} = s\%$, 且事务数不小于 $p (p = s\% * n)$ 次, Num 为 TID 中的元素个数。

改进算法的具体步骤如算法1^[7-10]。

算法1. 改进的 Apriori 算法

- 1) 重新构造数据库:
 - 对数据记录进行扫描,
 - 搜索出包含 t_1 的事务, 并生成 a_1 ;
 - 搜索出包含 t_2 的事务, 生成 a_2 ;
 -
 - 搜索出包含 t_k 的事务, 生成 a_k ;
- 2) 生成新格式的数据库为 $A(a_1 - a_k)$, 将 Num 小于 p 的项去掉以实现数据库的简化;
- 3) 计算 a_1 和 a_2 的交集, 若交集项目数小于 p 则舍去。以此类推分别计算 a_1 和 a_3, a_1 和 a_4, \dots, a_1 和 a_k 的交集; a_2 和 a_3, a_2 和 a_4, \dots, a_2 和 a_k 的交集; $\dots a_{k-1}$ 和 a_k 的交集, 保留的事务生成新的事务集 $B(b_1 - b_k)$;
- 4) 计算出与存储的事务集对应的项目集 $B_i \cup B_j$, 并存储为新的项目集;
- 5) 将新项目集整理即可得频繁二项集, 重复 3), 4) 两步, 再去掉重复的事务集和项目集, 就可得到频繁三项集与频繁四项集。
- 6) 检查结果, 重复上述步骤, 得到更多的频繁项集。

2.4 改进的 Apriori 算法与原算法对比实验

(1) 实验 1

对同一组数据分别用原算法和改进算法在不同的最小支持度下计算其算法用时, 其中最小支持度分别为 0.1、0.2、0.3, 同时保持置信度为 0.6。此次使用的数据规格是 100 条记录, 4 个属性。得到的结果如图 2 所示。

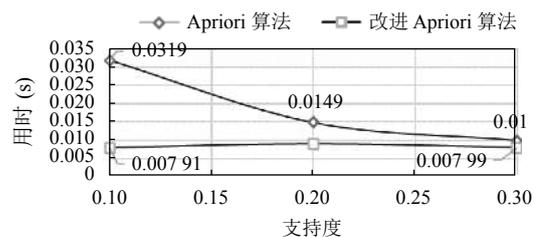


图2 原算法与改进算法不同支持度下用时对比图

从这个实验可以看出改进的算法在不同支持度下性能都比较稳定,而原算法只有在支持度比较高时性能才稍好;同时改进算法的每次用时都比原算法少一个数量级.

(2) 实验 2

本实验分别使用原算法和改进算法在不同数据量下计算其用时.此次使用的数据规格分别是:10条记录,4个属性;100条记录,4个属性;200条记录,4个属性,保证最小支持度为0.1,最小置信度为0.6;两种算法用时如图3所示.

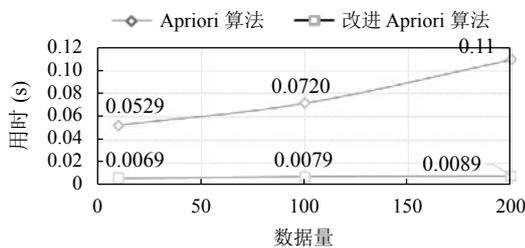


图3 原算法与改进算法不同数据量用时对比图

从这个实验可以看出,在数据量增加的情况下,原算法用时增加比较明显,而改进算法的用时基本没有变化;同时改进算法的每次用时都比原算法少一个数量级.

综合两个实验可以看出,本文提出的改进 Apriori 算法在复杂度和用时上均优于原算法,在不同的支持度和数据量下均能保持良好性能.

3 数据及离散化处理

3.1 路面数据来源 LTPP

路面长期使用性能研究计划(LTPP)是北美知名的公路研究计划之一,它对世界各国道路质量的提升和路面的研究提供的数据支持.研究人员使用各种路面类型的标准数据收集程序和协议来收集路面性能数据.本文数据就是来源于LTPP数据库.

在LTPP数据库中提取并进行清洗处理得到州代号为50的州2019年的每日车辙深度、温度和降水量的数据再以月份为单位进行平均得到月度数据,如表1所示.

在数据库中提取并清洗得到同一地方2008年到2019年4380天的车流量、平均温度和降水量的数据,然后以年为单位对每年的数据进行平均得到路面平整度和其影响因素(年平均日交通量(AADTT)和年平均温度以及年总降雨量)的数据如表2.

表1 车辙深度与影响因素数据

年份	月份	车辙深度(cm)	平均温度(度)	降水量(mm)
2019	1	0.92	2.8	64.8
2019	2	0.84	14.8	192.5
2019	3	1.31	14.2	201.3
2019	4	12	15.8	131.8
2019	5	4.46	20.4	65.4
2019	6	10.27	2.6	145.1
2019	7	4.15	25.8	179.6
2019	8	0.97	26.1	154.1
2019	9	0.67	22.8	218.9
2019	10	0.84	11.6	83.5
2019	11	5.23	12.4	58
2019	12	3.97	3.9	128.9

表2 路面平整度及其影响因素

年份	路面平整度(mm)	AADTT(辆)	年平均温度(°C)	年总降雨量(mm)
2008	0.719	1300	17.3	800
2009	0.717	1440	17.4	582
2010	0.721	1430	16.8	478.7
2011	0.718	1390	17.2	412
2012	0.537	720	15.6	437.4
2013	0.539	760	15.6	446.1
2014	1.008	2440	16.3	870.9
2015	1.113	2630	16.1	595.9
2016	0.863	1430	18	630.8
2017	0.856	1550	19.4	1095
2018	0.538	540	18.1	894.7
2019	0.567	810	17.4	1003.2

3.2 数据离散化

由于本文使用的数据数据量较大且复杂,为了简化数据结构通常需要使用数据离散化技术对数据进行离散化处理.连续数据的离散化就是将数据值域划分为若干个离散的区间,之后用不同的符号代表落在每个子区间中的属性值.数据挖掘中使用离散化后的数据得到的结果更加的简便、紧凑、准确,也更加方便被我们使用和重用^[11-13].本文使用K-means算法对数据进行离散化处理,处理流程如图4.

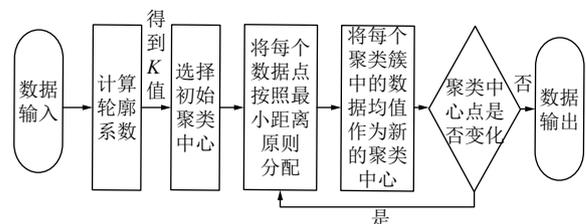


图4 K-means 算法离散化数据流程

对于一个数据点 X_i 的轮廓系数计算方法:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2)$$

其中, $S(i)$ 为轮廓系数, $a(i)$ 表示数据点 X_i 到同一簇内其他数据点不相似程度的平均值, $b(i)$ 表示数据点 X_i 到其他簇的平均不相似程度的最小值, $a(i)$ 、 $b(i)$ 的计算方法:

$a(i)=\text{average}(\text{数据点 } X_i \text{ 到所有它属于的簇中其它点的距离})$

$b(i)=\min(\text{数据点 } X_i \text{ 到所有其他簇的点的平均距离})$

轮廓系数越大表示数据聚类的效果越好。

以下以车辙深度数据以及其影响因素的离散化为例, 对车辙深度数据以及影响因素计算轮廓系数结果如图 5。

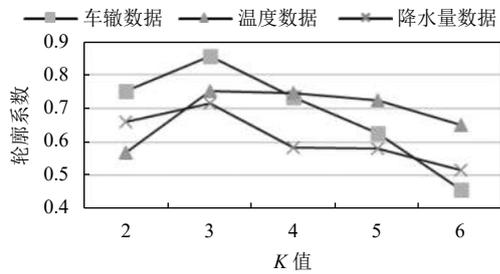


图 5 不同 K 值下车辙深度及其影响因素轮廓系数

通过对比可得, 对于这 3 种数据, 在 $K=3$ 时轮廓系数均达到最大值, 此时每种数据的聚类效果达到最佳。对车辙深度数据基于参数 $K=3$ 进行离散化处理结果如图 6。

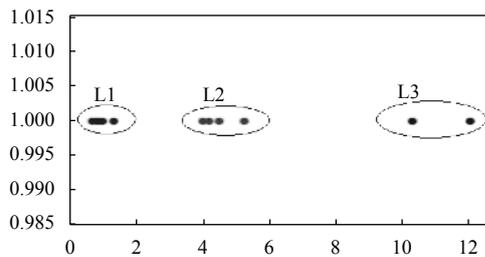


图 6 车辙深度数据离散化结果图

由车辙数据离散化结果图可将车辙深度数据分为 3 类。用 L1、L2 和 L3 表示第 1 类、第 2 类和第 3 类。L1 的聚类中心为 $[0.925, 1]$, L2 的聚类中心为 $[4.45, 1]$, L3 的聚类中心为 $[11.13, 1]$, 3 个等级的案例数分别为 6, 4, 2。

对车辙深度数据基于参数 $K=3$ 进行离散化处理, 结果如图 7。

由平均温度数据离散化结果图可将平均温度数据分为 3 类。用 T1 和 T2 和 T3 分别表示第 1 类、第 2

类和第 3 类。T1 的聚类中心为 $[3.1, 1]$, T2 的聚类中心为 $[13.76, 1]$, T3 的聚类中心为 $[23.775, 1]$, 3 个等级的案例数分别为 3, 5, 4。

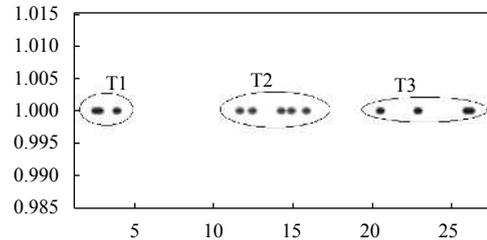


图 7 平均温度数据离散化结果图

对降水量数据基于参数 $K=3$ 进行离散化处理, 结果如图 8。

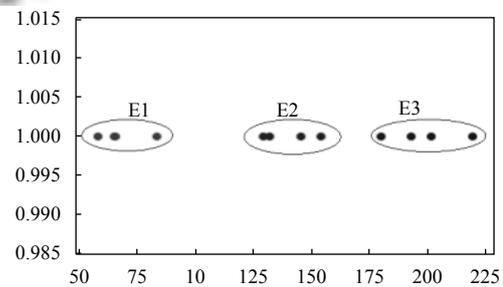


图 8 降水量数据离散化结果图

由降水量离散化结果图可将降水量数据分为 3 类。用 E1、E2、E3 分别表示第 1 类、第 2 类和第 3 类。E1 的聚类中心为 $[67.93, 1]$, E2 的聚类中心为 $[139.98, 1]$, E3 的聚类中心为 $[198.08, 1]$, 3 个等级的案例数分别为 4, 4, 4。

将以上 3 种数据的离散化结果整理如表 3。

表 3 车辙深度数据及其影响因素离散化汇总表

年份	月份	车辙深度	平均温度	降水量
2019	1	L1	T1	E1
2019	2	L1	T2	E3
2019	3	L1	T2	E3
2019	4	L3	T2	E2
2019	5	L2	T3	E1
2019	6	L3	T1	E2
2019	7	L2	T3	E3
2019	8	L1	T3	E2
2019	9	L1	T3	E3
2019	10	L1	T2	E1
2019	11	L2	T2	E1
2019	12	L2	T1	E2

同理使用同样的方法分别对平整度及其影响因素进行 K 值确定和离散化处理。将年平均日交通量分为 3 个等级 M1、M2、M3, 将路面平整度分为 4 个等级

N1、N2、N3、N4, 将年平均温度分为 5 个等级 P1、P2、P3、P4、P5, 将年总降水量分为 4 个等级 O1、O2、O3、O4, 其中等级越大表示程度越高或越严重, 可以得到表 4.

表 4 平整度及其影响因素离散化汇总表

STATE_CODE	YEAR	路面平整度	AADTT	年平均温度(°C)	年总降雨量(mm)
50	2008	N2	M2	P3	O3
50	2009	N2	M2	P3	O2
50	2010	N2	M2	P3	O1
50	2011	N2	M2	P3	O1
50	2012	N1	M1	P1	O1
50	2013	N1	M1	P1	O1
50	2014	N4	M3	P2	O3
50	2015	N4	M3	P2	O2
50	2016	N3	M2	P4	O2
50	2017	N3	M2	P5	O4
50	2018	N1	M1	P4	O3
50	2019	N1	M1	P3	O4

4 基于改进 Apriori 算法的路面状况影响因素关联规则挖掘

4.1 环境因素与路面车辙深度关联规则挖掘

利用改进的 Apriori 算法取最小支持度 *min-sup* 为 10%, 最小置信度 *min-con* 为 60%, 首先进行环境因素数据和路面车辙深度数据的关联规则挖掘^[4]. 使用改进的 Apriori 算法处理离散化数据可以得到各个频繁项集和其对应的支持度和置信度如表 5.

表 5 各个频繁项集和其对应的支持度

项集类型	项集	支持度
频繁一项集	L2	0.3333
	L3	0.1667
	E2	0.3333
	E3	0.3333
	T3	0.3333
	T2	0.4167
	L1	0.5
	T1	0.25
	E1	0.3333
	频繁二项集	{E2, L3}
{T2, L1}		0.25
{T3, L2}		0.1667
{E3, T2}		0.1667
{E1, L1}		0.1667
{E1, L2}		0.1667
{E2, T1}		0.1667
{E1, T2}		0.1667
{T3, E3}		0.1667
{E3, L1}		0.25
频繁三项集	{T3, L1}	0.1667
	{E3, T2, L1}	0.1667

强关联规则如表 6.

将表 6 结果转化成柱状图如图 9.

表 6 置信度大于最小置信度的强关联规则

关联规则	置信度
L3->E2	1.0
T2->L1	0.6
T1->E2	0.6667
E3->L1	0.75
{T2, L1}->E3	0.6667
{E3, L1}->T2	0.6667
{E3, T2}->L1	1.0

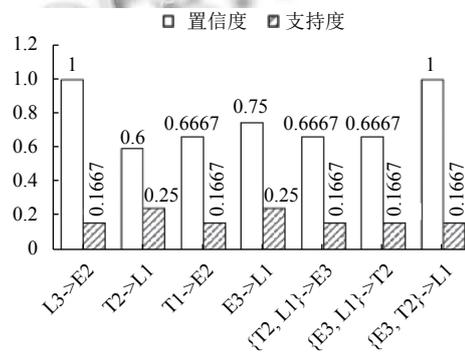


图 9 强关联规则支持度和置信度柱状图

针对上述实验结果. 在本文的研究中路面状况是因变量, 而影响因素则是对目标造成影响自变量. 所以在环境影响因素与路面车辙的关联规则挖掘中, 应该以车辙的深度作为分析的目标函数, 在上述支持度和置信度都满足条件的强关联规则中筛选出以车辙深度作为目标的关联规则如表 7.

表 7 强关联规则

强关联规则(A->B)	置信度	支持度
T2->L1	0.6	0.25
E3->L1	0.75	0.25
E3, T2->L1	1.0	0.166

针对上述 3 则强关联规则, 根据前文规定的车辙等级: 等级 1 为轻微的车辙问题, 等级 2 为一般的车辙问题, 等级 3 为严重的车辙问题. 对于温度等级 1 为低温, 等级 2 为温度一般, 等级 3 为高温情况; 对于降水量等级一为降水量稀少, 等级 2 为降水量一般, 等级 3 为降水量严重.

分析以上强关联规则, 可以得出如下结论:

规则 1. 当某地降温度为 T2 等级即处于 [11.6, 15.8] 范围时相对容易出现 L1 轻微情况的车辙问题.

规则 2. 当某地降水量为 E3 等级即处于 [179.6, 218.9] 时, 路面容易也出现轻微的车辙问题.

规则 3. 当某地降水量为 E3 等级即处于 [179.6, 218.9] 时, 同时温度为 T2 等级即处于 [11.6, 15.8] 时, 更容易出现 L1 等级的轻微车辙问题. 也就是当降水量为 E3 时, 轻微的车辙问题更容易在温度为 [11.6, 15.8] 之间发生. 或者当温度为 T2 等级即 [11.6, 15.8] 时, 轻微的车辙问题更容易在降水量为 [179.6, 218.9] 的地方发生.

4.2 环境交通因素对路面平整度关联规则挖掘

利用改进的 Apriori 算法取最小支持度 $min-sup$ 为 20%, 最小置信度 $min-con$ 为 60%, 进行影响因素和路面平整度数据的关联规则挖掘. 得到各个频繁项集和其对应的支持度和置信度如表 8.

表 8 频繁多项集

多项集类型	多项集	支持度
频繁一项集	M1	0.3333
	O2	0.25
	O1	0.3333
	M2	0.5
	P3	0.4167
	O3	0.25
频繁二项集	N1	0.3333
	N2	0.3333
	{N1, M1}	0.3333
	{P3, N2}	0.3333
频繁三项集	{P3, M2}	0.3333
	{N2, M2}	0.3333
	{P3, N2, M2}	0.3333

其中的强关联规则如表 9.

将表 9 结果转化成柱状图如图 10.

表 9 满足最小支持度的关联规则及其置信度

规则	置信度
N1→M1	1.0
M1→N1	1.0
N2→P3	1.0
P3→N2	0.8000
P3→M2	0.8000
M2→P3	0.6667
N2→M2	1.0
M2→N2	0.6667
{P3, N2}→M2	1.0
{N2, M2}→P3	1.0
{P3, M2}→N2	1.0
M2→{P3, N2}	0.6667
N2→{P3, M2}	1.0
P3→{N2, M2}	0.8000

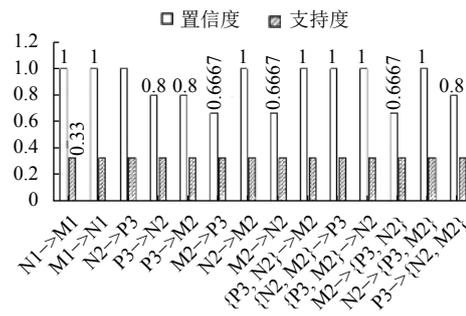


图 10 强关联规则支持度和置信度柱状图

根据因变量为 N 可以筛选出强关联规则如表 10.

表 10 强关联规则及其置信度和支持度

规则	强关联规则	置信度	支持度
1	M1→N1	1.0	0.333
2	P3→N1	0.799	0.333
3	M2→N2	0.666	0.333
4	P3, M2→N2	1.0	0.333

规则 1 和规则 2 是关于 N1 的强关联规则. 根据前文等级设定, 规则 1 说明当年平均日交通量为 M1(辆)等级即处于 [540, 810] 之间时容易出现 N1 等级轻微情况的路面平整度问题. 规则 2 说明当年平均温度 (°C) 为 P3 等级即处于 [16.8, 17.4] 相对较低的温度时, 路面易出现轻微的平整度问题.

规则 3 和规则 4 是关于两条关于 N2 的关联规则, 规则 3 说明当年平均日交通量 (辆) 为 M2 即处于 [1300, 1550] 相对较高时, 路面更容易出现 N2 等级的一般的平整度问题. 规则 4 说明在年平均温度 (°C) 为 P3 等级即处于 [16.8, 17.4] 相对较低的温度时, 在平均日交通量 (辆) 为 M2 即处于 [1300, 1550] 的路面更易出现一般的平整度问题.

4.3 道路养护管理建议

根据以上分析结果, 结合我国现状提出几点合理的养护管理建议:

1) 在对于不同气候条件的地方可以采取不同的措施来预防养护路面的破损情况. 如当气温较高时可以适当地洒水降温来降低路面温度, 降水量较大时可以适当地去帮助排水.

2) 加强对交通流量大的路段的日常保养, 及时修复损坏部分, 使公路及其沿线设施的各部分均保持完好、整洁、美观, 保障行车安全、舒适、畅通.

3) 采取正确的工程技术措施, 周期性地进行中修, 延长公路的服役年限, 以节省资金.

4) 对原标准过低或留有缺陷的路线、构造物、路面结构、沿线设施进行改善和补建, 逐步提高公路的使用质量、服务水平和抗灾能力。

5) 无论是关于道路修补的原材料, 还是施工的技术都应加强监督力度, 同时对于路面的检测验收也要严格采用国际统一标准。

5 总结

本文首先探讨了路线损坏情况的主要影响因素, 针对 Apriori 算法耗时及复杂度高的缺点, 提出一种不产生候选集的方法来产生频繁集的改进 Apriori 算法, 通过对比实验证明, 改进的算法在不同情况下都能保持良好的性能。之后使用改进的算法对路面损坏情况指标及其影响因素进行关联分析, 得到之间的强关联规则, 基于强关联规则得出不同环境下路面损坏情况的主要影响因素。相比于传统的分析方法, 本文使用的关联规则算法不仅能够对路面状况的影响因素进行定性分析而且能够定量地指出在不同情况下路面出现不同损坏情况的可能性, 能够对路面养护提供科学可靠的支持, 可为路面养护部门提供合理的养护建议与数据支撑。

参考文献

- 1 尹蕊, 王治, 王连伟, 等. 公路路面病害成因分析及防治措施. 河南建材, 2018, (5): 73-76, 81. [doi: 10.3969/j.issn.1008-9772.2018.05.042]
- 2 王萌. 高寒地区机场道面病害及其防治研究 [硕士学位论文]. 广汉: 中国民用航空飞行学院, 2018.
- 3 牛猛. 关联规则的基本研究. 河北工程大学学报(社会科学版), 2016, 33(2): 114-117.
- 4 张绍阳, 马玉兰, 王选仓. 基于关联分析的路面病害成因确定方法. 中国公路学报, 2008, 21(2): 98-103. [doi: 10.3321/j.issn:1001-7372.2008.02.019]
- 5 李丽苹. 基于预防性养护概念的水泥砼路面板底脱空处治技术研究 [硕士学位论文]. 南京: 东南大学, 2015.
- 6 陈井霞. 基于矩阵的关联规则挖掘改进算法. 电子技术与软件工程, 2014, (6): 210.
- 7 Zhang CB, Zhao Y, Li TT, *et al.* A post mining method for extracting value from massive amounts of building operation data. *Energy and Buildings*, 2020, 223: 110096. [doi: 10.1016/j.enbuild.2020.110096]
- 8 Huang CX, Huang X, Fang Y, *et al.* Sample imbalance disease classification model based on association rule feature selection. *Pattern Recognition Letters*, 2020, 133: 280-286. [doi: 10.1016/j.patrec.2020.03.016]
- 9 Liu JY, Shi DL, Li GN, *et al.* Data-driven and association rule mining-based fault diagnosis and action mechanism analysis for building chillers. *Energy and Buildings*, 2020, 216: 109957. [doi: 10.1016/j.enbuild.2020.109957]
- 10 Nie MX, Li YL. Application of association rules in analysis of pavement performance attenuation factor. *Proceedings of the 2019 2nd International Conference on Mechanical, Electronic and Engineering Technology*. Xi'an, China. 2019. 506-512.
- 11 邱欣, 洪皓珏, 杨青, 等. 基于 APRIORI-GBDT 算法的沥青路面路表温度预测. 公路交通科技, 2019, 36(5): 1-10, 19.
- 12 梁斌. 路面损坏计算方法和评定标准研究. 北方交通, 2020, (5): 71-74.
- 13 龙小勇, 蔡良才, 沈勇, 等. 基于关联规则挖掘的水泥混凝土道面综合病害关系研究. 西华大学学报(自然科学版), 2019, 38(5): 1-11. [doi: 10.3969/j.issn.1673-159X.2019.05.001]
- 14 Gopalakrishnan K, Agrawal A, Ceylan H, *et al.* Knowledge discovery and data mining in pavement inverse analysis. *Transport*, 2013, 28(1): 1-10. [doi: 10.3846/16484142.2013.777941]