Hans Journal of Data Mining 数据挖掘, 2027, 17(1), 1-9
Published Online January 2027 in Hans. https://www.hanspub.org/journal/hjdm
https://doi.org/10.12677/hjdm.2027.171001

Hans汉斯

城市快速路横断面案例的聚类识别与关联规则挖掘研究

朱　敏[1,2]，许璐瑶[1]，张财喜[3]

[1]萍乡学院工程与管理学院，江西　萍乡
[2]萍乡市海绵城市工程技术研究中心，江西　萍乡
[3]萍乡市工程咨询管理顾问有限责任公司，江西　萍乡

收稿日期：2026年10月2日；录用日期：2026年10月26日；发布日期：2026年11月2日

摘　要

针对城市快速路更新案例信息分散、工程规律难以提取的问题，本文对横断面案例进行聚类和关联规则挖掘，并利用快速路交通数据验证车道配置规律。结果表明，横断面案例可划分为5类典型模式，综合轮廓系数为0.534；地面辅道与非机动车道、高架主线和25~35 m横断面宽度之间存在较稳定的关联关系；交通量增加时，8车道和10车道方案的比例明显提高。研究结果可为快速路更新案例整理和初步方案分析提供参考。

关键词

城市快速路更新，横断面案例，聚类分析，关联规则，工程知识发现，数据挖掘

Clustering Identification and Association Rule Mining of Urban Expressway Cross-Section Cases

Min Zhu[1,2], Luyao Xu[1], Caixi Zhang[3]

[1]School of Engineering and Management, Pingxiang University, Pingxiang Jiangxi
[2]Pingxiang Engineering Technology Research Center for Sponge City, Pingxiang Jiangxi
[3]Pingxiang Engineering Consulting & Management Co., Ltd., Pingxiang Jiangxi

Received: October 2, 2026; accepted: October 26, 2026; published: November 2, 2026

文章引用: 朱敏, 许璐瑶, 张财喜. 城市快速路横断面案例的聚类识别与关联规则挖掘研究[J]. 数据挖掘, 2027, 17(1): 1-9. DOI: 10.12677/hjdm.2027.171001

Abstract

Urban expressway renewal cases are characterized by scattered information and difficulty in extracting engineering regularities. To address these issues, this study conducts clustering and association rule mining on cross-section cases and validates lane configuration patterns using expressway traffic data. The results show that the cross-section cases can be classified into five typical patterns, with an overall silhouette coefficient of 0.534. Relatively stable associations are identified among ground-level frontage roads, non-motorized vehicle lanes, elevated mainlines, and cross-section widths of 25~35 m. In addition, as traffic volume increases, the proportions of 8-lane and 10-lane schemes increase significantly. These findings provide a reference for the systematic compilation of expressway renewal cases and for preliminary scheme analysis.

Keywords

Urban Expressway Renewal, Cross-Section Cases, Cluster Analysis, Association Rules, Engineering Knowledge Discovery, Data Mining

Copyright © 2027 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

城市快速路是城市综合交通系统的重要组成部分。随着既有道路设施逐渐老化、交通需求持续增长以及建成区空间约束不断增强，快速路更新已经成为城市更新和交通基础设施优化的重要内容。与新建道路相比，快速路更新通常需要在既有建成环境、规划红线和现状道路条件下进行，设计人员需要同时考虑交通功能、道路等级、车道规模、空间宽度、慢行设施和周边用地等因素[1]，快速路建设形式通常包括地面、高架和地下等类型[2]。在实际工程中，快速路更新方案形成于可行性研究报告、规划论证文件、工程设计图纸和专家咨询意见之中。这些资料积累了大量具有参考价值的历史案例，但相关信息通常以文本描述、工程参数和横断面图纸等形式分散保存。设计人员开展新项目研究时，往往依靠个人经验查找相似案例并进行人工比较，历史案例中道路结构和功能设施之间的组合规律难以系统提取。

数据挖掘为工程案例中的经验发现提供了新的技术路径。聚类分析能够根据对象的属性相似性识别潜在群组，适合用于发现工程案例中的典型模式。关联规则挖掘能够从大量记录中提取特征共现关系，常用于发现项目属性之间的组合规律。Agrawal 等提出的关联规则方法为从事务型数据中识别高频项集和规则关系提供了基础[3]。针对同时含有数值属性和类别属性的数据，Huang 提出的 k-prototypes 方法扩展了 k-means 对混合型数据的处理能力[4]。轮廓系数则可用于比较不同聚类数下的聚类质量[5]。

在快速路更新研究方面，已有成果主要集中于更新度评价、多属性方案比较、知识图谱构建和案例推荐等方向[6]-[8]。已有研究分别从快速路更新度评价[6]、改扩建方案多属性比较[7]以及工程知识图谱和案例推理等角度开展了探索[8]。与此同时，工程领域的本体建模和知识组织研究也为异构工程信息的结构化表达提供了参考[9] [10]。前期研究已经尝试利用知识图谱组织快速路更新知识，但较少从横断面工程数据本身出发，系统识别典型设计模式，并分析道路组成、空间宽度、车道数和设施配置之间的关联关系。

基于此，本文以城市快速路更新横断面案例为研究对象，开展以下工作：一是对历史横断面数据进行清洗和结构化编码，建立面向数据挖掘的工程特征体系；二是采用混合型数据聚类方法识别快速路横

断面的典型模式；三是利用关联规则挖掘横断面宽度、红线宽度、车道数、道路组成和设施配置之间的组合关系；四是利用独立交通数据验证高峰小时交通量、设计速度和车道数之间的关系。本文关注历史工程案例中的模式识别和知识发现，不再重复开展文本分类、图匹配和综合方案推荐研究。

2. 数据来源与特征构建

2.1. 样本背景与描述性统计

本研究案例数据来源于真实工程项目调研及中国知网公开案例资料，选取的道路样本分布于全国 19 座城市。数据筛选严格遵循相关原则，样本道路以地面道路与高架主线为主。样本断面宽度及车道数量分布如表 1 所示，可见数据集覆盖范围较广，涵盖窄断面地下、高架主线至大宽度综合断面等多种工程场景。

Table 1. Descriptive statistics of cross-sectional cases
表 1. 横断面案例描述性统计

指标	统计结果
道路组成	高架快速路 71 条(65.1%)；地面快速路 26 条(23.9%)；地下快速路 12 条(11.0%)；地面辅道 42 (38.5%)
双向车道数	4 车道 18 (16.5%)；6 车道 75 (68.8%)；8 车道 14 (12.8%)；10 车道 2 (1.8%)
规划红线宽	均值 60.08 m；中位数 60.00 m；范围 32~140 m
横断面宽	均值 44.57 m；中位数 42.00 m；范围 13.75~140 m

2.2. 数据来源

本研究对获取的道路样本数据进行分类。第一类为城市快速路横断面数据，共 109 条记录，来源于前期收集的快速路更新案例资料。每条记录包括规划红线宽度、车道数、快速路结构类型、道路组成、横断面宽度、横断面布置说明和横断面设施组成等字段。第二类为道路特征文本数据，共 159 条记录，包括道路沿线特征描述和建设形式标签。建设形式分为地面快速路、地下快速路和高架快速路三类。该数据用于说明案例库的整体构成和建设形式分布，不参与本文的文本分类模型训练。第三类为交通需求与车道配置数据，共 171 条记录，包括设计速度、高峰小时交通量和车道数。本文使用快速路横断面案例中观察到的车道配置规律进行独立验证。本研究采用主要横断面数据、辅助文本数据和独立交通验证数据的分析结构。

本研究构建“核心横断面数据 + 辅助文本数据 + 独立交通验证数据”的分层分析架构。研究涉及的横断面案例主要变量含义及类型定义如表 2 所示。

Table 2. Main variables of cross-section cases
表 2. 横断面案例主要变量

变量	含义	类型
RW	规划红线宽度	数值型
NTL	双向车道数	数值型
RCS	快速路结构类型	类别型
RC	道路组成	类别型
MLW	横断面宽度	数值型
CSDS	横断面设计方案说明	文本型
CW	横断面设施组成	类别及列表型

2.3. 数据清洗与特征处理

横断面数据中的 RCS 包括地面快速路、地下快速路和高架快速路，RC 包括地下主线、地面主线、地面辅道和高架主线。CW 字段原始形式为设施名称列表，本文将其转换为设施是否出现的二元变量，包括人行道、非机动车道、绿化带、机动车道、分隔带、护栏和路侧带等。

原始数据中存在中央分隔带、中间分隔带等表达差异。本文对同义设施名称进行统一，将中央分隔带和中间分隔带统一为分隔带，将中间护栏和护栏统一为护栏，以减少字段表达差异对规则挖掘的影响。

关联规则方法要求将工程数据转换为离散项集。RW 划分为小于 40 m、40~50 m、50~60 m、60~70 m、70~80 m、80~100 m 和 100 m 以上七个区间。MLW 划分为小于 25 m、25~35 m、35~45 m、45~55 m、55~70 m、70~100 m 和 100 m 以上七个区间。车道数保留 4 车道、6 车道、8 车道和 10 车道四种取值。

3. 研究方法

3.1. 混合型工程数据聚类

由于横断面案例同时包含数值变量和类别变量，本文采用 k-prototypes 思路进行聚类。数值变量采用标准化后的 RW 和 MLW，类别变量采用 RCS、RC 和 NTL。案例 x_i 与聚类中心 z_j 之间的距离表示为：

$$d\left(x_i, z_j\right) = \Sigma\left(x_{ip} - z_{jp}\right)^2 + \gamma\Sigma\delta\left(x_{iq}, z_{ip}\right) \tag{1}$$

式中，第一项为数值变量距离，第二项为类别变量不一致数量，δ 在两个类别取值不同时为 1，否则为 0，γ 为类别变量权重参数。聚类过程中，数值变量使用簇内均值更新，类别变量使用簇内众数更新。本文测试聚类数 2~6 和类别权重参数 0.5、1.0、2.0 的组合，并以轮廓系数和各簇样本规模确定最终参数。

3.2. 关联规则挖掘

将每一条横断面记录视为一个事务，将结构类型、道路组成、车道数、红线宽度区间、横断面宽度区间和设施配置转换为事务项。关联规则表示为 A→B，其中 A 为前件，B 为后件。

支持度表示同时包含 A 和 B 的案例比例，置信度表示在包含 A 的案例中同时包含 B 的比例，提升度表示规则置信度相对于后件总体出现概率的提升程度。提升度大于 1，说明 A 和 B 之间存在正向关联。本文设置最小支持度为 5.5%、最小置信度为 60%，并优先保留支持度不低于 8%、置信度不低于 65%且提升度大于 1.5 的规则。对于由字段定义直接产生的机械关系，不将其作为核心工程发现。

在支持度由 4.6%提高到 9.2%、置信度由 55%提高到 70%时，候选规则数量逐步减少，但平均置信度提高，平均提升度保持在约 2.0 以上，说明主要工程关联并非由单一阈值组合产生，具体数据如表 3 所示。

Table 3. Threshold sensitivity analysis of association rules
表 3. 关联规则阈值敏感性分析

最小支持度	最小置信度	候选规则数	平均置信度	平均提升度
4.6%	55%	358	79.0%	2.17
5.5%	60%	319	82.0%	2.20
7.3%	65%	219	82.0%	2.04
9.2%	70%	132	85.0%	1.99
4.6%	55%	358	79.0%	2.17

3.3. 交通数据验证

研究利用 171 条快速路记录建立浅层决策树模型，以设计速度和高峰小时交通量为输入，以车道数为输出。采用 5 折交叉验证评价规则的稳定性，并比较不同树深度下的分类准确率。该验证不用于替代规范计算，而是检验交通需求与车道数之间的基本关系。

4. 横断面案例聚类结果

聚类数量确定

表 4 汇总了不同类别权重 γ 和聚类数 k 下的综合轮廓系数。$\gamma = 2.0$、$k = 5$ 时取值最高(0.534)，且五个簇的样本数为 25、33、36、7 和 8，最小簇规模为 7，因此选定该组参数。

Table 4. Clustering parameter comparison
表 4. 不同聚类参数下的综合轮廓系数

γ\k	2	3	4	5	6
0.5	0.526	0.443	0.430	0.426	0.461
1.0	0.458	0.394	0.407	0.436	0.452
2.0	0.439	0.389	0.474	0.534	0.516

如表 5 所示，模式 1 共 25 条记录，主要为地面主线案例，其中地面快速路占 24 条。该模式平均横断面宽度为 56.53 m，平均车道数为 6.64。案例中的机动车道和人行道出现比例较高，部分案例同时设置中央分隔设施，反映了以地面主线通行功能为主、兼顾基本城市道路服务功能的横断面特征。

Table 5. Overall results of clustering
表 5. 聚类结果总体情况

聚类模式	样本数	平均 RW/m	平均 MLW/m	平均 NTL
模式 1	25	60.16	56.53	6.64
模式 2	33	57.06	25.19	5.82
模式 3	36	54.06	49.25	5.61
模式 4	7	99.43	86.74	6.86
模式 5	8	65	29.13	5.75

模式 2 共 33 条记录，全部为高架主线案例。该模式平均横断面宽度为 25.19 m，明显低于模式 1 和模式 3，护栏出现比例较高，而人行道、非机动车道和绿化带出现比例较低。这说明高架主线主要承担机动车快速通行功能，地面慢行和绿化设施通常由其他道路空间承担。

模式 3 共 36 条记录，全部为地面辅道组成。该模式平均横断面宽度为 49.25 m，非机动车道出现比例达到 83%，人行道出现比例约为 64%。与高架主线型模式相比，地面辅道更强调对沿线地块和慢行交通的服务功能。

模式 4 仅包含 7 条记录，但红线宽度和横断面宽度均明显高于其他模式，平均红线宽度为 99.43 m，平均横断面宽度为 86.74 m。这类案例通常具有较宽的规划控制条件，可同时布置主线、辅道、慢行交通空间和分隔设施，可称为大宽度综合型模式。

模式 5 包含 8 条记录，全部为地下主线案例，7 条记录采用 6 车道方案。其平均横断面宽度为 29.13

m，设施以机动车道、结构边界和中间分隔设施为主，人行道、非机动车道和绿化带出现比例较低，体现了地下快速路与地面公共空间相分离的特征。

总体来看，聚类结果同时反映了道路组成、车道数和空间宽度的组合差异。地面主线和地面辅道虽然均属于地面道路，但两者在横断面宽度和慢行设施配置方面表现出明显差异。

5. 关联规则挖掘结果

5.1. 宽度和车道数关联规则

表 6 给出了部分具有工程解释意义的关联规则。从宽度关系看，红线宽度 50~60 m 且设置人行道的案例，其横断面宽度落入 45~55 m 区间的概率较高，提升度达到 4.19。红线宽度 60~70 m 且设置分隔带的案例，其横断面宽度处于 55~70 m 区间的概率也明显高于总体水平。

从道路结构看，高架主线与 25~35 m 横断面宽度之间具有较高关联度。高架主线且为六车道的案例中，87.5%的记录横断面宽度处于 25~35 m 区间。这与高架快速路主要承担机动车快速通行功能、地面慢行空间相对较少的工程特征一致。

Table 6. Association rules of cross-section engineering characteristics
表 6. 横断面工程特征关联规则

前件	后件	支持度	置信度	提升度
RW 为 50~60 m 且有人行道	MLW 为 45~55 m	8.30%	69.20%	4.19
RW 为 60~70 m 且有分隔带	MLW 为 55~70 m	10.10%	68.80%	4.16
MLW 为 45~55 m 且为六车道	RW 为 50~60 m	10.10%	78.60%	3.43
RW 为 60~70 m 且为高架主线	MLW 为 25~35 m	13.80%	93.80%	3.1
高架主线且为六车道	MLW 为 25~35 m	19.30%	87.50%	2.89
RW 为 60~70 m 且为地面辅道	设置非机动车道	11.90%	100.00%	2.73
地面辅道且为六车道	设置非机动车道	18.30%	76.90%	2.1
MLW 为 55~70 m	RW 为 60~70 m	12.80%	77.80%	2.12

5.2. 道路组成与设施配置规则

地面辅道与非机动车道之间呈现较稳定的关联关系。红线宽度为 60~70 m 且道路组成为地面辅道的案例，均配置了非机动车道。地面辅道且为六车道的案例中，非机动车道出现比例达到 76.9%。这说明地面辅道并非单纯承担机动车分流功能，还需要服务沿线地块出行和慢行交通。与高架主线相比，地面辅道更可能同时配置人行道、非机动车道和分隔带。

5.3. 规则的工程应用

关联规则的作用不是直接生成最终设计方案，而是为案例检索和经验判断提供参考。当规划红线宽度、道路组成和车道数已基本确定时，可以根据高提升度规则快速缩小横断面宽度的合理范围。例如，对于红线宽度为 60~70 m、采用地面辅道且需要配置六车道的项目，可以优先检查横断面宽度 55~70 m、同时配置非机动车道的历史案例。由此，关联规则反映的是历史案例中的统计共现关系，不代表严格的因果关系。具体方案仍需结合交通预测、道路规范、结构条件、管线条件和施工组织进行复核。

6. 交通需求与车道配置的独立验证

6.1. 高峰小时交通量分布

本数据中共有 171 条快速路车道配置记录。按照高峰小时交通量划分，当交通量低于 1500 时，4 车道和 6 车道占比较高，8 车道案例很少；当交通量处于 2500~3500 区间时，6 车道仍占主要比例，但 8 车道案例开始增加；当交通量超过 5000 时，8 车道和 10 车道的比例明显提高，4 车道案例基本消失。

6.2. 决策树验证结果

如表 7 所示，验证结果表明，浅层决策树能够识别交通量与车道数之间的基本分段关系，但模型准确率尚不足以替代工程设计计算。这是因为车道配置还受到道路等级、交叉口间距、周边用地、服务水平目标和规划预留条件等因素影响。数据的验证结果主要支持以下判断：交通需求是车道数配置的重要因素，且高峰小时交通量增加通常伴随着车道数增加；但仅使用设计速度和交通量不足以完整解释所有工程设计结果。

Table 7. Validation results of expressway lane numbers
表 7. 快速路车道数验证结果

决策树深度	平均准确率	标准差
2	69.60%	1.20%
3	68.40%	2.90%
4	69.60%	3.50%
5	66.70%	6.20%

7. 讨论

7.1. 对案例知识库建设的意义

传统快速路案例库通常以项目名称、建设地点和方案文本作为主要索引，工程人员需要通过关键词或项目经验查找资料。本文将横断面案例转换为结构类型、道路组成、车道数和空间宽度等标准化属性，并进一步提取设施配置关系，使案例库能够从资料存储向知识发现转变。聚类结果可以用于建立案例库的模式标签，关联规则可以作为案例检索时的辅助条件。

7.2. 对设计院工程经验整理的意义

快速路横断面设计经验往往分散在不同项目和不同设计人员的工作成果中。本文挖掘出的规则能够以较为直观的形式呈现历史经验。例如，地面辅道与非机动车道之间的关联关系，可以为慢行设施配置提供经验参考；高架主线与较窄横断面之间的关联关系，可以帮助设计人员快速识别适合高架结构的横断面范围。规则的实际应用仍应与道路等级、交通预测、规划红线、地形条件和施工条件结合。

7.3. 与前期研究的区别

前期研究主要围绕文本需求识别、工程参数转换、图结构案例匹配和候选方案评价展开。本文不再进行文本分类和图匹配，而是直接对结构化横断面数据进行聚类和关联规则挖掘，研究目标是识别典型案例模式和工程属性关系。虽然本文使用了前期案例资料中的部分数据，但研究问题、变量组织方式、分析方法和研究结论均不同。前期研究关注如何为新项目推荐候选方案，本文关注历史横断面案例中存

在什么样的统计模式和工程关联。

7.4. 研究局限

横断面数据样本规模相对有限，部分模式的样本数较少，大宽度综合型和地下主线型模式的统计稳定性需要更多案例进行验证。部分工程设施名称来自横断面文字说明，虽然进行了同义词统一，但不同项目的记录深度存在差异，可能影响设施关联规则的稳定性。关联规则反映统计共现关系，不能据此推断工程因果关系。规则还需要结合更多城市、不同道路等级和实际设计项目进行外部验证。

8. 结论

本文以城市快速路更新横断面案例为对象，利用聚类分析和关联规则挖掘方法，研究历史工程案例中的典型模式和属性关联关系。快速路横断面可划分为 5 类典型模式，分别表现为地面主线型、高架主线型、地面辅道型、大宽度综合型和地下主线型，综合轮廓系数为 0.534。

不同道路组成对应明显不同的空间配置特征。高架主线型案例平均横断面宽度约为 25.19 m，地面主线型和地面辅道型案例的平均横断面宽度分别约为 56.53 m 和 49.25 m，地下主线型案例平均横断面宽度约为 29.13 m。

关联规则挖掘表明，红线宽度、横断面宽度、道路组成、车道数和设施配置之间存在较稳定的组合关系。地面辅道与非机动车道之间具有较强关联，高架主线和六车道与 25~35 m 横断面宽度之间也具有较高关联度。

快速路交通数据的验证结果表明，交通量增加时，8 车道和 10 车道方案的比例明显提高。浅层决策树的平均验证准确率约为 69.6%，说明交通需求能够解释部分车道配置差异，但不能替代完整的工程设计计算。

聚类结果和关联规则可以为快速路更新案例库的分类整理、工程经验提取和初步方案分析提供支持。后续研究应继续扩大案例来源，补充道路等级、交叉口间距、交通组织、用地条件和建设时序等变量，并对挖掘出的规则开展跨地区和实际项目验证。

基金项目

本研究由江西省高校人文社会科学研究项目 GL24226 与江西省社会科学规划项目 26GL67D 资助。

参考文献

[1] 中华人民共和国住房和城乡建设部. CJJ 129-2009 城市快速路设计规程[S]. 北京: 中国建筑工业出版社, 2009.

[2] 吴作军, 王麒翔. 城市快速路建设形式研究——以成都市武侯大道(二环路-绕城高速)快速化改造为例[J]. 城市道桥与防洪, 2021(7): 20-24+9.

[3] Agrawal, R., Imieliński, T. and Swami, A. (1993) Mining Association Rules between Sets of Items in Large Databases. *ACM SIGMOD Record*, **22**, 207-216. https://doi.org/10.1145/170036.170072

[4] Huang, Z. (1998) Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery*, **2**, 283-304. https://doi.org/10.1023/a:1009769707641

[5] Kaufman, L. and Rousseeuw, P.J. (1990) Finding Groups in Data: An Introduction to Cluster Analysis. Wiley. https://doi.org/10.1002/9780470316801

[6] Zhu, M., Zhou, W., Hu, M., Du, J. and Yuan, T. (2023) Evaluating the Renewal Degree for Expressway Regeneration Projects Based on a Model Integrating the Fuzzy Delphi Method, the Fuzzy AHP Method, and the TOPSIS Method. *Sustainability*, **15**, Article 3769. https://doi.org/10.3390/su15043769

[7] Zhu, M., Zhou, W. and Yuan, T. (2023) Urban Expressway Renewal Strategy Based on Knowledge Graphs. *Applied Sciences*, **13**, Article 4666. https://doi.org/10.3390/app13084666

[8] Pan, X., Xu, M., Jia, J., Xia, Y. and Chen, Q. (2022) Evaluation Model of Highway Reconstruction and Expansion

Scheme Based on Multiple Attribute Decision Making. *Advances in Civil Engineering*, **2022**, Article ID: 3764557. https://doi.org/10.1155/2022/3764557

[9] Hu, X., Xia, B., Skitmore, M. and Chen, Q. (2016) The Application of Case-Based Reasoning in Construction Management Research: An Overview. *Automation in Construction*, **72**, 65-74. https://doi.org/10.1016/j.autcon.2016.08.023

[10] Farghaly, K., Soman, R.K. and Zhou, S.A. (2023) The Evolution of Ontology in AEC: A Two-Decade Synthesis, Application Domains, and Future Directions. *Journal of Industrial Information Integration*, **36**, Article 100519. https://doi.org/10.1016/j.jii.2023.100519