

DOI:10.11992/tis.201610015

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20170227.2217.028.html>

基于决策树分类的可拓建筑策划预测方法

郭强^{1,2,3}, 邹广天^{1,2,3}

(1. 哈尔滨工业大学建筑学院, 黑龙江哈尔滨 150006; 2. 哈尔滨工业大学建筑计划与设计研究所, 黑龙江哈尔滨 150006; 3. 黑龙江省寒地建筑科学重点实验室, 黑龙江哈尔滨, 150006)

摘要: 为提升建筑师在策划过程中科学预测的能力, 提出了一种基于决策树分类的可拓建筑策划预测方法。首先, 运用数据采集软件批量采集互联网中的建筑案例数据, 将数据预处理后存储至建筑案例库中; 其次, 通过评价特征选取、评价信息元集生成、决策树构建等步骤, 获得决策树模型; 最后, 运用该模型预测当前策划项目的性能指标是否满足要求, 并给出不满足要求情况下性能指标变换的途径。案例检验表明, 该方法能有效提高建筑师运用互联网数据的能力, 能够挖掘决策树分类知识, 从而加速计算机辅助可拓建筑策划的进程。

关键词: 可拓建筑策划; 决策树; 分类; 指标预测; 可拓变换

中图分类号: TP18; TU18 **文献标志码:** A **文章编号:** 1673-4785(2017)01-0117-07

中文引用格式: 郭强, 邹广天. 基于决策树分类的可拓建筑策划预测方法[J]. 智能系统学报, 2017, 12(1): 117-123.

英文引用格式: GUO Qiang, ZOU Guangtian. Prediction methods for extension architecture programming based on decision tree classification[J]. CAAI transactions on intelligent systems, 2017, 12(1): 117-123.

Prediction methods for extension architecture programming based on decision tree classification

GUO Qiang^{1,2,3}, ZOU Guangtian^{1,2,3}

(1. School of Architecture, Harbin Institute of Technology, Harbin 150006, China; 2. Architectural Planning and Design Institute, Harbin Institute of Technology, Harbin 150006, China; 3. Heilongjiang Cold Region Architectural Science Key Laboratory, Harbin 150006, China)

Abstract: To improve the prediction ability of architects, a prediction method for extension architecture programming (EAP) based on decision tree classification was proposed. First, the architectural case data from the Internet were obtained by data acquisition software, and stored in an architectural case database after data preprocessing. Second, through evaluation characteristics selection, evaluation information element set generation and decision tree construction, the decision tree model was discovered. Then, the performance indicators of the current project were predicted using this model, providing transformation approaches if the result did not satisfy the requirement. This study indicates that the proposed method can effectively improve an architects ability to use Internet data and mine decision tree classification knowledge, thus accelerating the process of computer aided EAP.

Keywords: extension architectural program; decision tree; classification; indicators prediction; extension transformation

可拓建筑策划(extension architectural program, EAP)是将可拓创新方法引入建筑策划领域,通过问题界定、可拓分析、可拓变换、策略评价等一系列程

序来生成创新策略的方法,其目的是指导建筑设计创新^[1]。可拓建筑策划过程中急需对场地、规模、功能、空间、形象、技术等方面的性能指标进行科学预测的方法。

随着计算机辅助建筑设计(CAAD)和信息技术

收稿日期: 2016-10-13. 网络出版日期: 2017-02-27.

基金项目: 国家自然科学基金项目(51178132).

通信作者: 邹广天. E-mail: zoug@hit.edu.cn.

的迅速发展,互联网中的建筑案例已成为建筑师不可或缺的数据资源。案例辅助建筑设计已经引起建筑学界的广泛关注。魏力恺构建并开发了基于建筑空间关系原型的建筑案例检索系统 Architable^[2],孟洁提出了基于案例推理(case-based reasoning)的建筑方案设计流程^[3],段正励等提出了基于 ArcGIS 软件的杭州市工业建筑遗产案例库构建与综合分析方法^[4],张颀等提出了基于建筑信息模型的建筑空间拓扑关系提取插件和案例检索方法^[5],孙澄等提出了基于神经网络的严寒地区办公建筑节能性能预测与形态设计方法^[6]。然而,上述方法均未涉及互联网中的海量案例,也缺乏比较深入的案例分析方法。因此,研究互联网中案例数据获取及从中挖掘知识的方法将具有重要意义。

决策树是一种常用的分类预测工具。它通过建立一个模型或分类器来预测类别,其优势是需要的数据准备很少,不需要任何领域知识,既擅长处理连续型变量,也擅长处理建筑案例中常见的离散型变量。基于决策树知识的可拓知识挖掘方法是决策树分类和可拓创新方法相交叉的结果。首先利用递推的基本思想构造决策树,并从中提取规则,然后分析当前面对的矛盾问题,利用这些规则预测可拓变换的实施效果,进而筛选出可行的变换^[7]。该方法能够预测可拓建筑策划项目的性能指标(通常难以通过计算获得)是否满足策划要求,并给出不满足要求情况下性能指标变换的途径。相关研究还有赵燕伟等提出的基于改进 BP 神经网络的可拓分类器构建方法^[8];李亚飞等提出的基于决策树分类的云南省迪庆地区景观类型预测方法^[9];王茂军等^[10]提出的基于决策树法的北京城市居民通勤距离模式挖掘方法。针对互联网中的海量建筑案例,我们提出了基于决策树分类的可拓建筑策划预测方法。

1 建筑案例数据的获取和预处理

1.1 数据获取

可拓建筑策划分类的对象是建筑案例库中的案例,用物元^[11]描述为

$$M = \begin{bmatrix} O, & c_1, & v_1 \\ & c_2, & v_2 \\ & \vdots & \vdots \\ & c_n, & v_n \end{bmatrix} = (O, C, V) \quad (1)$$

式中: c_1, c_2, \dots, c_n 为建筑案例的特征; v_1, v_2, \dots, v_n 为建筑案例特征对应的量值。案例特征应尽可能反映建筑的全部信息,分为建筑所处环境特征和建

筑物自身特征。前者包括建筑基地的社会、人文、交通、环境、景观、经济等方面,后者包括基本信息、设计构思、场地、功能、空间、形象、技术、评价等方面(见表1)。两者通过空间方位属性相关联,可以形成每栋建筑物的完整数据。

表1 建筑案例数据特征

Table 1 Characteristics of the architectural case data	
数据类型	数据特征
人口	人口数量、密度、结构比例等,人口职业、就业、拥有机动车数量、通勤出发时间、出行时长、出行方式等
经济	区域居民人均收入、居民消费状况等
交通	城市道路等级、交通拥堵状况、公共交通类型及站点分布等
环境	地形、地貌、土壤等地质特征,水质、水情等水文特征,气温、湿度、降水、风向、空气质量指数(AQI)等
设施	教育、商业、办公服务设施等
基本信息	项目名称、项目类型、项目性质、项目等级、高度类别、项目位置、项目造价、建筑师、设计团队、设计单位、委托单位、设计风格、设计时间、建成时间
设计构思	设计出发点、设计应对策略
场地	场地形状、地形呼应方式、建筑群体组合方式
功能	主体功能、附属功能
空间	空间类型、首层平面形状、标准层平面形状
形象	表皮处理方式、立面开窗方式、虚实对比程度
技术	结构类型、主体材料、辅助材料、主被动技术
评价	评价内容、评价等级
技术指标	规划面积、总建筑面积、容积率、地上层数、地下层数、建筑高度、建筑密度、停车位数量

数据获取是运用数据采集软件将互联网中的案例数据获取到建筑案例库的过程。鉴于智能性和稳定性,本文选用火车采集器,从政府门户网站、建筑策划与设计机构网站、建筑案例共享平台、各类百科等网站来采集案例。首先需要设置网址采集规则、内容采集规则、导入数据库规则和其他规则,数据采集结果以二维关系数据表形式存储在 MySQL 软件中,多个数据表之间以主键和外键相关联。另外,火车采集软件能够制定采集计划来完成建筑案例库数据的动态更新。

1.2 数据预处理

在分类与预测之前,需要进行数据预处理,包括数据转换、数据集成、数据归约等。

数据转换是参照项目名称-空间位置对照表、建筑中英文对照表、建筑同义词对照表、计量单位转换表,对案例数据进行转换处理,解决数据语义

多样性问题。

数据集成是将来自多个数据源的数据进行合并,以提高分类的速度和精确性。为统一格式,在数据集成中起关键作用的空间方位属性用(E, N)来表示,例如:哈尔滨工业大学建筑学院的空间方位属性,记作($126^{\circ}37'53.85''E, 45^{\circ}45'02.16''N$)。

数据集成时经常出现属性值缺失、数据不一致、数据存在噪声等问题,这些问题将对预测结果产生重大影响。噪声数据常用回归分析、离群点分析等方法来进行光滑处理^[12]。缺失值通过以下方式来处理:1)当缺失值恰好为分类特征时,直接删除该条数据;2)使用全局常量,如 Unkown 来填充;3)使用属性所有值的中心度量(如均值或中位数)来填充;4)使用其他演算推理的方式来批量填充。

数据值域不一致往往将造成预测结果与意义难以解释,常用数据归约的方法进行标准化处理。

2 可拓建筑策划分类预测方法

可拓建筑策划分类预测的核心是决策树算法,利用信息熵的原理,选择具有最高信息增益的特征作为分裂特征,递归地构建决策树的分支。它包括模型构建、模型检验、模型应用 3 个环节,具体分为评价特征选取、评价信息元集生成、决策树模型构建和当前策划项目指标预测 4 个步骤,如图 1 所示。

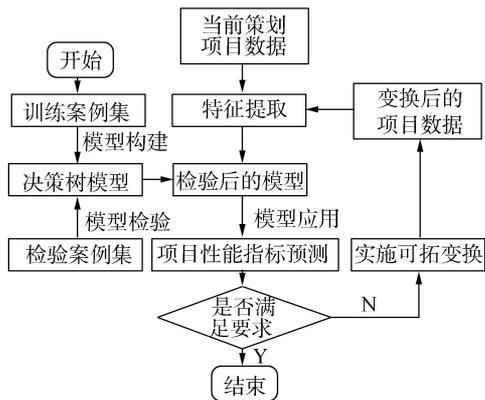


图 1 基于决策树分类的 EAP 预测方法流程图

Fig.1 Flow chart of the predict method for EAP based on decision tree classification

2.1 根据建筑专业知识选取评价特征

通过前期访谈、现场调研、模式构想、方案试做等环节,建筑师已经确定当前建筑策划项目的基本特征。现需要对它的性能指标进行预测,检验是否满足策划要求。首先确定要预测的性能指标,在决策树分类时将这些指标称为分类特征,如建筑风格、空间开放性、流线合理性、技术可行性、环境协调性等。确定分类特征后,根据建筑专业知识,选

取与之密切相关的评价特征,如与建筑风格相关的评价特征(包括屋顶类型、立面色彩、建筑材料等)。有时某些评价特征可能并未出现在建筑案例表中,需要根据表中数据计算,如容积率可通过规划用地范围内建筑面积总和与用地面积的比值计算得到。

2.2 选取目标数据生成评价信息元集

根据可拓建筑策划的预测要求,从建筑案例库中选取相关案例,应尽可能选择评价等级较高的数据,以保证分类结果的准确性。将取出的案例删除其他特征,只保留分类特征和评价特征,得到建筑案例的评价信息元集,记作

$$I = \{I_j\} = \{(O, C_j, V_j), j = 1, 2, \dots, s\} \quad (2)$$

该集合表示:对象 O 关于特征 C_j 的量值为 V_j 构成的信息元集合^[7]。

将建筑案例评价信息元集随机地分为两组:一组是训练案例集,用于构建决策树模型,一般取建筑案例总量的 80%~90%;另一组是检验案例集,取剩下的部分,用于检验模型的准确性。模型准确率是指检验案例集中被正确分类的建筑案例数量与该集中建筑案例总数量的比值。

2.3 构建决策树提取分类规则知识

决策树算法是将训练集递归地分裂成较小的子集,借助分裂准则来确定在每个结点上哪个属性能够产生最佳拆分^[12]。用 $E(I)$ 表示对建筑案例评价信息元集 I 做出正确类别判断所需的信息量,记作

$$E(I) = - \sum_{i=0}^m p_i \log_2(p_i) \quad (3)$$

式中: p_i 是 I 中任意数据属于类 C_i 的概率,用 $|C_i|/|I|$ 表示。

假设按评价特征 A 对 I 中的建筑案例进行划分,且特征 A 具有 k 个量值(v_1, v_2, \dots, v_k)。

2.3.1 特征 A 为离散型特征

若特征 A 为离散型特征,如建筑平面类型、功能类型、空间组合形式等,将信息元集 I 分为 k 个子集(I_1, I_2, \dots, I_k),以特征 A 为根进行分类的信息熵定义为

$$E_A(I) = - \sum_{i=0}^m \frac{|I_j|}{|I|} \times E(I) \quad (4)$$

式中 $\frac{|I_j|}{|I|}$ 为是信息元集 I 中第 j 个子集的权重。

以特征 A 为根进行分类的信息增益定义为

$$G(A) = E(I) - E_A(I) \quad (5)$$

为方便理解,将 300 个建筑案例组成的案例集作为信息元集,将建筑风格作为分类特征来说明决策树构建的原理。式(3)中, $E(I)$ 表示未开始确定

分裂特征时对 300 个案例进行分类所需的期望信息量。假定特征 A 是屋顶类型,可能取值为[坡屋顶、平屋顶、穹顶、其他类型屋顶],那么 $|I_2|/|I|$ 表示案例集中平屋顶的案例个数与案例总数的比。式(4)中, $E_A(I)$ 表示以屋顶类型为分裂点还需要的信息量。相应地,式(5)中 $G(A)$ 表示以屋顶类型进行划分我们得到了多少信息量。信息增量值越高,表明该特征对于分类的影响越大。根据式(4)、式(5)可以求出其他特征(如建筑材料)的信息增益,选择信息增益最大的特征作为分裂特征。

2.3.2 特征 A 为数值型特征

若特征 A 为数值型特征,如建筑面积、建筑层数等。首先将 A 的值按递增顺序排序,将每对相邻值的中点看作可能的分裂点,对于特征 A 的每个可能分裂点,计算 $E_A(I)$,选择最小期望信息需求的点作为 A 的分裂点。 I_1 是满足 $A \leq$ 分裂点的建筑案例集合,而 I_2 是满足 $A \geq$ 分裂点的建筑案例集合。

决策树模型采用贪心算法,以自顶向下递归方式构建,直到达到下列条件之一终止:

- ① 结点处的评价信息元属于同一个类;
- ② 没有剩余特征用来进一步划分;

③ 给定的分支中没有信息元,即该子集为空,该结点不需要构建。

以上算法均在 RapidMiner7.2 软件中进行编辑,决策树模型的剪枝和准确性检验通过设置 X-Validation 命令的相关参数来完成,经过多次反复训练在很大程度上提升了模型的鲁棒性。

决策树模型构建完毕后,从根结点到叶结点的每条路径建立一个规则,以 IF-THEN 形式的分类规则表示,并用支持度(support)和置信度(confidence)进行评估。这些规则将形成分类知识库,作为建筑性能指标预测的依据。规则 R 可表示为 $R: A \Rightarrow B$ [confidence, support]

以可拓建筑策划中的规则 R_1 为例, R_1 可表示为

R_1 : 屋顶类型=坡屋顶 \wedge 屋面材料=瓦片 \Rightarrow
建筑风格=欧式 (80%, 40%)

上述规则表示,屋顶类型为坡屋顶并且屋面材料为瓦片的建筑案例中,建筑风格为欧式的案例占到 80%。同时满足坡屋顶、屋面材料为瓦片、欧式风格的建筑案例占到所有案例的 40%。

2.4 预测当前策划指标获取可拓变换途径

提取当前建筑策划项目的评价特征,将其导入决策树模型,获得预测结果。若预测结果满足策划要求,则表明当前建筑策划方案是合理的,可以进

入后续的建筑策划书生成与表达环节。

若预测结果不能满足策划要求,则需要改变当前建筑策划方案的某些特征,可拓学提供了 5 种基本变换,即置换变换、增删变换、扩缩变换、分解变换和复制变换,还有多特征的组合变换等。在可拓建筑策划中,可拓变换 T 定义为

$$T(O_i, c_i, v_i') = (O_i, c_i, v_i')$$

针对多特征的变换很容易产生组合爆炸,因此,需要对变换的结果进行取舍,获得可行的解变换,获取解变换的步骤如下。

1) 利用上述的决策树模型对变换后的项目数据进行预测,了解实施变换以后产生怎样的效果,并将预测结果中满足策划要求的变换提取出来,形成当前策划问题的解变换集。

2) 通过计算变换难度,对解变换集进行排序和评优。变换难度通过变换系数来定义:

① 对于只存在类别差异,而不存在等级差异的建筑特征,如建筑平面类型、建筑材料类型等,无论变换为哪个值,变换系数都定义为 1;

② 对于存在等级差异的建筑特征,如建筑评价等级,当变换为降低等级时,无论降低几个等级,变换系数定义都为 1;当提高一个等级时,变换系数定义为 1;当提高两个等级时,变换系数定义为 2,以此类推。

对于多个变换而言,其变换难度^[7]按照每个变换系数加 1 后累乘再减 1 进行计算,定义为

$$\text{变换难度} = \prod (\text{变换系数} + 1) - 1$$

对于建筑师而言,根据变换难度对解变换集进行排序后,排名越靠前的变换,可操作性越强,越容易产生最优策略。

3 实验与分析

为了减少建筑建成后给周边环境带来的影响和破坏,在可拓建筑策划中需要预先评估建筑与周边环境的协调程度(即建筑环境协调度)。因此,选取某建筑策划项目的建筑环境协调度预测为例,将火车采集器作为互联网案例数据抓取工具,将 MySQL 软件作为数据存储工具,将 RapidMiner 作为数据预处理和决策树构建工具,具体步骤如下。

1) 选取评价特征。首先将建筑环境协调度定义为分类特征,可能取值为差、一般、好,记作 Y 。根据建筑专业知识选取周边环境类型、总建筑面积、建筑主体材料、窗墙比、建筑与环境的形体穿插渗透关系(即穿插渗透程度),并将这些参数作为评价特征,分别记作 c_1, c_2, c_3, c_4, c_5 。为简化计算的难

度,对评价特征的取值作出如下规定。

周边环境类型的可能取值为街道型、广场型、滨水型、植被型;总建筑面积的取值为原案例数据表中的值;建筑主体材料对原案例数据表中的材料进行归纳,可能取值为天然型、人工型、混合型;窗墙比将原案例数据表中的窗墙比数值进行归纳,可能取值为墙绝对主导型、墙主导型、窗墙对等型、窗主导型、窗绝对主导型,分别记作 I、II、III、IV、V;

穿插渗透程度根据建筑与环境的穿插渗透关系进行评估,可能取值为无渗透、局部渗透、完全渗透。

2)生成评价信息元表。通常直接从建筑案例库中选择案例数据。目前尚无这方面的数据,因此运用火车采集器抓取了 ArchDaily 网站最新发布的 300 个建筑案例,并导入 MySQL 中。经过转换评价特征值,填补缺失值,对不一致数据进行纠错后,最终得到建筑案例评价信息元表(见表 2)。

表 2 建筑案例评价信息元表

Table 2 Table of the building case evaluation information element

序号	建筑名称	周边景观类型 c_1	总建筑面积 c_2/m^2	建筑主体材料 c_3	窗墙比 c_4	穿插渗透程度 c_5	建筑环境协调度 Y
1	冰岛远足休息小屋	植被型	100	天然型	II	完全渗透	好
2	Linaires 科技园教学楼	广场型	4 800	人工型	I	局部渗透	差
3	巴厘岛 DRA 住宅	滨水型	1 013	天然型	II	完全渗透	好
4	双子住宅	植被型	743	混合型	V	无渗透	好
5	南京国际青年文化中心	街道型	465 000	人工型	III	无渗透	差
6	清华大学就业指导中心	街道型	21 000	天然型	III	局部渗透	好
7	Pink 住宅	植被型	3 000	天然型	III	局部渗透	好
...
300	悉尼邮轮码头	滨水型	4 600	混合型	V	无渗透	一般

3)构建决策树模型。首先,将评价信息元表导入 RapidMiner 软件,定义好分类特征后,运用 X-Validation 命令完成决策树模型的构建、检验与评估。该命令将评价信息元表随机分为 10 等份,依次将其中 9 份作为训练集,另外 1 份作为检验集,对模型进行 10 次评估,基本保证了该模型的鲁棒性。经评估,该模型准确率为 84.60%,满足要求。其次,从上述模型中提取出决策树分类规则。由图 2 和图 3 可知,穿插渗透程度是影响决策树分类最显著的特征。

```

穿插渗透程度=完全渗透
| 总建筑面积>35 750: 一般(好=0, 差=1,一般=3)
| 总建筑面积>35 750
| | 周边环境类型=广场型
| | | 总建筑面积>4 500
| | | | 建筑主体材料=人工型: 一般(好=0, 差=0,一般=2)
| | | | 建筑主体材料=天然型: 好(好=3, 差=0,一般=0)
| | | | 总建筑面积≤4 500: 好(好=19, 差=0,一般=2)
| | | 周边环境类型=植被型: 好(好=26, 差=0,一般=3)
| | | 周边环境类型=滨水型
| | | | 建筑主体材料=人工型: 一般(好=0, 差=0,一般=2)
| | | | 建筑主体材料=天然型: 好(好=3, 差=0,一般=0)
| | | | 建筑主体材料=混合型: 一般(好=0, 差=0,一般=2)
| | | 周边环境类型=街道型
| | | | 窗墙比= I :一般(好=0, 差=0,一般=2)
| | | | 窗墙比= II :好(好=2, 差=0,一般=0)
| | | | 窗墙比= III :一般(好=1, 差=0,一般=4)
| | | | 窗墙比= IV :好(好=2, 差=0,一般=0)
穿插渗透程度=局部渗透
| 总建筑面积>7 150
| | 建筑主体材料=人工型
| | | 总建筑面积>97 175.500: 一般(好=0, 差=0,一般=2)

```

图 3 提取的决策树分类规则(部分)

Fig.3 The decision tree classification rules extracted

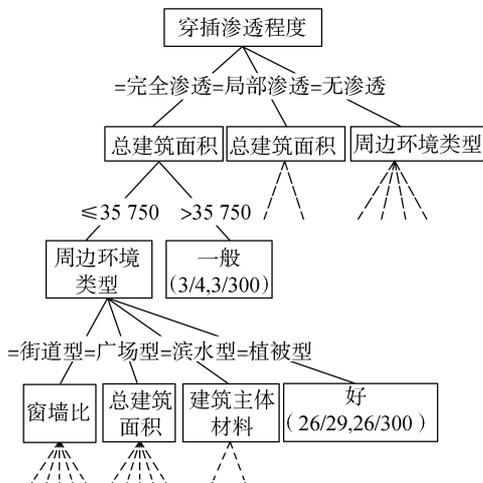


图 2 构建的决策树模型简图

Fig.2 The chart of decision tree model

4)应用决策树模型。从当前建筑策划方案中提取该建筑的 5 项特征,其中周边景观类型 c_1 为街道型,总建筑面积 c_2 为 4 000 m^2 ,建筑主体材料 c_3 为混合型,窗墙比 c_4 为 III 型,穿插渗透程度 c_5 为无渗透,需要对建筑环境协调度 Y 进行预测。将该建筑数据导入上述决策树模型,预测结果为“差”,不满足策划要求。

为使预测结果变为“好”,需要对该建筑的特征实施可拓变换。由于基地条件限制,总建筑面积在

建筑立项时已经确定,不能改变,因此,需要对特征 c_1, c_3, c_4, c_5 实施可拓变换,具体包括:

$$\begin{aligned}
 T_{11}(O_i, c_1, \text{街道型}) &= (O_i, c_1, \text{广场型}), \\
 T_{12}(O_i, c_1, \text{街道型}) &= (O_i, c_1, \text{滨水型}), \\
 T_{13}(O_i, c_1, \text{街道型}) &= (O_i, c_1, \text{植被型}), \\
 T_{31}(O_i, c_3, \text{混合型}) &= (O_i, c_3, \text{天然型}), \\
 T_{32}(O_i, c_3, \text{混合型}) &= (O_i, c_3, \text{人工型}), \\
 T_{41}(O_i, c_4, \text{III}) &= (O_i, c_4, \text{I}), \\
 T_{42}(O_i, c_4, \text{III}) &= (O_i, c_4, \text{II}), \\
 &\dots
 \end{aligned}$$

$$T_{52}(O_i, c_5, \text{无渗透}) = (O_i, c_5, \text{完全渗透})$$

以上是针对单一特征值的变换,还包括多特征组合变换,最终共产生 179 (即 $4 \times 3 \times 5 \times 3 - 1$) 种变换,对应着 179 个变换后的策划项目数据。

利用决策树模型对这些数据进行预测。经测试,预测结果为好、一般、差的项目数据分别为 69 条、77 条、33 条。这 69 条数据对应的变换,即为当前策划问题的解变换。将项目数据按照变换难度重新进行排序,排名前 5 位的数据将作为最优策略提供给建筑师,用于指导策划方案修改(见表 3)。

表 3 可拓变换排序结果(前 5 位)

Table 3 The result of extension transformation sorting (the top five)

序号	建筑环境协调度 Y	周边环境类型 c_1	建筑主体材料 c_3	窗墙比 c_4	穿插渗透程度 c_5	可拓变换	变换难度
1	好	街道型	天然型	III	局部渗透	$T_{31} \cap T_{51}$	3
2	好	街道型	混合型	II	完全渗透	$T_{42} \cap T_{52}$	3
3	好	街道型	混合型	IV	完全渗透	$T_{43} \cap T_{52}$	5
4	好	广场型	人工型	III	完全渗透	$T_{11} \cap T_{31} \cap T_{51}$	7
5	好	植被型	天然型	II	无渗透	$T_{13} \cap T_{31} \cap T_{42}$	7

4 结论

基于决策树分类的可拓建筑策划预测方法是充分发挥可拓创新方法和决策树分类方法的优势,对当前策划项目的性能指标进行科学预测的方法。

1)通过提出的建筑案例数据获取、数据预处理、决策树构建、分类预测、变换筛选等步骤,建筑师能够从互联网中的海量建筑案例中提取出决策树分类规则,并给出建筑性能指标的变换途径。

2)案例检验表明,该方法是可行的,具有较强的可操作性,能有效解决可拓建筑策划研究与应用的预测难题。

3)将计算机领域的数据采集、决策树分类技术引入建筑学领域,能有效提升建筑师运用互联网数据的能力,加快计算机辅助可拓建筑策划的进程。

参考文献:

[1] 连菲. 可拓建筑策划的基本理论与应用方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2010: 25-27.
LIAN Fei. The study on basic theory and applying methods of extension architectural program [D]. Harbin: Harbin Institute of Technology, 2010: 25-27.

[2] 魏力恺. 基于 CBR 和 HTML5 的建筑空间检索与生成研

究[D]. 天津: 天津大学, 2013: 108-114.
WEI Likai. Architectural spatial retrieval and generating based on CBR and HTML5 [D]. Tianjin: Tianjin University, 2013: 108-114.

[3] 孟洁. 基于案例推理的建筑方案设计流程研究[D]. 哈尔滨: 哈尔滨工业大学, 2014: 56-76.
MENG Jie. Research on architectural scheme design process using case-based reasoning[D]. Harbin: Harbin Institute of Technology, 2014: 56-76.

[4] 段正励, 刘抚英. 杭州市工业遗产综合信息数据库构建研究[J]. 建筑学报, 2013, 33(10): 45-48.
DUAN Zhengli, LIU Fuying. Study on the comprehensive information database of industrial heritage in Hangzhou[J]. Architectural journal, 2013, 33(10): 45-48.

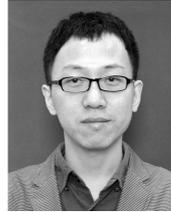
[5] 张颖, 李昌华, 李智杰. 基于拓扑特征的建筑信息模型检索方法[J]. 计算机应用研究, 2016, 33(3): 916-921.
ZHANG Jie, LI Changhua, LI Zhijie. Building information model retrieval based on topological features[J]. Application research of computers, 2016, 33(3): 916-921.

[6] 孙澄, 韩昀松. 光热性能考虑下的严寒地区办公建筑形态节能设计研究[J]. 建筑学报, 2016(2): 38-42.
SUN Cheng, HAN Yunsong. A study on energysaving design of office building forms in the severe cold region regarding daylighting and thermal performance [J]. Architectural journal, 2016(2): 38-42.

- [7] 杨春燕, 李小妹, 陈文伟, 等. 可拓数据挖掘方法及其计算机实现[M]. 广州: 广东高等教育出版社, 2010: 206-214.
- [8] 赵燕伟, 任设东, 陈尉刚, 等. 基于改进 BP 神经网络的可拓分类器构建[J]. 计算机集成制造系统, 2015, 21(10): 2807-2815.
ZHAO Yanwei, REN Shedong, CHEN Weigang, et al. Extension classifier construction based on improved BP neural network [J]. Computer integrated manufacturing systems, 2015, 21(10): 2807-2815.
- [9] 李亚飞, 刘高焕, 黄翀. 基于决策树分类的云南省迪庆地区景观类型研究[J]. 资源科学, 2011, 33(2): 328-334.
LI Yafei, LIU Gaohuan, HUANG Chong. Exploring landscapes based on decision tree classification in the Diqing region, Yunnan province [J]. Resources science, 2011, 33(2): 328-334.
- [10] 王茂军, 宋国庆, 许洁. 基于决策树法的北京城市居民通勤距离模式挖掘[J]. 地理研究, 2009, 28(6): 1516-1527.
WANG Maojun, SONG Guoqing, XU Jie. Data mining on commuting distance mode of urban residents based on the analysis of decision tree[J]. Geographical research, 2009, 28(6): 1516-1527.

- [11] YANG Chunyan, CAI Wen. Extensics: theory, method and application[M]. Beijing: Science Press, 2013: 20-27.
- [12] HAN Jiawei, KAMBER M, PEI Jian. 数据挖掘: 概念与技术[M]. 范明, 孟小峰, 译. 3 版. 北京: 机械工业出版社, 2012: 55-65.

作者简介:



郭强,男,1985 年生,博士研究生,中国人工智能学会会员,主要研究方向为可拓建筑学、可拓建筑策划数据挖掘,参加完成国家自然科学基金项目 1 项,参编《中国原创学科——可拓学发展报告 2016》。



邹广天,男,1960 年生,教授,博士生导师,博士,中国人工智能学会可拓学专业委员会副主任、中国建筑学会建筑师分会建筑策划专业委员会副主任、中国环境行为学会副会长,主要研究方向为建筑计划学、可拓建筑学、建筑设计创新学、环境行为心理学。主持完成国家自然科学基金项目 2 项,出版专著 1 部,主编与参编论文集多部,发表学术论文 200 余篇,被 EI、CSSCI 检索论文 20 余篇。

2017 第 13 届中国智能系统会议

The Chinese Intelligent Systems Conference (CISC)

中国智能系统会议是由中国人工智能学会智能空天系统专业委员会发起的系列学术会议,其宗旨是为本领域的专家学者、研究生以及工程技术人员提供一个学术交流的平台,以推动我国智能系统相关理论、技术与应用的发展。第 13 届中国智能系统会议(CISC'2017) 将于 2017 年 10 月 14—15 日在黑龙江省牡丹江市召开。本次会议由中国人工智能学会主办,中国人工智能学会智能主天系统专业委员会协办,牡丹江师范学院与北京航空航天大学承办。会议论文集将由 Springer 出版社在 Lecture Notes in Electrical Engineering 系列正式出版, EI 收录。热忱欢迎海内外广大同仁踊跃投稿并出席本届会议,交流学术成果。

主办单位:中国人工智能学会

协办单位:中国人工智能学会智能主天系统专业委员会

承办单位:牡丹江师范大学、北京航空航天大学

会议主席:贾英民,北京航空航天大学;杨敬民,牡丹江师范学院

会议网站:<http://sias.buaa.edu.cn/info/1007/1132.htm>