

针对参变量数据元集的可拓建筑策划分类知识挖掘方法

邹广天, 张斯, 郭强, 丁俐娟

哈尔滨工业大学建筑学院; 哈尔滨工业大学建筑计划与设计研究所, 哈尔滨 150006

摘要 为解决可拓分类知识挖掘方法难以满足可拓建筑策划数据及问题的特殊性需求, 将可集成参变量数据元集的可拓建筑策划数据作为研究对象, 提出适用于可拓建筑策划的可拓分类知识挖掘流程与方法, 即数据的预处理、选取评价特征并确定权重、建立关联函数并划分区间、计算获取可拓分类知识。拓展可拓建筑策划数据挖掘在方法层面的研究, 以方便获取重要的可拓建筑策划分类知识, 为可拓建筑策划提供决策支持, 为计算机辅助建筑策划提供新思路。

关键词 可拓建筑策划; 可拓数据挖掘; 可拓分类知识; 参变量数据元集

中图分类号 TU18

文献标志码 A

doi 10.3981/j.issn.1000-7857.2014.36.006

Methods for Extension Architectural Programming Classification Knowledge Mining on Parametric Data Set

ZOU Guangtian, ZHANG Si, GUO Qiang, DING Lijuan

School of Architecture, Harbin Institute of Technology; Architectural Planning and Design Institute, Harbin Institute of Technology, Harbin 150006, China

Abstract Extension architectural programming classification knowledge mining on parametric data set is an application of extension data mining in the extension architectural programming area. However, the existing methods for classification knowledge mining are difficult to meet the particularity demands of extension architectural programming. The paper takes extension architectural programming data which can be set into parametric data set as a research object, and puts forward a procedure and certain methods for classification knowledge mining suitable for extension architectural programming, including pretreatment of data, selecting evaluation characteristics and determining the weights, establishing the correlation function and dividing the interval, and calculating and accessing the extension classification knowledge. The paper aims to expand the method-level research on extension architectural programming data mining to facilitate access to important extension architectural programming classification knowledge, and provides a new idea for computer-aided architectural programming.

Keywords extension architectural programming; extension data mining; extension classification knowledge; parametric data set

建筑策划是针对建筑设计项目进行的策划性前期准备工作^[1]。建筑策划需借助各种信息收集方法与工具对相关信
息收集并分析, 揭示出业主、用户、建筑师及社会大众的价值
体系, 调查出场地、气候对建筑的影响, 为设计提供科学的依
据^[2]。20世纪90年代中国已基本建立了建筑策划的理论体

系, 并有众多理论研究成果。但由于计算机辅助策划的研究
进展缓慢, 造成建筑策划方法研究严重滞后于理论研究^[3]。

可拓建筑策划是将可拓学用于建筑策划的策划方法^[4],
由于可拓学易于将矛盾及创新问题转换为计算机可操作、可
处理的特点^[5], 可拓建筑策划便为计算机辅助策划提供了新

收稿日期: 2014-10-23; 修回日期: 2014-11-20

基金项目: 国家自然科学基金项目(51178132)

作者简介: 邹广天, 教授, 研究方向为建筑计划学、可拓建筑学、建筑设计创新学、环境行为心理学等, 电子信箱: zoug@hit.edu.cn; 张斯(共同第一作者),

硕士研究生, 研究方向为可拓建筑学, 电子信箱: zhang-si@163.com

引用格式: 邹广天, 张斯, 郭强, 等. 针对参变量数据元集的可拓建筑策划分类知识挖掘方法[J]. 科技导报, 2014, 32(36): 43-47.

思路。可拓建筑策划数据挖掘可为可拓建筑策划提供决策支持,但目前尚缺少相关的方法研究。本文的研究内容是可拓建筑策划数据挖掘方法研究的基础问题之一。

1 相关概念

1.1 可拓建筑策划分类知识挖掘

可拓建筑策划数据挖掘是可拓建筑策划和可拓数据挖掘相结合的产物,运用可拓数据挖掘技术,从可拓建筑策划相关的数据中挖掘与解决矛盾和创新问题有关的可拓知识。

可拓分类知识的挖掘方法是在变换下的分类。就是通过挖掘得到挖掘对象属于哪个可拓分类的知识,其结果是按综合关联度 K 的值将对象分为与 5 个域相对应的 5 类:正质变类 D_+ , 负质变类 D_- , 正量变类 D_+ , 负量变类 D_- 及拓界类 D_0 [6]。因此分类结果根据变换而有所不同。

可拓建筑策划分类知识的挖掘方法是可拓建筑策划数据挖掘方法之一,它是用可拓分类挖掘技术寻找与可拓建筑策划有关的可拓分类知识,发现在可拓建筑策划中主动实施或自然发生变换的作用和效果,辅助创新性策略生成,从而提供决策支持。

1.2 可拓建筑策划中的参变量数据元集

在可拓建筑策划中与可拓分类知识相关的数据并非是一成不变的,当条件改变时,事、物、关系及空间的某些性质将随之发生变化。可拓建筑策划数据随时间、地域、空间属性等具体条件而变化,引入条件变量 t 可得到可拓建筑策划参变量数据元,记作

$$g(t) = (O(t), c, v(t))$$

将这些由某一条件变量 t 引起的参变量数据元集合起来可形成参变量数据元集,记作

$$\{g(t)\} = \begin{bmatrix} O_1(t), & c, & v_1(t) \\ O_2(t), & c, & v_2(t) \\ \vdots & \vdots & \vdots \\ O_n(t), & c, & v_n(t) \end{bmatrix} = (\{O(t)\}, c, V(t))$$

对于参变量数据元集中的任一数据,在某一条条件变量 t_i 变化时可记作

$$g(t_i) = \begin{bmatrix} O_g(t_i), & c_{g1}, & v_{g1}(t_i) \\ & c_{g2}, & v_{g2}(t_i) \\ & c_{g3}, & v_{g3}(t_i) \\ & \vdots & \vdots \\ & c_{gn}, & v_{gn}(t_i) \end{bmatrix} = (O_g(t_i), C_g, V_g(t_i))$$

当条件变量为时间时,建筑的风格、材质、结构形式、空间组织方式等将随之改变;当条件变量为地域时,建筑的材质、院落布局、功能组合方式、空间限定形式等将随之改变;当条件变量为空间属性时,建筑的功能、形式、结构、形态等将随之改变;当条件变量为其他建筑策划条件时,建筑本体及建筑空间中与该条件产生强相关的特征量值也将随之改变。

2 挖掘流程

对可拓建筑策划分类知识进行挖掘的基本思路是在保持样本最大相似性的基础上进行挖掘。通过对可拓建筑策划数据的可拓分析,可归纳出与变换有关的特征,并构建其参变量数据元集。对集合中任一数据元,都可表述为在某一条件下由参变量控制的特征及量值,对参变量施加变换则相应量值发生改变,应用可拓分类知识的挖掘方法可获得与变换有关的可拓分类知识。

由此,将可拓分类知识挖掘方法的一般流程进行梳理,并结合可拓建筑策划数据形式各异、问题最优解不唯一的特性,建立针对参变量数据元集的可拓建筑策划分类知识挖掘流程(图1)。



图1 参变量数据元集的可拓建筑策划分类知识挖掘流程

Fig. 1 Flowchart of extension architectural programming classification knowledge mining on parametric data set

3 数据的预处理

可拓建筑策划数据挖掘的对象是可拓建筑策划数据。可拓建筑策划数据是对与可拓建筑策划相关的客观事物直观描述,其记录形式多为文本或图像。就数据的结构而言,大部分可拓建筑策划数据属于非结构化数据,少部分属于半结构化数据。而现有的数据挖掘技术更适合挖掘结构化数据,因此在数据预处理阶段的任务是按挖掘目标分别完成数据集成与清理,以及数据由非结构及半结构到结构化数据的数据变换 [7]。

3.1 数据集成与清理

可拓建筑策划数据分布在建筑理论研究文献、建筑设计规范、建筑设计资料集、设计案例、实态调查资料中。在进行建筑策划时,建筑师会从期刊杂志、门户网站上收集建筑大师柯布西耶作品的资料,这些资料从不同角度对作品进行分

析,内容常发生重合,还会存在对同一建筑师、同一作品有不同的称呼,如“柯布西耶”也作“柯比西埃”,“流水别墅”也作“考夫曼住宅”等。在数据集成阶段需对资料中的有价值数据进行统一格式并清除冗余的数据。

抽取各个数据源中的相关数据,进行合并处理。再对数据进行同义词的定义,以对数据中关于同一对象的数据进行重组整合。统一数据格式,将数据用基元表示法表示为对象、特征、量值的有序三元组,确定数据集合中的参变量,构建可拓建筑策划参变量数据元集。在数据集成的过程中,还需完成数据清理,删除相似重复记录、不完整数据、错误数据。

3.2 数据变换

数据变换是将基元表示的数据变换为结构化数据形式,即行数据,可用二维表表达。对任一基元表示的可拓建筑策划数据可按对象 O 、特征 c 、量值 v 建立二维数据表(表1)。

表1 可拓建筑策划二维数据

Table 1 Extension architectural programming two-dimensional data table

数据元编号	事物/动作/关系/空间名	c_1	c_2	...	c_n
g_i	O	v_{11}	v_{21}	...	v_{n1}
		v_{12}	v_{22}	...	v_{n2}
		v_{13}	v_{23}	...	v_{n3}
		\vdots	\vdots	\vdots	\vdots
		v_{1n}	v_{2n}	...	v_{nn}

对于可拓建筑策划文本数据,可按语段的语义完成基元表示,再按对象、特征、量值建立二维数据表。例如“位于东京表参道的TOD's旗舰店是由伊东丰雄设计的”,可变换为表2。根据其他数据源提供的数据信息,可在此基础上继续填充有关TOD's旗舰店的信息。

表2 TOD's旗舰店数据

Table 2 TOD's flagship store data

数据元编号	建筑名称	设计师	区位	功能	...
g_i	TOD's旗舰店	伊东丰雄	东京表参道	商业	...
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

可拓建筑策划图像数据与文本数据的最主要区别在于图像中的线与块蕴含着丰富的信息,这些信息是多角度的,按需可进行不同数据变换。

图像数据中的“线”,常表示“由A流向B”。在可拓建筑策划中常用的功能分区图,其中的人流流线(图2)可按关系元表示法进行变换(表3)。

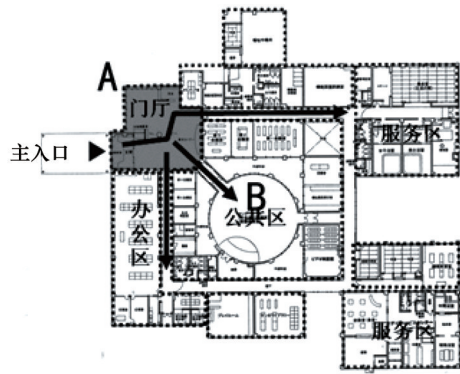


图2 某老年活动中心平面功能分区

Fig. 2 Plane function zoning diagram of senior activity center

表3 老年活动中心平面功能分区数据

Table 3 Plane function zoning data table of senior activity center

数据元编号	图名	关系名	前项	后项
g_i	老年活动中心平面功能分区	人流位置关系	门厅	办公区
		人流位置关系	门厅	公共区
		人流位置关系	门厅	服务区

图像数据中的“块”,常表示“区域的功能”。图像中还包含与其相关的其他特征信息,在数据变换时应注意对块进行编码,再对块进行信息的提取。如在某户型的平面图中表达了各区块的功能及面积,经数据变换后可写作表4。在数据变换时,若无表示房间功能的文字信息,可对平面中的家具布置进行识别,如布置了床可识别为卧室,布置了洗手池及便器可识别为卫生间等。

表4 某户型数据

Table 4 Apartment data

数据元编号	图名	块编号	功能	使用面积/m ²
g_i	某户型CAD图	A	客厅	23.37
		B	主卧	13.33
		C	次卧	12.04
		D	厨房	6.97
		E	卫生间	4.66

经数据变换后的二维数据表在挖掘时,可取其中的部分使用,也可取全体作为一个单元使用。如可对住宅建筑在不同地域的户型特征进行挖掘,也可对同一地区住宅建筑户型随时间演变的特征进行挖掘。

4 选取评价特征并确定权重

将参变量数据元中与挖掘目标有关的特征作为评价特征,科学、合理地选取评价特征是挖掘成果科学、有效的保障之一。该阶段包括选取评价特征和确定各特征的权重值。确定评价特征权重值是计算综合关联度所做的准备工作。

4.1 选取评价特征

在建筑策划领域中,并无确定的原理或公式可以作为评判一个建筑好坏的标准,这个“好”的标准不唯一,且“好”是多角度的。如贝聿铭设计的卢浮宫改扩建项目玻璃金字塔是堪称经典的建筑作品,而盖里设计的毕尔巴鄂古根汉姆博物馆虽然同为博物馆建筑却与前者截然不同,没有对称规整的布局,亦无沉稳庄重的造型,但它还是成为了“火焰般燃烧”的建筑经典。

在进行可拓建筑策划时,应针对设计条件,设计条件存在的关联选取评价特征。对设计条件进行可拓分析,可视选取原始设计条件的特征与经可拓分析得到的特征共同作为评价特征。例如,拟策划一博物馆,原始设计条件包括了建筑面积、必要的功能、展品信息、基本场地条件等内容。在选取评价特征时,不仅需包含这些特征,还应包括通过可拓分析得到建筑室内外面积、地上及地下面积、功能组合方式、融合的新功能、展品对空间形态及物理环境的需求、场地中的特殊条件、场地与建筑结合方式等一并作为评价特征。这些经拓展分析、共轭分析得到的评价特征都极易成为创新性策略生成的突破口。

对设计条件通过进行可拓分析而选取评价特征,是可拓建筑策划数据挖掘的一个重要特征。经发散分析可拓展更多的可能性,经相关分析可获取设计条件及评价特征的关联性,经蕴含分析可取得设计条件与设计需求的蕴含关系,经可拓分析可得出设计条件的伸缩取值范围,进而发现新的利于解决矛盾的评价特征。在解决矛盾问题时,与一般的建筑策划数据挖掘相比,可拓建筑策划数据挖掘的结果因为这样的可拓分析而更具有优势、更易于实现创新。

4.2 确定评价特征的权重值

对某一挖掘目标而言,选取的评价特征一般都不是唯一的,各评价特征与挖掘目标的关联程度不同。因而需对各评价特征赋予权重。例拟策划一观演建筑,选取评价特征为功能、流线、采光、场地布局及空间感受,则需根据观演建筑的特点对上述评价特征赋予权重。

在可拓建筑策划中,对评价特征的权重赋值主要取决于对各方意见的分析,即设计者、建筑使用者及建筑投资者的需求。为获取这些意见,面向不同目标采用不同方法,对设计者及投资者可采取会谈的方法,对使用者可采取问卷调查、虚拟现实等方法^[8]。通过收集汇总的资料可初步获取设计意见,建筑设计者、建筑使用者及建筑投资者对设计需求的重要程度常有数种不同的排序方式,即几组不同的评价特征权重值。对这几组不同的权重值可根据模糊意见集中法及层次分析法加以确定^[8]。

5 建立关联函数并划分区间

选取得到的评价信息元需首先单独计算各评价特征的关联函数值,然后进行规范化处理,再依据上一步骤得出的权重值计算综合关联度。基于对可拓建筑策划实际问题的可拓展分析,建立关联函数及划分区间^[9]。

通过可拓分析,对可拓建筑策划实际问题中评价特征的最优情况进行抽象,通常可归为为极值在满意区间内,如“拟建设一建筑面积为 1500 m² 的活动中心,面积可上下浮动 5%”。在该情况下,可选取最优点在区间中点且无公共端点的初等关联函数,其表达式^[9]为

$$k(x) = \frac{\rho(x, X_0)}{D(x, X_0, X)}$$

若极值在满意区间外,如“拟建一特殊人群服务站,公共区交通距离在满足使用要求前提下尽量短捷”。在该情况下,可选取最优点不在区间中点且无公共端点的初等关联函数,其表达式^[9]为

$$k(x) = \frac{\rho(x, x_0, X_0)}{D(x, X_0, X)}$$

除此之外,还有一些情况适合选取离散型关联函数,相应的选取方法及公式见文献[5]。

由可拓建筑策划数据建立参变量数据元集所选取的条件变量 t 不同,则评价特征不同,相应量值的满意范围与可接受范围亦不同。建立关联函数时,在满足区间套原则^[9]前提下,可结合设计条件、设计意见、相关规范及经验数据取得恰当的区间值,并分别确定满意区间及可接受区间。在可拓建筑策划领域中,具体参数的设定与产品销售、客户满意度等其他问题一样,需要根据具体问题做具体分析,并应对设计需求中排序靠前的评价特征做细致分析,谨慎取值。

6 计算获取可拓分类知识

按照已选定的关联函数及区间,计算对于每个评价特征的关联值,再按权重值计算综合关联度,计算关联积、关联差,以获取关于该参变量数据元集的可拓分类类别。

在可拓建筑策划数据分类挖掘中,正质变知识是价值最高的目标知识。正质变知识在可拓建筑策划中可理解为寻找有什么策划要素会在设计条件改变后从不符合要求到满足要求,还可以考察哪些设计条件可在可拓建筑策划中得到拓展。

根据可拓集的划定,满足变换前后关联数的样本属于正质变域即正质变类 D_+ ,记变换前评价信息元为 D_i ,变换后的评价信息元为 D_i' ,则满足变换前后综合关联度的条件^[9]为

$$K(D_i) \leq 0 \quad K'(D_i') > 0$$

通过计算满足变换前后关联数的样本数 $|E_+|$ 、负域的样本数 $|E_-|$ 、拓界的样本数 $|E_0|$ 及样本总数 $|U|$ 可获得关于该条可拓知识的支持度与可信用度^[9],即:

$$I_1 = \left(\frac{|E_-| + |E_0|}{|U|}, \frac{|E_+|}{|E_-| + |E_0|} \right)$$

如对某一地区的集合住宅户型中卧室与客厅两种功能空间的组织方式进行挖掘,对入住率、房间使用率、入住者满意程度等进行收集并挖掘,可得到不同组合方式的可拓分类知识。其中,由组合方式A变为组合方式B,通过挖掘可得施加该变换的可拓分类知识。在该地区进行集合住宅建筑策划时,利用关于该变换的正质变知识创新性地施加变换生成策略,可使策划项目的入住率、房间使用率、入住者满意程度更高。同时,在挖掘时若选定邻里交往模式、公共活动区绿化、设施等作为评价特征,可获取在这些角度进行变换的可拓分类知识,最终辅助创新性策略的生成。

在参变量数据元集的可拓建筑数据分类挖掘中,不同变换作用下对同一条可拓建筑策划数据的可拓分类结果不同。可拓分类结果具有可转化性,其转化程度取决于对设计条件的变换程度。因此,对可拓分类后不同类数据的可拓变换方法应有所不同。对正量变类 D_+ 数据,可施加其他变换,以满足设计需求;对负质变类 D_- 、负量变类 D_0 及拓界类 D_0 数据可先利用正量变知识转化到正域,再施加其他变换,最终满足设计需求。

7 结论

通过建立可拓建筑策划参变量数据元集,有针对性地对建筑策划中的条件变量进行挖掘,可获取利于解决策划矛盾、实现策划创新的可拓分类知识,为可拓建筑策划提供决策支持,并为计算机辅助策划提供新思路。

基于可拓学、建筑策划、数据挖掘3个学科的交叉研究,为挖掘方法的基础理论研究,挖掘流程中部分子环节的操作方法还需要进一步深入研究,并应用于实例以检验其可行性。

参考文献(References)

- [1] 庄惟敏. 建筑策划导论[M]. 北京: 中国水利水电出版社, 2000: 8-17.
Zhuang Weimin. Architectural programming guide[M]. Beijing: China

Water Power Press, 2000: 8-17.

- [2] 徐强. 建筑策划信息系统熵变模型构建研究[J]. 情报科学, 2011, 29(12): 1820-1824.
Xu Qiang. Research on the entropy change model construction for the construction plan information system[J]. Information Science, 2011, 29(12): 1820-1824.
- [3] 苏实. 从建筑策划的空间预测与评价到空间构想的系统方法研究[D]. 北京: 清华大学, 2011.
Su Shi. From space prediction and evaluation to the systematic method study of space conceiving in architectural programming[D]. Beijing: Tsinghua University, 2011.
- [4] 连菲, 邹广天. 化解矛盾, 寻求创新——可拓建筑策划理论探析[J]. 新建筑, 2011(3): 79-81.
Lian Fei, Zou Guangtian. Solving contradiction and seeking innovation—exploration and analysis of extensible architectural programming theory [J]. New Architecture, 2011(3): 79-81.
- [5] 杨春燕, 蔡文. 可拓学[M]. 北京: 科学出版社, 2014: 14-17.
Yang Chunyan, Cai Wen. Extenics[M]. Beijing: Science Press, 2014: 14-17.
- [6] 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: 47-55.
Yang Chunyan, Li Xiaomei, Chen Wenwei, et al. Extension data mining methods and computer-implemented[M]. Guangzhou: Guangdong Higher Education Press, 2010: 47-55.
- [7] 汪伟, 邹璇, 詹雪. 论数据挖掘中的数据预处理技术[J]. 煤炭技术, 2013, 32(5): 152-153.
Wang Wei, Zou Xuan, Zhan Xue. Discussion on data preprocessing technology in data mining[J]. Coal Technology, 2013, 32(5): 152-153.
- [8] 张冉. 基于可拓学的参数化建筑设计研究[D]. 哈尔滨: 哈尔滨工业大学, 2011.
Zhang Ran. Study of parametric building design based on extenics[D]. Harbin: Harbin Institute of Technology, 2011.
- [9] 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.
Lian Fei. The study on basic theory and applying methods of extension architectural program[D]. Harbin: Harbin Institute of Technology, 2010.

(责任编辑 赵业玲)

·学术动态·



2014年光学薄膜前沿国际会议在上海召开

2014年10月20—24日,由国家自然科学基金委员会、中国光学学会主办的“2014年光学薄膜前沿国际会议”在上海召开。

美国亚利桑那大学教授 Angus Macleod、德国夫琅和费应用光学与精密机械研究所教授 Norbert Kaiser、法国菲涅尔研究所教授 Claude Amra、美国劳伦斯利弗莫尔国家实验室教授 Christopher Stolz、浙江大学教授刘旭、同济大学教授王占山共同担任会议主席。来自欧洲、北美等10多个国家和地区的光学薄膜领域的60余位外国专家,来自国内高校、科研院所的150余位专家和技术人员与会,围绕设计与分析、纳米微结构与材料、薄膜制备与控制、测试与表征、极紫外与软X射线、薄膜前沿应用、高功率激光薄膜7个专题进行了交流。

详见中国科协网 <http://www.cast.org.cn/n35081/n35548/n38620/16064013.html>。