



(12)发明专利

(10)授权公告号 CN 106951611 B

(45)授权公告日 2020.08.07

(21)申请号 201710131029.7

G06F 30/20(2020.01)

(22)申请日 2017.03.07

G06K 9/62(2006.01)

(65)同一申请的已公布的文献号

审查员 石爽

申请公布号 CN 106951611 A

(43)申请公布日 2017.07.14

(73)专利权人 哈尔滨工业大学

地址 150006 黑龙江省哈尔滨市南岗区西  
大直街92号

(72)发明人 孙澄 张冉 韩昀松 董琪

曲大刚

(74)专利代理机构 哈尔滨市阳光惠远知识产权

代理有限公司 23211

代理人 刘景祥

(51)Int.Cl.

G06F 30/13(2020.01)

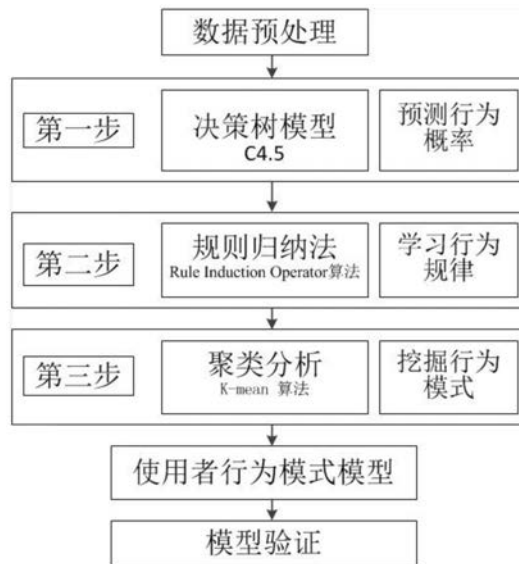
权利要求书2页 说明书12页 附图4页

(54)发明名称

一种基于使用者行为的严寒地区建筑节能设计优化方法

(57)摘要

本发明提出了一种基于使用者行为的严寒地区建筑节能设计优化方法,针对严寒地区特殊的气候条件,利用调研数据分析得到促动节能行为的建筑设计参数组;应用数据挖掘技术中的机器学习算法对行为实测数据的学习得到更为准确的使用者随机行为模式;通过行为模式优化传统预测模型,并利用基于贝叶斯理论的高斯过程矫正预测模型,得到优化节能性能设计参数组;耦合促动节能行为和优化节能性能参数组得到优化的参数集合;最后结合优化设计参数组和能耗预测模型,建立严寒地区建筑节能设计新流程。本发明能够降低不确定性对预测模型的影响,更加优化参数的筛选过程、更加准确的节能效果预测、更理性的建筑节能设计流程以及更准确的预估建筑设计结果。



1. 一种基于使用者行为的严寒地区建筑节能设计优化方法,其特征在于:

步骤1、针对严寒地区气候区,进行数据的采集,通过抽样调查和试点实验获取建立模型的基础数据;

步骤2、根据数据采集的结果建立严寒地区建筑空间典型模型组,分析使用者的属性特征建立促动节能行为的建筑设计参数组,得出能够导向节能行为的建筑空间类型及参数值域;

步骤3、在RapidMiner平台上应用数据挖掘技术中的决策树模型建立使用者行为基础数据库,利用规则归纳算法对决策树模型进行修剪以得到行为模型规则,再通过聚类分析法得到使用者随机行为模式模型,其结果直接作为参数输入能耗模拟平台Energyplus,并对模型结果进行验证,再通过Matlab平台进行多项逻辑回归初步建立建筑能耗预测模型;

步骤4、采用机器学习技术中贝叶斯理论支持下的高斯过程对模型进行校正和优化从而建立严寒地区建筑能耗优化预测模型;

步骤5、模型校正后进行数据校验,通过与实例对比,论证预测模型的预测能力和在建筑节能设计中应用的可行性,并通过回归分析得到可生成节能的建筑设计参数值域,建立优化节能性能设计参数组;

步骤6、耦合促动节能行为和优化节能性能的建筑设计参数组,得到控制办公建筑能耗的空间类型及参数值域,建立办公建筑节能设计优化参数组;

步骤7、以参数选择和设计流程为基础设计严寒地区办公建筑节能优化策略,以优化的参数为基础、以优化的预测模型为能耗性能评价工具设计得到优化的建筑节能设计方案;

所述步骤1具体为:由抽样调查数据得到建筑空间特征、使用者的属性及特征和促动使用者各类行为的建筑及其环境要素,由试点实验得到使用者行为的实测数据;

所述步骤3具体为:

步骤3.1、利用C4.5算法建构决策树模型,并应用决策树模型建立行为基础数据库,预估使用者室内活动概率;

步骤3.2、通过规则归纳法的Rule Induction Operator算法对决策树进行修剪,去除决策树中对分类不添加有用信息的节点,学习行为模式的规律;

步骤3.3、应用聚类分析法的K-means算法对行为模式进行最后的分类,并使其能够形成使用者行为配置文件;

严寒地区使用者耗能行为及影响要素调查主要包括三方面:

第一方面为受试者的基本信息;

第二方面是受试者回答问卷时所在建筑的基本信息;

第三方面为受试者所在建筑空间所在状态。

2. 根据权利要求1所述的方法,其特征在于:所述抽样调查采用现场实测、问卷调查及访谈方式进行,其中所述现场实测内容包括建筑形态及空间几何参数以及非建筑几何参数,所述问卷调查包括对严寒地区使用者基本信息的抽样调查和严寒地区使用者耗能行为及影响要素的调查。

3. 根据权利要求2所述的方法,其特征在于:所述决策树对一组数据分隔为预定义的种类,并对给定的数据进行描述、分类及定义,以建立一个分类模型,基于输入属性得到预测价值的目标属性,决策树模型的每个内部节点都对应一个预测属性,内部节点的分支数目

与相应的预测属性的可能值的数目相等。

4. 根据权利要求1所述的方法,其特征在於:决策树模型的建立分为两个步骤,即学习和分类;在学习过程中,将数据集中的记录数据自动且随机分为两组,即训练组和校验组;在分类过程中,通过交叉验证法对决策树的准确性进行验证以估测学习过程的性能表现。

5. 根据权利要求4所述的方法,其特征在於:在交叉验证过程中,数据集被划分成大小相等的 $k$ 个子集,在 $k$ 个子集中,一个子集作为测试数据集,其余 $k-1$ 个子集作为训练数据集,交叉验证过程重复 $k$ 次, $k$ 个子集中的每一个子集都作为测试数据, $k$ 次迭代的 $k$ 次结果将取平均值,得到最后唯一的估计值;其中 $k$ 取10。

## 一种基于使用者行为的严寒地区建筑节能设计优化方法

### 技术领域

[0001] 本发明属于建筑节能技术领域,特别涉及一种基于使用者行为的严寒地区建筑节能设计优化方法。

### 背景技术

[0002] 建筑节能设计是涉及多领域、多学科的交叉研究。在数字技术迅速发展,建筑性能化设计趋向的背景下,建筑节能设计不仅与建筑能耗预测研究相关,也与数字模拟技术及机器学习技术的发展紧密相联。

[0003] 目前,建筑能耗预测方法主要有时间序列法,傅里叶级数模型(Fourier series models),回归模型,人工神经网络(artificial neural network models)及模糊逻辑理论(fuzzy logic models)等。建筑能耗预测模型虽有初步的发展,但其预测结果仍与实际能耗间有巨大差异,而使用者行为是造成这一差异的重要因素。既有研究表明,因使用者行为所造成既有建筑能耗预测模型预测值与实际值的差距,影响模型的预测精度。从而在预测模型与节能建筑设计过程结合时,影响节能建筑设计性能。因此,建立使用者行为模式模型是建筑设计阶段更为精确预测建筑能耗的关键。

[0004] 建筑性能模拟在建筑设计初期阶段,应用模拟对建筑方案进行调试和操作,具有巨大的应用潜力。建筑性能化设计的潜力虽然已被广泛认可,但在实际的建筑方案中因其与真实数据的差异性,常常不能提供建筑节能设计的最佳方案;而在为设计者提供对建筑系统性能的认知方面,也因此差异性的存在,不能够真实反馈建筑设计方案对性能的影响。这一差异性,是由于建筑性能模拟中输入参数量大,以使用者行为参数为代表的非线性、离散型及不确定性等多类型参数更增加了影响要素的复杂性。因而,在建筑性能模拟中,如何针对不确定性的参数进行有效分析和整合,通过新手段新方法修正不确定性和随机性对模拟结果的影响,是优化建筑性能化设计的预测精度,增强其在建筑实际项目中的应用能力的核心问题之一。

[0005] 模拟技术对使用者行为的描述与刻画还很不完善,随着节能工作与实践的深入,逐渐暴露出很多问题,例如:难以准确反映实际人行为对建筑用能水平的影响,模拟结果常常偏离实测数据;难以准确反映实际人行为对系统性能表现的影响,一些节能技术的模拟评估结论与实际观测到的能耗状况不相符,而一些技术手段带来的性能提升和节能潜力在模拟结果中又得不到有效反映;在基于模拟技术的宏观能耗研究、能耗定额与节能标准规范中,由于缺乏具有实际代表性的参考使用模式,导致一些性能指标与实际情况严重脱节等等。这些问题影响了模拟技术的可靠性与可信度,极大地限制了其在节能工作中进一步发挥作用。并且,由于建筑围护结构和系统的优化,建筑节能技术性能标准越发严格,节能系统逐渐的广泛应用,直接增加了使用者行为要素对建筑能耗的影响。这一变化进一步增大了预测值和实际值间的差距,而此差异性将导致性能预测的偏差,最终影响建筑设计结果达到预测的性能表现。

[0006] 随着机器学习技术的繁荣,建筑研究人员开始尝试利用机器学习的方法解决节能

建筑性能化设计的瓶颈问题。也有研究者针对如使用者行为模式等无法量化的问题,利用机器学习的数据挖掘技术得到较为优化的使用者行为模型。

[0007] 在建筑节能设计研究方面,秦蓉等人采用神经网络算法提取热环境参数分布CFD模拟的关键信息,实现了神经网络与热过程模拟耦合迭代的中庭热环境全年动态模拟。喻伟等人建立基于GA-BP网络的建筑能耗和室内热舒适状况的预测模型,提升了预测精度,使建筑师在设计阶段能够简单且准确地获得设计建筑的能耗和室内舒适状况。孙澄应用GA-BP神经网络建模技术,以建筑形态参量为输入参量,以耗热量指标为输出参量,建构严寒地区办公建筑采暖能耗预测模型。目前研究多数利用神经网络进行了对热环境及能耗水平进行模拟,其结果取得了一定程度的优化,但神经网络的过拟合问题,和精度较低问题,建筑性能及能耗预测模型仍需要进一步的优化和修正。

[0008] 在使用者行为模式研究方面,王闯等人利用马尔科夫链建立室内使用者移动模型,该方法可有效反映建筑中各房间使用者状况随时间的随机变化规律,并为室内发热量及设备使用行为的模拟计算提供了基础。Simona D'Oca利用数据挖掘技术建立了更为准确的办公建筑开关窗行为模型,使建筑耗能模型更为合理,并进一步得出了使用者行为对建筑能耗使用,室内热环境及使用者工作效率的影响。

[0009] 建筑的运行状况及能耗与人的行为密切相关。在系统形式灵活可控的情况下,由于生活作息、偏好习惯及个体行为行为表现的差异,建筑能耗大不相同。人的行为与气象条件、围护结构、系统设备一样,是影响建筑能耗水平的重要因素。在室外气象、围护结构、设备系统形式等确定的情况下,建筑使用者对能耗相关设备及建筑构件的调节和控制,决定了建筑的总体能耗。在目前的模拟分析工作中,往往更多地侧重于气象参数、围护结构和设备系统方面,对人行行为大多采用较为简化的方式进行描述,无法有效体现出人行行为的各种显著影响,因此在很多情况下,造成了模拟分析结果与实际情况偏差很大,影响了分析、预测和评估建筑能耗的准确性。由于既有建筑能耗及性能模拟平台中,不能够对建筑使用者行为进行较为准确和具体的刻画,模拟结果难以反映建筑方案实际建成后的能耗水平和建筑性能表现,造成了预测结果和实际数据具有较大偏差。由于对使用者特征的不详细描述,造成了某些建筑技术措施在实际建筑里的性能表现在模拟计算中得不到合理反映,模拟评估结论易与实测不相符。由于对使用者行为的促动因素,仍缺乏与建筑设计要素相关的深入研究,不能够为建设促动使用者节能行为的建筑节能设计提供理论支持。

## 发明内容

[0010] 本发明的目的在于克服现有技术的缺点与不足,提供一种基于使用者行为的严寒地区建筑节能设计优化方法,解决现有技术中存在的问题。

[0011] 本发明的目的通过以下技术方案实现:一种基于使用者行为的严寒地区建筑节能设计优化方法,

[0012] 步骤1、针对严寒地区气候区,进行数据的采集,通过抽样调查和试点实验获取建立模型的基础数据;

[0013] 步骤2、根据数据采集的结果建立严寒地区建筑空间典型模型组,分析使用者的属性特征建立促动节能行为的建筑设计参数组,得出能够导向节能行为的建筑空间类型及参数值域;

[0014] 步骤3、在RapidMiner平台上应用数据挖掘技术中的决策树模型建立使用者行为基础数据库,利用规则归纳算法对决策树模型进行修剪以得到行为模型规则,再通过聚类分析法得到使用者随机行为模式模型,其结果直接作为参数输入能耗模拟平台Energyplus,并对模型结果进行验证,再通过Matlab平台进行多项逻辑回归初步建立建筑能耗预测模型;

[0015] 步骤4、采用机器学习技术中贝叶斯理论支持下的高斯过程对模型进行校正和优化从而建立严寒地区建筑能耗优化预测模型;

[0016] 步骤5、模型校正后进行数据校验,通过与实例对比,论证预测模型的预测能力和在建筑节能设计中应用的可行性,并通过回归分析得到可生成节能的建筑设计参数值域,建立优化节能性能设计参数组;

[0017] 步骤6、耦合促动节能行为和优化节能性能的建筑设计参数组,得到控制办公建筑能耗的空间类型及参数值域,建立办公建筑节能设计优化参数组;

[0018] 步骤7、以参数选择和设计流程为基础设计严寒地区办公建筑节能优化策略,以优化的参数为基础、以优化的预测模型为能耗性能评价工具设计得到优化的建筑节能设计方案。

[0019] 进一步地,所述步骤1具体为:由抽样调查数据得到建筑空间特征、使用者的属性及特征和促动使用者各类行为的建筑及其环境要素,由试点实验得到使用者行为的实测数据。

[0020] 进一步地,所述步骤3具体为:

[0021] 步骤3.1、利用C4.5算法建构决策树模型,并应用决策树模型建立行为基础数据库,预估使用者室内活动概率;

[0022] 步骤3.2、通过规则归纳法的Rule Induction Operator算法对决策树进行修剪,去除决策树中对分类不添加有用信息的节点,学习行为模式的规律;

[0023] 步骤3.3、应用聚类分析法的K-means算法对行为模式进行最后的分类,并使其能够形成使用者行为配置文件。

[0024] 进一步地,所述抽样调查采用现场实测、问卷调查及访谈方式进行,其中所述现场实测内容包括建筑形态及空间几何参数以及非建筑几何参数,所述问卷调查包括对严寒地区使用者基本信息的抽样调查和严寒地区使用者耗能行为及影响要素的调查。

[0025] 进一步地,所述决策树对一组数据分隔为预定义的类,并对给定的数据进行描述、分类及定义,以建立一个分类模型,基于输入属性得到预测价值的目标属性,决策树模型的每个内部节点都对应一个预测属性,内部节点的分支数目与相应的预测属性的可能值的数目相等。

[0026] 进一步地,决策树模型的建立分为两个步骤,即学习和分类;在学习过程中,将数据集中的记录数据自动且随机分为两组,即训练组和校验组;在分类过程中,通过交叉验证法对决策树的准确性进行验证以估测学习过程的性能表现。

[0027] 进一步地,在交叉验证过程中,数据集被划分成大小相等的k个子集,在k个子集中,一个子集作为测试数据集,其余k-1个子集作为训练数据集,交叉验证过程重复k次,k个子集中的每一个子集都作为测试数据,k次迭代的k次结果将取平均值,得到最后唯一的估计值;其中k取10。

### 附图说明

- [0028] 图1是基于机器学习的使用者行为模式建构方法示意图；  
 [0029] 图2是决策树模型分支流程图；  
 [0030] 图3是决策树模型生成过程图；  
 [0031] 图4是数据集群图；  
 [0032] 图5是先验信息、样本信息与后验信息的关系图。

### 具体实施方式

[0033] 下面将结合本发明实施例中的附图对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0034] 本发明提供一种基于使用者行为的严寒地区建筑节能设计优化方法:

[0035] 步骤1、针对严寒地区气候区,进行数据的采集,通过抽样调查和试点实验获取建立模型的基础数据;

[0036] 首先,针对严寒地区气候区,进行数据的采集,通过抽样调查和试点实验获取建立模型的基础数据。由调研数据得到建筑空间特征,使用者的属性及特征,促动使用者各类行为的建筑及其环境要素;由试点实验得到使用者行为的实测数据。其中,建筑空间特征及使用者的属性为进一步研究促动使用者行为要素及行为模式提供了事实依据和理论支持。

[0037] 1) 抽样调查

[0038] 抽样调查是使用者行为及其影响因素研究中经常使用的数据收集方法。对已建成建筑进行抽样调查,包括建筑基本信息和使用者特征,得到严寒地区某一类建筑的典型模型数据组 and 使用者属性及动态特征。并通过调研将得到使用者行为的主要促动因素及其影响程度,为建立促动使用者节能行为的建筑空间模式及参数阈值提供数据支持。抽样调查以地域原则、规模原则、类型原则为原则,以覆盖更多建筑样本,满足样本分布符合贝尔形状曲线,通过样本来反应总体。抽样调查采用现场实测、问卷调查及访谈方式进行。通过现场实测获得所调研建筑的基本设计参数值;通过问卷调查主要收集使用者的基本社会属性、对建筑空间环境的认知情况、及其耗能行为与节能行为等信息;通过访谈使问卷的内容与结构更合理化。

[0039] ①现场实测现场实测的主要内容包括建筑形态及空间几何参数(表1)及非建筑几何参数(表2),分析实测得到的数据可以得到各类建筑的典型模式 and 设计参数值域,将作为建立典型建筑模型组的依据,作为模拟中的建模的信息基础。

[0040] 表1建筑形态及空间几何设计参数

类别	参数	
[0041] 建筑形态几何参数	建筑朝向	
	建筑高度	建筑首层高度; 建筑标准层高度; 建筑层数

[0042]	建筑开窗	建筑各朝向开窗数量；首层窗高；首层窗宽； 标准层窗高；标准层窗宽
	建筑空间几何参数	建筑平面形式简图；空间模式； 建筑平面设计尺寸；主要空间尺寸

[0043] 表2非建筑几何参数

	类别	参数
[0044]	建筑围护结构热工性能	外墙面导热系数；屋顶导热系数； 外墙面构造形式
	建筑透明介质热工性能	各朝向窗导热系数
	建筑设备热工性能	建筑采暖系统；建筑空调系统； 采暖期天数；灯具数量； 灯具类型；照明分区； 灯具照明效率；灯罩隔热系数

[0045] ②问卷调查问卷调查主要包括两个部分，即严寒地区使用者基本信息抽样调查和严寒地区使用者耗能行为及影响要素调查。主要采用网络调查方式，调研样本应符合贝尔形状曲线。调研内容和结构需合理设置，并通过信度和效度的检验。在正式调研前，采用预调研和访谈的形式使调研的内容更具有合理性。

[0046] 严寒地区使用者基本信息及行为特征调查主要包括人员所在空间类型，空间规模，最多在室人数，采暖方式，开关窗的原因，各季节开窗次数，开关灯的原因，各季节开灯时长的差异，使用空间是否采用遮阳，如果有遮阳类型，各季节使用频率及改变遮阳设备状态的原因。通过这部分的问卷调查的一定气候条件下特定建筑类型的空间特征、使用者的基本信息及影响行为的主要因素，以作为进行下一阶段使用者行为及其影响要素调查问卷设计的依据，并为制定试点实验实测方案及后续模拟工作构建建筑空间模型提供数据支持。

[0047] 严寒地区使用者耗能行为及影响要素调查主要内容主要包括三方面：

[0048] 第一方面为受试者的基本信息，主要为回答问卷人员的年龄、性别、学历、收入、职业等。

[0049] 第二方面是受试者回答问卷时所在建筑的基本信息，包括受试者的所在建筑地点，建筑年龄，建筑空间模式及尺寸，开窗形式、数量及尺寸，供暖方式等。

[0050] 第三方面为受试者所在建筑空间所在状态，包括回答问卷的时候是否开启了窗、灯及遮阳，此时对热感觉、空气质量等指标的评价，开关灯、开关窗、改变遮阳状态的主要原因。受试者在接受调查时的室内环境感受及近14天的感受，在近14天耗能行为相关问题。

[0051] 在调研后分析以下与建筑状态（窗，制冷，照明及遮阳情况）有关的参数，及各参数与建筑状态的关系。受试者调研时刻的热感觉；受试者调研时刻感知的照度；受试者调研时刻感知的室内空气质量；受试者调研时刻感知的噪声水平；调研时刻的室外温度、风速、室



外水平太阳辐射;调研当日的太阳辐射时间;受试者年龄、性别;室内人数、平均年龄及性别构成;所在建筑空间面积;供暖及制冷方式。

[0052] 2) 试点实验

[0053] 试点实验是在实测使用者行为经常使用的记录方法。由于严寒地区办公建筑在气候、地域的特殊性,多数建筑均采用集中供热,且室内无温控构件,在本发明中不对温控行为进行实测及研究。依据初步调研仅有较少建筑采用了遮阳设备,较难对人员的遮阳行为进行统计,本发明仅在抽样调查中对遮阳行为进行相关数据的统计。因而本发明分别对一类建筑的不同平面类型建筑空间的开窗行为、照明行为与使用者作息三方面,对于行为的开启程度、开启时间和有重要影响的行为节点进行记录。实测样本量应包括不同朝向的各建筑类型,且比例均衡。测量需保证一定的测量时间,并应覆盖不同季节。

[0054] 在实测时,应用hobo UX90-005m、hobo UX90-006m感应器对照明开关情况、室内照度变化及人员在室情况进行记录,并利用hobo U12-012对室内照度情况进行实测;应用hobo U90-001m感应器对开关窗行为进行检测。

[0055] 所采用设备主要实测开窗行为、照明行为、使用者在室情况(模拟中的重要参数输入部分)及室内照度变化,均为长期监测型设备,在考虑在未来可进行更长期监测的前提下,采用大内存型号,如表3所示。所得数据在本方法中利用机器学习算法中的数据挖掘技术形成行为模式模型。

[0056] 表3使用者行为试点实验设备概况及应用要点

	设备名称	测试目的	输出数据模式	数据应用方法
[0057]	hobo UX90-001m	开窗行为	开关窗机械变化次数	形成使用者行为模式模型
	hobo UX90-005m	小规模空间(<5m) 照明行为 使用者在室情况	开关灯次数 使用者在室人数及时间	
	hobo UX90-006m	大规模空间(<12m) 照明行为 使用者在室情况	开关灯次数 使用者在室人数及时间	
[0058]	hobo U12-012	室内照度变化 室内温度(℃) 相对湿度(%)	室内照度变化 室内温度(℃) 相对湿度(%)	影响行为要素研究; 照明行为监测的辅助测量

[0059] 步骤2、根据数据采集的结果建立严寒地区建筑空间典型模型组,分析使用者的属性特征建立促动节能行为的建筑设计参数组,得出能够导向节能行为的建筑空间类型及参数值域;

[0060] 步骤3、在RapidMiner平台上应用数据挖掘技术中的决策树模型建立使用者行为基础数据库,利用规则归纳算法对决策树模型进行修剪以得到行为模型规则,再通过聚类分析法得到使用者随机行为模式模型,其结果直接作为参数输入能耗模拟平台Energyplus,并对模型结果进行验证,再通过Matlab平台进行多项逻辑回归初步建立建筑能耗预测模型;

[0061] 1) 促动节能行为设计参数组

[0062] 应用广义加性模型的逻辑回归对建筑行为及其各影响要素的分析。因无法预设各行为影响因素与使用者行为的关系为线性相关,采用光滑的非线性函数来描述其关系,利用似然比检验在5%水平进行显著性分析。利用统计学软件R进行以上过程的分析。并最终得到开窗、关灯行为等节能行为较大概率的建筑空间的各设计参数值域。

[0063] 2) 使用者行为模式模型

[0064] 以调研及实测数据为基础,提出基于机器学习的使用者随机行为模式建构方法,如图1所示。主要分为三个步骤:首先,利用C4.5算法建构决策树模型,并应用其建立行为基础数据库,预估使用者室内活动概率;然后,通过规则归纳法的Rule Induction Operator算法对决策树进行修剪,去除决策树中对分类不添加有用信息的节点,学习行为模式的规律;最后,应用聚类分析的K-means算法对行为模式进行最后的分类,并使其能够形成链接常用模拟软件Energyplus使用者行为配置文件。应用机器学习技术对使用者行为进行挖掘,这一方法通过机器的自学习过程,更理性地定义行为模式,与传统方法相比,能够更深入、准确的刻画使用者行为模式。行为模式建构过程在可视化数据挖掘平台Rapidminer进行,该平台具有良好的适用性和应用性,其图形界面易于操作。目前,无论在住宅或是办公建筑方面,对于建筑行为模式模型的验证与校验研究数量很少。2009年Haldi和Robinson提出运用交叉验证法对行为模型进行校验。交叉验证法是评价模型预测性能的最常用方法之一,他是通过将数据进行分割,一部分数据用于拟合模型,一部分用于对模型数据的校验和修正。在其他研究中,均采用Haldi和Robinson的方法进行验证,并论证其可能性。Schweiker等人利用此方法验证了多个开窗行为模型的准确率;Langevin等人运用此方法验证了基于实地调研建立的某空调制冷办公空间模型的准确性。本发明也采用Haldi和Robinson于2009年提出的行为模型校验方法对所建立的行为模型进行验证。

[0065] 数据挖掘一般是指从大量的数据中通过算法搜索隐藏于其中信息的过程。数据挖掘通常与计算机科学有关,并通过统计、在线分析处理、情报检索、机器学习、专家系统和模式识别等诸多方法来实现上述目标。

[0066] 在本发明中利用决策树模型建立使用者行为基础数据库,利用规则归纳算法对决策树模型进行修剪,以得到行为模型规则,再通过聚类分析法得到使用者行为模式。

[0067] ①决策树模型

[0068] 决策树是在已知各种情况发生概率的基础上,通过构成决策树来求取净现值的期望值大于等于零的概率,是直观运用概率分析的一种图解法,属于有监督学习。

[0069] 由于决策树是一种分支流程图图形的分类模型,其数据表示易于理解和操作。在本发明中利用决策树对一组数据分隔为预定义的类,并对给定的数据及进行描述、分类及定义,以建立一个分类模型,如图2所示,基于一定的输入属性(预测属性)得到预测价值的目标属性(标签属性)。树模型的每个内部节点都对应一个预测属性,内部节点的分支数目

与相应的预测属性的可能值的数目相等。

[0070] 决策树模型的建立分为两个步骤,即学习和分类,如图3所示。在学习过程中,将数据集中的记录数据自动且随机分为两组:训练组及校验组。然后,利用决策树相关算法形成决策树模型。在本发明中,将使用C4.5算法,应用预测性分析和数据挖掘软件RapidMiner建立决策树模型。

[0071] 在构建决策树时,C4.5算法通过评估“增益”(收益率)来处理未知属性值记录的数据集。增益值的计算方法如下:

$$[0072] \quad Gain(\vec{y}, j) = Entropy(\vec{y} - (j|\vec{y})) \quad (1)$$

[0073] 其中,

$$[0074] \quad Entropy(y) = -\sum_{j=1}^n (y_j/y) \log (y_j/y) \quad (2)$$

$$[0075] \quad Entropy(j/y) = (y_j/y) \log (y_j/y) \quad (3)$$

[0076] 其中,Entropy为熵,y为数据集,j为分割参数。

[0077] 这个过程使用熵(entropy)衡量数据的无序性,并最大的提升增益。在分类过程中,通过交叉验证法对决策树的准确性进行验证,以估测的学习过程的性能表现。在交叉验证过程中,数据集被划分成大小相等的k个子集。在k个子集中,一个子集作为测试数据集,其余k-1个子集作为训练数据集。交叉验证过程重复k次,k个子集中的每一个子集都作为测试数据。k次迭代的k次结果将取平均值,得到最后唯一的估计值。在本发明的验证中k的取值拟为10。当达到一定精度时,决策树能够对新的数据集进行分类和预测。

[0078] ②规则归纳算法

[0079] 关联规则的一般形式为  $X \Rightarrow Y$ , 其中X、Y表示数据集中的实例(即数据库中的记录)包含的项或项集。项表示为一个“属性-值”对,项集是多个项的连接。度量关联规则的参数是指提升X和Y的共同概率,表示为  $\sigma(XUY)$ 。支持度揭示了规则的重要性。置信度(Confidence)表示为实例在包含X的前提下,也包含Y的条件概念,即  $\sigma(XUY) / \sigma(X)$ , 它表示规则的可信程度。

[0080] 规则归纳法是一种在数据集上生成规则集的分类数据挖掘技术。规则归纳法具有对一阶逻辑易于理解及表示的特点,并且容易获得先验信息。各种规则归纳算法在机器学习和应用数据挖掘文献领域中广泛应用。这些算法以似然模型预测方法为代表,可用于预测、分类和回归及大数据的描述性建模。在本发明中通过规则归纳法对决策树进行修剪,去除决策树中对分类不添加有用信息的节点。并通过将过拟合的和过特定的树模型转化为一般模型,来提升不可见数据集的预测能力。

[0081] 本发明中,信息增益(information gain)已被用来作为选择属性和数值分裂的规则归纳的标准参数。在决策树模型中进行这一类似过程,所有的属性的熵(公式1)进行计算,选择具有最小熵属性参数进行分裂。利用Rule Induction Operator算法应用到对应的数据集进行反复的生长和修剪,目标为指导错误率低于50%。

[0082] ③聚类分析法

[0083] 聚类分析是对于静态数据分析的一门技术,在包括机器学习的许多领域受到广泛应用,如数据挖掘,模式识别,图像分析以及生物信息。聚类是把相似的对象通过静态分类

的方法分成不同的组别或者更多的子集(subset),这样让在同一个子集中的成员对象都有相似的一些属性,常见的包括在坐标系中更加短的空间距离等。一般把数据聚类归纳为一种非监督式学习。

[0084] 类分析是将数据合并成不同的集群的过程,其中,一类数据在同一个群组的情况下具有很高的相似性,另一类在不同的集群情况下具有较低的相似性。集群之间的相似性通常是基于群组间距离计算的,距离通过欧几里得距离度量来获得:

$$[0085] \quad d(a,b) = d(b,a) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \dots + (b_n - a_n)^2} \quad (4)$$

[0086] 其中,  $a = (a_1, a_2, \dots, a_n)$ ,  $b = (b_1, b_2, \dots, b_n)$

[0087]  $a, b$ 是 $n$ 维欧氏空间中的两个点。

[0088]  $k$ -均值聚类是一种用于数据挖掘中的聚类分析的矢量量化方法。鉴于该算法的简单性,它是一种广泛使用的分类技术。假定一个数据集 $D$ ,包含一个数 $n$ 的记录或实例,群组数 $k$ 需指定。每个集群都与一个质心(中心点)相联系,这个质点是各集群点的平均距离,每个点被分配到最接近的质心的群集。

[0089] 通过对聚类距离性能来评价聚类模型的性能。在本发明中,利用Davies-Bouldin Index (DBI) 指数来评价性能。DBI指数是指群组内的平均距离和群组外距离之间的比值,通过以下公式进行运算。

$$[0090] \quad E = \frac{1}{n} \sum_{i=1}^n \max_{i \neq j} \left[ \frac{R_i + R_j}{M_{ij}} \right] \quad (5)$$

[0091] 其中, $n$ 为群组数,

[0092]  $R_j$ 为通过平均每个群组对象和群组中心之间的距离,所得到的群 $i$ 和群 $j$ 之间的平均距离,

[0093]  $M_{ij}$ 为各群组中心距离。

[0094] 根据以上公式,较小的DBI值表明聚类算法的性能更好。低DBI指标的群组,代表聚类产生低集群内距离(集群高相似度)和高群组间距离(群组低相似度), $k=n$ 算法,其中 $k$ 为群组数, $n$ 为样本点,能够产生适宜的聚类,如图4所示。

[0095] 3) 建筑能耗预测模型

[0096] 利用拉丁超立方抽样(Latin Hyper-cube Sampling)从典型模型组中抽取建筑样本进行能耗模型计算,使少量的能耗样本计算能够代表严寒地区建筑的能耗情况,降低随机性对结果的影响。模拟过程通过EnergyPlus 8.0来完成,拉丁超立方样本和模拟过程通过jEPlus来协调。概率密度函数(Probability Density Function)及变异系数均方根(Root-Mean-Square Deviation)用于衡量模拟结果与实际能耗的差距。

[0097] 应用多变量逻辑回归分析建立使用者行为模式与建筑设计要素和能耗的定量关系,所得数学模型通过F检验及可决系数(R-squared)对其拟合度等进行判断,均匀分布的伪随机整数用于修正不确定参数的范围和数值。

[0098] 步骤4、采用机器学习技术中贝叶斯理论支持下的高斯过程对模型进行校正和优化从而建立严寒地区建筑能耗优化预测模型;

[0099] 步骤5、模型校正后进行数据校验,通过与实例对比,论证预测模型的预测能力和在建筑节能设计中应用的可行性,并通过回归分析得到可生成节能的建筑设计参数值域,

建立优化节能性能设计参数组；

[0100] 采用机器学习技术中贝叶斯理论支持下的高斯过程，对模型进行校正和优化，以减少随机性及不确定性对模型预测精度的影响，提升能耗预测模型的预测能力，最终建立严寒地区建筑能耗优化预测模型。在模型矫正后进行数据校验，通过与实例对比，论证预测模型的预测能力，和在建筑节能设计中应用的可行性。并通过回归分析得到可生成较节能的建筑设计参数值域，建立优化节能性能设计参数组。

[0101] 1) 贝叶斯学习

[0102] 贝叶斯学习方法是基于贝叶斯定理和贝叶斯假设发展起来的一种系统的统计推断和决策方法。贝叶斯定理是贝叶斯学习理论运用概率表示所有形式的不确定性来实现学习和推理过程的重要工具。

[0103] 贝叶斯定理是关于随机事件A和B的条件概率(或边缘概率)的一则定理，是基于先验信息、总体信息及样本信息的统计理论。贝叶斯定理能够将一个事件的先验概率和后验概率联系在一起，同时充分运用先验信息以及已知的样本数据来确定这个事件的后验概率。

[0104] 总体信息，即总体分布或总体所属分布族给我们的信息。样本信息，即从总体抽样的样本给我们提供的信息。这是最“新鲜”的信息，并且数量越多越好。先验信息，即在抽样之前有关统计问题的一些信息，先验信息主要来源于经验和历史资料。基于上述三种信息(总体信息、样本信息和先验信息)进行的统计推断被称为贝叶斯统计学。贝叶斯统计中的两个基本概念是先验分布和后验分布。先验分布：总体分布参数 $\theta$ 的一个概率分布。后验分布：根据样本分布和未知参数的先验分布，求出的在样本已知下，未知参数的条件分布。两者关系：先验分布 $\pi(\theta)$ 是指能反映在抽样前时人们对 $\theta$ 有一定的认识，后验分布 $\pi(\theta/x)$ 是能反映在抽样后人们对 $\theta$ 的认识。之间的差异可以看作是是由于样本出现后对 $\theta$ 认识作出一定的调整。因此，用总体信息和样本信息对先验分布 $\pi(\theta)$ 作调整后的结果可以认为是后验分布 $\pi(\theta/x)$ 。

[0105] 设 $\theta$ 为连续性的随机变量，涉及 $\theta$ 样本信息设为 $x$ ， $\pi(\theta)$ 为先验密度， $\pi(\theta/x)$ 为后验密度， $p(x/\theta)$ 为似然函数，则连续参数的贝叶斯模型表示为：

$$[0106] \quad \pi(\theta/x) = \frac{p(x/\theta)\pi(\theta)}{\int p(x/\theta)\pi(\theta)d\theta} \quad (6)$$

[0107] 这个在样本 $x$ 给定下，式(6)即称为 $\theta$ 的后验分布。它集中体现了总体、样本和先验等三种信息中有关 $\theta$ 的所有信息，而又是排除了一切与 $\theta$ 无关的信息之后得出的结果。故基于此，后验分布 $\pi(\theta/x)$ 对 $\theta$ 统计推断显得更为有效，也是最合理的。图5表示了先验信息、样本信息与后验信息的关系。

[0108] 在使用贝叶斯定理过程中，通常的学习过程如下：首先对要学习模型的权空间或函数空间做一定的先验假设，然后对其进行贝叶斯推断，同时利用已经假定的先验分布 $p(\theta)$ 和得到的观测样本数据相结合的似然，从而得到后验密度。贝叶斯学习中经常用到的先验分布有无信息先验分布和共轭先验分布。在无法通过以往任何的知识来帮助我们确定 $p(\theta)$ 的情况下，贝叶斯提出了一个原则，即我们可以使用均匀分布来作为 $p(\theta)$ 。在这种情况下，参数在它的变化范围内取各个值的概率是平等的，这种确定先验分布的原则就是贝叶

斯假设。

[0109] 2) 高斯过程

[0110] ①基础理论高斯过程指的是一组随机变量的集合,这个集合里面的任意有限个随机变量都服从联合高斯分布。高斯过程任意的有限变量集合的分布都是高斯型,即对任意整数 $n \geq 1$ 及任意的一组随机变量 $t_1, \dots, t_n \in T$ ,与其对应的t时刻的过程状态 $f(t_1), f(t_2), \dots, f(t_n)$ ,的联合概率分布服从n维高斯分布。

[0111] 高斯过程是基于统计学习理论和贝叶斯理论发展起来的一种机器学习方法,适于处理高维度、小样本和非线性等复杂回归问题,且泛化能力强,与神经网络、支持向量机相比,高斯过程具有容易实现、超参数自适应获取、非参数推断灵活以及输出具有概率意义等优点。在统计学和机器学习两个领域中,共享通用一些基本理论和算法。统计学的一个基本关注点在于对数据和模型关系之间的理解,而机器学习的目标主要是用于更准确的预测和对学习算法行为的理解上。机器学习是一种黑盒算法,而统计学更希望得到模型的理论解释。高斯过程模型在某种层面将统计学和机器学习联系到了一起。高斯过程模型包括贝叶斯线性模型、样条模型、合适条件下的神经网络,并且其也与支持向量机有密切的联系。

[0112] ②应用方法本发明在建立优化能耗预测模型时通过高斯过程进行随机过程以确定能耗的概率。这一随机过程能够处理所有可能的随机参数的集合,并形成函数的广义概率分布。根据概率理论,高斯过程可以很容易地处理许多近似正态分布随机变量。这一过程遵循贝叶斯理论,通过观察到的数据修改之前的分布,以实现目标分布。这一推理方法已被用于建筑能耗模型的参数矫正。根据输出的类型,回归和分类都通过高斯过程来确定。回归分析用来处理实际数据的连续输出,而分类用于考量离散输出分类的标签。

[0113] 首先,将优化的随机样本作为训练数据。其次,形成如协方差函数等高斯过程的先验信息。目前,有多种协方差函数适合本发明,但定义协方差函数的超参数的适宜值的会产生诸多问题。虽然数值在过程中已被优化,但超参数的先验分布需要被预定义。在本发明中,采用最广泛使用的平方指数协方差函数(Squared Exponential)。这种协方差函数必然需要两个超参数:长度规模和幅度。长度尺度的逆演示了在这个过程中的输入的相关性,而幅度表示未知函数值的方差。在本发明中应用高斯过程来完成超参数的分析。最后,高斯过程通过基于nested Expectation Propagation(nested EP)算法的多项Probit模型建构。相比蒙特卡罗马尔科夫链(Markov Chain Monte Carlo),nested EP同样也具有微小不精确性,但计算时间较短。以上过程通过MATLAB 2014a的GP-Stuff来完成。

[0114] 步骤6、耦合促动节能行为和优化节能性能的建筑节能设计参数组,得到控制办公建筑能耗的空间类型及参数值域,建立办公建筑节能设计优化参数组;

[0115] 步骤7、以参数选择和设计流程为基础设计严寒地区办公建筑节能优化策略,以优化的参数为基础、以优化的预测模型为能耗性能评价工具设计得到优化的建筑节能设计方案。

[0116] 本发明通过调研使用者与建筑的互动关系,在考虑使用者随机行为前提下,建立反映建筑能耗水平的预测模型,耦合促动使用者节能行为的建筑要素,得到建筑节能设计优化参数阈值,提出严寒地区建筑节能设计优化设计流程及设计策略,能够得到更理性化的建筑节能设计结果,增强建筑性能化设计的应用性。

[0117] 在理论层面,本发明综合使用者耗能及节能行为与建筑设计因素对建筑能耗性能

的影响,建立在行为学、心理学、建筑学、数学、计算机科学等多学科交叉的基础上,获得多重视角下的研究成果。构建了基于使用者行为节能建筑优化设计方法,为建筑节能设计带来设计方法上的启发。

[0118] 在实践应用层面,本发明利用机器学习算法降低行为模型及预测模型建立过程中的不确定因素影响,优化建筑性能模拟结果,提升预测模型精确度和性能模拟的实际应用性,使建筑师能够在建筑设计阶段进行更有效的能源规划和建筑方案设计。为建筑师提供基于使用者行为的办公建筑节能设计优化参数阈值及设计流程,提出设计建议和要求,有利于建筑师进行更合理化、人性化的设计。

[0119] 本发明在理论与应用实践两方面,为建筑节能设计提供了更理性化的新思路。通过对严寒地区建筑使用者随机行为中几个重要节点的研究与分析,综合考虑人、建筑与环境的相互作用与影响;并应用机器学习算法降低随机性和不确定性,提升建筑能耗预测模型精确度,优化建筑性能化设计结果;从通过适宜的建筑环境促动使用者的节能行为这一角度,提出办公建筑节能设计优化设计流程及策略;为严寒地区建筑的节能设计提供理论支持,也为建筑设计师提供具体指导与建议。

[0120] 以上对本发明所提供的一种基于使用者行为的严寒地区建筑节能设计优化方法,进行了详细介绍,本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本发明的限制。

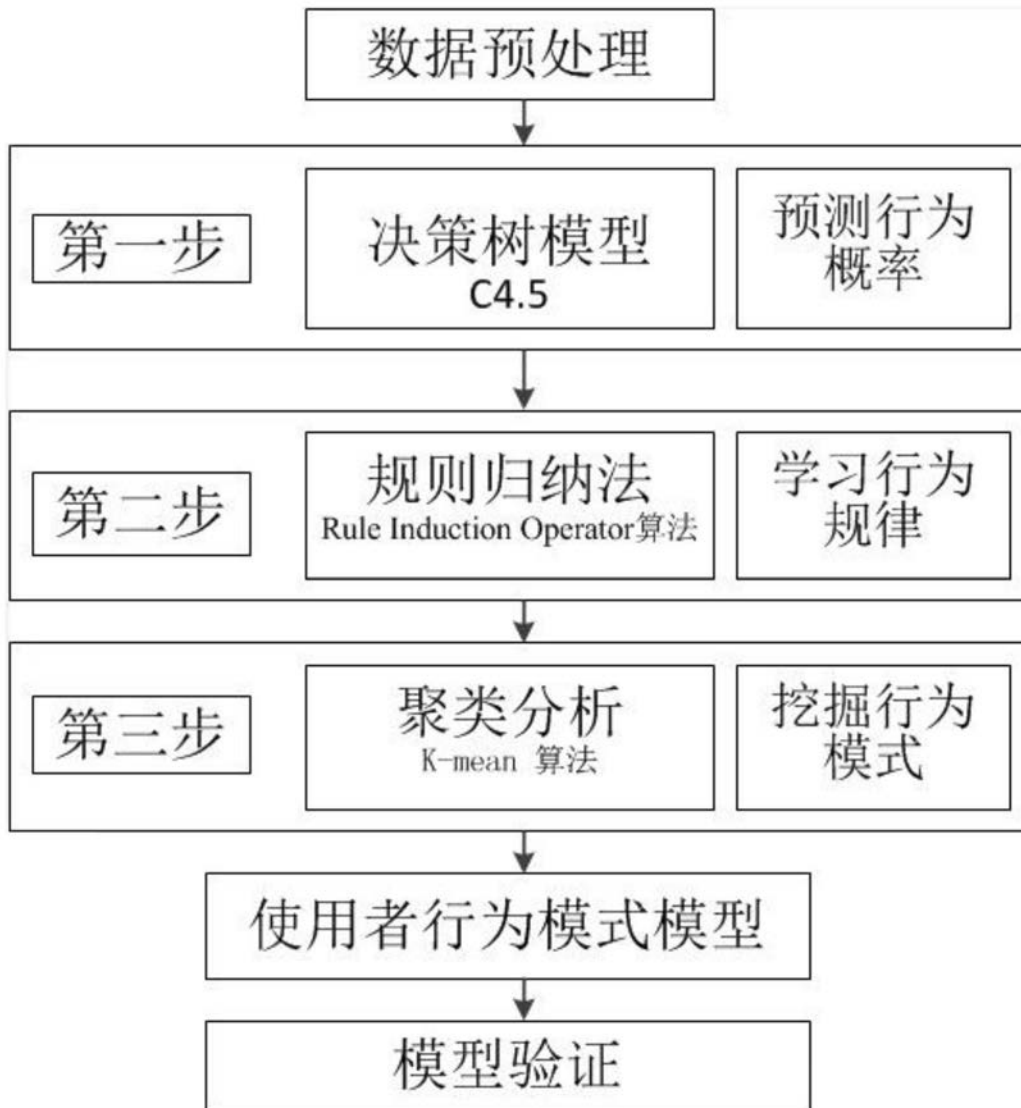


图1



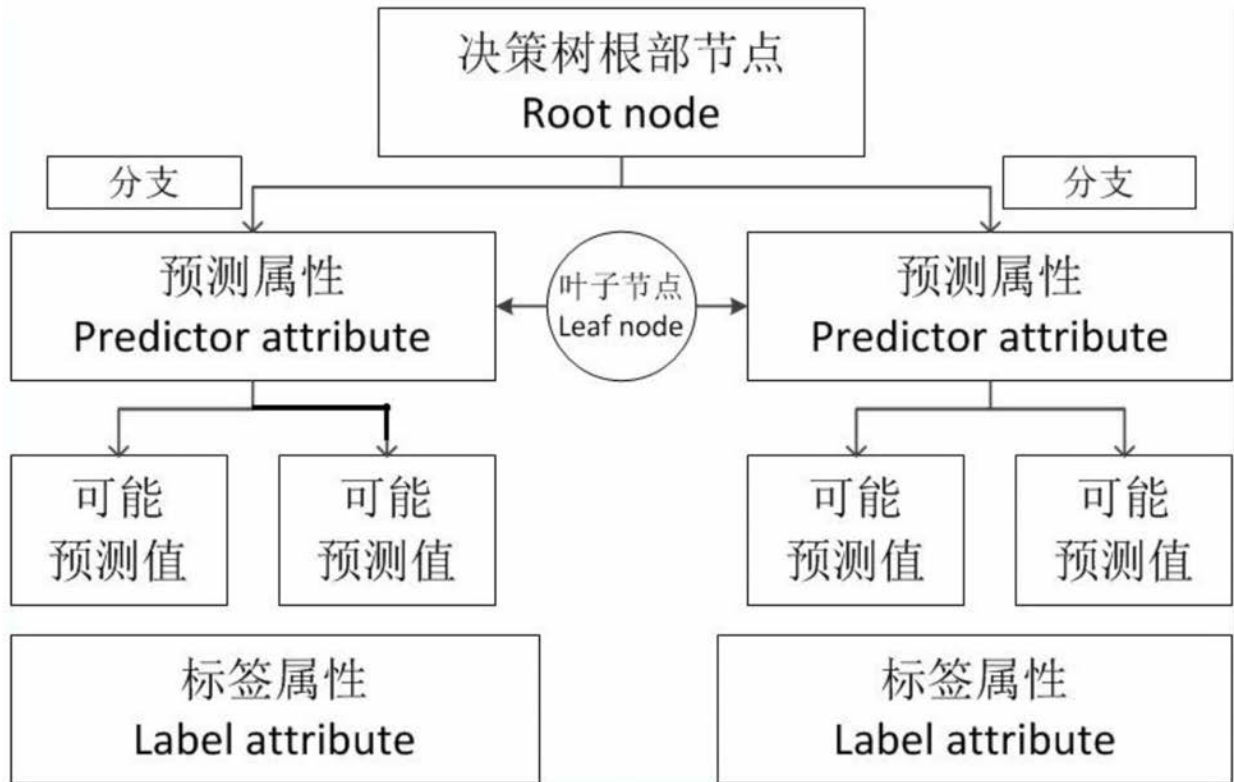


图2

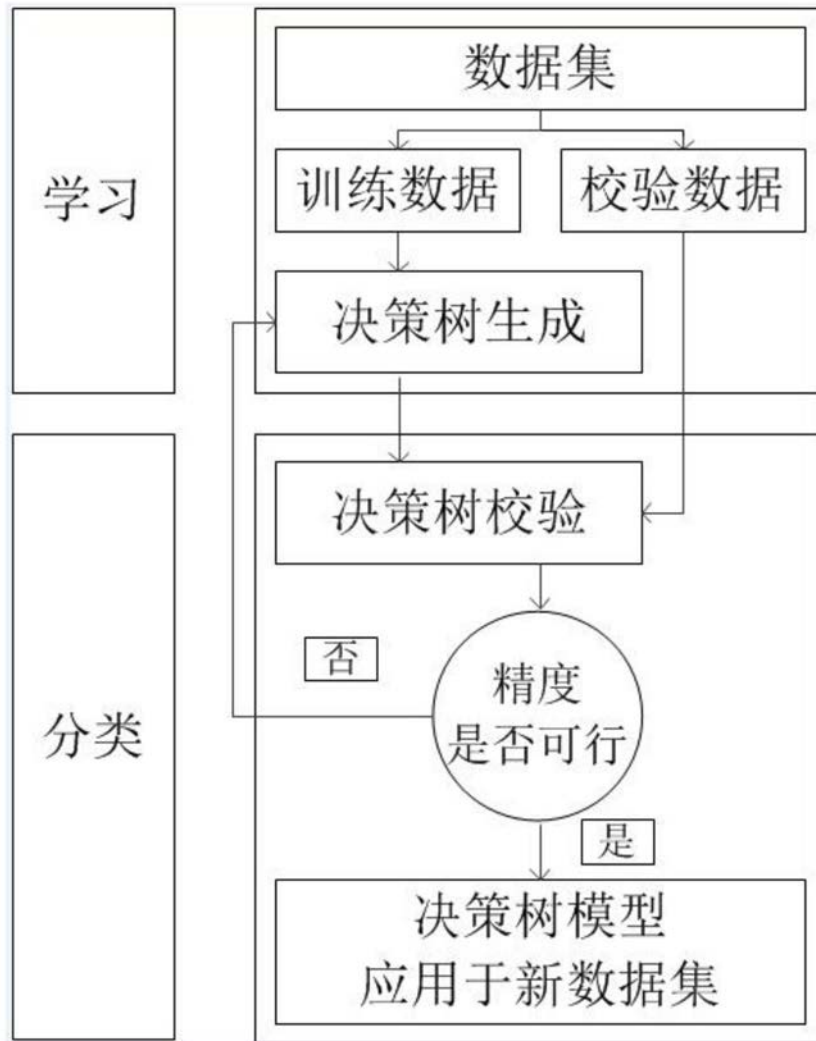


图3

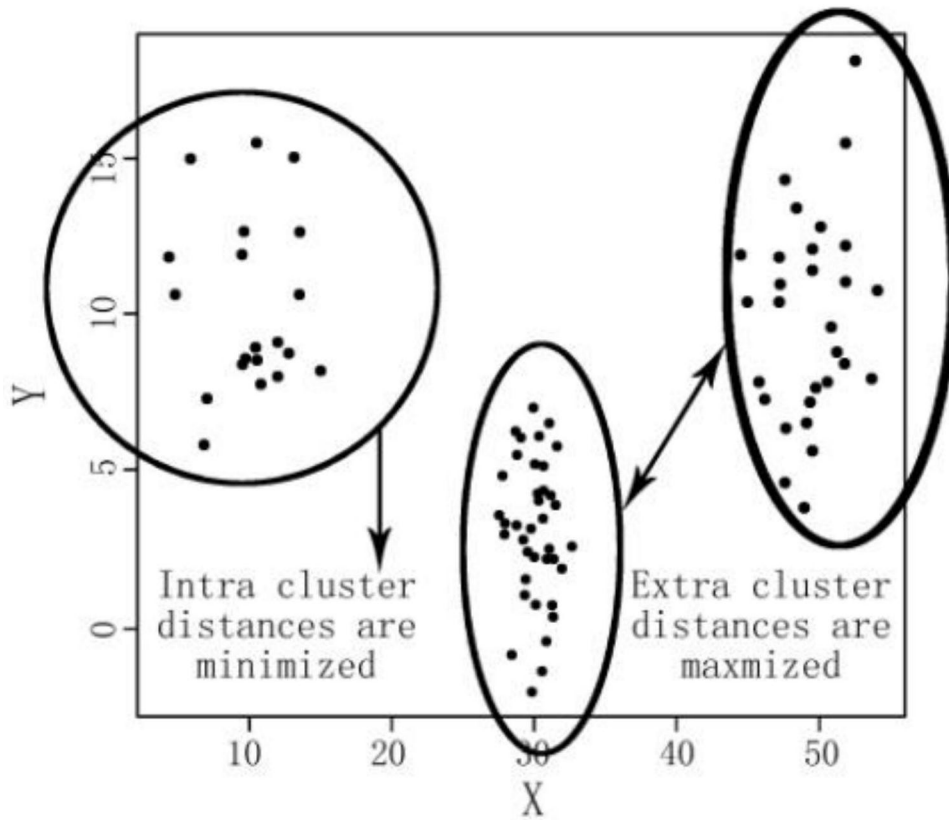


图4



图5