

数据挖掘技术及其在建筑节能中的应用^①

张运楚^{1,2}, 韩怀宝¹, 杨红娟^{1,2}, 杨崇涛¹, 王兆斌¹

¹(山东建筑大学 信息与电气工程学院, 济南 250101)

²(山东省智能建筑技术重点实验室, 济南 250101)

摘要: 人类社会中的每次技术进步都会催生一系列新的产品和服务, 但同时也导致资源和能源消耗的剧增. 技术的进步虽然提高了资源和能源的利用效率, 但这种人均能耗不断递增的发展模式不可持续. 建筑节能除了关注供应侧的能效外, 合理的引导需求侧用能是实现建筑节能的关键. 要实现建筑节能模式由供应侧到需求侧的转变, 就必须恰当描述特定室内环境下的用能特征, 才能从需求侧评估建筑能耗的合理性, 进而精确辨识能源浪费的原因. 建筑物自动化系统和物联网技术的快速发展与普及, 获取了大量特定室内环境下的用能特征数据, 利用数据挖掘技术可以从这些低密度价值的建筑运维数据中萃取节能线索和策略. 本文简述了数据挖掘技术, 综述了各种挖掘方法在建筑节能中的应用, 并展望了发展趋势.

关键词: 数据挖掘; 算法; 能耗分析; 建筑节能; 关联挖掘; 分类; 聚类; 神经网络

引用格式: 张运楚, 韩怀宝, 杨红娟, 杨崇涛, 王兆斌. 数据挖掘技术及其在建筑节能中的应用. 计算机系统应用, 2017, 26(9): 151-157. <http://www.c-s-a.org.cn/1003-3254/5869.html>

Data Mining Technology and Its Application in Building Energy Efficiency

ZHANG Yun-Chu^{1,2}, HAN Huai-Bao¹, YANG Hong-Juan^{1,2}, YANG Chong-Tao¹, WANG Zhao-Bin¹

¹(School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China)

²(Shandong Provincial Key Laboratory of Intelligent Buildings Technology, Jinan 250101, China)

Abstract: Every technological advance in the development of human society will lead to a series of new products and services, but at the same time, it will lead to a sharp increase in the consumption of resources and energy. Although the progress of technology has improved the efficiency of resource and energy use, the development mode of increasing per capita energy consumption is not sustainable. In addition to the supply side of energy efficiency, a reasonable guide to demand side can be the key used to achieve building energy efficiency. To achieve the change of building energy-saving mode from the supply side to the demand side, we must properly describe the energy by using characteristics in a specific indoor environment, and assess the reasonableness of building energy consumption from the demand side, then identify accurately the reason of energy waste. Along with the rapid development of building automation system and IOT technology, a large amount of building energy consumption data with specific indoor environment features are acquired, then we can use data mining technology to extract energy saving clues and strategies from these low density value building daily operation data. This paper briefly introduces the data mining technology, and summarizes the application of various mining methods in building energy saving, and prospects of its development trend.

Key words: data mining; algorithm; building energy consumption analysis; building energy efficiency; association mining; classification; clustering; neural network

① 基金项目: 国家自然科学基金青年基金(61303087)

收稿时间: 2016-12-06; 采用时间: 2016-12-22

随着建筑行业的快速发展,建筑能耗问题越来越突出.建筑能耗在总能耗中的占比不断上升,对能源造成巨大压力.建筑运行能耗作为建筑能耗的主要组成部分,浪费尤为严重.据统计,建筑运行使用能耗占建筑生命周期总能耗的80%~90%^[1,2].因此,实行建筑节能是发展资源节约型社会、实现经济的可持续发展的必要途径.如何从庞大的数据与信息中快速、高效的提取出蕴藏在其中的有用知识,发现能源使用过程中的漏洞,使建筑用能更加合理高效,已成为行业面临的关键问题.

常用的建筑能耗数据分析方法主要是通过模拟软件建模的方法定性、定量的分析建筑能耗指标,如DeST、EnergyPlus、DOE-2、TRNSYS、PKPM-CHEC等能耗模拟软件,该方法需要丰富的专业知识并建立精确的数学模型.但是,模拟假定的参数并不能反映实际运行过程中的能耗,而且很多实际情况难以模拟^[3].基于数据挖掘(Data Mining, DM)的能耗数据分析方法可以在没有假设前提的情况下挖掘信息,其对于知识的发现是未知的、非直觉的、甚至是违背常理的.随着建筑能耗数据的大量累积,数据挖掘在能耗分析中的作用越发凸显.

数据挖掘是一门交叉性学科,是从大数据中通过算法发现潜在有用知识的过程.它通过统计学、机器学习、专家系统、模式识别等方法 and 数据库技术搜索隐藏的信息和规律,在工业、商业等各个领域已得到广泛应用.本文首先介绍了数据挖掘技术的主要方法,分析了数据挖掘技术在建筑节能中的典型应用,最后展望了数据挖掘在建筑节能中的应用前景.

1 数据挖掘概述

数据挖掘也常称为数据库知识发现(Knowledge Discovery in Database, KDD),是指从大量不完全、含噪、模糊、随机的数据中,提取隐含在其中事先未知的潜在有用信息和知识的过程^[4].简言之,数据挖掘就是从存放在数据库的大量数据中挖掘有用知识的过程.不同于专家系统仅仅对已有知识进行学习,数据挖掘可以从数据库中发现隐含的新知识,其挖掘的知识是未知的.通过数据挖掘可以发现数据库中数据间隐藏的内在关系,其挖掘的知识可用于历史数据评价、发展趋势预测、过程控制、信息管理并对决策产生影响.与一般数据分析处理技术相比,数据挖掘有以下特点^[5]:

(1)海量性,数据挖掘的数据源是海量的;(2)可用性,数据挖掘的知识是有用的,挖掘结果可用于行为指导和趋势预测;(3)相对性,数据挖掘不是发现新的自然科学定理或数学公式,它发现的知识的相对的,有特定条件约束的,面向特定领域的.

数据挖掘是数据库技术与机器学习技术的结合,许多方法与机器学习方法相似.常用的数据挖掘方法主要包括关联分析、线性回归、分类分析、聚类分析、神经网络方法等.

1.1 关联挖掘(Association Mining)

关联挖掘也称关联规则挖掘(Association Rule Mining),是一种用于发现数据项目集间的依赖关系和关联关系的数据挖掘方法.关联规则是一种描述事件同时出现的规律和模式的方法,其描述了事件间同时出现的概率关系,推理所得结论并不具备因果关系.关联挖掘方法通过设定最小支持度和最小置信度阈值来判定数据或信息间的关系,高于该阈值的关联规则是我们感兴趣的.关联分析代表性算法有Apriori算法,FP-Tree算法及后期的一些改进算法^[6].

1.2 多元线性回归模型(Multivariable Linear Regression Model)

多元线性回归模型用于描述多个自变量与因变量间的相关关系.实际问题中,因变量的变化往往受到多个自变量的影响,当自变量与因变量间呈线性关系时,可用线性回归模型研究.线性回归模型常用于对历史数据的分析、预测未来趋势等.

1.3 分类(Classification)

分类属于一种特殊的预测方法,通过对训练集中样本特征或属性的提取、训练,建立特征的描述函数或模型,通过模型预测新数据所属的类别.常用的分类方法包括决策树法(Decision Tree)、支持向量机(Support Vector Machine, SVM)等.

决策树是经典的分类方法,它通过自顶向下递归的方式构造决策树,是以实例为基础的归纳学习方法.决策树是一个预测模型,代表对象属性和对象值间的映射关系,其结果为单一输出.ID3算法、C4.5算法是有代表性的决策树算法^[7].支持向量机是一种通过寻求结构风险最小化来提高学习机泛化能力的分类方法,通过在特征空间中找出最优分类超平面使类间隔最大化.SVM分类结果由少数支持向量决定,因此具有较好的适应能力与鲁棒性,分类准确度也较高.常用的支

持向量机有 LS-SVM, 模糊 SVM, 多分类 SVM 等^[8].

1.4 聚类(Clustering)

聚类是将具有较高相似度的类或簇聚集在一起, 相异的类或簇分离开来的过程. 相似度是基于描述对象的属性值来计算的, 距离是常用的度量方式. 同一簇中的对象具有较高的相似度, 可以被当作一个整体对待. 聚类属于无监督学习方法, 按聚类原理的不同通常可将聚类分为: 划分方法、层次方法、基于密度的方法、基于网格的方法、基于模型的方法等^[9].

1.5 神经网络(Neural Network)

神经网络是模拟人脑工作过程而抽象出的数学模型. 神经网络是由多个神经元(节点)组成的信息处理系统, 每个神经元都是基本处理单元节点, 节点间的连接代表连接该信号的加权值或权重值. 该过程可看作 m 维输入到 n 维输出的映射, 神经网络的性能可由输出的期望值与实际值间的差来衡量. 常用的神经网络有 BP(Back Propagation)网络, RBF(Radial Basis Function)网络等.

2 数据挖掘在建筑节能中的应用

公共建筑内设备种类及数量繁多, 产生海量的能耗数据. 随着国家出台相关规定需对建筑用能建立能耗监测系统, 对空调、照明插座、动力等实行分项计量后, 积累了大量的建筑能耗数据. 量大、高维是这些数据的主要特点, 传统的数据分析方法不能很好的发现其中蕴藏的知识, 而数据挖掘技术能很好的解决此类问题. 用数据挖掘发现建筑能耗知识的过程可分为四步: (1) 数据收集和预处理; (2) 建立能耗数据挖掘模型; (3) 挖掘结果评估; (4) 模型的应用.

2.1 关联分析的应用

挖掘能耗数据背后隐藏的能耗使用特征是实现建筑节能的重要方向. 关联挖掘是一种典型的关联分析方法, 可以有效的发现能耗数据间的联动关系及推理影响能耗指标产生的因素. 关联挖掘方法通过最小支持度(Min_sup)和最小置信度(Min_conf)阈值来判定数据或信息间的关系, 阈值由用户事先设定. AM 过程即为寻找满足高于 Min_sup 和 Min_conf 的强关联规则过程. 该过程可分为两步: 1) 找出所有频繁项集, 即满足的 $\text{sup}(X) \geq \text{Min_sup}$ 所有项集; 2) 找出强关联规则, 即频繁项集即中满足 $\text{conf}(X) \geq \text{Min_conf}$ 的关联规则.

2.1.1 Apriori 的应用

Apriori 是一种典型的迭代型关联挖掘方法, 最早由

R Agrawal^[10]提出, 后来的很多算法都是基于该思想. Apriori 方法通过迭代搜索频繁项集发现强关联规则. 苏江夺^[11]将 Apriori 算法应用到校园分项用电系统数据分析中, 分析某分项能耗变化与其他分项能耗变化间的耦合关系. 首先将校园分项能耗分为 8 个分项, 通过阈值比例和数据转化将能耗浮点数据转化成变化趋势型的符号数据(上升 \uparrow , 下降 \downarrow 和不变 $-$); 然后设定最小支持度 Min_sup、最小置信度 Min_conf 和其他相关配置参数; 最后通过校园平台分项能耗记录和 Apriori 算法搜索频繁项集, 发现了多条强关联规则(如结果显示空调末端用电、照明插座用电与其他分项用电存在耦合关系), 推理结论可为校园能源的管理改善提供参考依据. 张春杰^[12]通过 Apriori 算法生成的强关联规则发现了午休前后空调系统存在的不合理运行问题, 改善这些问题可以实现节能且提高空调系统运行效率.

将 Apriori 算法应用到建筑能耗数据分析中, 优势是思路简单、过程循序渐进, 缺点是如果数据库太大需要扫描次数太多, 且迭代过程中产生的候选集较大导致运算量大, 运行时间较长.

2.1.2 FP-growth 的应用

针对 Apriori 算法固有的缺点, Han J 等人^[13]提出了 FP-growth 算法, 该算法通过构建 FP_tree 找到频繁项集. FP-growth 算法与 Apriori 算法相比的改进: 1) 主要扫描两次数据库, 避免了多次扫描浪费时间; 2) 无需产生候选集, 仅需构造 FP_tree 即可产生频繁模式.

S Lin^[14]等将 FP-growth 算法用于挖掘建筑分项数据间的关联. 数据集合包括建筑总能耗(低、中、高), 空调能耗(低、中、高), 空调状态(开、关), 温度(低、中、高), 以天为单位生成记录, 通过 30 条记录产生强关联规则. 结果表明空调系统能耗跟建筑总能耗、空调系统分单元间能耗、温度与空调系统能耗、温度与建筑总能耗间存在相关关系. Yu Z^[15]通过 FP-growth 算法发现住宅建筑中的不节能行为. 试验中的项集包括各种负荷(热水、灯、厨房、冰箱、娱乐、家务等)、室外气象参数(温度、相对湿度、风速、光照情况)、建筑自身参数(面积、热损失系数)等, 通过产生的关联规则发现用户在日常生活中的不节能行为. 类似的研究还有通过 FP-growth 算法发现空调系统新风处理机组运行中的耗能行为^[16].

2.2 多元线性回归的应用

多元线性回归就是根据已知变量, 利用最小二乘法求回归系数, 建立自变量与因变量间线性关系的过

程. 线性回归是一种广泛应用的节能分析模型, 可用于能耗预测、能耗影响因素分析等方面. 对于自变量的选择通常遵循几个原则:

- (1) 自变量与因变量间成显著的线性相关关系;
- (2) 自变量间应具有互斥性, 或者相关程度很小;
- (3) 自变量数据容易测量.

文献[17]利用线性回归分析寻找能耗数据与影响因素间的关系. 首先通过研究寻找与电能耗变化最为紧密的影响因素(室内外温差、日均开灯时长、人均照度、使用人数等)建立回归模型; 然后验证模型的准确性, 移除非线性变量. 该模型分析结果可结合未来影响因素指标预测未来一段时间的电能需求量. 文献[18]提出一种基于多元线性回归的建筑能耗分析与预测模型, 因变量选择为天然气和电能耗, 自变量为建筑类型、年代、占地面积、居住人数等参数, 通过模型分析出对能耗影响显著的因素. 文献[19,20]在建筑初始设计阶段通过回归模型定量预测建筑能耗, 结果可为节能建筑设计提供参考依据.

2.3 分类的应用

分类属于典型的有监督学习方法, 其过程可分为两步: 1) 训练: 训练集→特征提取→训练→分类器; 2) 分类: 新数据→特征提取→分类.

2.3.1 决策树的应用

决策树是一种树形结构, 由根节点、内部节点和叶节点组成. 内部节点表示属性, 每条分支表示一个输出测试, 叶节点代表实例所属类别. 决策树通过计算熵(entropy)和信息增益确定属性间的关系构建决策树, 信息增益最大的属性作为分支节点属性. 文献[21]将决策树方法用于建筑用能需求建模. 首先将属性按区间划分, 如温度(high/low)、热损失系数(high/low)、室内人员数量(是否 ≤ 2)、能耗来源(供热等)等, 输出为能耗密度(high/low), 然后通过公式计算样本数据的信息熵、信息增益, 信息增益最大的属性作为分支节点判断属性, 通过逐级向下计算直到输出结束, 构建的决策树可用于预测建筑能耗密度及能耗需求. 文献[22]通过建立能耗模式判定树与离群点分析, 判断实时能耗数据的模式, 监测能耗是否存在异常.

2.3.2 SVM 的应用

SVM 具有泛化能力强、全局最优、对维数不敏感等优点, 可成功应用于负荷预测. 经典的 SVM 方法是一种二分类方法, 而在实际的数据挖掘应用中, 一般

要解决的是多分类问题, 所以后期提出了 SVM 改进方法来解决实际问题, 最小二乘支持向量机(LS-SVM)就是一个典型的改进算法.

LS-SVM 是 SVM 的一种变形算法, 通过非线性变换将负荷样本映射到高维空间, 在该空间上寻找负荷预测回归函数并建立 LS-SVM 模型, 从而进行负荷预测. 文献[23]将 LS-SVM 用于空调负荷预测中. 首先通过确定与负荷密切相关的因素, 选择输入变量类型; 然后将与预测量密切相关的因素(历史负荷数据、气象因素等)从输入空间映射到特征空间构造最优线性回归函数, 利用结构风险最小化原则并通过引入 Lagrange 函数求目标函数最小值训练 LS-SVM 模型参数, 最后通过测试样本验证该模型的准确性. 针对原始输入变量存在的明显粗糙性, 文献[24]通过粗糙集(Rough Sets, RS)属性简约寻找与预测负荷密切相关且最少的输入变量, 提高了该算法的运行效率. 类似基于 SVM 的改进方法在建筑节能中的应用还有回归型支持向量机^[25], 灰色模型支持向量机^[26]等.

2.4 聚类的应用

聚类分析法的优势在于不需要考虑分类结构, 只需考虑数据本身的特征, 即通过数据间的相似度实现分类. 聚类结果可用于能耗预测、基准评价、运行优化等. 常用的聚类方法有 K-均值(K-means)算法、Chameleon 算法、DBSCAN 算法等.

文献[27]使用合成聚类方法(Agglomerative Hierarchical Clustering, AHC)分析一周中相似能耗状况. 聚类流程为: 特征向量生成→特征向量转换→分成 7 类→移除离群点→找出相似类. 基本思想为: 首先将每个特征向量看成一个单独的类, 然后利用制定的规则进行合并, 直到类的数目减到指定的数目为止. 特征向量为可以描述能耗情况的参数, 如日均能耗量, 日能耗峰值. 文中通过异化系数描述类之间的不相似度并作为是否聚为同一类的标准; 通过终止规则决定是否合并最近的类. 聚类结果可用于发现能耗异常点并制定节能控制策略. 文献[28]采用基于 Weka 平台的 Chameleon 算法建立面向公共建筑能耗运行优化的聚类模型. 首先使用 K-均值把与某个点最近的 k 个点捡起来形成若干小簇, 然后通过簇间的互联性和紧密性决定是否合并簇. 将该方法应用到商场和办公建筑的能耗聚类中, 求得每个簇的平均能耗和能耗分布率并将其作为节能评价指标, 可为节能提供决策依据.

与其他数据挖掘方法相比,聚类不需要太多的专家领域知识和复杂的公式计算,只从数据自身处理的角度考虑,且能获得不错的分析效果.聚类没有一个客观的正确与否评价标准.分类或关联分析可通过测试数据检验其正确性,但由于聚类属于无监督学习,分成几类是可变的,可以通过是否能够有效应用于数据挖掘实例中来对其评估.

2.5 神经网络的应用

神经网络的非线性适应性信息处理能力、并行处理、容错性等优势使其克服了其他智能方法的缺陷,已被广泛应用于能源预测等方面.将神经网络用于分析建筑能耗指标的基本思路为:首先通过特征提取,找出对能耗指标较为敏感的参数作为神经网络的输入量,能耗指标作为输出;然后划分训练样本集并对网络训练,当达到所设置的训练精度时训练结束.训练完成后的网络对于每一个输入量都可将其划归到最接近的类中.

BP神经网络是一种常用的方法,已广泛应用于能耗指标评价和能耗预测等领域中.文献[29]系统性阐述了基于神经网络的建筑节能评价体系,包括建筑节能因素分析及神经网络在建筑节能评价中的应用.文献[30]将BP神经网络用于构建公共建筑能耗分析模型.首先选取温度、湿度、天气特征、是否是正常工作日作为敏感输入量,建筑能耗值为输出;设置网络连接权值、阈值、误差函数、目标值、学习次数等参数;训练中通过不断调整参数直到达到设置的精度为止.实验结果表明基于BP神经网络应用于公共建筑能耗分析准确率较高.文献[31]采用BP神经网络进行建筑负荷预测,输出量为空调年耗电量和实际单位面积耗电量,结果证明了BP神经网络用于建筑能耗预测的可行性.

其他的神经网络方法在建筑节能指标分析中也有广泛应用.文献[32]采用基于粒子群优化的RBF神经网络方法设置日能耗报警阈值,完善了能耗监测平台的报警功能;文献[33]提出了一种基于PCA(Principal Component Analysis)和RBF的建筑能耗预测模型,按贡献率提取主成分并作为神经网络输入,结果表明该方法可有效提高预测精度;文献[34]将能耗基准评价法与人工神经网络结合用于预测能源利用指数EUI(Energy Use Index),发现了与EUI变化密切相关的因素.

基于神经网络的数据挖掘方法的优点:1)非线性拟合能力强,能映射复杂的非线性关系;2)对噪声数据

有较强的鲁棒性;3)准确率高,学习规则简单;4)学习能力强,分布处理能力强.缺点:1)由于不能观察学习过程,因此欠缺解释推理过程和输出结果的能力;2)对冗余和不完整数据挖掘结果存在异常;3)容易训练过度;4)系统参数多,学习时间长.常用的基于神经网络的挖掘方法有基于自组织神经网络的数据挖掘、基于模糊神经网络的数据挖掘等.

3 数据挖掘在建筑节能中的发展趋势

除去无法调控的室外气象条件之外,影响建筑运行能耗的因素大致可分为两类:一类是与使用人员相关的需求侧,主要包括室内人员密度及其变化,室内人员对建筑空间的使用时间和行为,以及室内人员对室内环境的要求和自主调控意识等;另一类是与建筑系统相关,通常把建筑本体以及为建筑提供室内舒适环境的各系统(照明系统、空调供暖和通风系统等)看作一个整体,称为建筑系统,这个系统的目的是为人类活动提供舒适的室内环境,其影响建筑运行能耗的因素主要包括建筑本体空间形态、维护结构热工性能、系统设备的能效等性能参数,还包括各系统根据室内需求做出的各种运行调控策略,调控策略是否得当影响供应系统的整体效率.

要实现建筑节能模式由供应侧到需求侧的转变,就必须恰当描述特定室内环境下的用能特征,才能从需求侧评估建筑能耗的合理性,进而精确辨识能源浪费的原因.这就需要解决以下三个问题:(1)能耗感知(energy consumption sensing),获取必要的空间粒度和时间粒度能耗数据,得到室内空间能量踪迹.(2)环境感知(environments sensing),获取温度、湿度、噪声、空气质量、照度等影响用户舒适感的各种室内环境状态参数,以及室外气象参数.(3)情景感知(context awareness),采集能源消耗时室内特定空间的人员分布及行为状态,获取用户踪迹(Occupants footprint).

由于存在既要满足众多使用者对舒适的需求,同时又要满足能耗最小化这一相互矛盾的目标,使得建筑能耗管理问题变得愈发复杂.通过获取能耗、环境和情景这三类室内空间状态,将能量踪迹和用户踪迹信息融合,可以为能耗统计、能源审计、节能管理、节能改造和行为节能等提供有效的技术支持.

随着数据挖掘技术的发展,将数据库技术、人工智能、统计学等应用到建筑节能中有广阔的发展前景.

由于数据挖掘中单个算法具有缺点和局限性,对于数据挖掘算法的改进和多算法组合可以弥补传统算法的缺陷,也是数据挖掘技术的发展方向。另外,通过对电力公司的能源数据以及其他各类第三方数据进行深入分析和挖掘,结合行为科学、云计算技术、大数据分析,进而为用户提供一整套适合于其生活方式的节能建议,也是一种发展趋势。

4 结语

本文首先介绍了数据挖掘技术,针对应用较广泛的挖掘方法作了介绍,然后介绍了数据挖掘典型方法在建筑节能中的应用,最后对数据挖掘技术在建筑节能中的应用发展作了展望。

随着建筑设备和建筑耗能的增加,建筑能耗数据海量量化是不可避免的,如何从大数据中挖掘出有用的信息,为建筑节能提供参考依据是建筑行业的发展方向,同时为数据挖掘技术提供了用武之地。数据挖掘和建筑节能技术在我国的发展还较为落后,将数据挖掘技术应用到建筑节能中还有很长的路要走。

参考文献

- 1 清华大学建筑节能研究中心. 中国建筑节能年度发展研究报告 2015. 北京: 中国建筑工业出版社, 2015.
- 2 Ramesh T, Prakash R, Shukla KK. Life cycle energy analysis of buildings: An overview. *Energy and Buildings*, 2010, 42(10): 1592–1600. [doi: 10.1016/j.enbuild.2010.05.007]
- 3 李骥, 邹瑜, 魏峥. 建筑能耗模拟软件的特点及应用中存在的问题. *建筑科学*, 2010, 26(2): 24–28, 79.
- 4 Han J, Kamber M. 数据挖掘: 概念与技术. 范明, 孟小峰译. 北京: 机械工业出版社, 2007.
- 5 张敏, 李亚非. 数据挖掘综述. *江苏科技信息: 学术研究*, 2010, (2): 83–84.
- 6 王爱平, 王占凤, 陶嗣干, 等. 数据挖掘中常用关联规则挖掘算法. *计算机技术与发展*, 2010, 20(4): 105–108.
- 7 杜丽英. 基于数据挖掘的决策树算法分析. *吉林建筑工程学院学报*, 2014, 31(5): 48–50.
- 8 汪海燕, 黎建辉, 杨风雷. 支持向量机理论及算法研究综述. *计算机应用研究*, 2014, 31(5): 1281–1286.
- 9 贺玲, 吴玲达, 蔡益朝. 数据挖掘中的聚类算法综述. *计算机应用研究*, 2007, 24(1): 10–13.
- 10 Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 1993, 22(2): 207–216. [doi: 10.1145/170036]
- 11 苏江夺. 基于关联规则挖掘的节能系统用电分项关联耦合分析的研究与实现[硕士学位论文]. 广州: 华南理工大学, 2013.
- 12 张春杰. 基于 Apriori 算法的校园空调的节能研究[硕士学位论文]. 杭州: 杭州电子科技大学, 2014.
- 13 Han JW, Pei J, Yin YW. Mining frequent patterns without candidate generation. *ACM SIGMOD Record*, 2000, 29(2): 1–12. [doi: 10.1145/335191]
- 14 Lin SF, Hao C, Li DD, *et al.* Comprehensive energy efficiency analysis of buildings based on FP-growth association rules. *International Journal of Electrical Energy*, 2015, 3(3): 197–202.
- 15 Yu Z, Haghghat F, Fung BCM, *et al.* A methodology for identifying and improving occupant behavior in residential buildings. *Energy*, 2011, 36(11): 6596–6608. [doi: 10.1016/j.energy.2011.09.002]
- 16 Yu Z, Haghghat F, Fung BCM, *et al.* A novel methodology for knowledge discovery through mining associations between building operational data. *Energy and Buildings*, 2012, 47: 430–440. [doi: 10.1016/j.enbuild.2011.12.018]
- 17 聂子航, 于学军. 基于多元线性回归的办公建筑电力能耗评估预测模型的设计. *电子设计工程*, 2016, 24(3): 40–43, 46.
- 18 樊丽军. 基于多元线性回归模型的建筑能耗预测与建筑节能分析. *湘潭大学自然科学学报*, 2016, 38(1): 123–126.
- 19 Asadi S, Amiri SS, Mottahedi M. On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. *Energy and Buildings*, 2014, (85): 246–255.
- 20 Hygh JS, DeCarolis JF, Hill DB, *et al.* Multivariate regression as an energy assessment tool in early building design. *Building and Environment*, 2012, 57: 165–175. [doi: 10.1016/j.buildenv.2012.04.021]
- 21 Yu Z, Haghghat F, Fung BCM, *et al.* A decision tree method for building energy demand modeling. *Energy and Buildings*, 2010, 42(10): 1637–1646. [doi: 10.1016/j.enbuild.2010.04.006]
- 22 卿晓霞, 肖丹, 王波. 能耗实时监测的数据挖掘方法. *重庆大学学报*, 2012, 35(7): 133–137. [doi: 10.11835/j.issn.1000-582X.2012.07.023]
- 23 唐莉, 唐中华, 靳俊杰. 最小二乘支持向量机(LS-SVM)在短期空调负荷预测中的应用. *建筑节能*, 2013, (2): 56–58.
- 24 耿艳, 韩学山, 韩力. 基于最小二乘支持向量机的短期负荷预测. *电网技术*, 2008, 32(18): 72–76.
- 25 李建, 陈烈, 茅林明. 支持向量机在建筑能耗预测中的应用. *建筑节能*, 2014, (12): 77–80. [doi: 10.3969/j.issn.1673-7237.2014.12.020]

- 26 唐杰明, 刘俊勇, 杨可, 等. 基于灰色模型和最小二乘支持向量机的电力短期负荷组合预测. 电网技术, 2009, 33(3): 63-68.
- 27 Seem JE. Pattern recognition algorithm for determining days of the week with similar energy consumption profiles. *Energy and Buildings*, 2005, 37(2): 127-139. [doi: [10.1016/j.enbuild.2004.04.004](https://doi.org/10.1016/j.enbuild.2004.04.004)]
- 28 刘文凤. 数据挖掘在公共建筑能耗分析中的应用研究[硕士学位论文]. 重庆: 重庆大学, 2010.
- 29 卢德华. 基于人工神经网络的建筑节能体系综合评价. 建设科技, 2007, (9): 64-65.
- 30 刘长俊. 基于遗传 BP 神经网络技术的大型公建能耗分析模型的研究与应用[硕士学位论文]. 南宁: 广西大学, 2012.
- 31 蒋毅. 建筑能耗的统计平台及其基于 BP 神经网络预测方法的研究[硕士学位论文]. 广州: 华南理工大学, 2012.
- 32 徐聪麒. 基于粒子群优化的 RBF 网络在节约型校园能耗监管平台中的应用研究[硕士学位论文]. 南宁: 广西大学, 2012.
- 33 戴坤成, 王贵评, 赵超. 基于 PCA 与 RBF 的建筑能耗预测建模. 福州大学学报(自然科学版), 2015, 43(4): 512-516. [doi: [10.7631/issn.1000-2243.2015.04.0512](https://doi.org/10.7631/issn.1000-2243.2015.04.0512)]
- 34 Yalcintas M. An energy benchmarking model based on artificial neural network method with a case example for tropical climates. *International Journal of Energy Research*, 2006, 30(14): 1158-1174. [doi: [10.1002/\(ISSN\)1099-114X](https://doi.org/10.1002/(ISSN)1099-114X)]