



WILEY  
ISTE



Data Mining and Machine Learning  
in Building Energy Analysis

# 建筑能耗

## 分析中的

# 数据挖掘与机器学习

[法] 弗雷德里克·马古尔斯 (Frédéric Magoulès)

赵海祥 (Hai-Xiang Zhao)

史晓霞 陈一民 庄俊华 樊清

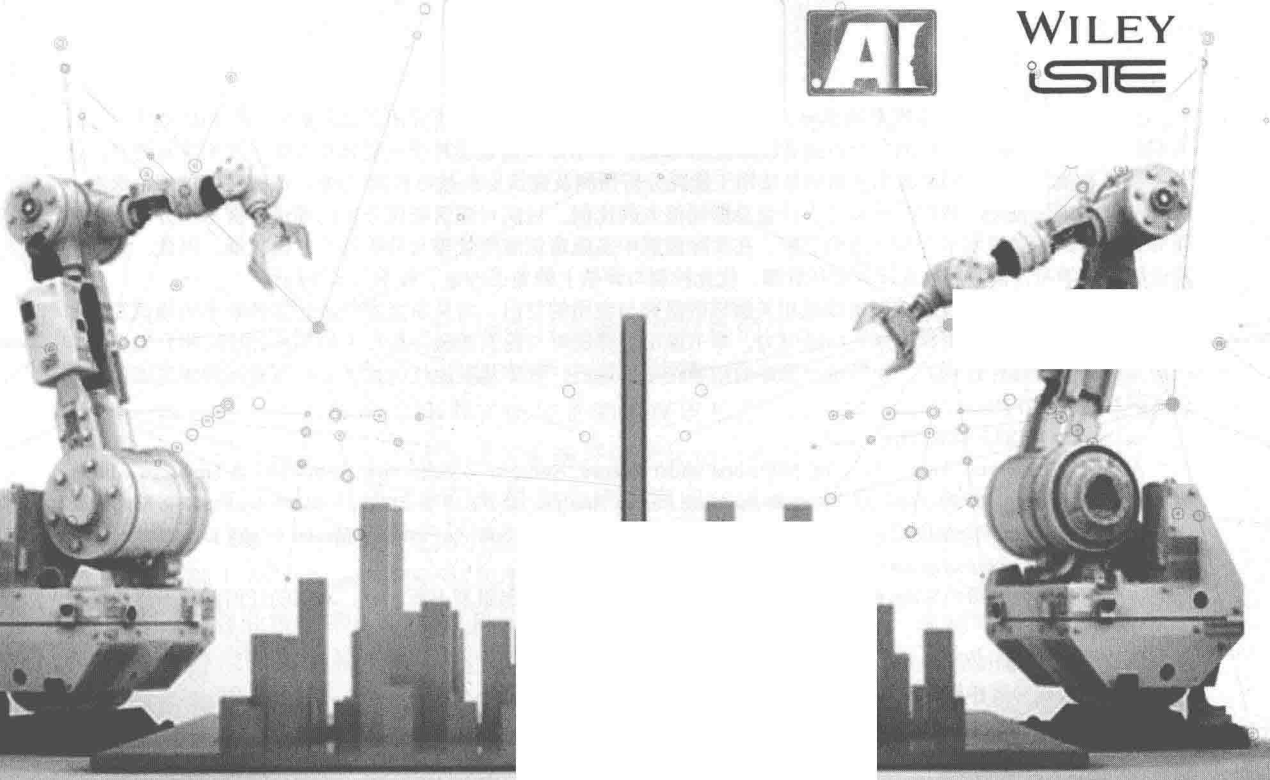
著  
译



机械工业出版社  
CHINA MACHINE PRESS

AI

WILEY  
ISTE



Data Mining and Machine Learning  
in Building Energy Analysis

# 建筑能耗

常州大学图书馆  
藏书章

## 分析中的

# 数据挖掘与机器学习

[法] 弗雷德里克·马古尔斯 (Frédéric Magoulès)

赵海祥 (Hai-Xiang Zhao)

史晓霞 陈一民 庄俊华 樊清

著  
译



机械工业出版社  
CHINA MACHINE PRESS

建筑的能源性能受很多因素的影响,本书针对建筑的复杂特性,重点研究用最新的数据挖掘和机器学习方法来对建筑能耗进行精确的预测、分析或者故障检测/诊断。本书涉及建筑能耗分析的建模及用于模型降阶与并行计算的技术和相关算法,同时提出了新的算法用于能耗分析预测及建筑能耗故障检测/诊断,既有一定的理论深度,又有较好的应用宽度。我国建筑能耗占社会总能耗很大的比例,目前对建筑能耗分析的理论、技术和方法所做的研究工作与国际先进水平有相当大的差距,在实际建筑中实施建筑能耗管理与分析的水平也较低。因此,当前特别迫切需要学习并借鉴国外在建筑能耗管理、优化控制与评估上的先进理论、技术与实施经验。

本书很好地填补了我国在建筑能耗相关领域的研究与应用的空白,对从事能源管理和能源效率的知识发现和数据挖掘研究领域的工程师有很大的吸引力,本书提出的算法对与建筑能耗分析相关的工程领域的研究生有很好的借鉴作用,对设计建筑的工程师也有很好的指导作用。最后,对于建筑能耗的预测分析对我国的建筑能耗管理也有很好的促进作用。

Copyright © ISTE Ltd 2016

All Right Reserved. This translation published under license. Authorized translation from English language edition, entitled *Data Mining and Machine Learning in Building Energy Analysis*, ISBN: 978-1-84821-422-4, by Frédéric Magoulès and Hai-Xiang Zhao, Published by John Wiley & Sons. No part of this book may be reproduced in any form without the written permission of the original copyrights holder.

本书中文简体字版由Wiley授权机械工业出版社独家出版。未经出版者书面允许,本书的任何部分不得以任何方式复制或抄袭。

版权所有,翻印必究。

北京市版权局著作权合同登记 图字:01-2017-5452号。

## 图书在版编目(CIP)数据

建筑能耗分析中的数据挖掘与机器学习 / (法) 弗雷德里克·马古尔斯等著; 史晓霞等译.  
—北京: 机械工业出版社, 2018.7

书名原文: *Data Mining and Machine Learning in Building Energy Analysis*

ISBN 978-7-111-60267-5

I. ①建… II. ①弗… ②史… III. ①数据采集—应用—建筑能耗—能耗计算 ②机器学习—应用—建筑能耗—能耗计算 IV. ①TU111.19-39

中国版本图书馆CIP数据核字(2018)第136286号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑: 顾谦 责任编辑: 顾谦

责任校对: 刘雅娜 责任印制: 孙炜

北京中兴印刷有限公司印刷

2018年7月第1版第1次印刷

184mm×240mm·7印张·153千字

0 001—3 000册

标准书号: ISBN 978-7-111-60267-5

定价: 45.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

网络服务

服务咨询热线: 010-88361066

机工官网: [www.cmpbook.com](http://www.cmpbook.com)

读者购书热线: 010-68326294

机工官博: [weibo.com/cmp1952](http://weibo.com/cmp1952)

010-88379203

金书网: [www.golden-book.com](http://www.golden-book.com)

封面无防伪标均为盗版

教育服务网: [www.cmpedu.com](http://www.cmpedu.com)

# 译者序 |

当前，能源与环境的压力迫使全社会重视所存在的问题。建筑能耗占当前社会总能耗很大比例，这些能耗不仅涉及建筑材料生产与建筑施工的过程，而且涉及建筑全寿命周期运行，由于需要满足各种功能需要所带来的能耗更是占到了建筑能耗的绝大部分。对建筑能耗进行分析、评价和对建筑设备与设施进行优化运行无疑成为建筑节能降耗的重要措施。

本书以当前的建筑能耗分析和管理为研究方向，以最新的数据挖掘和机器学习为主要研究方法，详细、全面地介绍了建筑能耗分析中的模型类型，对现存模型进行了对比分析，并在此基础上利用 EnergyPlus 软件对建筑能耗进行了数据的采集。本书在现有工作的基础上，对人工智能模型中的 RDP (Recursive Deterministic Perceptron, 递归确定性感知器) 方法进行改进用于故障检测与诊断，研究了支持向量机 (Support Vector Machine, SVM) 的模型降阶和并行计算，并提出了新方法用于建筑能耗分析的预测与管理，为建筑低能耗的实现提供了一定的理论指导和评估依据。

我国是能耗大国，建筑能耗占社会总能耗的比例很大。当前，我国极为重视在全社会各个领域的节能减排工作，特别是在建筑领域，相继出台了大型公共建筑节能监测等法律法规。但目前国内对建筑能耗分析的理论、技术和方法所做的研究工作与国际先进水平还有相当大的差距，在实际建筑中实施建筑能耗管理与分析的水平也较低。

因此，当前特别迫切需要学习并借鉴国外在建筑能耗管理、优化控制与评估上的先进理论、技术与实施经验。本书所涉及的建筑能耗的预测、故障检测与诊断等方法具有新颖性和先进性，对从事建筑能耗分析的研究人员和相关工程技术人员都有很好的启发性，可为我国高校、科研机构在建筑能耗分析与管理的研究、教学提供有益的参考，为企业单位进行低能耗建筑的设计、实施、运行维护提供借鉴。

北京建筑大学陈一民翻译了本书的第 1 章，庄俊华翻译了本书的第 2 章，史晓霞翻译了本书的第 3~5 章和第 7 章，樊清翻译了本书的第 6 章。全书由史晓霞进行统稿。

本书在翻译过程中，得到了北京建筑大学胡玉玲、刘慧、刘辛国、李壮举、谭志、张俊红等人的支持与帮助，在此表示深深的谢意！

最后感谢本书的责任编辑顾谦先生，他的严谨与辛勤工作使本书的出版成为可能。

需要注意的是，书中的矢量、矩阵等，为保持原书形式，并未修改为黑斜体，请读者注意。译者在翻译过程中参阅了不同领域的大量文献，力求使本书的翻译更为准确，但由于译者水平有限，书中难免有翻译不准的情况，恳请广大读者予以批评指正，谢谢。

译者

2018年4月



# | 原书序

建筑的能源性能受到很多因素的影响，比如环境气象条件，建筑结构和特点，居住者及其行为，加热、通风和空调系统部件的分段操作。这些复杂的特性使得很难对建筑能耗进行精确的预测、分析或者故障检测 / 诊断。

本书将重点研究使用最新的数据挖掘和机器学习方法来解决这些问题。本书首先对目前建筑能耗的预测、分析或者故障检测 / 诊断方法进行了回顾，包括详细的和简化的工程方法、统计方法和人工智能方法。然后提出了一种模拟单个或者多个建筑能耗分布的方法。基于这些数据集，对用于预测的 SVM 模型进行训练和测试。大量的实验结果证明，这些模型具有较高的预测精度和鲁棒性。然后利用 RDP 对建筑能耗进行故障检测和诊断。在实验中，RDP 模型显示了很高的检测能力，这里也提出了一种基于 RDP 模型演化用于故障诊断的新方法。因为特征子集的选择严重影响模型的性能，根据获得数据的可行性和两种滤波方法评价下所提供的得分来选择最佳特征。实验结果证明了所选子集的有效性，表明了所提特征选择方法能保证模型的精确性并减少了计算用时。预测建筑能耗的一个挑战是当数据量很大时，加快模型训练。为了解决这个问题，本书提出了一种基于分解方法的 SVM 的高效并行实现。并行化是在训练过程中最耗时的部分进行的。底层并行使用一个共享内存的 MapReduce 范例，使得系统特别适用于多核和多处理器系统。实验结果表明，与 Libsvm 相比，最初的实现提供了一个很高的加速，并且在计算时间和存储要求方面优于最先进的 Pismv 消息传递接口（Message Passing Interface, MPI）实现。

本书的目的旨在使用具体的例子来解释和说明用于解决建筑能耗预测、分析或故障检测 / 诊断的数据挖掘和机器学习的最新技术。遵循教学式的方法，在适当时逐渐引入数学和计算概念，并在需要时加深理解。本书主要面向与建筑能耗分析相关的工程领域的研究生，它也可能引起从事能源管理和能源效率的知识发现和数据挖掘研究领域的工程师的兴趣。作者已经使用本书部分内容很多年，在武汉科技大学（Wuhan University of Science and Technology）（中国）、巴黎萨克雷大学（University Paris Saclay）（法国）、法国国立工艺学院（Conservatoire National des Arts et Métiers）（法国）、巴黎中央理工大学（École Centrale des Arts et Manufactures）（法国）、南锡国立科学技术工程师学院（École Supérieure des Sciences et Technologies de l'Ingénieur de Nancy）（法国）、杜伊斯堡-埃森大学（University Duisburg-Essen）（德国）、中央大学（Chuo University）（日本）、同志社大学（Doshisha University）（日本）、庆应义塾大学（Keio University）（日本）和电气通信大学（University of Electro Communications）（日本）人工智能的研讨会上进行演讲和讲座。

Frédéric Magoulès

Hai-Xiang Zhao

2015 年 11 月

# 原书前言 |

建筑能耗在所有能耗中占有相当大的比例，因此建筑节能在能源领域是一个很重要的课题。建筑的能源系统很复杂，因为它受到很多因素的影响，如环境气象条件，建筑特性，居住者及其行为，加热、通风和空调（Heating, Ventilating and Air-Conditioning, HVAC）系统这些分段部分的运行。这些复杂特性使得想快速、准确地进行建筑能耗的预测或者故障检测 / 诊断很困难。

人工智能模型在解决复杂问题上引起了广泛的关注。在本书中，回顾了最近为解决这些问题而开发的模型，包括详细的和简化的工程方法、统计方法和高效的人工智能方法。然后，对单个和多个建筑的能耗分布通过测量进行了确定或者进行了仿真。基于这些数据集，对用于预测的 SVM 模型进行了训练和测试。大量实验结果表明，这些模型具有较高的预测精度和鲁棒性，然后利用 RDP 神经网络模型来对建筑能耗进行故障检测和诊断，并提出了一种故障诊断的新方法。它是基于 RDP 模型的评价，其中每种方法都能检测一种设备故障。怎样选择特征子集将影响模型性能。这里根据获得数据的可行性和两种滤波方法评价下所提供的得分，来选择最佳特征。实验结果证明了所选子集的有效性，且表明所提出的特征选择方法能保证模型精度并减少计算用时。最后，预测建筑能耗最困难的挑战之一是当数据集很大时对模型训练的加速。本书为了解决这样的问题，基于分解算法提出了一种 SVM 的高效并行实现方法。并行化在最费时的训练工作中被执行，即更新梯度矢量  $f$ 。内部问题使用一个序列最小优化求解器进行处理。潜在的并行由一个共享内存 MapReduce 范例来处理，使得系统尤其适用于多核的和多处理器系统。实验结果表明，与 Libsvm 相比，这里的实现提供了一个高速增长，并且在速度和存储要求上也优于最先进的 Pism MPI 实现。

本书结构如下：

第 1 章中，首先介绍了建筑能源的概念，并介绍了本书的背景和动机、问题和挑战。然后，回顾了最近开发的有关预测建筑能耗的模型和方法。这些先前的研究包括用合适的模型解决各级能源分析问题、优化模型参数、处理输入以获得更好的性能、简化问题及对不同模型比较。最后，总结每种模型的优缺点。即使这一章不能详尽说明，但它包含了建筑能源分析领域的书目分析和研究历史。因此，在第一次阅读时可以省略本章，但读者可以在阅读后面的章节时参考它作为进一步阅读或者作为补充信息源。

第 2 章提出了如何获得历史能耗概况。列出了常见的方法并解释了为什么本书中选择用 EnergyPlus 软件产生这些数据。另外，开发了合适的接口来控制 EnergyPlus 仿真多建筑，多建筑的多样性来自不同结构特征、围护结构材料、居住者等。本章也介绍了数据的不确定性和校准。

第 3 章介绍了一种最流行的人工智能即人工神经网络（ANN）的原理。介绍了该模型的机理和一些常用的扩展，如前馈神经网络、径向基神经网络、递归神经网络和 RDP。最后，介绍了这些模型的应用。第 3 章也介绍了另一种高能力人工智能模型，即 SVM，包

括原始形式、对偶形式、正规化和核方法，加上扩展包括支持向量回归（Support Vector Regression, SVR），一类、多类和直推式 SVM（transductive SVM）及应用。这些模型在如自动文本分类、计算生物学及图像处理等许多应用领域中显示了优越性。

第4章对SVR模型用于预测建筑能耗进行了评价。首先，针对制定实验提出了几个重要的问题，包括实验的示意流程、软件和硬件环境、所需的格式化数据、模型选择和模型性能评估。然后，应用SVR来预测冬季区域供暖（最近2天评估，剩下的训练）、预测一年的电力消耗（任选48h用于测试，其他的用于训练）。接下来，测试SVR的鲁棒性。用三个数据集训练模型，即1月、1~4月和1~8月，其他每个月用来测试模型。然后，用99栋大楼的能耗训练模型，在一栋全新的大楼测试。最后，利用RDP神经网络模型来检测和诊断建筑能耗故障，并提出了一种故障诊断的新方法。它基于几个RDP模型的评价，其中每一个模型都能检测特定设备的故障。在实验中的诊断方法成功地诊断了冷机故障。它也能够根据故障的可能性对可能的故障源进行降序排列。

第5章针对SVR应用于建筑能耗提出了一种新的特征选取方法，并解释了相关的工作、算法和实现，用大量实验证明了该方法的有效性，根据实际应用的可行性和预报器的实用性来选择特征。最后一个标准通过两种滤波方法来评价：梯度引导特征选择和相关系数法。为了评价所提出的方法，使用了三个训练集来评价特征选取前后模型的性能变化。实验结果表明，所选子集能够提供更具竞争力的预报器。在不损失模型性能的前提下减少特征数量，可以使模型更容易用于实践，在某些情况下也可以改善性能。例如，对于50座建筑的数据采用径向基函数和多项式核，模型精度提高且学习成本明显降低。将机器学习方法应用于建筑能耗预测时，该工作是选择最佳特征子集的第一指南。

第6章提出了一种适用于多核系统的原始的SVR和SVC并行算法。介绍了相关工作、分解二次规划问题（Quadratic Problem, QP）求解器、SVC五个基准数据集的实现细节及比较实验和结果分析。然后，介绍了并行SVR及其在建筑能耗预测中的应用。这里所提出的方法是基于一种分解方法，对变量进行迭代优化。并行化是在简单而实用的编程框架MapReduce下进行编程的，共享高速缓存用来存储核函数矩阵的列，这个实现尤其适用于多核和多处理器系统。这里也用大量的实验测试了SVC和SVR，结果表明新方法在解决大规模问题时很有效。在顺序实现方面，它实现了高的加速比。结果也表明，实现性能在训练速度和存储要求上是优于先进的并行实现的。

本书通过总结所提出的创新方法和调查建筑能耗分析的未来得出结论。

# 目 录 |

译者序

原书序

原书前言

## 第 1 章 建筑能耗分析概述 //1

- 1.1 简介 //1
- 1.2 物理模型 //2
- 1.3 灰色模型 //3
- 1.4 统计模型 //4
- 1.5 人工智能模型 //5
  - 1.5.1 神经网络 // 5
  - 1.5.2 SVM // 7
- 1.6 现有模型比较 //8
- 1.7 小结 //9

## 第 2 章 建筑能源分析的数据采集 //10

- 2.1 简介 //10
- 2.2 调查或问卷调查 //10
- 2.3 测量 //12
- 2.4 仿真 //14
  - 2.4.1 仿真软件 // 15
  - 2.4.2 仿真过程 // 16
- 2.5 数据不确定性 //19
- 2.6 校准 //20
- 2.7 小结 //21

## 第 3 章 人工智能模型 //23

- 3.1 简介 //23

### 3.2 ANN //24

- 3.2.1 单层感知器 // 24
- 3.2.2 前馈神经网络 // 25
- 3.2.3 RBF 网络 // 26
- 3.2.4 RNN // 27
- 3.2.5 RDP // 28
- 3.2.6 神经网络的应用 // 30

### 3.3 SVM //31

- 3.3.1 SVC // 31
- 3.3.2  $\varepsilon$ -SVR // 34
- 3.3.3 一类 SVM // 36
- 3.3.4 多类 SVM // 37
- 3.3.5  $\nu$ -SVM // 38
- 3.3.6 直推式 SVM // 39
- 3.3.7 二次型问题求解器 // 40
- 3.3.8 SVM 的应用 // 46

### 3.4 小结 //47

## 第 4 章 建筑能耗分析中的人工智能 //48

- 4.1 简介 //48
- 4.2 建筑能耗预测中的 SVM //48
  - 4.2.1 能耗预测定义 // 48
  - 4.2.2 实际问题 // 49
  - 4.2.3 SVM 用于预测 // 52



#### 4.3 神经网络用于故障检测和诊断 //56

4.3.1 故障描述 // 58

4.3.2 故障检测中的 RDP // 58

4.3.3 故障诊断中的 RDP // 61

#### 4.4 小结 //63

### 第5章 SVM 的模型降阶 //64

#### 5.1 简介 //64

#### 5.2 模型降阶概述 //64

5.2.1 包装器方法 // 65

5.2.2 滤波器方法 // 65

5.2.3 嵌入式方法 // 66

#### 5.3 模型降阶用于能耗 //66

5.3.1 简介 // 66

5.3.2 算法 // 67

5.3.3 特征集描述 // 68

#### 5.4 独栋建筑能耗的模型降阶 //69

5.4.1 特征集选择 // 69

5.4.2 实验评价 // 71

#### 5.5 多栋建筑能耗的模型降阶 //71

#### 5.6 小结 //73

### 第6章 SVM 的并行计算 //74

#### 6.1 简介 //74

#### 6.2 并行 SVM 概述 //74

#### 6.3 并行二次问题求解器 //75

#### 6.4 基于 MPI 的并行 SVM //77

6.4.1 信息传递接口编程模型 // 77

6.4.2 Psvm // 79

6.4.3 Psvm // 79

#### 6.5 基于 MapReduce 的并行 SVM //80

6.5.1 MapReduce 编程模型 // 80

6.5.2 缓冲技术 // 81

6.5.3 稀疏数据表示 // 82

6.5.4 MRPsvm 和 Psvm 的比较 // 82

#### 6.6 基于 MapReduce 的并行 $\epsilon$ -SVR //84

6.6.1 实施方面 // 84

6.6.2 能耗数据集 // 85

6.6.3 建筑能耗预测评价 // 86

#### 6.7 小结 //88

### 第7章 建筑能耗分析的总结与展望 //89

### 参考文献 //91

# 第 1 章

## 建筑能耗分析概述

### 1.1 简介

在欧洲，建筑能耗占总能耗的 40%，占 CO<sub>2</sub> 排放量的 36%<sup>[EUR 10]</sup>。图 1.1 给出了法国 1990~2009 年 20 年间每个部门的年能耗。工业部门从 30% 降为 25%，运输基本稳定大约 30%，但是住宅能耗则从 37% 增长到 41%。通过这些年来建筑能耗的增长率，可以预计未来这个比率还会继续增加。因此建筑能耗预测对提高建筑能源利用率具有重要意义，进而节约能源并减少环境影响。

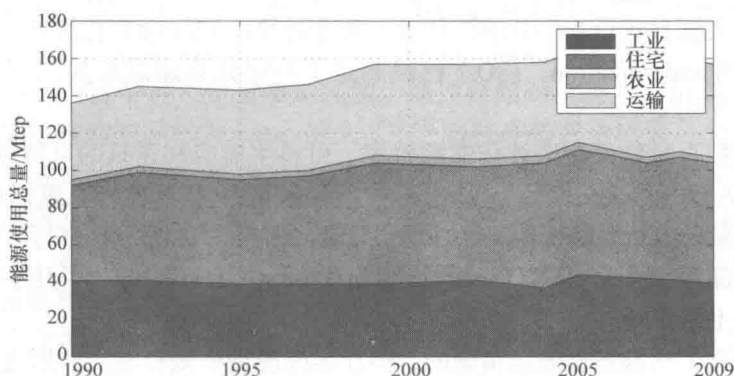


图 1.1 法国各部门的年能耗 (来源: 文献 [COM 11])

但是，建筑能源系统相当复杂，因为能源类型和建筑类型差别很大。在文献中，考虑的主要能源形式为热/冷负荷、热水和电力消耗。最常考虑的建筑类型是办公室、住宅和工程建筑，从小房间到大屋不等。一个建筑的能源行为受很多因素影响，如气象条件（特别是全球温度）、建筑结构及所使用物理材料的热特性、居住者及其行为、分段部分 [如 HVAC] 和照明系统。

由于能源系统的复杂性，准确的能耗预测是非常困难的。近年来，为了能耗预测，已经提出了大量不管是精细的还是简单的方法，并被应用到一系列问题中。这项研究工作已经在设计新建筑，或者现代建筑的运行和改造过程中进行，从一个建筑的子系统到区域或国家级的建模。通过对各影响因素的深入分析或者通过考虑主要因素近似使用，可以对整个建筑或分段部分进行预测。一个有效和高效的模型一直是研究和工程界的目标。

下面对建筑能耗建模与预测相关的最近工作进行了回顾（更多细节参看文献 [ZHA 12b] 和上面的参考文献）。在这个应用中使用的方法包括工程的、统计的和人工智能的方法。使用最广泛的人工智能方法是人工神经网络（ANN）和支持向量机（SVM）。在 2003 年和 2010 年，Krarti 和 Dounis 提供了人工智能方法用于建筑能源系统的两个综述 [KRA 03, DOU 10]。本书接下来主要集中于预测应用。为了进一步丰富内容，为读者提供各种预测方法更完整的视图，本节也回顾了工程和统计的方法。此外，也有一些混合方法，将以前某些模型结合起来优化预测性能（参看文献 [YAO 05, WAN 06, KAR 06, LIA 07]）。接下来，将描述问题、模型及相关问题，如数据预处理 / 后处理及这些模型比较。

## 1.2 物理模型

工程方法利用物理原理来计算整个建筑层或分段部分的热动力和能量行为。在过去的 50 年中，这些方法得到了充分的发展。这些方法可以分成两类：详细的综合方法和简化方法。综合方法使用非常复杂的物理功能或热动力学，一步一步，利用建筑或环境信息，如外部气候条件、建筑构造、运行、实用率表（utility rate schedule）和 HVAC 设备作为输入，来精确地计算建筑所有部件的能耗。在本节，重点将放在模型和应用的全局视图上，而计算过程的细节远远超出了本章的目的。更多细节读者可参考文献 [CLA 01]。特别地，对于 HVAC 系统，在文献 [MCQ 05] 中详细介绍了能源计算。国际标准化组织（the International Organization for Standardization, ISO）已经制定了一个计算建筑及其部件空间加热和制冷能源使用的标准 [ISO 08]。

已经开发了数百种软件用来评估能源效率、可再生能源和建筑的可持续性，如 DOE-2、EnergyPlus、BLAST 和 ESP-r<sup>[SIM 11]</sup>。其中一些已经被广泛用于建筑能源标准的制定和对建筑能耗与节能测量的分析。在文献 [ALH 01, CRA 08] 中对这些工具进行了调研。为向读者提供信息，美国能源部（the U.S. Department of Energy, DoE）保存了几乎所有能源仿真工具的清单，并在不断更新。

虽然这些仿真工具都是有效且精确的，但在实际应用中则存在一些困难。因为这些工具都是基于物理原理，所以想获得一个精确的仿真，则需要建筑和环境参数作为输入数据。一方面，这些参数对许多组织来说是不可能的，例如在一个大的建筑中每个房间的信息总是很难获得，精确输入的缺乏将导致仿真精度很低。另一方面，操作这些工具通常需要繁琐的专家工作，使得执行困难。因此，有些研究者已经提出了更简单的模型来替代某些应用。

Al-Homoud<sup>[ALH 01]</sup> 回顾了两种简化的方法：一种是度日法，这里只分析一个指标，即度日，这种稳态法适用于以围护结构能源为主的小型建筑的能耗估算；另一种方法是 bin，也称为温度频率法，它可以用来为以内生负荷占主导或者负荷与室内 / 室外温差不成线性关系的大型建筑建模。

气象条件是决定建筑能源使用的重要因素，这些有多种形式，包括温度、湿度、太阳辐射和风速，并且随时间而变化。为了简化建筑能量计算中的气象条件进行了一些研究。White 和 Reichmuth<sup>[WHI 96]</sup> 试图利用月平均气温来预测每月建筑能耗。此预测比标准程序更精确，通常利用加热和制冷度日或温度 bin（温度箱）。Westphal 和 Lamberts<sup>[WES 04]</sup> 根据一些

天气变量,包括月平均最高气温和最低气温、大气压力、云量和相对湿度,预测了非住宅建筑的年加热和制冷负荷。与复杂的仿真工具(如ESP、BLAST和DOE2)相比,在低质量围护结构的建筑方面,他们的结果显示了较好的精度。

除了天气条件外,建筑特点也是决定能源性能的另一个重要且复杂的因素。

Yao和Steemers<sup>[YAO 05]</sup>开发了一种预测每日能耗情况的简单方法,用于住宅建筑的可再生能源系统设计。总的建筑能耗定义为家电、热水和空间供暖等几个组件的总和。对每个组件,采用特定的建模方法。例如,为了对电器设备建模,使用来自大量统计数据平均终端消耗;而在对空间加热需求进行建模时,利用一个简化的物理模型。因为平均值随季节变化,所以这种方法一次只能预测一个季节的能源需求。

通过采用这种分—和(divide-and-sum)概念,Rice等人<sup>[RIC 10]</sup>简化了每段计算来解释系统级建筑能耗。在“更新ASHRAE/ACCA住宅加热和制冷负荷计算程序和数据”(RP-1199)项目中,Barnaby和Spitler<sup>[BAR 05b]</sup>提出了一种住宅负荷系数法,这是一种简单的方法,可以通过手动完成。不同来源的负荷贡献分别计算,然后再相加。Wang和Xu<sup>[WAN 06]</sup>简化了建筑的物理特性来实现预测。对于建筑围护结构,基于频率特性分析,利用易于获得的物理细节来确定模型参数。对于各种内部组件,利用集中热质量的热网络来表示内部质量。基于运行数据,利用遗传算法来辨识模型参数。Yik等人<sup>[YIK 01]</sup>利用详细的仿真工具获得不同类型建筑的冷负荷曲线,提出了一种将这些详细仿真结果组合的一个简单模型来确定一个建筑的同时冷负荷。

校准是建筑能源仿真中另一个重要的问题。通过仔细调节输入,仿真可以精确地匹配现实中特定建筑被仿真的能源行为。Pan等人<sup>[PAN 07]</sup>将校准仿真总结为一个建筑能源分析方法,并将其应用于分析一个高层商业建筑的能源使用。经过多次重复校准步骤后,在预测特定建筑的真实能源使用问题时,该能源模型显示了较高的精度。文献[RED 06]提供了校准仿真的详细论述。由于校准是一项费时、费力的工作,可以看到,利用一个详细的工程方法进行精确地仿真具有高度的复杂性。

注意到,在简化的模型和精细的模型之间没有明显的界限。还可以利用一些综合的工具如EnergyPlus<sup>[CRA 01]</sup>做简化仿真。正如Al-Homoud<sup>[ALH 01]</sup>所建议的,如果目的是研究趋势、比较系统或替代方案,那么简化的分析方法可能足够了。相反,建筑和子系统的详细能源分析和生命周期成本分析,则采用更综合的工具将会更合适<sup>[ALH 01]</sup>。

### 1.3 灰色模型

当一个系统的信息部分已知时,称之为一个灰色系统。当只有不完整的或不确定的数据时,可以用灰色模型来分析建筑物能源行为。

1999年,Wang等人<sup>[WAN 99]</sup>将一个灰色模型应用于预测建筑热湿系统,预测精度相当高。Guo等人<sup>[GUO 11]</sup>利用一个改善的灰色系统来预测住宅建筑中热泵热水器的能耗。他们评估了数据采样间隔对预测精度的影响,发现最好的间隔是4周。该模型不需要输入数据,预测误差在正常范围内。Zhou等人<sup>[ZHO 08]</sup>通过将两个天气预报模块整合成一个简化的建筑热负荷模型,进行了冷负荷的在线预测。这在文献[WAN 06]中进一步发展:一个是温度/相对

湿度预测，通过利用一个改进的灰色模型获得；另一个是太阳辐射预测，利用回归模型得到。实验结果表明，只要在训练过程中使用了来自第一个模块的预测天气数据，那么简化的热网络模型性能就会得到改善。

## 1.4 统计模型

在建筑能耗方面，已经广泛考虑统计模型，包括回归模型，如具有额外输入的自回归模型（AutoRegressive model with eXtra inputs, ARX）、自回归积分移动平均（AutoRegressive Integrated Moving Average, ARIMA）、具有额外输入的自回归积分移动平均（AutoRegressive Integrated Moving Average with eXtra inputs, ARIMAX）和有条件的需求分析（Conditional Demand Analysis, CDA）。

统计回归模型简单地将能源消耗或能源指数与影响变量联系起来。这些经验模型是由历史性能数据发展起来的，这意味着在训练模型之前，必须收集足够的历史数据。对回归模型的大量研究已经应用于以下情况：第一种情况是在一个或几个气象参数简化变量的基础上来预测能源使用；第二种情况是预测有用的能源指数；第三种是估计能源使用的重要参数，如总热损失系数、总热容和增益因数，这些在分析建筑或子系统的热行为方面都是有用的。

在一些简化的工程模型中，利用回归模型将能耗与天气变量联系起来，获得能源特征<sup>[BAU 08, WES 99, PFA 05]</sup>。Bauer 和 Scartezzini<sup>[BAU 98]</sup>提出了一种回归方法，通过处理内部和太阳能增益同时处理冷热计算。Ansari 等人<sup>[ANS 05]</sup>通过增加建筑围护结构每个部件的冷负荷来计算建筑的冷负荷。各分段冷负荷是内外温差的简单回归函数。Dhar 等人<sup>[DHA 98, DHA 99]</sup>以室外干球温度为唯一天气变量，建立了商业建筑的冷热负荷模型，提出了一种新的基于温度的傅里叶级数模型来表示冷热负荷随时间和温度的非线性关系。如果湿度和太阳数据也可用，他们建议利用广义傅里叶级数模型，因为它具有更多的工程相关性和更高的预测能力。此外，考虑将干球温度作为模型开发的唯一变量，Lei 和 Hu<sup>[LEI 09]</sup>评估了回归模型用于预测在炎热夏季和寒冷冬季区域办公楼改造项目中的节能。结果表明，单变量线性模型用于冷、热天气情况下的能源使用建模是充分和实用的。Ma 等人<sup>[MA 10]</sup>将多元线性回归和自回归方法综合来预测大型公共建筑的每月能耗。在 Cho 等人<sup>[CHO 04]</sup>的工作中，开发了 1 天、1 周和 3 个月测量的回归模型，预测的年能耗分别是 100%、30% 和 6%。这些结果表明，测量周期的长度严重地影响着基于温度的回归模型。

关于能量指数的预测，Lam 等人<sup>[LAM 10]</sup>利用主成分分析（Principle Component Analysis, PCA）开发了与全球太阳辐射和干湿球温度相关的气候指数  $Z$ 。他们发现， $Z$  具有与所仿真的冷负荷、HVAC 和建筑能源使用情况趋势相同的特点。这种趋势是通过线性回归分析得到的相关性分析得到的。该模型是基于 1979~2007 年的数据开发的。Ghiaus<sup>[GHI 06]</sup>通过使用能给出干球温度和热损失这两个变量关系的分位数 - 分位数图的第一和第三、四分位数之间的距离，开发了一种鲁棒回归模型，使干球温度与热损失关联起来。

Jiménez 和 Heras<sup>[JIM 05]</sup>利用 ARX 估计建筑部件的  $U$  和  $g$  值。Kimbara 等人<sup>[KIM 95]</sup>开发了一种 ARIMA 模型，实现了在线预测。首先从过去的负荷数据导出模型，然后用其预测下



一天的负荷曲线。ARIMAX 模型也用于控制商业建筑的高峰用电需求<sup>[HOF 98]</sup>和预测建筑的电力需求<sup>[NEW 10]</sup>。在文献 [NEW 10] 中, Newsham 和 Birt 特别强调居住者的影响, 这显然可以提高模型的精度。

Aydinalp-Koksal 和 Ugursal<sup>[AYD 08]</sup> 建议当预测国家级建筑能耗时, 考虑一种叫作 CDA 的基于回归的方法。在他们的实验比较中, CDA 表现出与神经网络和工程方法一样的精确预测能力, 但它更容易开发和使用。然而, CDA 模型的缺点是缺乏细节和灵活性, 它需要大量的输入信息。CDA 也被用于住宅能耗分析的早期工作<sup>[LAF 94]</sup>。

## 1.5 人工智能模型

### 1.5.1 神经网络

在建筑能源预测中 ANN 是最广泛使用的人工智能模型。这类模型很好地解决了非线性问题, 是解决这一复杂问题的有效方法。在过去的 20 年中, 研究者们将 ANN 应用于分析各种情况下各种类型的建筑能耗, 如冷/热负荷、电力消耗、分段部分的运行和优化及使用参数的估计。在本节中, 将回顾过去的研究, 并根据所处理的应用将它们分组。此外, 在最后强调模型优化, 如输入数据的预处理、ANN 和其他模型之间的比较。

在 2006 年, Kalogirou<sup>[KAL 06]</sup> 做了 ANN 在建筑能源应用方面的概述, 包括太阳能热水系统、太阳辐射、风速、房间内气流分布、能耗预测、室内空气温度和 HVAC 系统分析。

Kalogirou 等人<sup>[KAL 97]</sup> 利用 BP (反向传播) 神经网络预测建筑所需热负荷。通过 225 栋大楼的能耗数据, 其中很大一部分是从小空间到大房间, 对模型进行训练。Ekici 和 Aksoy<sup>[EKI 09]</sup> 使用相同的模型预测三栋建筑的热负荷, 利用瞬态一维导热有限差分法来计算训练和测试数据集。Olofsson 等人<sup>[OLO 98]</sup> 对瑞典北部一些小型独栋住宅的年供热需求进行了预测。后来, Olofsson 和 Andersson<sup>[OLO 01]</sup> 开发了一个对独栋住宅有很高预测率的神经网络, 它基于短期 (通常为 2~5 周) 测量数据预测长期的能量需求 (年供热需求)。

在文献 [YOK 09] 中, Yokoyama 等人使用 BP 神经网络预测了一个建筑中的制冷需求。在他们的工作中, 提出了一种叫作模型修整方法的全局优化方法来辨识模型参数。Kreider 等人<sup>[KRE 95]</sup> 记录了每小时能耗数据的递归神经网络的结果, 在只知道天气和时间戳的情况下, 来预测将来建筑物的供热和制冷能源需求。基于同样的递归神经网络, Ben-Nakhi 和 Mahmoud<sup>[BEN 04]</sup> 预测了三栋办公楼的冷负荷。利用 1997~2000 年的冷负荷数据进行模型训练, 利用 2001 年的数据进行模型测试。Kalogirou<sup>[KAL 00]</sup> 利用神经网络对被动式太阳能建筑 (这里不使用机械和电加热设备) 的能耗进行了预测。考虑不同区域天气对能耗的影响, Yan 和 Yao<sup>[YAN 10]</sup> 利用 BP 神经网络对用采暖度日和冷制度日表示的不同气候带的建筑的冷/热负荷进行了预测, 并以这两种能源测量作为输入变量的一部分来训练神经网络。

在建筑用电量预测的应用中, 早期的研究<sup>[JOI 92]</sup> 已经成功利用神经网络来预测工程中心建筑每小时的电力消耗及冷却水和热水。Nizami 和 Al-Garni<sup>[JAV 95]</sup> 尝试了一种简单的前馈神经网络, 将电能消耗与居住者人数和气象数据联系起来。González 和 Zamarreño<sup>[GON 05]</sup> 利用一种特殊的神经网络预测了短期电力负荷, 这种网络将部分输出作为反馈。相反, Azadeh

等人<sup>[AZA 08]</sup>预测了能源密集型制造业的长期年电力消耗,结果表明当能耗浮动很大时,神经网络很适合解决这个问题。Wong 等人<sup>[WON 10]</sup>采用一个神经网络对亚热带气候下具有照明控制的办公楼进行了能耗预测。模型输出包括每天的制冷、供热、电力照明和整个建筑的电力使用。

ANN 也用于对分段组件的行为,大多是对 HVAC 系统进行分析和优化。Hou 等人<sup>[HOU 06a]</sup>预测了一个建筑中的空调负荷,这是 HVAC 系统优化控制的关键。Lee 等人<sup>[LEE 04]</sup>利用广义回归神经网络对一个建筑的空气处理装置进行故障检测和诊断。Aydinalp 等人<sup>[AYD 02]</sup>说明了神经网络可以用来估计家电、照明和空间制冷(ALC)能耗,也是关于加拿大住宅区的社会经济因素对这些能耗影响的很好的模型。在他们接下来的工作中,将神经网络模型成功地开发来估计同一地区的空间和生活热水采暖能耗<sup>[AYD 04]</sup>。

在文献[BEN 02]和文献[BEN 04]中,利用广义递归神经网络控制空调装置并对公共建筑和办公建筑的 HVAC 储热进行优化。Yalcintas 等人<sup>[YAL 05]</sup>利用神经网络预测热带气候下一个建筑的制冷机组的能源使用情况。后来,他们使用一个三层前馈神经网络预测设备改造中的节能<sup>[YAL 08]</sup>。Gouda 等人<sup>[GOU 02]</sup>利用一个多层前馈神经网络,采用易测量的输入,包括室外温度、太阳辐射度、供暖阀门位置和建筑室内温度来预测内部温度。

建筑能源性能参数可以通过神经网络进行估计。在文献[OLO 99, OLO 02, LUN 02, LUN 04]中,作者估计了总热量损失系数、总热容和增益系数,这对可靠的能量需求预测是很重要的。该方法是基于对简单数据、室内/室外温差、所提供的热量和可用热量进行训练的神经网络模型的分析。Kreider 等人<sup>[KRE 95]</sup>记录了递归神经网络用于小时能耗数据的结果,他们还记录了从建筑数据训练的网络中找到热阻  $R$  和热容  $C$  的结果。Zmeureanu<sup>[ZME 02]</sup>提出了一种利用广义递归神经网络方法评估现有屋顶单元的性能系数。Yalcintas 提出了一种基于 ANN 的热带气候下建筑能源的基准测试技术,重点是预测加权能源使用指数。可选的建筑种类繁多<sup>[YAL 06, YAL 07]</sup>。

训练模型的数据可以通过现场测量、调查、计费采集或者仿真获得。原数据可能含有噪声或者无用变量,因此在模型开发前应进行清洗和降维。关于数据预处理技术的研究有很多。González 和 Zamarreño<sup>[GON 05]</sup>利用神经网络的两个阶段预测了短期电力负荷。第一层预测天气变量,第二层预测能源使用,并以第一层输出作为输入。Yokoyama 等人也利用同样的两阶段技术预测了冷负荷<sup>[YOK 09]</sup>。首先从数据中剔除趋势和周期变化,然后将转换后的数据作为模型训练的主要输入。另外的输入,包括空气温度和相对湿度,被认为使用预测值。在这项工作中,他们对能源需求预测的影响也进行了调查。

Ben-Nakhi 和 Mahmoud<sup>[BEN 04]</sup>预测了下一天的冷负荷分布,用一个单一变量——室外干球温度对模型进行了训练。Ekici 和 Aksoy<sup>[EKI 09]</sup>在没有考虑气候变量的情况下预测了建筑热负荷。网络训练仅有三个输入:透明度比、建筑朝向和绝缘厚度。Kreider 和 Haberl<sup>[KRE 94]</sup>用最近的数据作为输入来预测最近的未来,利用递归神经网络来预测更远的未来。Yang 等人<sup>[YAN 05]</sup>利用累积窗和滑动窗的方法训练神经网络,实现在线建筑能耗预测。滑动窗将输入样本限定在很小的范围内。

Olofsson 等人<sup>[OLO 98]</sup>在预测年供热需求前利用 PCA 对变量进行降维。在后面的工作中,

他们基于短期测量数据实现了长期能源需求预测<sup>[OLO 01]</sup>。Kubota 等人<sup>[KUB 00]</sup>使用遗传算法进行了变量提取和选择,然后开发了模糊神经网络用于建筑能耗负荷预测。这里,变量提取意味着将原始变量转化为有意义的信息,作为模糊推理系统的输入。Hou 等人<sup>[HOU 06a]</sup>将粗糙集和神经网络结合预测空调负荷。利用粗糙集理论找到影响负荷的相关因素,并以此作为预测冷负荷的神经网络的输入。Kusiak 等人<sup>[KUS 10]</sup>通过一个神经网络与五个多层感知器(MLP)组合对建筑物每天的蒸汽负荷进行了预测,在几个案例研究中,它优于其他九种数据挖掘算法,包括分类和回归树(Classification And Regression Trees, CART)、穷举卡方自动交互检测(exhaustive CHI-squared Automatic Interaction Detection, CHAID)、提高树(boosting tree)、多元自适应回归样条(Multivariate Adaptive Regression Splines, MARS)、随机森林(random forest)、SVM、MLP 和 k-近邻(k-Nearest Neighbors, k-NN)。变量选择采用相关系数矩阵算法和提高树算法。Karatasou 等人<sup>[KAR 06]</sup>研究了在每小时能源负荷预测方面统计过程如何改进神经网络模型。在输入预处理和模型选择过程应用统计方法,如假设检验、信息标准和交叉验证。实验结果表明,预测精度可与文献中记录的最好结果相媲美。

神经网络的输出可能并不完全如人们所期望的。Kajl 等人提出了一种模糊逻辑通过对神经网络结果进行后处理来校正输出。模糊辅助允许用户确定几个建筑参数对年和月能耗的影响<sup>[KAJ 96, KAJ 97]</sup>。

在文献中进行了神经网络与其他预测模型之间的比较。Azadeh 等人<sup>[AZA 08]</sup>证明了神经网络特别适用于能耗波动很大的制造业的年度用电量预测。通过方差分析,它优于传统的非线性回归模型。Aydinalp 等人<sup>[AYD 02]</sup>证明了对于估计加拿大住宅区 ALC 能耗及社会经济因素对能耗的影响方面,神经网络比工程模型的预测性能更好。后来,在文献[AYD 08]中对 ANN 与 CDA 方法进行了比较。在这项工作中,发现在解决与 ANN 模型面临相同问题时,CDA 具有很好的解决能力,而且 CDA 更容易开发和使用。Neto<sup>[NET 08]</sup>对预测建筑能耗的神经网络模型和精细的工程方法进行了比较,两种模型都表现出了很高的预测精度,但在短期预测方面 ANN 模型比工程模型略好。

### 1.5.2 SVM

SVM 越来越多地应用于研究和工业。即使很少的训练数据,它们仍是解决非线性问题的很有效的模型。过去的5年中,人们对将 SVM 模型应用于建筑能源分析进行了大量研究。

首先,Dong 等人<sup>[DON 05a]</sup>将 SVM 应用于预测热带区域四个建筑的月电力消耗。对三年数据进行了训练,并将所得模型用于1年数据来预测当年的房东效用(landlord utility)。结果表明,SVM 在这一问题上表现很好。

Lai 等人<sup>[LAI 08]</sup>将该模型应用于一栋建筑的1年电力消耗。变量包括天气变化。在实验中,模型是从一年的性能中得出,然后测试3个月的行为。他们也测试了该模型每天的基础数据集,以验证短期内该方法的稳定性。另外,他们对一定的历史性能手动增加了扰动,通过检查贡献权重的变化利用该模型检测扰动。

Li 等人<sup>[LI 09]</sup>采用 SVM 预测一座办公楼的每小时冷负荷,发现支持向量回归机的性能

优于传统的 BP 神经网络。Hou 和 Lian<sup>[HOU 09]</sup> 也采用 SVM 预测 HVAC 系统的冷负荷。结果表明 SVM 优于 ARIMA 模型。

Li 等人<sup>[LI 10a]</sup> 利用 BP 神经网络、径向基神经网络、广义递归神经网络和 SVM 方法对建筑的年度电力消耗进行了预测。他们发现，与其他模型相比较，广义递归神经网络和 SVM 更适用于这类问题。并且，在所有预测模型中，SVM 的性能最好。模型以 59 栋大楼的数据进行了训练，并以 9 栋大楼做测试。

将物理模型和 SVM 结合，Liang 和 Du<sup>[LIA 07]</sup> 提出了一种经济、有效的用于 HVAC 系统的故障检测和诊断方法。通过使用一个四层 SVM 分类器，用少量的训练样本可以很快且精确地辨识出正常情况和三种可能故障。三种主要故障是循环挡板卡住 (recirculation damper stuck)、冷却线圈结垢/块 (cooling coil fouling/block) 和送风风机速度下降 (supply fan speed decreasing)。指标是供气和混合气体温度、出口水温和阀门控制信号。

进行了前或后处理模型训练的研究。在预测建筑制冷负荷训练 SVM 前，Lv 等人<sup>[LV 10]</sup> 采用 PCA 减少变量。Li 等人<sup>[LI 10c]</sup> 采用了改进的 PCA，称为核主成分分析。Li 等人<sup>[LI 10b]</sup> 根据样本的相似性，采用了一个模糊 C 均值聚类算法来聚类样本。然后，他们对每个样本利用模糊隶属度函数来表示其对模型的贡献。在后处理阶段，Zhang 和 Qi<sup>[ZHA 09]</sup> 在利用 SVM 预测完建筑热负荷后，利用马尔可夫链进行进一步的区间预测。

## 1.6 现有模型的比较

通过以上的描述和分析，显然需要大量的计算来评估建筑能源系统，从子系统到建筑级，甚至区域级或国家级。表 1.1 对回顾性研究工作进行了简要总结。考虑到问题和模型的区别，省略了工程方法，因为它们可以解决所有问题。在某些情况下，每个模型都有其自身的优点。

表 1.1 建筑能耗预测常用方法综述

问题	统计	ANN	SVM
供热 / 制冷	[BAU 98, ANS 05] [DHA 99, DHA 98]	[KAL 97, EKI 09, OLO 98] [OLO 01, YAN 10, YOK 09] [KRE 95, BEN 04, KAL 00]	[LI 09, HOU 09] [LV 10, ZHA 09]
电力	[MA 10, HOF 98] [AZA 08, NEW 10]	[JOI 92, GON 05, AZA 08] [WON 10, AZA 08]	[DON 05a, LAI 08] [LI 10a]
简化	[DHA 98, DHA 99] [LEI 09]	[BEN 04, EKI 09, OLO 98] [KUB 00, KUS 10]	
系统级	[ANS 05, LEI 09] [MA 10, CHO 04]		
子系统		[HOU 06a, LEE 04, AYD 02] [AYD 04, BEN 02, BEN 04] [YAL 05, YAL 08, GOU 02]	
能源参数	[JIM 05]	[OLO 99, OLO 02, LUN 02] [LUN 04, KRE 95, ZME 02]	
能源指标	[LAM 10, GHI 06]	[YAL 06, YAL 07]	
数据预 / 后处理	[CHO 04, NEW 10]	[KAJ 96, KAJ 97, KRE 94] [YAN 05, KAR 06, KUS 10]	[LI 10c, LV 10] [ZHA 09, LI 10b]