计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2019,28(1):156−162 [doi: 10.15888/j.cnki.csa.006752]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 Q-Learning 算法的建筑能耗预测①

陈建平, 陈其强, 胡　文, 陆　悠, 吴宏杰, 傅启明

(苏州科技大学 电子与信息工程学院, 苏州 215009)
(江苏省建筑智慧节能重点实验室, 苏州 215009)
(苏州市移动网络技术与应用重点实验室, 苏州 215009)
通讯作者: 傅启明, E-mail: fqm_1@126.com

摘　要: 提出一种基于 Q-learning 算法的建筑能耗预测方法. 通过将建筑能耗预测问题建模为一个标准的马尔科夫决策过程, 利用深度置信网对建筑能耗进行状态建模, 结合 Q-learning 算法, 实现对建筑能耗的实时预测. 通过美国巴尔的摩燃气和电力公司公开的建筑能耗数据进行测试实验, 结果表明, 基于本文所提出的模型, 利用 Q-learning 算法可以实现对建筑能耗的有效预测, 并在此基础上, 基于深度置信网的 Q-learning 算法具有更高的预测精度. 此外, 实验部分还进一步验证了算法中相关参数对实验性能的影响.

关键词: 强化学习; 建筑能耗预测; Q-learning; 深度置信网

引用格式: 陈建平,陈其强,胡文,陆悠,吴宏杰,傅启明.基于 Q-Learning 算法的建筑能耗预测.计算机系统应用,2019,28(1):156–162. http://www.c-s-a.org.cn/1003-3254/6752.html

Prediction of Building Energy Consumption Based on Q-Learning

CHEN Jian-Ping, CHEN Qi-Qiang, HU Wen, LU You, WU Hong-Jie, FU Qi-Ming

(College of Electronics and Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China)
(Jiangsu Key Laboratory of Building Intelligent Energy Saving, Suzhou 215009, China)
(Suzhou Key Laboratory of Mobile Network Technology and Application, Suzhou 215009, China)

Abstract: This study proposed a building energy consumption prediction method based on Q-learning algorithm. By modeling the building energy consumption prediction problem as a standard Markov decision process, combining with the deep belief network to model the state, we use Q-learning algorithm to achieve the real-time prediction of the building energy consumption. Based on the building energy consumption data published by Baltimore Gas and Electric Power Company of the United States, the proposed model were tested and the results show that the Q-learning algorithm can be used to predict the building energy consumption successfully. Moreover, deep belief network can improve the prediction accuracy effectively. In addition, some experimental results further verify the influence of related parameters on experimental performance.

Key words: reinforcement learning; building energy consumption prediction; Q-learning; deep belief network

1　引言

建筑作为能耗占比最大的领域, 虽然自身拥有巨大的节能潜力, 但是, 随着经济的高度发展, 建筑面临的高能耗低能效的问题也日益严峻. 近年来, 我国在建

① 基金项目: 国家自然科学基金 (61502329, 61876121, 61772357, 61750110519, 61672371, 61602334, 61472267); 江苏省重点研发计划 (BE2017663); 江苏省高校自然科学研究项目 (18KJB520045); 江苏省建设系统科技指导项目 (2017ZD005)
Foundation item: National Natural Science Foundation of China (61502329, 61876121, 61772357, 61750110519, 61672371, 61602334, 61472267); Key Research and Development Program of Jiangsu Province (BE2017663); Natural Science Research Program of Higher Education, Jiangsu Province (18KJB520045); Science and Technology Guidance Program of Construction System, Jiangsu Province (2017ZD005)
收稿时间: 2018-07-31; 修改时间: 2018-08-27; 采用时间: 2018-08-30; csa 在线出版时间: 2018-12-26

筑节能领域取得了明显进展, 但从能耗预测的角度看, 建筑能耗预测仍然存在很多不足之处[1]. 构建建筑能耗预测模型是预测建筑未来时刻能耗、在线控制能耗以及获取能耗运行最优策略的前提和核心[2–4]. 但是, 建筑具有面积大、能耗大和能耗复杂等特点, 并且建筑自身是一个包含多种系统, 设备相互连接的复杂非线性系统. 因此, 研发精度高、适应性强的能耗预测模型并非是件容易的事. 从建筑自身来说, 其能耗受到多种外界因素的影响, 例如外界气候、建筑物自身结构、内部设备运行特点、人员分布动态特征等, 这些因素使得建筑物能耗变得更加复杂, 也加剧了能耗预测的难度. 近年来, 国内外许多业界学者和专家的主要关注点在于如何在提高建筑能耗预测的准确性并简化能耗预测模型的同时, 实现在线控制及优化建筑能耗. 建筑的完整生命周期包括很多环节, 如设计、建造、运行、维护等, 其中每个环节运用的节能方法或者技术都能够对实现节能的目标产生重要影响, 因此, 能耗预测在建筑节能中就显得势在必行. 与此同时, 建筑在自身运行中产生的一大批真实能耗数据被搁置或者直接丢弃, 并没有真正实现任何价值, 对节能而言, 又是一种资源浪费.

强化学习在智能建筑领域, 尤其是在建筑节能问题上已经引起国内外相关学者的广泛关注. Dalamagkidis 等人提出设计一种线性强化学习控制器, 可监督控制建筑热舒适度、空气质量、光照需求度、噪音等, 与传统 Fuzzy-PD 相比, 其效果更优[5]; Yu 等人提出了一种用强化学习在线调整低能耗建筑系统的监督模糊控制器的无模型方法, 其中用 Q-learning 算法监控建筑物的能源系统[6]; Bielskis 等人利用强化学习方法构建室内照明控制器, 通过强化学习方法自适应调节照明系统, 进而节约能源消耗[7]; Li 等人提出一种多网格 Q-learning 方法, 通过近似建筑环境模型求解近似节能优化策略, 并将初始策略用于精确建筑模型, 在线学习最优控制策略, 加快算法求解实际问题中的收敛速度[8]; Liu 等人提出基于强化学习 Q-learning 算法监督控制建筑热质量, 进而节约能源消耗[9]; Yang 等人提出基于强化学习方法的建筑能耗控制方法, 该控制方法运用表格式 Q 学习和批量式 Q 学习在 Matlab 平台上实现建筑能耗控制, 实验结果表明, 该方法较其他方法多节约百分之十的能耗[10]; Zamora-Martínez 等人给出一种利用位置环境在线预测建筑物能耗的方法, 该方法从一个完全随机的模型或者一个无偏的先验知识获得模型参数, 并运用自动化技术使得房屋适应未来的温度条件, 达到节能的效果[11]; Nijhuis 等人提出一种基于公开可用的数据开发住宅负载模型, 该模型运用强化学习中蒙特卡罗算法, 基于时间使用规律对家庭居住房屋进行建模, 该模型中主要影响能耗的相关因素为天气变量、领域特征和人员行为数据, 通过对 100 多个家庭每周的用电量进行验证, 实验结果表明, 该方法的预测性能较其他类似方法更精确[12]; Liesje Van Gelder 等人基于强化学习中蒙特卡罗方法提出一种整合影响建筑能耗诸多不确定因素的概率分析和设计方法, 该方法可以合并原模型, 取代原始模型, 并检查潜在情景的优化结果[13].

本文利用 DBN 将建筑能耗初始状态映射至高维特征空间, 结合强化学习中 Q-learning 算法, 将输出的状态特征向量作为 Q-learning 算法的输入, 实现对建筑能耗的预测. 实验表明, 运用强化学习进行能耗预测是可行的, 并且改进后的能耗预测方法精度更高, 这充分说明了强化学习在建筑能耗预测领域具有很大的研究潜力.

2 相关理论

2.1 马尔可夫决策过程

能成功保留所有相关信息的状态信号就是具有马尔科夫性 (Markov Property) 的, 而只要具有马尔可夫性的强化学习问题就被称为马尔可夫决策过程 (Markov Decision Process, MDP). 马尔可夫性可作如下定义: 假设强化学习问题中, 状态和奖赏值的数量都是有穷的, 在问题中, 学习器 (Agent) 与环境交互, 在t时刻执行动作后, 会在$t+1$时刻获得一个反馈, 在最普通的情况中, 这个反馈可能是基于前面发生的一切, 因此, 这种环境动态性可以通过概率分布来定义, 如公式 (1) 所示:

$$\Pr\{x_{t+1}=x', r_{t+1}=r|x_t,u_t,r_t,\cdots,r_1,x_0,u_0\} \quad (1)$$

然而, 如果状态信号也是有马尔可夫性的, 那么 Agent 在$t+1$时获得的环境反馈就只取决于 Agent 在t时刻的状态和动作. 在此情况下, 这种环境动态性可以通过公式 (2) 来定义:

$$\Pr\{x_{t+1}=x', r_{t+1}=r|x_t,u_t\} \quad (2)$$

即当且仅当对所有的x', r, 以及历史 $x_t,u_t,r_t,x_{t-1},u_{t-1},\cdots,r_1,x_0,u_0$, 有公式 (1) 与公式 (2) 相等, 则状态是具有马尔可夫性的, 即 Agent 在下一时刻获得的反馈只与上一时刻的状态和动作相关, 这种情况下, 环境和问题作为一个整体也是具有马尔可夫性的.

MDP 可以用来对强化学习问题进行建模, 其通常被定义为一个四元组, $M=(X,U,R,T)$, 其中X表示状态集合; U表示动作集; R表示奖赏函数, $R(x,u)$是指 Agent 在状态x时采取动作u所获得的回报值; T是状态转移函数, $T(x,u,x')$是指 Agent 在状态x下采取动作u后转移到状态x'的概率.

强化学习的最终目标是要学习到一个能够获得最大期望累计奖赏的最优策略, 并利用该策略进行决策. 然而, 由于最终计算获得的最优策略可能是一个动作, 也可能是某个动作被选择的概率, 因此, 策略被分为确定策略 (deterministic policy) 和随机策略 (random policy) 两种. 其中, 确定策略$\bar{h}:X\to U$表示 Agent 在某一状态下执行某一动作, 例如$u=\bar{h}(x)$表示 Agent 在状态x下执行动作u; 随机策略$\tilde{h}:X\times U\to[0,1]$表示 Agent 在某一状态下执行某一动作的概率, 例如$P(u|x)=\tilde{h}(x,u)$表示 Agent 在状态x下执行动作u的概率. 在本文中, 策略直接用h表示, 策略是每个状态$x\in X$和动作$u\in U$在状态x下执行动作u的概率$h(x,u)$的映射. 假设当前时刻为k, 当前状态为x_k, 策略为h, 而 Agent 依据当前状态x_k和策略h执行动作u_k后, 在$k+1$时刻, Agent 通过环境反馈, 获得的立即奖赏为$R(x_k,u_k,x_{k+1})$. Agent 在强化学习问题中, 不断地重复上述过程, 并且与环境不断交互, 学习到最优策略, 并达到获取最大期望累计奖赏的目的.

对 Agent 在给定一个状态或者状态动作对时, 为了评估该状态或者状态动作对的好坏程度, 在强化学习中给出值函数的定义. 几乎所有强化学习算法都是通过值函数对策略进行评估, 而值函数有状态值函数$V^k(x)$和动作值函数$Q^k(x,u)$两种. 其中, $V^k(x)$表示 Agent 在当前状态x下遵循策略h的期望回报; 而$Q^k(x,u)$表示 Agent 在当前状态动作对(x,u)下遵循策略h所能获得的期望回报. $V^k(x)$和$Q^k(x,u)$是相应 Bellman 公式的不动点解, 如公式 (3) 和公式 (4) 所示:

$$V^h(x)=\sum_{u\in U}h(x,u)\sum_{x'\in X}T(x,u,x')[R(x,u,x')+\gamma V^h(x')] \tag{3}$$

$$Q^h(x,u)=\sum_{x'\in X}T(x,u,x')[R(x,u,x')+\gamma\sum_{u'\in U}h(x',u')Q^h(x',u')] \tag{4}$$

其中, γ是折扣因子. 最优策略h^*表示能够获得最大期望累计奖赏的最优策略, 而最优策略所相应的最优值函数和最优动作值函数$V^*(x)$和$Q^*(x,u)$如公式 (5) 和 (6) 所示:

$$V^*(x)=\max_{u\in U}\sum_{x'\in X}T(x,u,x')[R(x,u,x')+\gamma V^*(x')] \tag{5}$$

$$Q^*(x,u)=\sum_{x'\in X}T(x,u,x')[R(x,u,x')+\gamma\max_{u'\in U}Q^*(x',u')] \tag{6}$$

上述两个公式也被称作最优 Bellman 公式.

2.2 Q-learning 算法

Q-learning 是一种经典的离策略算法, 其更新准则: $Q(s,a)\leftarrow Q(s,a)+\alpha[r+\gamma Q\max(s',a')-Q(s,a)]$, 即 Q-learning 是利用基于行为策略所选择的实际动作来更新目标策略 Q 值的. Q-learning 算法具体流程如算法 1 所示[14].

算法 1. Q-learning 算法

1. 随机初始化$Q(s,a)$
2. Repeat(for each episode)
3. 初始化s
4. Repeat(for each step of episode)
5. 利用从Q中得到的策略在s中选择a
6. 采取动作a, 得到r, s'
7. $Q(s,a)\leftarrow Q(s,a)+\alpha[r+\gamma\max Q(s',a')-Q(s,a)]$
8. $s\leftarrow s'$
9. 直到s是终止状态
10. 直到Q最优

3 基于 Q-learning 算法的建筑能耗预测方法

3.1 建筑能耗状态表示

DBN 可以应用在多种领域的问题中, 例如执行非线性维数减少、图像识别、视频序列和动作捕捉数据等问题. 此外, DBN 可以根据不同的抽象层将学习任务分解成相应的子问题.

DBN 由很多限制玻尔兹曼机堆叠在一起, DBN 是一个时滞神经网络, 主要分为可视层$\mathbf{v}$和隐层$\mathbf{h}$, 每一层之间存在相关链接, 但每一层内的单元之间不存在相互链接. 隐层每个单元的作用主要在于获取可视层单元的输入数据所具有的高阶数据的特征, 因此, 由可视层和隐层链接配置的能量被定义为:

$$E(\mathbf{v},\mathbf{h})=-\sum_{i,j}v_ih_jW_{ij}-\sum_i v_ia_i-\sum_j h_jb_j \tag{7}$$

其中, i表示可视层节点, j表示隐层节点, w_{ij}表示第i个可视单元与第j个隐层单元之间的权重. 此外, v_i和v_j表示第i个可视单元与第j个隐层单元的状态, a_i和b_j表示可视层与隐层的偏置向量. $\sum_{i,j}v_ih_jW_{ij}$表示隐层和可视单元

之间的输出与它们的相关权重的乘积和; $\sum_i v_i a_i$和$\sum_j h_i b_j$分别表示可视层和隐层的输出. RBM 定义了一个联合的概率$p(\mathbf{v},\mathbf{h})$, 覆盖了隐层和可视层.

$$p(\mathbf{v},\mathbf{h})=\frac{e^{-E(\mathbf{v},\mathbf{h})}}{Z} \tag{8}$$

其中, $Z=\sum \mathbf{v},\mathbf{h} e^{-E(\mathbf{v},\mathbf{h})}$表示归一化常数. 为了确定一个由状态$v$表示的数据点的概率, 通常使用边缘概率$p(v)=\sum_h p(\mathbf{v},\mathbf{h})$对隐藏层的状态进行求和. 对于可见层或者隐藏层任意给定点的输入都可以用上述方程计算概率. 为了确定模型中的条件概率, 这些值被进一步用于执行推理. 为了使模型的可能性最大化, 必须要计算出关于权重对数的梯度. 式 (7) 中第一项的梯度在一些代数运算之后可以写成:

$$\frac{\partial \log(\sum_h \exp(-E(v,h)))}{\partial W_{ij}}=v_i\cdot p(h_j=1|v) \tag{9}$$

在 RBM 中隐藏层和可视层被激活的概率可以用下面的公式来表示:

$$p(h_j=1|\mathbf{v})=\sigma(b_j+\sum_i v_i W_{ji}) \tag{10}$$

$$p(v_j=1|\mathbf{h})=\sigma(a_j+\sum_j h_i W_{ji}) \tag{11}$$

其中, $\sigma(\cdot)$表示 sigmoid 函数. 此外, 为了学习一个 RBM, 在训练数据中, 权重向量更新公式如下:

$$\Delta W_{ij}=\frac{\partial \log(p(\mathbf{v},\mathbf{h}))}{\partial W_{ij}}=\langle v_i h_j\rangle_0-\langle v_i h_j\rangle_\infty \tag{12}$$

其中, $\langle v_i h_j\rangle_0$表示样本数据分布的期望, $\langle v_i h_j\rangle_\infty$表示模型分布下的期望.

总的来说, 一个深度信念网络是由一个任意的数字给出的. 其中, 可视层x(输入向量) 和l层隐层h^k的之间的联合分布的定义如公式 (13) 所示:

$$p(\mathbf{x},\mathbf{h}^1,\cdots,\mathbf{h}^k)=\prod_{k=0}^{l-2} p(\mathbf{h}^k|\mathbf{h}^{k+1})p(\mathbf{h}^{l-1},\mathbf{h}^l) \tag{13}$$

其中, $p(\mathbf{h}^k|\mathbf{h}^{k+1})$指可视层在 RBM 第$k$层的隐层单元的条件分布概率; $p(\mathbf{h}^{l-1},\mathbf{h}^l)$指在顶层的 RBM 中可视层和隐层的联合分布概率.

如图 1 所示, 一个 DBN 包含 1 个可视层和 3 个隐层, 其中, $v(i)$层是可视层; $h_1(j)$, $h_2(k)$和$h_3(l)$是隐层. 可视层的每个单元代表真实值, 隐层的每个单元代表 2 进制的神经元. DBN 可以通过贪心无监督的方法进行训练, 通过从下到上的顺序分别训练其中的每一个 RBM, 使用隐层的输出作为下一个 RBM 的输入, 直到最后一个 RBM 被训练结束. 此外, DBN 通过在模型的底层修改初始状态以此推断出最顶层的隐藏层, 从而将从环境中获取的初始状态映射到二值状态空间.

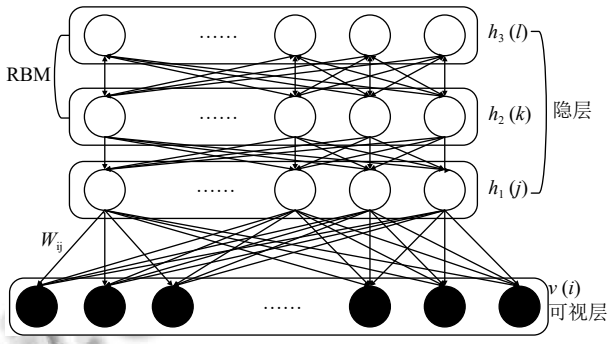


图 1 DBN 框架图

3.2 环境建模

从历史数据可知, 每个时刻测量的能耗值在区间[1,5]. 将连续三个时刻的历史能耗作为 DBN 的输入, 用于表示当前的状态, 并利用 DBN 对当前状态进行特征提取. 经过 DBN 特征提取后, 将最终计算出的输出的状态表示值作为强化学习 Q-learning 算法的输入, 即状态集X. 在 Q-learning 算法中, 动作U与状态X是对应的, 即$U\in[1,5]$, 通过在区间[1,5]上选择最优动作, 达到预测下一时刻能耗值的目的.

假设预测的下一时刻的能耗值用E^{t+1}表示, 实际的能耗值用e^{t+1}, 则两者之间的误差为$E=|E^{t+1}-e^{t+1}|$. 奖赏被建模为一个负值的变量, 相当于预测结果的惩罚值, 如公式 (14) 所示:

$$r=-E \tag{14}$$

即当预测的能耗值与实际能耗值越接近, 获得的r值就越大, 反之, 获得的r值就越小.

3.3 基于 Q-learning 算法的建筑能耗预测算法

DBN 将从环境中获取的初始状态映射到一个二值状态空间, 并且将获得的状态作为 Q-learning 算法的输入, 基于 Q-learning 算法的建筑能耗预测算法具体流程如算法 2 所示.

算法 2. 基于 Q-learning 算法的建筑能耗预测算法

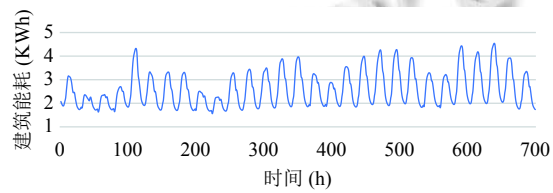
1. 初始化 DBN, 输入状态集X
2. For each RBM in DBN
3. Repeat
4. For each $x\in X$
5. 令 RBM 可见层$RBM^{visible}=x$
6. 执行 RBM, 训练出 RBM 的参数
7. 更新 RBM 的权重及各个节点的偏置, 固定 RBM 的参数

8. End for
9. Until converge
10. For each $x \in X$
11. 令 RBM 可见层$RBM^{visible}=x$
12. 推断出RBM^{hidden}
13. 将RBM^{hidden}作为下一个 RBM 的可见层
14. End for
15. End for
16. 将最后计算出的X作为 Q-learning 算法的状态集
17. 随机初始化$Q(s,a)$; $s \in X$
18. Repeat(for each episode)
19. 初始化s
20. Repeat(for each step of episode)
21. 利用从Q中得到的策略在s中选择α
22. 采取动作α, 得到r, s'
23. $Q(s,a) \leftarrow Q(s,a)+\alpha[r+\gamma \max Q(s',a')-Q(s,a)]$
24. $s \leftarrow s'$
25. 直到是终止状态
26. 直到最优

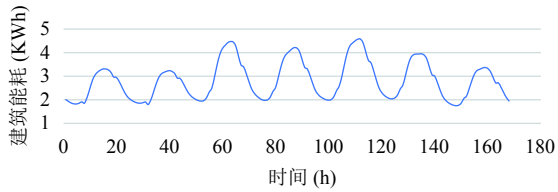
4 实验及结果分析

4.1 实验设置

为了验证本文所提出的建筑能耗预测方法的有效性, 本章节采用的实验数据是美国巴尔的摩燃气和电力公司记载的建筑能耗数据, 具体时间为 2007 年 1 月至 2017 年 12 月. 本节以其中 General Service(< 60 kW)部分数据为实验数据集, 如图 2 所示, 子图 1 表示 2017 年 9 月共 30 天的能耗数据, 子图 2 展示的是 2017 年 9 月 23 日至 9 月 29 日共一周 7 天的能耗数据, 图 2 中数据采集的步长为 1 次/1 h. 在所有实验中, 数据集分为两部分, 一部分用于模型的训练, 一部分用来评估该能耗预测方法的性能, 学习速率$\alpha = 0.4$, 折扣因子$\gamma = 0.99$.



(a) 2017年9月1日~2017年9月30日



(b) 2017年9月23日~2017年9月29日

图 2 能耗实际值

4.2 实验结果分析

图 3 和图 4 主要展示了 Q-learning、基于 DBN 的 Q-learning 算法对一个星期的建筑能耗预测值与实际值的对比图, 横坐标表示时间, 纵坐标表示建筑能耗. 在实验过程中, 每个算法都被独立执行 20 次, 图中的数据即 20 次实验的平均值. 从两幅图中可以看出, 两种算法都可以预测出未来一周的建筑能耗值. 因此, 用 DBN 构建能耗动态模型, 并采用 Q-learning 算法进行建筑能耗预测的方法是可行的. 此外, 从图中可以清晰地看出, 改进的基于 DBN 的 Q-learning 算法的能耗预测准确性较经典 Q-learning 算法更高, 主要原因是通过 DBN 构造高维特征向量, 进一步提高函数逼近器的泛化能力, 提高算法预测的准确性.

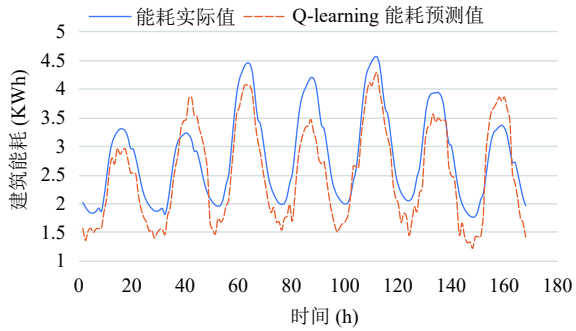


图 3 Q-learning 算法能耗预测值与实际值对比

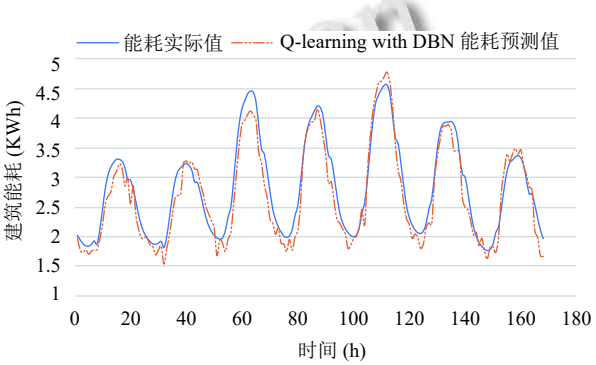


图 4 基于 DBN 的 Q-learning 算法能耗预测值与实际值对比

表 1 主要表示了 DBN 隐藏层神经元个数不同时, 不同算法对能耗预测的性能分析. 表格中的数据表示建筑能耗实际值与预测值的均方根误差, 表格中的数据是算法被独立执行 20 次的平均值. 从表 1 还可以看出相同算法在不同隐藏层神经元的个数下, 算法的性能也不一致, 当隐藏层神经元个数为 5, 10, 20, 50, 100 时, 建筑能耗的预测值与实际值的均方根误差分别

为 0.325, 0.225, 0.122, 0.127, 0.138. 由此可以看出, 神经元个数越少, 预测的准确性越差, 而神经元个数越多时, 预测的准确性越好, 但是当神经元数量足够多时, 预测的准确性几乎保持一致, 甚至准确性变差. 由此可见, 为了提高建筑能耗预测的准确性, 选择合适的隐藏层神经元个数是有必要的, 由表 1 可知, 本文中, 隐藏层神经元个数取 20.

表 1 神经元个数对基于 DBN 的 Q-learning 算法预测性能的影响

神经元数量	5	10	20	50	100
RMSE value	0.325	0.225	0.122	0.127	0.138

表 2 主要表示了不同 α 值以及在不同数据步长对基于 DBN 的 Q-learning 算法预测性能的影响分析. 表格的第一行表示 α 的不同取值, 表格的第一列表示数据的步长, 即每个数据之间的时间间隔分别为 1 h, 1 day, 1 week 和 1 month. 表格中的数据表示建筑能耗实际值与预测值的均方根误差, 都是算法被独立执行 20 次的平均值. 由表 2 可以清晰地知道, 当数据步长为 1 h 时, 尽管 α 的取值在不断变化, 建筑能耗的预测值与实际值的均方根误差总是比较稳定, 预测的准确性较高; 当数据步长为 1 week 时, α 取值越大, 建筑能耗的预测值和实际值的均方根误差越小, 预测的准确性相对较低; 而当数据步长为 1 week 和 1 month 时, α 取值越大, 建筑能耗的预测值和实际值的均方根误差越大, 预测的准确性更低. 同样的, 当 $\alpha(\alpha\geq 0.4)$ 取值一致时, 数据的步长越小, 建筑能耗的预测值和实际值的均方根误差越小, 预测的准确性越高; 数据的步长越大, 建筑能耗的预测值和实际值的均方根误差越大, 预测的准确性越低. 综上所述, 为了最大化能耗预测的准确性, 这里我们选取数据步长为 1 h, α 取值我们选取 0.4.

表 2 不同α值及不同数据步长对基于 DBN 的 Q-learning 算法预测性能的影响

数据步长	α 值					
	0.2	0.3	0.4	0.5	0.6	0.7
1 h	0.136	0.132	0.122	0.119	0.129	0.135
1 day	1.233	1.156	0.985	0.912	0.685	0.843
1 week	1.312	1.114	1.109	1.112	1.698	1.723
1 month	1.205	2.209	2.352	2.417	2.423	2.436

5 结束语

本文提出一种基于 Q-learning 算法的建筑能耗预测模型. 该模型通过深度置信网自动提取特征, 并利用贪心无监督的方法自下而上地训练深度置信网中的每一个 RBM. 所提出的模型将隐层的输出作为下一个 RBM 的输入, 实现对能耗状态的预处理, 并以此构建高维状态向量. 此外, 该模型将能耗预测问题建模为一个标准的马尔可夫决策过程, 将深度置信网的输出状态向量作为 Q-learning 算法的输入, 利用 Q-learning 实现对能耗的实时预测. 为了验证模型的有效性, 本文采用美国巴尔的摩燃气和电力公司记载的建筑能耗数据进行测试实验, 实验结果表明, 所提出的模型可以有效地预测建筑能耗, 并且基于 DBN 的 Q-learning 算法较传统的 Q-learning 算法有较高的预测精度. 此外, 本文还进一步分析了相关参数对算法性能的影响.

本文主要对单一固定建筑能耗进行预测, 下一步, 将考虑对多样变化的建筑能耗进行预测和迁移研究, 同时不断完善模型, 更好地实现建筑能耗预测, 进一步达到建筑节能的目的.

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-2014. 北京: 中国建筑工业出版社, 2014.

2 Zhao HX, Magoulès F. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews, 2012, 16(6): 3586–3592. [doi: 10.1016/j.rser.2012.02.049]

3 Kim BG, Zhang Y, van der Schaar M, *et al*. Dynamic pricing and energy consumption scheduling with reinforcement learning. IEEE Transactions on Smart Grid, 2016, 7(5): 2187–2198. [doi: 10.1109/TSG.2015.2495145]

4 Dalamagkidis K, Kolokotsa D. Reinforcement learning for building environmental control. In: Reinforcement Learning. Weber C, Elshaw M, Mayer NM, eds. London: IntechOpen, 2008. [doi: 10.5772/5286]

5 Dalamagkidis K, Kolokotsa D, Kalaitzakis K, *et al*. Reinforcement learning for energy conservation and comfort in buildings. Building and Environment, 2007, 42(7): 2686–2698. [doi: 10.1016/j.buildenv.2006.07.010]

6 Yu Z, Dexter A. Online tuning of a supervisory fuzzy controller for low-energy building system using reinforcement learning. Control Engineering Practice, 2010, 18(5): 532–539. [doi: 10.1016/j.conengprac.2010.01.018]

7 Bielskis AA, Guseinoviene E, Dzemydiene D, *et al*. Ambient lighting controller based on reinforcement learning components of multi-agents. Elektronika ir Elektrotechnika, 2012, (5): 79–84.

8 Li BC, Xia L. A multi-grid reinforcement learning method for energy conservation and comfort of HVAC in buildings. Proceedings of 2015 IEEE International Conference on Automation Science and Engineering. Gothenburg, Sweden. 2015.

9 Liu S, Henze GP. Investigation of reinforcement learning for building thermal mass control. Proceedings of the First National IBPSA-USA Conference. Boulder, CO, USA. 2004.

10 Yang L, Nagy Z, Goffin P, *et al*. Reinforcement learning for optimal control of low exergy buildings. Applied Energy, 2015, 156: 577–586. [doi: 10.1016/j.apenergy.2015.07.050]

11 Zamora-Martínez F, Romeu P, Botella-Rocamora P, *et al*. On-line learning of indoor temperature forecasting models towards energy efficiency. Energy and Buildings, 2014, 83: 162–172. [doi: 10.1016/j.enbuild.2014.04.034]

12 Nijhuis M, Gibescu M, Cobben JFG. Bottom-up Markov chain Monte Carlo approach for scenario based residential load modelling with publicly available data. Energy and Buildings, 2016, 112: 121–129. [doi: 10.1016/j.enbuild.2015.12.004]

13 Van Gelder L, Janssen H, Roels S. Probabilistic design and analysis of building performances: Methodology and application example. Energy and Buildings, 2014, 79: 202–211. [doi: 10.1016/j.enbuild.2014.04.042]

14 Sutton RS, Barto AG. Reinforcement learning: An introduction. Cambridge: MIT Press, 1998.