

基于知识-数据混合驱动的综合能源系统多元负荷预测方法^①

王力成^{②*} 王子非* 邓宝华* 凌 锋* 张有兵*

(* 浙江工业大学信息工程学院 杭州 310023)

(** 之江实验室人工智能研究院 杭州 311121)

摘 要 当前的综合能源系统(IES)负荷预测方法几乎都是单一的数据驱动方法,忽略了 IES 中的能量耦合关系。此外,在现有的研究中,数据驱动方法的训练数据主要集中于历史负荷、气象等影响因素,较少考虑可再生能源出力以及不同能源供应给 IES 负荷预测结果带来的影响。针对上述问题,本文提出一种知识-数据混合驱动的 IES 多元负荷预测方法。该方法首先通过解析模型对 IES 中的能量耦合特性知识进行描述,并利用该知识模型对原始样本数据进行重构。然后将重构后的新样本数据作为数据驱动模型的训练样本,并使用基于随机森林算法的特征选择方法和 Dropout 技术提高模型的泛化能力。最后采用某 IES 工业园区的实际数据对本文所提方法的有效性进行了验证。仿真结果表明,该方法相较于传统单一数据驱动模型具有更好的预测效果和较高的可靠性。

关键词 综合能源系统(IES); 知识-数据混合驱动; 能量耦合特性; 多元负荷预测; 随机森林算法; Dropout 技术

0 引 言

负荷预测作为保障能源系统安全稳定运行的前提,可以为系统的优化调度提供数据参考^[1]。目前对于负荷预测的研究主要针对单一能源系统,仅侧重于某一类能量的有效利用,不能充分发挥出多种能源互补的优势,同时也无法保证综合能源系统(integrated energy system, IES)整体的安全经济运行。而 IES 的多元负荷预测一方面可为系统运行规划提供先验知识,极大提高系统的灵活调度能力,另一方面 IES 能够充分利用多种能源,因此 IES 的多元负荷预测成为当前研究的一大热点。当前,由于综合能源系统所采集到的数据信息包含了大量的历史负荷数据、环境因素数据和气象因素数据等,且这些数据的随机性和复杂程度较高,因此,传统应用于综合能源系统的负荷预测方法难以达到令人满意的

效果^[2]。

近年来,深度学习由于具有“特征提取能力强、计算效率高”等特点^[3-4],因此受到国内外学者的广泛研究。文献[5,6]分别基于循环神经网络和径向基函数神经网络较好地实现了负荷的短期预测,但算法程序的实现较为复杂,且模型简单,无法对复杂冗余数据进行有效处理。文献[7-9]分别利用灰色关联分析法和改进的粒子群算法对深度学习算法做出了相应的改进,改进后算法的预测效果得到了一定程度上的提升。然而,由于深度学习方法与综合能源系统的内在物理结构是相互独立的,难以解释具有强因果关系的 IES 系统中各能源子网络之间存在的能量耦合特性。因此,单纯利用深度学习模型得到的负荷预测结果受到了一些专家和学者的质疑。

本文综合考虑到 IES 物理结构模型中包含的能

① 国家自然科学基金(51777193)和浙江省自然科学基金(LQ20E070002)资助项目。

② 男,1989年生,博士,副教授;研究方向:低惯量电力系统频率控制,深度强化学习在网络化系统中的应用,可再生新能源并网,分布式计算;联系人,E-mail: wanglicheng@zjut.edu.cn。

(收稿日期:2021-12-07)

量耦合特性知识以及深度学习模型具有高效处理海量复杂数据信息的能力,提出一种基于知识-数据混合驱动的 IES 多元负荷预测方法。首先基于原始样本数据库中的历史负荷数据、气象因素数据、可再生能源出力数据等信息建立解析模型,并通过建立好的知识模型计算得到各能源的供应值。然后将知识模型计算所得到的数据和原始样本数据进行重构,并将重构后新的样本数据作为数据驱动模型的训练数据进行训练。最后,在 IES 负荷预测模型的训练过程中,利用基于随机森林算法的特征选择方法和 Dropout 技术分别对冗余特征和神经网络进行处理,从而进一步提升该预测模型的泛化能力。算例结果表明,通过本文所提方法训练完备的 IES 多元负荷预测模型相较于传统单一数据驱动模型不仅具有较好的可解释性,还有助于改善 IES 多元负荷预测模型的预测效果。

1 综合能源系统的用能机理分析

在对综合能源系统中不同类型负荷进行预测之前,还需要充分分析 IES 内部各能源子系统之间的能源交互与存储,深入探索各种负荷的构成机理,进而获得历史负荷数据和其他影响因素下未来负荷的变化趋势。本节将从综合能源系统的物理结构和 IES 中的能量耦合特性 2 个方面对 IES 的多元负荷特性进行介绍。

1.1 综合能源系统的物理结构

截至目前,关于能源系统预测的研究工作主要还是集中于单一能源系统,而针对单一能源系统的预测工作忽略了当前综合能源系统中不同能源子系统之间日益耦合的特点。典型综合能源系统的物理结构模型框架如图 1 所示。

从能源类型的角度来看,综合能源系统主要包含了电力、天然气和热力等能源子网络。在电力网络中,除了一些常规的电气设备(如电力传输线、不同级别的用电负荷),还包含分布式可再生能源的发电设备(如风电、光伏)和电储能装置等;在天然气网络中,天然气管道、气储能装置和气动阀门成为了该网络的核心;热力网络则主要由热力管道、

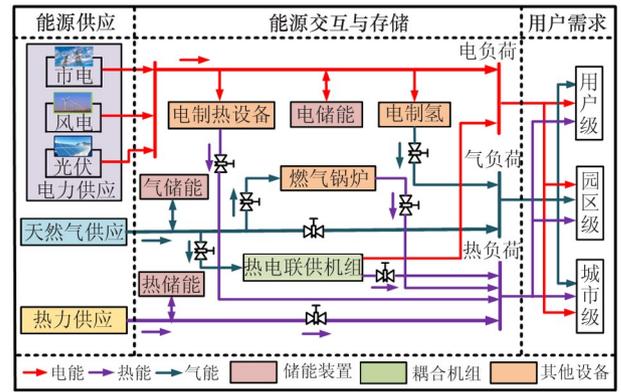


图 1 典型综合能源系统的物理结构模型

热力阀门、热负荷组成。从系统功能的角度来看,综合能源系统包括了能源供给、能源交互与存储以及负荷需求响应三大功能。例如,电力网络在消纳风电、光伏等可再生能源、提升电力资源优化配置的同时,还能通过电制热、电制氢等耦合设备将电能转换热能和气能,从而满足热负荷和气负荷的能源需求;天然气网络中的气源除了供给气负荷外,还可以通过热电联供机组产生电能和热能,用于供给电负荷和热负荷;热力网络中的热源不仅满足了热负荷需求,还能与热储能装置进行能量的交互,从而提高系统的灵活性。

综上所述,综合能源系统中各种负荷需求和其他多种异质能源紧密相关。因此,在对综合能源系统进行多元负荷预测时,还需考虑 IES 中存在的能量耦合特性、可再生能源出力的波动性以及不同能源供应等影响因素对 IES 多元负荷预测结果造成的影响。

1.2 IES 中的能量耦合特性

本文以 1.1 节中的典型综合能源系统为研究对象,对 IES 内部各能源子系统的能量耦合特性进行详细分析。在能源供给侧,IES 中各能源子系统通过从上级能源网络获得电力、天然气和热力的能源供应,并通过不同的能源转化设备实现电能、气能和热能的能源交互与存储,进而满足用户需求侧中不同类型的负荷需求。其中,能源供给侧的能量耦合关系可建模如下:

$$\begin{bmatrix} P^e \\ P^g \\ P^h \end{bmatrix} = \begin{bmatrix} \eta_T & \lambda_{\text{CHP}}^g \eta_{\text{CHP}}^e & 0 \\ \lambda_{\text{P2G}}^e \eta_{\text{P2G}} \eta_T & 1 & 0 \\ \lambda_{\text{EH}}^e \eta_{\text{EH}} \eta_T & \lambda_{\text{CHP}}^g \eta_{\text{CHP}}^h + \lambda_{\text{GB}}^g \eta_{\text{GB}} & 1 \end{bmatrix} \begin{bmatrix} E^e \\ E^g \\ E^h \end{bmatrix}$$

$$+ \begin{bmatrix} P^w + P^v + P_{st}^e \\ P_{st}^g \\ P_{st}^h \end{bmatrix} \quad (1)$$

其中, P^e 、 P^g 和 P^h 分别代表上级能源网络为用户需求侧提供的电功率、气功率和热功率; E^e 、 E^g 和 E^h 分别是上级能源网络获得的电能、气能和热能; P^w 、 P^v 为综合能源系统中风电和光伏的输出功率; P_{st}^e 、 P_{st}^g 和 P_{st}^h 分别为电、气、热网络中储能装置的充放电功率; η_T 、 η_{P2G} 、 η_{EH} 、 η_{GB} 、 η_{CHP}^e 、 η_{CHP}^h 分别代表电能的传输效率、电制氢效率、电制热设备效率、燃气锅炉效率、热电联供机组的产电和产热效率; λ_{CHP}^g 、 λ_{GB}^g 分别为供给热电联产机组和燃气锅炉的天然气分配系数; λ_{P2G}^e 、 λ_{EH}^e 分别为提供给电制氢和电制热设备的电能分配系数。综合考虑到电能和天然气分配系数的实际物理意义,有 $\lambda_{CHP}^g + \lambda_{GB}^g = 1$ 和 $\lambda_{P2G}^e + \lambda_{EH}^e = 1$ 。

在用户需求侧,从上级能源网络获得的电力、天然气、热力资源经由各种能源转化设备可满足不同规模级别的用户需求。为保障用户需求侧的正常供能,综合能源系统中的能量耦合关系式还需考虑各能源子网络的供需平衡,即: $P^e = L^e$; $P^g = L^g$; $P^h = L^h$ 。其中, L^e 、 L^g 和 L^h 分别代表 IES 的电负荷、气负荷和热负荷。

2 考虑 IES 能量耦合特性的知识-数据混合驱动模型

相较于传统单一数据驱动模型的 IES 负荷预测方法,本文所提方法综合考虑到综合能源系统物理结构模型中蕴藏的能量耦合特性以及深度学习具有高效处理海量高维复杂数据的能力,提出一种基于知识-数据混合驱动的综合能源系统多元负荷预测模型。其中,解析模型保留了综合能源系统中的能量耦合特性知识,在提高数据驱动模型可解释性的同时,还能充分挖掘数据驱动模型提取样本数据中重要信息的能力^[10-11],进而保证了本文所提方法的有效性。

2.1 基于能量耦合特性知识的解析模型

为进一步体现 IES 中的能量耦合特性,获得用

于本文所提方法中数据驱动模型训练的新数据样本,本节利用解析模型对从上级能源网络获得的各能源供应进行求解,求解过程如下。

先将式(1)转化为式(2)形式。

$$\begin{cases} P^e = \eta_T E^e + \lambda_{CHP}^g \eta_{CHP}^e E^g + P^w + P^v + P_{st}^e \\ P^g = \lambda_{P2G}^e \eta_{P2G} \eta_T E^e + E^g + P_{st}^g \\ P^h = \lambda_{EH}^e \eta_{EH} \eta_T E^e + \lambda_{CHP}^g \eta_{CHP}^h E^g + \lambda_{GB}^g \eta_{GB} E^g \\ \quad + E^h + P_{st}^h \end{cases} \quad (2)$$

再对式(2)进行化简可分别计算得到当前历史数据下电能、气能和热能的供应值,结果如式(3)~(5)所示。

$$E^e = \frac{L^g - P_{st}^g}{k_1 \eta_T} - \frac{L^g + k_1 (P^w + P^v + P_{st}^e - L^e) - P_{st}^g}{k_1 \eta_T (1 - k_1 k_2)} \quad (3)$$

$$E^g = \frac{L^g + k_1 (P^w + P^v + P_{st}^e - L^e) - P_{st}^g}{1 - k_1 k_2} \quad (4)$$

$$\begin{aligned} E^h = & L^h - \frac{k_3 \eta_T (L^g - P_{st}^g)}{k_1 \eta_T} \\ & + \frac{k_3 \eta_T [L^g + k_1 (P^w + P^v + P_{st}^e - L^e) - P_{st}^g]}{k_1 \eta_T (1 - k_1 k_2)} \\ & - \frac{(k_4 + k_5) [L^g + k_1 (P^w + P^v + P_{st}^e - L^e) - P_{st}^g]}{1 - k_1 k_2} \\ & - P_{st}^h \end{aligned} \quad (5)$$

其中, $k_1 = \lambda_{P2G}^e \eta_{P2G}$, $k_2 = \lambda_{CHP}^g \eta_{CHP}^e$, $k_3 = \lambda_{EH}^e \eta_{EH}$, $k_4 = \lambda_{CHP}^g \eta_{CHP}^h$, $k_5 = \lambda_{GB}^g \eta_{GB}$ 。

最后对解析模型计算所得到的电能、气能、热能和原始样本数据进行重构,重构后的新数据样本作为下一节数据驱动模型的训练样本。

2.2 基于数据驱动的深度神经网络模型

2.2.1 深度置信网络的算法原理

典型的深度置信网络 (deep belief network, DBN) 模型^[12-13]如图 2 所示,该模型由多个受限玻尔兹曼机 (restricted Boltzmann machine, RBM)^[14] 和一个反向传播 (back-propagation, BP) 网络^[15] 堆叠而成。其中, $\mathbf{X} = (x_1, x_2, \dots, x_n)$ 和 $\mathbf{Y} = (y_1, y_2, \dots, y_j)$ 分别为深度置信网络的输入样本和输出样本; \mathbf{W}_m 为第 $m-1$ 个和第 m 个隐含层上所有神经元连接的权重向量, $\mathbf{W}_m = (w_{m,1}, w_{m,2}, \dots, w_{m,k})$; \mathbf{B}_m 为

第 m 个隐含层上神经元的阈值向量, $\mathbf{B}_m = (b_{m,1}, b_{m,2}, \dots, b_{m,f})$ 。

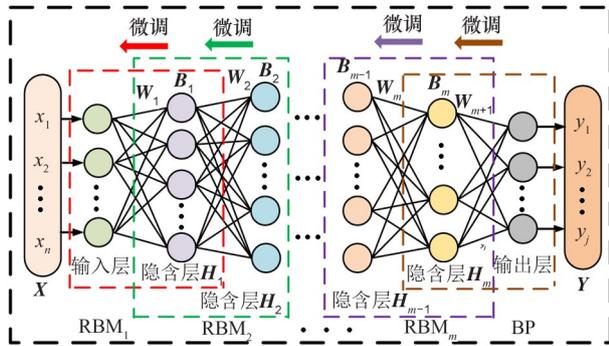


图2 典型的深度置信网络模型

RBM 是构成 DBN 的基本单元,每一个 RBM 中上一层神经元的输出可作为下一层神经元的输入,即图2中隐含层也可作为某个 RBM 的输入层(输入层亦可以被称为可视层)。以图2中的第1个受限玻尔兹曼机 RBM_1 为例,输入层 \mathbf{X} 和隐含层 \mathbf{H}_1 分别为 RBM_1 的可视层和隐含层,可视层神经元的输出作为隐含层 \mathbf{H}_1 的输入。其中, $\mathbf{H}_1 = (h_{1,1}, h_{1,2}, \dots, h_{1,e})$ 表示第1个隐含层的状态向量; $\mathbf{A}_1 = (a_{1,1}, a_{1,2}, \dots, a_{1,g})$ 为可视层神经元的阈值向量,并定义 $\lambda_1 = (\mathbf{W}_1, \mathbf{B}_1, \mathbf{A}_1)$ 为 RBM_1 的参数集合。

由于受限玻尔兹曼机是一种基于能量的概率生成模型,因此 RBM_1 的能量函数可被定义为^[16-17]

$$E(\mathbf{X}, \mathbf{H}_1 | \lambda_1) = - \sum_{c=1}^n (a_{1,c} x_c) - \sum_{l=1}^k (b_{1,l} h_{1,l}) - \sum_{c=1}^n \sum_{l=1}^k w_{1,c,l} x_c h_{1,l} \quad (6)$$

其中, $a_{1,c}$ 、 x_c 分别为 RBM_1 中可视层上第 c 个神经元的阈值和输入, $b_{1,l}$ 、 $h_{1,l}$ 分别为 RBM_1 中隐含层上第 l 个神经元的阈值和状态信息, $w_{1,c,l}$ 则为 RBM_1 中可视层上第 c 个神经元与隐含层上第 l 个神经元之间的权重。

当隐含层 $\mathbf{H}_1 = (h_{1,1}, h_{1,2}, \dots, h_{1,e})$ 的状态向量确定后,依据式(7)可计算得到可视层上第 c 个神经元被激活的概率。

$$P(x_c = 1 | \mathbf{H}_1) = s_f(a_{1,c} + \sum_{l=1}^k w_{1,c,l} x_c h_{1,l}) \quad (7)$$

其中, s_f 代表激活函数;0 和 1 分别表示神经元未激

活和激活的2种状态。

同理,一旦可视层 $\mathbf{X} = (x_1, x_2, \dots, x_n)$ 的状态向量被确定,依据式(8)可计算得到隐含层上第 l 个神经元的激活概率。

$$P(h_{1,l} = 1 | \mathbf{X}) = s_f(b_{1,l} + \sum_{c=1}^n w_{1,c,l} x_c) \quad (8)$$

其中, RBM_1 的参数集合 λ_1 可以通过最大对数似然函数的方法确定,计算公式如下:

$$\hat{\lambda}_1 = \arg \max_{\lambda_1} L(\lambda_1) = \arg \max_{\lambda_1} \sum_{t=1}^N \ln P(\mathbf{X}_t | \lambda_1) \quad (9)$$

其中, \mathbf{X}_t 表示第 t 个输入样本; N 为输入样本的总数。

在 DBN 的训练过程中,其他 RBM 的参数可依据上述 RBM_1 的参数计算公式求得,同时利用顶层的误差反向传播算法(back propagation, BP)回归层实现输出信息的拟合预测,并对 DBN 的网络参数进行微调。

2.2.2 改进后的深度置信网络

在综合能源系统中,由于 IES 的多元负荷与众多影响因素相关,因此相较于单一能源系统,其负荷预测模型的复杂程度势必更高,可能会导致模型在训练过程中出现“过拟合”的问题。为了避免上述问题的出现,本文采用 Dropout 技术^[18-19]对 DBN 神经网络进行了改进,改进后的 DBN 神经网络如图3所示。

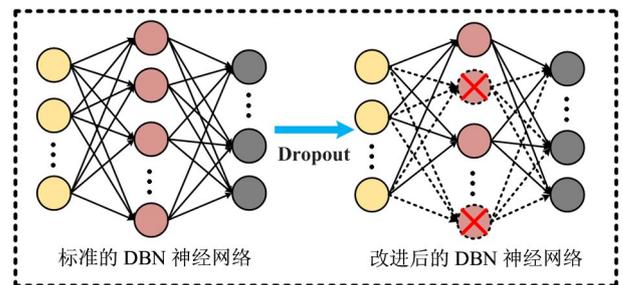


图3 改进后的 DBN 神经网络

其中,基于 Dropout 技术改进后的 DBN 网络参数训练流程如下。

(1) 随机删除 DBN 中隐含层上的部分神经元。即,隐含层上部分神经元的输入输出变为 0,其他神经元保持不变。

(2)小批量训练样本通过 Dropout 后的神经网络进行前向传播,获得 Dropout 处理前后的输出损失值。

(3)将步骤(2)中的损失值通过 Dropout 后的神经网络进行反向传播,并在剩余神经元的基础上按照随机梯度下降算法实现 DBN 的网络参数更新。

(4)恢复步骤(1)中被删除的神经元(此时,之前被删除神经元的参数保持不变,未被删除的神经元在步骤(1)和步骤(2)中已经更新),并换下一小批量训练样本执行步骤(1)至步骤(4)。

重复执行上述步骤,直至 DBN 模型训练结束。

3 IES 的多元负荷预测流程

本文所提的基于知识-数据混合驱动的综合能

源系统多元负荷预测流程如图 4 所示。该预测流程主要包含离线训练和在线评估两部分内容。在离线训练过程中,为了避免 IES 负荷预测模型出现过拟合、提高模型的泛化能力,本文先是利用随机森林算法^[20]对输入数据进行预处理,提取与预测结果相关性程度强的特征变量作为 IES 负荷预测模型的输入,以此来降低输入数据的复杂性;然后利用 Dropout 技术将深度置信网络中的部分神经元进行稀疏化处理,从而有效地提高了 IES 负荷预测模型的泛化能力,降低模型的过拟合风险;最后基于纵向对比方法对不同隐含层层数、同一隐含层上不同神经元个数下 IES 负荷预测模型的预测精度进行分析对比,从而获得 IES 负荷预测模型的最优网络架构,并以此进行模型的训练。

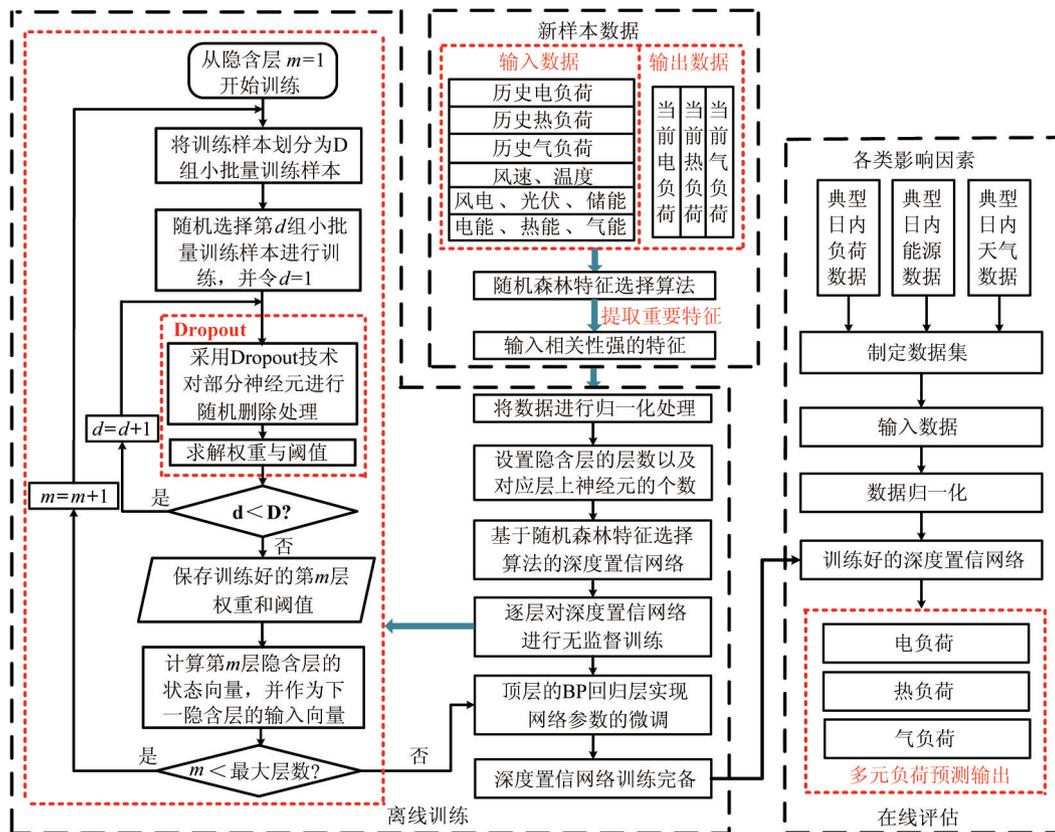


图 4 基于知识-数据混合驱动的综合能源系统多元负荷预测流程

3.1 输入变量和输出变量的选取

由于本文所提方法是一种知识-数据混合驱动的 IES 负荷预测模型,因此在该模型的训练过程中,输入样本数据不仅包含历史负荷、温度、风速、光伏

和风电等传统数据集,还包括 2.1 节计算得到的电能、气能、热能以及储能的充放电功率等数据信息;输出数据为当前时刻的负荷数据。为了避免某些特征对 IES 负荷预测模型的预测结果产生不好的影

响,本文在对 IES 负荷预测模型进行训练之前采用了随机森林算法对各种特征的相关性程度做了定量分析,以此筛选出与预测结果相关性程度强的输入特征。文章选取相关性程度大于 0.3 的特征作为输入变量,并选取当前的电、气、热负荷作为输出变量。其中,基于随机森林算法实现某种特征相关性程度的定量分析过程如下。

步骤 1 将袋外数据(out-of-bag, OOB)样本编号为 $r = 1, 2, 3, \dots, R$, 每个样本包含 p 个特征。

步骤 2 设置 $r = 1$, 基于训练样本建立决策树,并将袋外数据记为 V_b^{OOB} 。

步骤 3 利用决策树实现 V_b^{OOB} 的预测,并计算袋外数据的预测误差,记作 E_b^{OOB} 。

步骤 4 随机对 V_b^{OOB} 中的特征 X_s 加入噪声干扰,将受到干扰后的数据集记为 V_{bs}^{OOB} ,再次利用决策树完成 V_{bs}^{OOB} 的预测,并计算此时袋外数据的预测误差,记为 E_{bs}^{OOB} 。

步骤 5 对于 $r = 2, 3, \dots, R$, 循环执行步骤 2 ~ 4。

步骤 6 定量计算特征 X_s 的相关性程度 U_s^{OOB} :

$$U_s^{OOB} = \frac{1}{R} \sum_{i=1}^R (E_b^{OOB} - E_{bs}^{OOB}) \quad (10)$$

基于上述特征相关性程度的定量分析方法可以实现重要特征的筛选,降低了输入变量的维度,进一步提升了模型的泛化能力。

3.2 离线训练

步骤 1 将采集到的历史负荷、温度、风速、风电和光伏等原始样本数据、储能充放电功率数据以及 2.1 节计算所得的电能、热能、气能数据信息进行重构,获得新的样本数据库。

步骤 2 将新的样本数据分为训练集和测试集,同时为避免不同变量之间的单位差异给 IES 负荷预测结果带来的影响,本文对新的样本数据进行归一化处理,数据归一化的计算公式如下:

$$v_{base} = \frac{v_{real} - v_{min}}{v_{max} - v_{min}} \quad (11)$$

其中, v_{base} 、 v_{real} 分别为样本数据归一化后的基值和样本数据的实际值, v_{max} 、 v_{min} 分别为样本数据的最大、最小值。

步骤 3 逐层进行无监督训练。在对改进后的 DBN 神经网络进行训练时,首先将新的样本数据集划分为若干组的小批量样本数据,然后随机选择一组小批量样本数据用于 Dropout 技术改进后的 DBN 神经网络训练,从而获得 DBN 神经网络的权重和阈值参数。

步骤 4 通过 DBN 模型顶层的 BP 网络实现输出信息的预测拟合,并对模型的网络参数进行反向微调。

本文定义平均绝对百分比误差(mean absolute percentage error, MAPE)为 IES 负荷预测模型的预测精度指标,MAPE 的计算公式为

$$MAPE = \frac{1}{N} \sum_{h=1}^m \left| \frac{y_{real}(h) - y_{pre}(h)}{y_{real}(h)} \right| \times 100\% \quad (12)$$

式中, $y_{real}(h)$ 、 $y_{pre}(h)$ 分别为 h 时刻下各类负荷的实际值和预测值, N 为样本数据的总数。

3.3 在线预测

步骤 1 对于综合能源系统的规划调度人员,可根据量测采集到的典型日内负荷数据、能源数据和天气数据等信息,形成综合能源系统负荷预测模型的输入数据。

步骤 2 依据式(11)对输入数据进行相应的归一化处理。

步骤 3 将归一化处理后的数据输入到已经离线训练完备的深度置信网络模型中,获得对应的输出数据。

步骤 4 将输出数据进行反归一化处理,得到 IES 多元负荷的预测值,并基于式(12)计算得到各负荷的预测精度指标。

4 算例分析

为了验证本文所提方法的有效性,本文利用 Matlab 平台实现了基于知识-数据混合驱动的综合能源系统多元负荷预测。其中,实验仿真所用数据均采集于南昌某工业园区内的综合能源系统,该综合能源系统不仅包含了用于满足不同能源需求的热电联产机组、燃气锅炉、电制热设备等能源转换设

备,还配置了相应的电、热、气储能装置来提高系统运行灵活性和经济性。其中,该系统中能源的分配系数以及设备的能源转化效率可见表 1 和 2。

表 1 综合能源系统中不同能源的分配系数

参数	$\lambda_{\text{CHP}}^{\text{e}}$	$\lambda_{\text{GB}}^{\text{e}}$	$\lambda_{\text{P2G}}^{\text{e}}$	$\lambda_{\text{EH}}^{\text{e}}$
分配系数	0.6	0.4	0.5	0.5

表 2 综合能源系统中不同设备的能源转化效率

参数	η_{T}	η_{P2G}	η_{EH}	η_{GB}	$\eta_{\text{CHP}}^{\text{c}}$	$\eta_{\text{CHP}}^{\text{h}}$
转换效率	0.9	0.5	0.7	0.85	0.35	0.42

4.1 预测实例

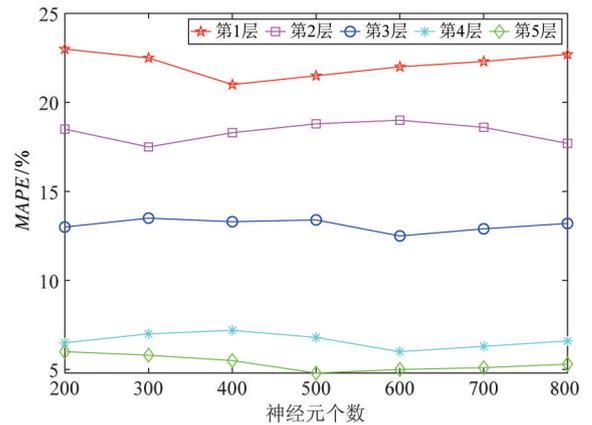
本文算例选取上述综合能源系统 2018 年一整年的小时级数据作为原始样本数据,将 2018 年 1 月 1 日至 2018 年 3 月 30 日的春季数据作为传统单一数据驱动模型的原始训练样本数据和知识-数据混合驱动模型的新训练样本数据的基础;并选取 2018 年 3 月 31 日这一典型日的对应信息作为测试样本数据,分别对 2 种模型进行预测分析。

为了进一步分析传统单一数据驱动模型和本文所提方法在输入变量选择上的差异,本文基于随机森林算法的特征选择方法分别对 2 种模型的输入特征进行了相关性程度的定量分析。上述 2 种模型的特征相关性重要程度分析结果分别如附录 A 和附录 B 所示。

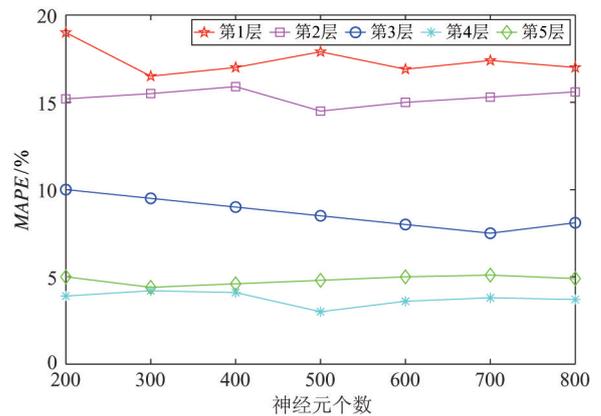
4.2 数据驱动模型的最佳网络架构选取

由于本文所提方法中的数据驱动模型是一种深度置信网络算法,而该算法的预测效果很大程度上依赖于网络架构的选取。因此,本文在实验仿真过程中采用了一种纵向对比方法对深度置信网络模型中隐含层的层数以及各隐含层上的神经元个数进行分析对比,以此来选取数据驱动模型的最佳网络架构。其中,传统单一数据驱动模型和本文所提方法中的数据驱动模型在不同网络架构下的 MAPE 值分别如图 5(a)和(b)所示。

由图 5(a)和(b)可知,在本算例中,传统单一数据驱动模型和本文所提方法中的数据驱动模型的最优隐含层层数分别为 5 层和 4 层、各隐含层上神



(a) 传统单一数据驱动模型



(b) 本文所提方法中的数据驱动模型

图 5 在不同网络架构下 2 种模型的预测效果

经元的最佳个数分别为 400、300、600、600、500 和 300、500、700、500、300。此外,由图 5 可知,本文所提方法中的数据驱动模型在相同的网络架构下的 MAPE 值都要小于传统单一数据驱动模型。

4.3 结果分析与对比

本文依据 4.2 节中实验仿真所获得的最佳网络架构分别对传统单一数据驱动模型和本文所提方法中数据驱动模型进行离线学习,从而得到训练完备的 IES 负荷预测模型。为了进一步分析网络架构对综合能源系统中不同负荷预测效果产生的影响,表 3 和表 4 给出了不同隐含层层数和神经元个数下传统单一数据驱动模型和本文所提方法的 MAPE_1 、 MAPE_2 和 MAPE_3 值,同时将 MAPE_1 、 MAPE_2 和 MAPE_3 分别记为电负荷、热负荷和气负荷的预测精度指标。其中,表 3、表 4 中 2 种模型的隐含层总数都设置为 5 层,对应隐含层上的神经元个数分别为 400、300、600、600、500 和 300、500、700、500、300。

表 3 传统单一数据驱动模型的各负荷预测结果

隐含层数	训练时间/s	评估时间/s	MAPE ₁ /%	MAPE ₂ /%	MAPE ₃ /%
1 层	12.28	0.96	24.81	23.50	24.12
2 层	16.30	1.37	19.30	17.46	18.07
3 层	21.45	1.82	13.60	12.49	13.12
4 层	27.61	2.40	7.06	6.14	6.58
5 层	31.23	2.64	5.93	5.01	5.22

表 4 本文所提方法的各负荷预测结果

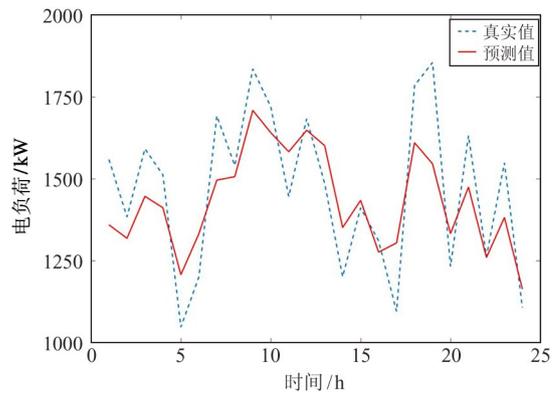
隐含层数	训练时间/s	评估时间/s	MAPE ₁ /%	MAPE ₂ /%	MAPE ₃ /%
1 层	14.48	1.01	17.06	16.27	16.53
2 层	17.26	1.48	15.95	15.09	15.36
3 层	24.27	1.97	9.04	8.55	8.84
4 层	29.54	2.56	3.99	3.27	3.66
5 层	34.36	2.78	5.15	4.43	4.82

从表 3 和表 4 中可以看出,无论是传统单一数据驱动模型还是本文所提方法,多隐含层的网络架构下的 IES 负荷预测结果都要明显优于单隐含层,且在相同数量隐含层的情况下,表 4 中 IES 的各负荷预测精度都要高于表 3。然而,神经网络中的隐含层数越多也不意味着 IES 的各负荷预测效果越好。如表 4 所示,本文所提方法中数据驱动模型的隐含层层数为 4 层时的 IES 多元负荷预测精度要高于其隐含层层数为 5 层的情况,且此时的电、热、气 3 种负荷的 MAPE₁、MAPE₂ 和 MAPE₃ 值最低,仅为 3.99%、3.27% 和 3.66%,此时的 IES 多元负荷预测模型已经达到非常令人满意的精度了。除此之外,相较于表 3 中的传统单一数据驱动模型,表 4 中本文所提方法的离线训练和在线预测时间几乎保持不变,且能满足 IES 负荷的在线预测要求。由此可见,相较于传统单一数据驱动模型,本文所提方法确实有助于提高综合能源系统中多元负荷的预测精度,且由于本文所提方法充分考虑了综合能源系统中的能量耦合特性知识,极大地保留了综合能源系统中非线性的强因果关系,因此文中所提方法具有更好的可解释性。

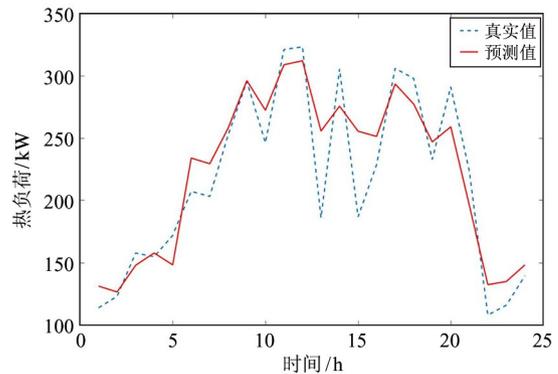
为进一步验证本文所提方法相较于传统单一数据驱动模型所体现出来的优越性,本算例将两者的网络架构都设置为最优,并利用测试样本对本文所提方法和传统单一数据驱动模型的泛化能力进行对

比分析。其中,传统单一数据驱动模型和本文所提方法的多元负荷预测结果如图 6、7 所示。

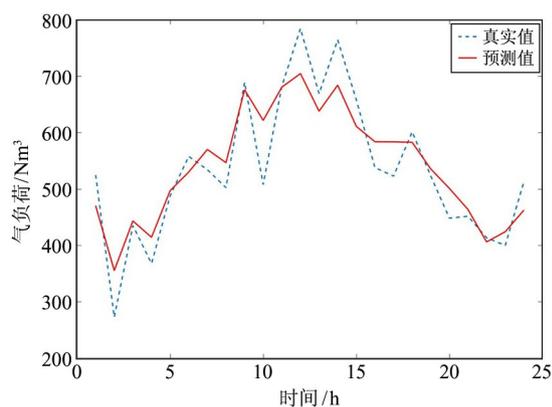
由图 6 和图 7 可知,传统单一数据驱动模型的电、热、气负荷预测结果曲线虽然总体上能够跟踪



(a) 传统单一数据驱动模型的电负荷预测结果

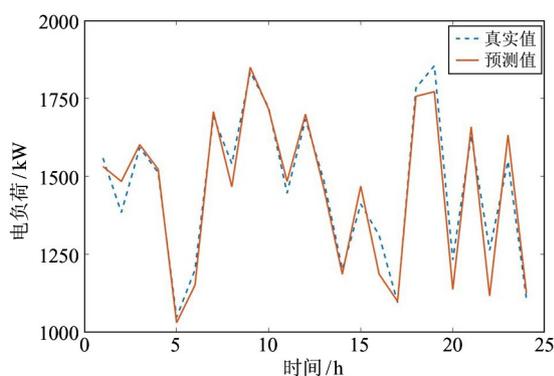


(b) 传统单一数据驱动模型的热负荷预测结果

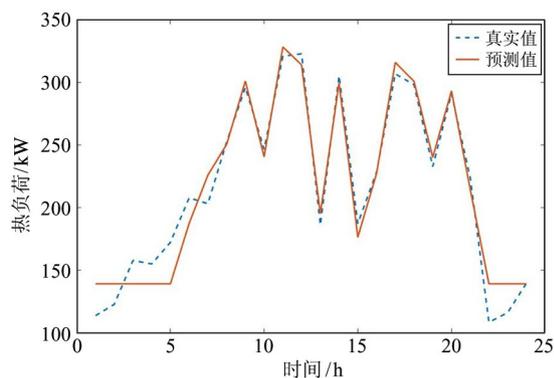


(c) 传统单一数据驱动模型的气负荷预测结果

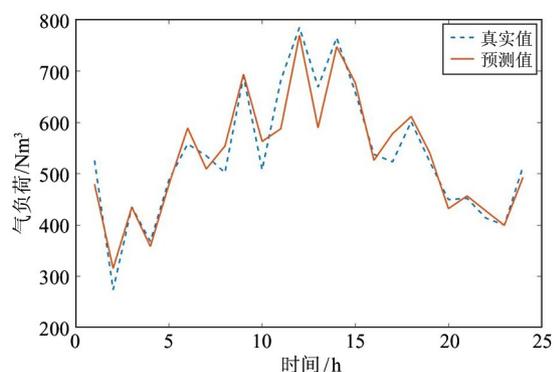
图6 传统单一数据驱动模型的多元负荷预测结果



(a) 本文所提方法的电负荷预测结果



(b) 本文所提方法的热负荷预测结果



(c) 本文所提方法的气负荷预测结果

图7 本文所提方法的多元负荷预测结果

实际值的变化趋势,但部分时段仍然存在较大的预测误差。而考虑了综合能源系统中能量耦合特性知识的本文所提方法下的电、热、气负荷预测结果曲线与实际值的变化趋势非常接近,且具有较高的预测精度。

综上所述,相比于传统单一数据驱动模型,本文所提方法不仅能有效地提高 IES 多元负荷的预测精度,还保留了综合能源系统中的能量耦合特性知识,保证了 IES 负荷预测结果的可靠性。

5 结论

本文以综合能源系统为研究对象,提出一种基于知识-数据混合驱动的综合能源系统多元负荷预测方法,其主要贡献可总结如下。

(1) 本文将综合能源系统中的能量耦合特性知识模型与数据驱动模型进行了有效的结合,通过对基于能量耦合特性知识模型计算得到的能源供应量和原始样本数据进行重构,可获得用于训练数据驱动模型的新样本数据。算例结果表明,通过本文所提方法训练完备的 IES 负荷预测模型相较于传统单一数据驱动模型不仅具有较好的可解释性,还有助于改善该负荷预测模型的预测效果。

(2) 本文考虑到综合能源系统中多元负荷与其他影响因素之间的相关性,利用随机森林算法对不同特征的相关性程度做了定量分析,从而提高了模型的计算效率。此外,为避免模型出现过拟合,文章利用 Dropout 技术对 DBN 神经网络进行了处理,进一步提高了模型的泛化能力。

(3) 在数据驱动模型的训练过程中,本文利用纵向对比方法可以较快地选择出具有最佳网络架构的数据驱动模型,避免了因网络架构随机性给模型预测结果带来的影响。

(4) 在之前的 IES 负荷预测研究中,通常是利用历史负荷、气象信息等数据作为训练样本。而本文在此基础上还考虑了风电、光伏等可再生能源出力、储能以及不同能源供应对未来负荷预测产生的影响,进而更接近综合能源系统的实际运行工况。

参考文献

[1] 朱继忠,董瀚江,李盛林,等. 数据驱动的综合能源系

- 统负荷预测综述[J]. 中国电机工程学报, 2021, 41(23):7905-7923.
- [2] 罗凤章,张旭,杨欣,等. 基于深度学习的综合能源配电系统负荷分析预测[J]. 高电压技术,2021,47(1):23-32.
- [3] 杨挺,赵黎媛,王成山. 人工智能在电力系统及综合能源系统中的应用综述[J]. 电力系统自动化,2019,43(1):2-14.
- [4] 程乐峰,余涛,张孝顺,等. 机器学习在能源与电力系统领域的应用和展望[J]. 电力系统自动化,2019,43(1):15-31.
- [5] 朱刘柱,王绪利,马静,等. 基于小波包分解与循环神经网络的综合能源系统短期负荷预测[J]. 电力建设,2020,41(12):131-138.
- [6] 翟晶晶,吴晓蓓,王力立. 基于径向基函数神经网络的综合能源系统多元负荷短期预测[J]. 电力需求侧管理,2019,21(4):23-27, 34.
- [7] 田浩含,张智晟,于道林. 基于改进 LSTM 的区域综合能源系统多元负荷短期预测研究[J]. 电力系统及其自动化学报,2021,33(9):130-137.
- [8] 田浩含,撤奥洋,于立涛,等. 基于 GRA-LSTM 神经网络的区域综合能源系统多元负荷短期预测模型[J]. 广东电力,2020,33(5):44-51.
- [9] 李守茂,戚嘉兴,白星振,等. 基于 IPSO-WNN 的综合能源系统短期负荷预测[J]. 电测与仪表,2020,57(9):103-109.
- [10] XU Y, DAI Y, ZHANG R. Extreme learning machine-based predictor for real-time frequency stability assessment of electric power systems [J]. Neural Computing and Applications, 2013, 22(3):1-4.
- [11] HSU C T, KANG M S, CHEN C S. Design of adaptive load shedding by artificial neural networks[J]. IET Generation Transmission and Distribution, 2005, 152(3):415-421.
- [12] 仇怡超,闻达,王晓茹,等. 基于深度置信网络的电力系统扰动后频率曲线预测[J]. 中国电机工程学报, 2019,39(17):5095-5104, 290.
- [13] 朱乔木,党杰,陈金富,等. 基于深度置信网络的电力系统暂态稳定评估方法[J]. 中国电机工程学报, 2018,38(3):735-743.
- [14] 郑智聪,王红,齐林海. 基于深度学习模型融合的电压暂降源识别方法[J]. 中国电机工程学报,2019,39(1):97-104, 324.
- [15] 杨茂,王金鑫. 考虑可再生能源出力不确定的孤岛型微电网优化调度[J]. 中国电机工程学报,2021,41(3):973-985.
- [16] 史佳琪,谭涛,郭经,等. 基于深度结构多任务学习的园区型综合能源系统多元负荷预测[J]. 电网技术, 2018,42(3):698-707.
- [17] 胡实,唐昊,吕凯,等. 考虑广义需求侧资源的深度置信网络短期负荷预测方法[J]. 控制理论与应用, 2023,40(3):493-501.
- [18] 赵荣臻,文云峰,叶希,等. 基于改进堆栈降噪自动编码器的预想事故频率指标评估方法研究[J]. 中国电机工程学报,2019,39(14):4081-4093.
- [19] 范海东,王立峰,孟瑜炜,等. 基于自编码网络与 dropout 机制的发电设备故障预警研究[J]. 电力信息与通信技术,2021,19(10):112-118.
- [20] 赵腾,王林童,张焰,等. 采用互信息与随机森林算法的用户用电关联因素辨识及用电量预测方法[J]. 中国电机工程学报,2016,36(3):604-614.

Multivariate load forecasting method of integrated energy system based on hybrid knowledge-data driven

WANG Licheng^{* **}, WANG Zifei^{*}, DENG Baohua^{*}, LING Feng^{*}, ZHANG Youbing^{*}
 (* College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)
 (** Institute of Artificial Intelligence, Zhijiang Laboratory, Hangzhou 311121)

Abstract

The current integrated energy system (IES) load forecasting methods are almost all single data-driven methods, ignoring the energy coupling relationship in IES. In addition, in the existing research, the training data of the data-driven method is mainly concentrated on historical load, weather and other influencing factors, and less consideration is given to the volatility of renewable energy output and the impact of various energy supplies on the IES load forecast results. In response to the above problems, this paper proposes a hybrid knowledge-data driven comprehensive energy system multiple load forecasting method. This method first describes the knowledge of energy coupling characteristics in IES through an analytical model, and uses the knowledge model to reconstruct the original sample data. Then use the reconstructed new sample data as the training sample of the data-driven model, and use the feature selection method based on the random forest algorithm and the Dropout technology to improve the generalization ability of the model. Finally, the effectiveness of the method is verified by the actual data of an IES industrial park. The simulation results show that this method has better prediction effect and higher reliability than the traditional single data-driven model.

Key words: integrated energy system (IES), hybrid knowledge-data driven, energy coupling characteristics, multivariate load forecasting, random forest algorithm, Dropout technology