

基于GAN的负荷数据修复及其在EV短期负荷预测中的应用

赵厚翔¹, 沈晓东¹, 吕林¹, 兰鹏¹, 刘俊勇¹, 刘代勇²

(1. 四川大学电气工程学院, 四川省成都市 610065; 2. 国网四川省电力公司德阳供电公司, 四川省德阳市 618000)

摘要: 随着电动汽车(EV)充电设施规模的不断扩大, EV充电数据可以更方便地获得。某些非人为因素会导致数据集中存在数据缺失和数据异常的问题, 阻碍了EV负荷预测精度的提高。因此, 文中在生成对抗网络(GAN)中采用用于插补的门控循环单元神经网络(GRUI)细胞来处理不完整负荷数据集中前后观测值间的不规则时滞关系, 提出了基于GRUI-GAN的数据插补模型来实现EV负荷数据修复。然后, 以带有Mogriker门控机制的长短期记忆网络进行EV负荷预测。最后, 实验结果表明了所提方法可以生成精度较高的新数据对缺失值进行插补, 并且经所提方法修复之后的数据有效提高了EV负荷预测精度。

关键词: 电动汽车(EV); 短期负荷预测; 数据插补; 生成对抗网络(GAN); 门控循环单元神经网络; 长短期记忆网络(LSTM网络)

0 引言

随着电动汽车(EV)的广泛普及, 电池中储存的电能为电网实施需求响应以及极端情况下的电网支撑提供了可能^[1]。然而EV的高普及率也会给电网带来众多不利影响^[2]。EV充电站作为EV与电网的连接点, 充电站的负荷预测精确度关乎电网的运行效率和安全。同时, EV作为一种电网资源, 利用智能充电提高EV利用效率也离不开精确的负荷预测。因此, 更加精确的EV负荷预测结果, 对于支持EV充电发展^[3]、促进EV高效使用^[4]以及保障电网安全至关重要。

随着智能充电设施的大规模部署, 可获得的实际EV历史充电数据为传统电力负荷预测中使用时间序列分析^[5-7]、机器学习^[6-9]、深度学习^[10]等算法进行预测提供了数据支撑。但这种基于数据驱动的预测算法的性能表现也一定程度依赖数据集质量, 高质量的数据集往往能提高数据驱动算法的预测精度^[11-12]。

由于终端数据采集器或者网络通信故障等原因, 常常使得历史负荷数据集中存在数据缺失和数据异常的问题^[13]。而不同的数据修复方法对预测精度的提高也有不同^[14], 现有文献中对于缺失值和

异常值的处理方法主要有3种: ①直接删除法^[15-16], 这种方法往往会忽略数据中的重要信息, 并且缺失率越高, 对于后续数据分析的影响越大; ②简单插补法, 通常采用均值插补、中位数插补以及其他数学插补方法^[17-18], 这种方法只关注数值之间的联系, 以统计学方法实现插补; ③基于机器学习的插补方法^[19-21], 这种方法一般包含基于最大似然期望最大化的插值、基于 k 近邻(k -NN)的插值和基于矩阵分解的插值等。然而, 这些方法几乎只关注数据本身的统计学特征, 忽视了时间序列数据中的时序特征, 插补精度较低。

生成对抗网络(generative adversarial network, GAN)是近年来新提出的一种生成式的学习算法。由于GAN在解决数据生成任务时的无模型重构优势^[22], 已有研究将其应用于电力系统数据生成场景。文献[23]将GAN应用于电力系统量测缺失数据重构, 并且取得了较好的精度。文献[24]以新的约束方法提高对抗损失函数的约束能力, 提升了GAN的重构精度。

因此, 为了在低质量的EV负荷数据集上获取更加精确的预测结果, 本文以基于用于插补的门控循环单元神经网络(gated recurrent unit neural network cells for imputation, GRUI)和GAN(下文简称GRUI-GAN)的数据插补模型对负荷数据中的空缺值(待插补的缺失值和异常值)进行精确插补, 提高数据质量。由于长短期记忆(long short-term memory, LSTM)网络在多种短期负荷预测场景下

收稿日期: 2021-02-01; 修回日期: 2021-04-19。

上网日期: 2021-07-07。

国家自然科学基金重点项目(U2066209); 中国南方电网有限责任公司重点科技项目(060400KK52190017)。

都取得了良好的预测效果^[10,18,25-26],本文以最近提出的 Mogrifier 门控机制^[27]进一步提升 LSTM 网络性能,以在经处理过的数据集上获取更加精确的 EV 短期负荷预测结果。

1 初始性数据分析

本文 EV 负荷数据集从某 35 kV 配电系统供电区域内 2019 年 1 月 1 日至 2019 年 8 月 31 日共 158 个充电桩的充电交易数据集中提取,充电交易数据形式及提取方法见附录 A 表 A1。由于该区域内 EV 渗透率较低,导致该区域内充电桩记录的数据并不连续,为便于算法验证,将数据进行聚合和重采样(聚合和重采样方法见图 A1),得到该区域以 30 min 为间隔的连续 EV 负荷数据集。通过初步整理可以发现,原始数据集中存在连续片段的数据缺失问题,这对于应用数据驱动算法来进行精确的负荷预测很不利。本文所使用的数据集数据缺失情况见表 A2。

同时,数据集中的异常值也会影响数据驱动算法的性能表现。为了快速识别数据中的异常值,本文选取了被广泛使用的箱形图异常值检测方法。该方法是基于统计的检测方法,通过确定数据中的上下四分位数以及四分位距来识别异常值,具体分析方法见文献[10]。根据文献[10]中箱形图分析过程,本文的箱形图分析结果以及异常值分布结果见附录 A 图 A2。

根据统计结果可知,原始数据集中,数据缺失点和数据异常点共 583 个,接近数据集的 5%,并且数据集中涉及许多连续片段的缺失。因此,可以认为原始数据集的质量较差。为了提高数据驱动算法在这种数据集上的预测表现,有必要研究一种方法来对空缺值进行精确插补以提高数据集质量。

2 基于 GRUI-GAN 的负荷数据修复方法

2.1 GAN

负荷数据中的数据插补问题实际就是一个数据生成问题,是通过某种方式生成数据来实现对空缺值的插补。而 GAN 是由 Goodfellow 等人^[28]受博弈论中二人零和博弈思想启发而开发的一种新的生成式学习算法。其基本结构由一个生成器和一个鉴别器构成,分别用于生成伪造数据和实现对伪造数据与真实数据的区分^[29]。通过生成器和鉴别器相互对抗训练,最终使得生成器能够生成与真实数据非常接近的新数据^[24]。但是,由于 GAN 存在模式崩溃的问题,使得传统 GAN 很难训练。而以

Wasserstein 距离来优化的 GAN 模型(下文简称 WGAN)^[23]可以有效摆脱 GAN 的模式崩溃问题并提高其学习稳定性。因此,在本文中所采用的 GAN 为 WGAN。

2.2 基于 GRUI-GAN 的数据插补模型

对于给定的一个包含空缺值的 d 维时间序列 $x=[x_0, x_1, \dots, x_{n-1}] \in \mathbf{R}^{d \times n}$ 以及对应的时间节点 $t=[t_0, t_1, \dots, t_{n-1}]$,完成空缺值的插补。首先,需得到空缺值在时间序列中的具体位置,为此定义空缺值位置矩阵 $S \in \mathbf{R}^{d \times n}$,以 0-1 变量来确定时间节点 t_i 是否存在观测值 $x_{i,j}$ 数据空缺。若有数据空缺,则对应位置矩阵的元素 $s_{i,j}=0$,若没有空缺,则 $s_{i,j}=1$ 。为了更加直观地理解,给出一个含空缺值的 3 维时间序列(每行代表每个特征量在 t_i 为 0、1、2、3 时刻的观测值)如式(1)所示。

$$\begin{cases} x = \begin{bmatrix} 1 & 5 & N_{\text{one}} & 8 \\ N_{\text{one}} & 7 & 6 & N_{\text{one}} \\ 4 & N_{\text{one}} & N_{\text{one}} & 2 \end{bmatrix} \\ S = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{cases} \quad (1)$$

其中, N_{one} 为数据空缺值。从上面这个例子中可以直观地看到,由于 N_{one} 的存在,使得 2 个相邻有效观测值之间的时滞存在变化,而这种时滞的变化也显示出了空缺值之间的分布特征。同时,不同的时滞对后续观测值的影响也会有所不同,如果观测值缺失了一段时间,其影响也应随时间衰减。为了记录 2 个相邻有效观测值之间的时滞,引入时滞矩阵 $\delta \in \mathbf{R}^{d \times n}$ 来描述时滞变化,其元素的表达式为:

$$\delta_{i,j} = \begin{cases} t_i - t_{i-1} & s_{i-1,j} = 1 \text{ 且 } i > 0 \\ \delta_{i-1,j} + t_i - t_{i-1} & s_{i-1,j} = 0 \text{ 且 } i > 0 \\ 0 & i = 0 \end{cases} \quad (2)$$

前述含空缺值的 3 维时间序列对应的时滞矩阵为:

$$\delta = \begin{bmatrix} 0 & 1 & 2 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 2 & 1 \end{bmatrix} \quad (3)$$

由于目前大多数的 GAN 框架主要应用于图像处理,对于时间序列的插补并不具有适应性。循环神经网络(RNN)是一种非常适用于处理时间序列的网络结构,但是传统的 RNN 在处理较长的时间序列时存在梯度爆炸和梯度消失问题。近年来,已经有很多的变体,比如 LSTM 网络,门控循环单元(GRU)等,可以有效克服这一问题。为了简化

GAN的构建,本文采用具有更高计算效率和可扩展性的GRU来构建生成器和鉴别器的基本网络结构。考虑到数据的不完整以及2次连续有效观测值之间的时滞变化很大,使得传统的GRU结构并不适用于数据插补的场景。因此,为了更加高效地处理这种不规则的时滞并从时滞中学习隐式信息,本文提出了基于GRU的GRUI,其基本结构见图1。

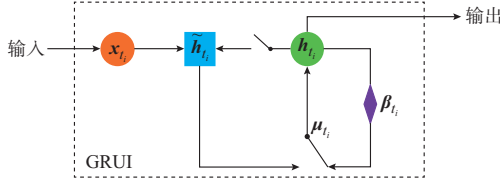


图1 GRUI结构图
Fig. 1 GRUI structure diagram

为了反映对观测值的影响,引入了时间衰减向量 β ,并且 β 的大小应与时滞矩阵 δ 相关, δ 越大, β 越小, β 中的元素 β_{t_i} 的表达式如下。

$$\beta_{t_i} = \frac{1}{e^{\max(0, W_\beta \delta_{t_i} + b_\beta)}} \quad (4)$$

其中, W_β 和 b_β 为需要学习的参数; $\delta_{t_i} = (\delta_{t_i,j})_{1 \times n}$ 。为了使 $\beta \in (0, 1]$,使用负指数函数来进行限制。此外,为了获取 δ 矩阵的相互作用, W_β 为全权重矩阵。在获得衰减向量之后,将逐元素乘以衰减向量 β 来更新GRU的隐藏状态 h_{t-1} 。因为在数据预处理时常采用归一化的方法,所以选择将 h_{t-1} 与衰减向量相乘来更新隐藏状态,使得更新之后的隐藏状态的值不至于过小。

GRUI的更新公式如下。

$$h'_{t-1} = \beta_{t_i} \odot h_{t-1} \quad (5)$$

$$\mu_t = \sigma(W_\mu[h'_{t-1}, x_t] + b_\mu) \quad (6)$$

$$r_t = \sigma(W_r[h'_{t-1}, x_t] + b_r) \quad (7)$$

$$\tilde{h}_t = \tanh(W_h[r_t \odot h'_{t-1}, x_t] + b_h) \quad (8)$$

$$h_t = (1 - u_t) \odot h'_{t-1} + u_t \odot \tilde{h}_t \quad (9)$$

式中: μ_t 为更新门; r_t 为重置门; $\tilde{h}_t$ 为隐藏状态候选集; x_t 为细胞中的当前输入; $\sigma(\cdot)$ 表示sigmoid激活函数; W_μ 、 W_r 、 W_h 、 b_μ 、 b_r 和 b_h 为训练参数; $\odot$ 表示逐元素相乘运算。

由于自编码器旨在重构目标样本,因此,本文在生成器中采用了自编码器和GRUI结构。为了强制自编码器学习到更加有用的信息,本文借鉴了降噪自编码器的思想。降噪自编码器的主要思想是删除一些原始样本(破坏原始样本)并重建原始样本。但

是,考虑到数据体量较少,如果进行删除样本,会使数据量进一步减少,不利于自编码器学习到更多的信息。因此,本文通过向原始样本中引入从正态分布 $N(0, 0.01)$ 中采样的随机噪声 η 来达到对原始样本破坏的效果。生成器的函数 $G(\cdot)$ 如下。

$$G(x + \eta) = x' \quad (10)$$

式中: x 为真实数据; x' 为伪造数据。

为了让生成器生成最接近 x 的完整时间序列 x' ,本文在生成器的损失函数中加入了平方误差损失函数,生成器的损失函数 $L_2(x)$ 如下。

$$L_2(x) = \|x \odot S - G(x + \eta) \odot S\|_2 \quad (11)$$

$$L_G = \lambda L_2(x) - D(x') \quad (12)$$

式中: λ 为控制鉴别器损失和方差损失权重的超参数; $D(\cdot)$ 为鉴别器的函数。

鉴别器的网络结构实际上是一个解码器,它是由GRUI层和全连接层组成。鉴别器的目的是区分伪造数据 x' 和真实数据 x ,其输出结果为生成的 x' 为真实数据的概率。因此,为了找到一组参数使得在输入真实数据 x 时输出较高的概率,而在输入伪造数据 x' 时输出较低的概率,将鉴别器的损失函数 L_D 定义为:

$$L_D = -D(x) + D(x') \quad (13)$$

生成器和鉴别器按式(14)进行min-max博弈对抗训练,最终达到平衡。

$$\min_G \max_D V(D, G) = E_{x \sim P_x(x)}(D(x)) - E_{x' \sim P_{x'}(x')}(D(x')) \quad (14)$$

式中: $E(\cdot)$ 为数学期望; $P_x(x)$ 和 $P_{x'}(x')$ 分别为真实数据 x 和伪造数据 x' 的概率分布。

对抗训练权重更新公式见文献[30],本文所提数据质量提升方法如下。

1)首先,采用-1值来代替原始数据中的空缺值,得到空缺值被标记的序列 x ,并根据-1值的分布情况构建相应的位置矩阵 S 和时滞矩阵 δ 。

2)在数据加载阶段,经zero-mean标准化之后将数据构建成[批样本数,序列长度,序列特征数]形式的3维数组输入GRUI-GAN模型中。

3)在生成器中,被标记并添加了噪声 η 的序列 x 及其时滞矩阵 δ 被输入到GRUI细胞中,在经由GRUI层处理之后会生成与输入序列对应的隐藏状态,该隐藏状态会进一步馈入全连接层中以生成低维向量 z 。然后, z 通过另一个全连接层馈入下一个GRUI层进行处理,该GRUI层的所有输出将被组合成新的完整序列 x' 。

4)而在鉴别器中,不完整序列 x 或者完整序列 x' 及其相应的时滞矩阵 δ 同样借助GRUI得到有效处理。鉴别器中GRUI层的最后一个隐藏状态被馈入全连接层中,全连接层的输出为被鉴别为真实数据的可能性。

5)经过生成器和鉴别器多轮的对抗训练,生成器将可以生成最接近 x 的新序列 x' ,然后使用 x' 去插补 x 中的空缺值。本文所提的插补框架如图2所示,数据插补公式如下。

$$x_{im} = x \odot S + (1 - S) \odot x' \quad (15)$$

式中: x_{im} 为插补完成之后的 x 。

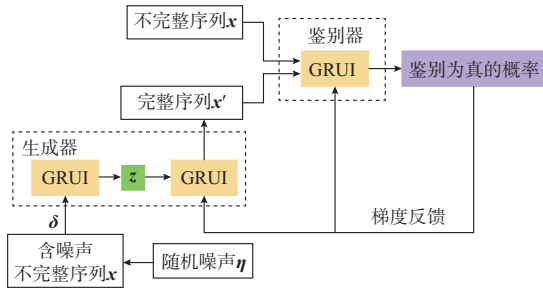


图2 基于GRUI-GAN的负荷数据插补框架
Fig. 2 Load data imputation framework based on GRUI-GAN

同时,为了提高生成器拟合原始序列的能力,在本文的GAN中使用了在1次迭代中生成器更新 N 次,鉴别器更新1次的机制。

3 基于GRUI-GAN的数据修复及Mogrifier LSTM网络的EV短期负荷预测方法

3.1 Mogrifier LSTM网络模型

传统电力短期负荷预测中,LSTM网络是一种在多种场景中广泛适用的算法^[10,18,25-26]。附录A图A3显示了LSTM网络的一般结构。在LSTM网络中,隐藏状态 h 和当前输入 x 都是相互独立地馈入LSTM网络细胞中,使得LSTM网络因缺少这种上下文联系而导致一些信息的丢失。为了解决这一问题,文献[27]在LSTM网络的基础上新增加了一种门控机制,借助这种机制,可以将隐藏状态 h 和当前输入 x 进行充分交互,极大地提高LSTM网络对于上下文信息的建模能力。

图3为具有5轮交互更新的Mogrifier LSTM网络结构,其中 $h_{t_i-1,0} = h_{t_i-1}$ 为输入当前LSTM网络中的隐藏状态 h , $x_{t_i-1} = x_{t_i}$ 为当前输入 x 。 $h_{t_i-1,0}$ 经过线性变换(图中虚线箭头)和 x_{t_i-1} 馈入sigmoid门

中(图中 σ)得到新的输入 $x_{t_i,1}$ 。同样,在下一轮更新中, $x_{t_i,1}$ 经过线性变换和 $h_{t_i-1,0}$ 馈入sigmoid门中得到新的隐藏状态 $h_{t_i-1,2}$ 。通过多轮交替更新的方式最终得到新的隐藏状态 $h_{t_i-1}^{new}$ 和新的输入 $x_{t_i}^{new}$,实现上下文信息的充分交互。

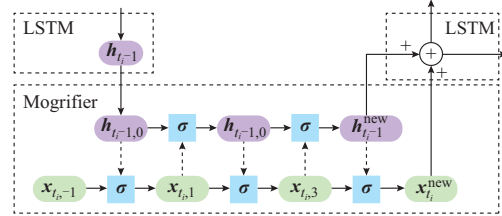


图3 Mogrifier LSTM网络的基本结构
Fig. 3 Basic structure of Mogrifier LSTM network

Mogrifier LSTM网络实际上是在LSTM网络结构的基础上对隐藏状态 h 和输入 x 进行了交叉运算,其运算公式如下:

$$x_{t_i,k} = 2\sigma(Q_k h_{t_i,k-1}) \odot x_{t_i,k-2} \quad k \in \{1, 3, 5, \dots, r\} \quad (16)$$

$$h_{t_i,k} = 2\sigma(R_k x_{t_i,k-1}) \odot h_{t_i,k-2} \quad k \in \{0, 2, 4, \dots, r-1\} \quad (17)$$

其中,当前输入 $x_{t_i} = x_{t_i-1}$,当前隐藏状态 $h_{t_i-1} = h_{t_i-1,0}$;下标 k 表示第 k 次交互更新对应的变量; $r \in \mathbb{N}$ 为交互更新的轮次,是一个超参数,需根据具体的应用场景进行调整,当 $r=0$ 时,则为原始LSTM网络结构; R_k 和 Q_k 为随机初始化矩阵,为了减少附加的模型参数,可将 R_k 和 Q_k 分解为低维矩阵 $R_{left,k}$ 、 $R_{right,k}$ 和 $Q_{left,k}$ 、 $Q_{right,k}$ 的乘积,即 $R_k = R_{left,k} R_{right,k}$ 和 $Q_k = Q_{left,k} Q_{right,k}$;为了保证经过多轮更新之后数值的稳定性乘以常数2。

经过交替更新得到新的隐藏状态 $h_{t_i-1}^{new}$ 和新的输入 $x_{t_i}^{new}$ 馈入LSTM网络中,按原始LSTM网络更新公式更新细胞状态 c 和隐藏状态 h :

$$c = f \odot c_{t_i-1} + i \odot j \quad (18)$$

$$h = o \odot \tanh c \quad (19)$$

$$f = \sigma(W^{fx} x_{t_i}^{new} + W^{fh} h_{t_i-1}^{new} + b^f) \quad (20)$$

$$i = \sigma(W^{ix} x_{t_i}^{new} + W^{ih} h_{t_i-1}^{new} + b^i) \quad (21)$$

$$j = \tanh(W^{jx} x_{t_i}^{new} + W^{jh} h_{t_i-1}^{new} + b^j) \quad (22)$$

$$o = \sigma(W^{ox} x_{t_i}^{new} + W^{oh} h_{t_i-1}^{new} + b^o) \quad (23)$$

式中: c_{t_i-1} 为前一时刻的细胞状态; W^{fx} 、 W^{fh} 、 W^{ix} 、 W^{ih} 、 W^{jx} 、 W^{jh} 、 W^{ox} 、 W^{oh} 和 b^f 、 b^i 、 b^j 、 b^o 分别为权重矩阵和偏置向量。

3.2 EV短期负荷预测框架

本文所采用的EV短期负荷预测框架与传统电力短期负荷预测框架大体相同,见附录A图A4。为提高原始数据集的质量,首先对原始数据中的缺失值和异常值进行检测与识别,并将缺失值和异常值用-1代替。根据识别结果按照2.2节所提方法建立位置矩阵 S 和时滞矩阵 δ 。然后,对替代后的原始数据采用zero-mean标准化方法进行标准化并构建为[批样本数,序列长度,序列特征数]的数组形式,然后输入GRUI-GAN模型中对空缺值进行插补。

在获得了质量较好的数据之后,使用数据分析方法抽取了数据集中的显著特征用于构建输入特征集,分析过程见附录B。根据分析结果,本文选取了时段类型、星期类型、温度、最高温度、最低温度作为本文的特征量,同时考虑到前后负荷的相互影响,前2个时刻的负荷值也作为本文的特征。因此,本文EV负荷预测模型的输入 x 包含前述7个特征量。 x 也需要经过标准化后构建为[批样本数,序列长度,序列特征数]的数组形式,并输入基于Mogriker LSTM网络的预测模型中进行预测。最后,经过全连接层输出预测结果 y ,其形式为[批样本数,预测步长,预测变量数]。

对于预测模型优化器的选择,常用的优化器一般有Adam、SGD、Adagrad以及RMSProp等。其中,Adam优化器以其良好的性能表现,在负荷预测中成为较主流的选择,因此,本文采用Adam优化器来优化训练过程。

在模型评估函数的选择方面,平均绝对百分比误差(MAPE)和均方根误差(RMSE)在负荷预测中经常被采用作为评估函数。由于本文数据集中EV负荷存在负荷值为0的情况,因此MAPE不能用来作为本文所提模型的误差函数。而平均绝对误差(MAE)相比于其他误差函数,对误差的鲁棒性更强。

因此,本文选择MAE和RMSE作为本文模型的评估函数,其表达式分别如下。

$$E_{MAE} = \frac{1}{M} \sum_{t=1}^M |y_t - y_{t,pre}| \quad (24)$$

$$E_{RMSE} = \sqrt{\frac{1}{M} \sum_{t=1}^M (y_t - y_{t,pre})^2} \quad (25)$$

式中: y_t 为 t 时刻的实际值; $y_{t,pre}$ 为 t 时刻的预测值; M 为样本数量。

4 实验及数据分析

4.1 数据及实验平台描述

本文所选EV充电负荷历史数据集为某35 kV供电区域内2019年1月1日至2019年8月31日的多个充电站实际交易数据。由于该区域较低的EV渗透率和充电交易的随机性,所记录的数据并不能以固定采样间隔覆盖每天所有时段,因此,为便于算法验证,本文通过将数据进行初步的聚合和重采样,得到该区域以30 min为间隔的EV历史负荷数据集。

本文硬件平台为Windows 10,64位操作系统,CPU为AMD Ryzen 5 3600 6核处理器,主频为3.59 GHz,内存为16 GB,软件平台为Anaconda3,Python3.6,所用库为TensorFlow1.7.0以及其他常用封装库。

4.2 EV负荷数据插补

在负荷数据插补实验中,本文选取了传统负荷预测中常用的几种空缺值插补方法(均值插补、分段线性插值法插补(用PLI表示)^[18]和 k -NN插补)作为对比方法以验证所提方法有效性。同时,为比较GAN框架下本文所提出的GRUI细胞结构相比于GRU细胞结构的优势,以GRU细胞构建的GRU-GAN插补方法也被用于本文插补实验的对比。

由于本文数据集自身为不完整数据,因此无法直接评估不同算法的空缺值插补精度。文献[23-24]中通过划分训练集和测试集来训练和测试GAN模型的实验方法在本文中并不适用。因此,本文将所有数据在所提出的GRUI-GAN模型中进行训练,即训练集为整个数据集。而经整个数据集训练之后的GRUI-GAN模型对数据集的插补结果的好坏只能借助后续负荷预测结果来评估。因此,为验证所提方法的有效性,本文插补实验包含如下2个部分。

1)从数据集中选取一部分完整数据(不包含缺失值且不考虑异常值),随机缺失不同比例的数据构建不完整数据进行实验,以验证在小数据集上采用全数据集训练之后模型的插补表现。

2)将整个不完整数据集重新按照上述实验方法获取,经过不同方法插补得到的不同完整数据集用于后续EV负荷预测,通过最终EV负荷预测结果间接验证所提插补方法的有效性。

在参数设置方面,本文主要通过观察生成器和鉴别器的损失函数图像来确定GAN的网络参数。当生成器和鉴别器的损失函数图像呈现出文献[22-24]中损失函数图像的形式,并且实现对抗训练最

终收敛,则该参数就被确定为本文的参数。本文以文献[29]的参数设置为基础,并结合本文的场景进行调参,调参结果为:隐藏层 GRUI 神经元个数为 10,学习率为 0.005,损失函数权重超参数 λ 为 1.2,训练次数设置为 3 000。

在第 1 部分实验中,本文将原始负荷数据中不包含缺失值(2019 年 5 月 26 日至 7 月 30 日)并且不考虑异常值的负荷数据作为完整样本。考虑到不同场景所导致的负荷数据缺失形式的不同,本文在随机缺失(模拟随机异常情况导致随机单个数据点数据的缺失)、片段缺失(模拟系统故障无法及时恢复导致整天数据缺失)和混合缺失(既包含随机缺失,也包含片段缺失,在该实验中比例为 6:4)3 种负荷数据缺失场景下进行了插补实验,并将插补结果与完整样本进行比较以均方误差^[29]来评估插补精度,如表 1 至表 3 所示。

表 1 随机缺失插补误差对比
Table 1 Imputation error comparison for random missing

缺失率/%	不同空缺值插补方法下的均方误差				
	GRUI-GAN	GRU-GAN	k-NN 插补	均值插补	PLI
5	0.039 4	0.040 7	0.042 9	0.047 9	0.089 2
10	0.060 0	0.069 8	0.082 4	0.091 1	0.185 7
15	0.108 8	0.117 5	0.130 8	0.143 3	0.279 1
20	0.135 5	0.140 0	0.156 4	0.189 1	0.332 9
25	0.157 8	0.162 0	0.1832	0.227 0	0.409 8
30	0.186 4	0.197 7	0.238 4	0.282 7	0.533 6

表 2 片段缺失插补误差对比
Table 2 Imputation error comparison for segment deletion

缺失天数	不同空缺值插补方法下的均方误差				
	GRUI-GAN	GRU-GAN	k-NN 插补	均值插补	PLI
1	0.017 8	0.018 9	0.028 1	0.023 6	0.043 0
2	0.021 0	0.022 7	0.035 1	0.040 9	0.073 1
3	0.027 2	0.028 5	0.041 7	0.046 4	0.101 1
4	0.047 9	0.049 9	0.077 1	0.086 7	0.140 3
5	0.050 1	0.052 0	0.084 5	0.090 9	0.175 8
6	0.064 5	0.066 2	0.098 4	0.106 4	0.226 8
7	0.067 5	0.069 9	0.103 1	0.117 8	0.274 2
8	0.081 1	0.083 8	0.138 6	0.141 2	0.324 4
9	0.082 4	0.085 9	0.185 1	0.199 6	0.366 8

从表 1 至表 3 的插补误差计算结果可以看出,所提基于 GRUI-GAN 的数据插补模型的插补精度在 3 种负荷缺失场景下都优于传统负荷预测中的负荷缺失值插补方法。特别是在片段缺失最严重的场景

表 3 混合缺失插补误差对比
Table 3 Imputation error comparison for mixed missing

缺失率/%	不同空缺值插补方法下的均方误差				
	GRUI-GAN	GRU-GAN	k-NN 插补	均值插补	PLI
5	0.036 0	0.038 5	0.040 2	0.045 0	0.077 1
10	0.065 6	0.068 7	0.087 5	0.104 2	0.189 6
15	0.108 1	0.113 5	0.134 0	0.154 2	0.282 4
20	0.119 4	0.123 8	0.144 5	0.181 4	0.364 8
25	0.168 0	0.175 9	0.199 3	0.237 3	0.467 7
30	0.181 1	0.194 5	0.232 5	0.295 7	0.545 7

下,所提插补方法的插补精度与传统方法的精度相比提升了 1 个量级。此外,由于在 GRU 细胞中引入了时间衰减向量 β 来学习空缺值导致的不规则时滞对观测值的影响,使用 GRUI 细胞构建 GRUI-GAN 在 3 种插补场景下的结果都比 GRU-GAN 结果更好。因此,可以认为,所提插补方法能够以较好的精度对原始负荷数据进行插补。同时,在附录 A 图 A5 中展示了缺失率为 5% 的混合缺失场景下(与本文缺失数据场景相似)一段序列的不同方法实际插补效果,可以看出,基于 GAN 的数据修复方法能够以较好的精度修复数据。

在第 2 部分实验中,本文采用 GRUI-GAN、k-NN 插补、均值插补以及 PLI 对原始负荷数据中的空缺值进行插补,以获取经 4 种不同方法修复之后的完整负荷数据,并在下一节中用于 EV 负荷预测。而最终修复结果的合理性,将根据在 4 个不同数据集上的负荷预测结果间接反映。

4.3 EV 负荷预测

特定的数据集选取合适的特征用于模型训练,将会有效提升模型的性能。因此,本文通过数据分析可视化方法从数据集中抽取了时段特征、日期类型和温度因素(温度数据从气象网站获取,不包含缺失值和异常值)作为典型特征构建输入特征集,具体分析过程见附录 B。

基于深度学习的负荷预测模型的预测精度受到模型结构的影响,而模型结构主要与其数据集规模有关,数据集规模较小而构建的神经网络模型较大容易造成过拟合,反之则容易造成欠拟合。因此,确定合适的网络参数也十分重要。本文首先参考现有文献中 LSTM 网络的参数设置,初步确定其他次要参数的取值。然后,根据文献[17]中的参数寻优方法确定主要参数的最优取值,在确定主要参数取值之后,再采用同样的寻优方法依次确定其他参数。最终,本文基于 Mogrifler LSTM 网络的预测模型的

主要参数设置为:隐藏层数为53,神经元个数为32,预测模型训练的选取时间步长为10,按批进行训练的每批样本数目为150,学习率为0.000 2,交互轮次为3,降维阶数为1。

为了验证本文所提预测模型的有效性,本文选取了文献中常见的几种短期负荷预测算法:LSTM网络、双向长短期记忆(Bi-LSTM)网络、GRU、支持向量机(SVM)、自回归差分移动平均(ARIMA)、季节性外部信息自回归差分移动平均(SARIMAX)作为对比实验。首先,在经GRUI-GAN方法修复之后的数据集上进行日负荷预测实验,并评估所选预测算法的预测误差。图4展示了所选算法在测试集(2019年8月11日至8月31日)上随机选取目标日的日负荷预测效果,根据本文所选评估指标计算的预测误差如表4所示。综合几种方法可视化结果以及预测误差可知,本文使用的Mogrifier LSTM网络算法和同样以改进LSTM网络上下文信息建模能力为目的Bi-LSTM网络算法在测试集上都取得了优于传统LSTM网络的良好表现。虽然在高峰时刻的预测精度上,本文所提方法略逊于Bi-LSTM网络算法,但是其余大部分时段,Mogrifier LSTM网络算法都有更好预测表现并且总体误差也更小。

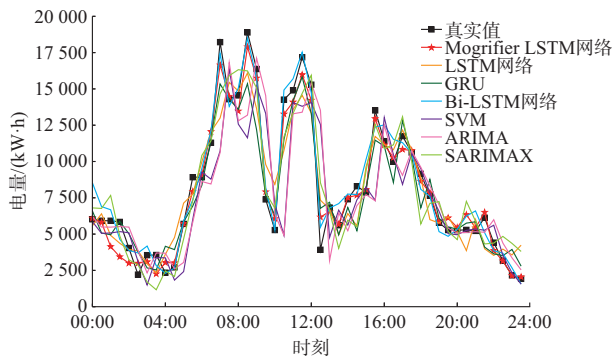


图4 某日预测结果对比

Fig. 4 Comparison of forecasting results on a day

表4 算法误差对比
Table 4 Algorithm error comparison

算法	$E_{MAE}/(\text{kW}\cdot\text{h})$	$E_{RMSE}/(\text{kW}\cdot\text{h})$
Mogrifier LSTM 网络	685.82	913.68
LSTM 网络	1 441.74	1 856.00
Bi-LSTM 网络	952.71	1 294.73
GRU	1 238.37	1 681.27
SVM	1 735.99	2 572.43
ARIMA	1 802.58	2 619.94
SARIMAX	1 703.29	2 215.89

此外,为了进一步评估不同负荷数据修复方法对预测结果的影响,本文将4.2节经不同方法修复之后的负荷数据重新用于训练模型,最终在整个测试集上评估了日负荷预测平均误差,评估结果如图5所示。从图中可以明显看出,本文所选的几种预测算法,在经GRUI-GAN修复之后的数据集上都取得了更高的预测精度。因此,可以认为经GRUI-GAN方法修复之后的数据集质量得到了有效改善,从而使得预测模型能够捕获更加精确的信息以提高预测结果精度。同时,根据图5显示结果可以看出,本文所使用的Mogrifier LSTM网络算法在测试集上可以保证一定的稳定性。

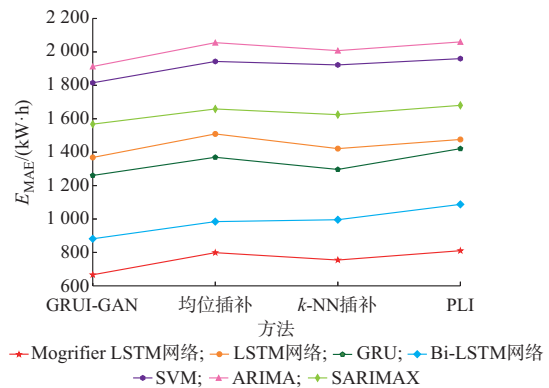


图5 不同插补方法对算法预测误差影响

Fig. 5 Influence of different imputation methods on forecasting error of algorithm

5 结语

面对EV负荷数据质量较差情况下的高精度负荷预测需求,本文提出了基于GRUI-GAN的负荷数据修复方法。通过在传统GRU细胞结构中引入时间衰减向量构建GRUI细胞结构来捕获不同时滞对观测值的影响。以这一结构构建的GAN可以处理不完整负荷序列,经过对抗训练生成较高精度的数据进行插补,从而提高负荷数据质量。然后,在预测算法模型方面,本文在传统LSTM网络中采用的Mogrifier门控机制以多轮的交互更新方式使LSTM网络上下文信息得到充分交互,进一步优化传统LSTM网络的性能。但在负荷峰值时段,Mogrifier LSTM网络的拟合精度仍有提高的空间,因此未来将会研究Mogrifier LSTM网络在负荷峰值时段的拟合精度提高方法,以进一步提高EV负荷预测精度。

附录见本刊网络版(<http://www.aeps-info.com/aeps/ch/index.aspx>),扫英文摘要后二维码可以阅读

网络全文。

参考文献

- [1] CROZIER C, MORSTYN T, MCCULLOCH M. The opportunity for smart charging to mitigate the impact of electric vehicles on transmission and distribution systems [J]. *Applied Energy*, 2020, 268: 114973.
- [2] WANG H J, WANG B, FANG C, et al. Charging load forecasting of electric vehicle based on charging frequency [J]. *IOP Conference Series: Earth and Environmental Science*, 2019, 237(6): 062008.
- [3] ARIAS M B, BAE S. Electric vehicle charging demand forecasting model based on big data technologies [J]. *Applied Energy*, 2016, 183: 327-339.
- [4] AL-OGAILI A S, TENGKU HASHIM T J, RAHMAT N A, et al. Review on scheduling, clustering, and forecasting strategies for controlling electric vehicle charging: challenges and recommendations [J]. *IEEE Access*, 2019, 7: 128353-128371.
- [5] AMINI M H, KARABASOGLU O, ILIĆ M D, et al. ARIMA-based demand forecasting method considering probabilistic model of electric vehicles' parking lots [C]// 2015 IEEE Power & Energy Society General Meeting, July 26-30, 2015, Denver, USA.
- [6] MAJIDPOUR M, QIU C, CHU P, et al. Modified pattern sequence-based forecasting for electric vehicle charging stations [C]// 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), November 3-6, 2014, Venice, Italy.
- [7] BUZNA L, DE FALCO P, KHORMALI S, et al. Electric vehicle load forecasting: a comparison between time series and machine learning approaches [C]// 2019 1st International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED), May 28-30, 2019, Cagliari, Italy.
- [8] XYDAS E S, MARMARAS C E, CIPCIGAN L M, et al. Electric vehicle load forecasting using data mining methods [C]// IET Hybrid and Electric Vehicles Conference 2013 (HEVC 2013), November 6-7, 2013, London, UK.
- [9] XU X B, LIU W X, ZHOU X, et al. Short-term load forecasting for the electric bus station based on GRA-DE-SVR [C]// 2014 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), May 20-23, 2014, Kuala Lumpur, Malaysia.
- [10] ZHU J C, YANG Z L, CHANG Y, et al. A novel LSTM based deep learning approach for multi-time scale electric vehicles charging load prediction [C]// 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), May 21-24, 2019, Chengdu, China.
- [11] 邓带雨, 李坚, 张真源, 等. 基于EEMD-GRU-MLR的短期电力负荷预测[J]. *电网技术*, 2020, 44(2): 593-602.
DENG Daiyu, LI Jian, ZHANG Zhenyuan, et al. Short-term electric load forecasting based on EEMD-GRU-MLR [J]. *Power System Technology*, 2020, 44(2): 593-602.
- [12] 冯丽, 邱家驹. 离群数据挖掘及其在电力负荷预测中的应用 [J]. *电力系统自动化*, 2004, 28(11): 41-44.
FENG Li, QIU Jiaju. Outlier data mining and its application in electric load forecasting [J]. *Automation of Electric Power Systems*, 2004, 28(11): 41-44.
- [13] 张帅, 杨晶显, 刘继春, 等. 基于多尺度时序建模与估计的电力负荷数据恢复 [J]. *电工技术学报*, 2020, 35(13): 2736-2746.
ZHANG Shuai, YANG Jingxian, LIU Jichun, et al. Power load recovery based on multi-scale time-series modeling and estimation [J]. *Transactions of China Electrotechnical Society*, 2020, 35(13): 2736-2746.
- [14] 蒋定国, 姚义振, 全秀峰, 等. 不同数据处理策略对Chl-a浓度预测精度的影响 [J]. *人民长江*, 2019, 50(4): 58-64.
JIANG Dingguo, YAO Yizhen, QUAN Xiufeng, et al. Research on influence of different data processing strategies on prediction accuracy of Chlorophyll-a concentration [J]. *Yangtze River*, 2019, 50(4): 58-64.
- [15] KAISER J. Dealing with missing values in data [J]. *Journal of Systems Integration*, 2014: 42-51.
- [16] SILVA L O, ZÁRATE L E. A brief review of the main approaches for treatment of missing data [J]. *Intelligent Data Analysis*, 2014, 18(6): 1177-1198.
- [17] 杨智宇, 刘俊勇, 刘友波, 等. 基于自适应深度信念网络的变电站负荷预测 [J]. *中国电机工程学报*, 2019, 39(14): 4049-4061.
YANG Zhiyu, LIU Junyong, LIU Youbo, et al. Transformer load forecasting based on adaptive deep belief network [J]. *Proceedings of the CSEE*, 2019, 39(14): 4049-4061.
- [18] 杨龙, 吴红斌, 丁明, 等. 新能源电网中考虑特征选择的Bi-LSTM网络短期负荷预测 [J]. *电力系统自动化*, 2021, 45(3): 166-173.
YANG Long, WU Hongbin, DING Ming, et al. Short-term load forecasting in renewable energy grid based on bi-directional long short-term memory network considering feature selection [J]. *Automation of Electric Power Systems*, 2021, 45(3): 166-173.
- [19] BATISTA G E A P A, MONARD M C. An analysis of four missing data treatment methods for supervised learning [J]. *Applied Artificial Intelligence*, 2003, 17(5/6): 519-533.
- [20] HASTIE T, MAZUMDER R, LEE J, et al. Matrix completion and low-rank SVD via fast alternating least squares [J]. *Journal of Machine Learning Research*, 2014, 16(1): 3367-3402.
- [21] MAZUMDER R, HASTIE T, TIBSHIRANI R. Spectral regularization algorithms for learning large incomplete matrices [J]. *Journal of Machine Learning Research*, 2010, 12(11): 2287-2322.
- [22] CHEN Y Z, WANG Y S, KIRSCHEN D, et al. Model-free renewable scenario generation using generative adversarial networks [J]. *IEEE Transactions on Power Systems*, 2018, 33(3): 3265-3275.
- [23] 王守相, 陈海文, 潘志新, 等. 采用改进生成式对抗网络的电力系统量测缺失数据重建方法 [J]. *中国电机工程学报*, 2019, 39(1): 56-64.
WANG Shouxiang, CHEN Haiwen, PAN Zhixin, et al. A reconstruction method for missing data in power system measurement using an improved generative adversarial network [J]. *Proceedings of the CSEE*, 2019, 39(1): 56-64.
- [24] 杨玉莲, 齐林海, 王红, 等. 基于生成对抗和双重语义感知的配

- 电网量测数据缺失重构[J]. 电力系统自动化, 2020, 44(18): 46-54.
- YANG Yulian, QI Linhai, WANG Hong, et al. Reconstruction of missing measurement data in distribution network based on generative adversarial network and double semantic perception [J]. Automation of Electric Power Systems, 2020, 44(18): 46-54.
- [25] KONG W C, DONG Z Y, JIA Y W, et al. Short-term residential load forecasting based on LSTM recurrent neural network [J]. IEEE Transactions on Smart Grid, 2019, 10(1): 841-851.
- [26] TAN M, YUAN S P, LI S H, et al. Ultra-short-term industrial power demand forecasting using LSTM based hybrid ensemble learning [J]. IEEE Transactions on Power Systems, 2020, 35(4): 2937-2948.
- [27] MELIS G, KOČISKÝ T, BLUNSOM P. Mogrifier LSTM [EB/OL]. [2021-03-01]. <https://arxiv.org/abs/1909.01792>.
- [28] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [J]. Advances in Neural Information Processing Systems, 2014: 2672-2680.
- [29] LUO Y, CAI X, ZHANG Y, et al. Multivariate time series imputation with generative adversarial networks [EB/OL]. [2021-03-01]. <https://dblp.uni-trier.de/rec/conf/nips/LuoCZXY18.html>.
- [30] HARADAL S, HAYASHI H, UCHIDA S. Biosignal data augmentation based on generative adversarial networks [C]// 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, July 18-21, 2018, Honolulu, USA.
- 赵厚翔(1997—),男,硕士研究生,主要研究方向:人工智能在电力系统中的应用。E-mail:1104337814@qq.com
- 沈晓东(1975—),男,通信作者,博士,副教授,主要研究方向:园区多能源系统优化等。E-mail:shengxd@scu.edu.cn
- 吕林(1963—),男,硕士,教授,硕士生导师,主要研究方向:配电网规划等。E-mail:lvlin@scu.edu.cn
- (编辑 鲁尔姣)

Load Data Restoration Based on Generative Adversarial Network and Its Application in Short-term Load Forecasting of Electric Vehicle

ZHAO Houxiang¹, SHEN Xiaodong¹, LYU Lin¹, LAN Peng¹, LIU Junyong¹, LIU Daiyong²

(1. College of Electrical Engineering, Sichuan University, Chengdu 610065, China;

2. Deyang Power Supply Company, State Grid Sichuan Electric Power Company, Deyang 618000, China)

Abstract: As the scale of electric vehicle (EV) charging facilities keeps expanding, EV charging data can be obtained more conveniently. Some non-human factors will lead to the problems of missing data and abnormal data in the data set, which hinders the improvement of EV load forecasting accuracy. Therefore, this paper uses the gated recurrent unit neural network cells for imputation (GRUI) in the generative adversarial network (GAN) to deal with the irregular time-delay relationship between the previous and later observation data in the incomplete load data set, and a data imputation model based on GRUI-GAN is proposed to restore the EV load data. Then, the long short-term memory (LSTM) network with the Mogrifier gating mechanism is used for EV load forecasting. Finally, the experimental results show that the proposed method can generate new data with high accuracy to interpolate missing values, and the data restored by the proposed method can effectively improve the accuracy of EV load forecasting.

This work is supported by the State Key Program of National Natural Science Foundation of China (No. U2066209) and China Southern Power Grid Co., Ltd. (No. 060400KK52190017).

Key words: electric vehicle (EV); short-term load forecasting; data imputation; generative adversarial network (GAN); gated recurrent unit neural network; long short-term memory network

