

引文格式:

秦红莲, 何玉林, 黄哲学. 基于神经网络的工厂用电状态识别方法 [J]. 集成技术, 2019, 8(4): 42-51.

Qin HL, He YL, Huang ZX. Neural networks-based factory power consumption state recognition [J]. Journal of Integration Technology, 2019, 8(4): 42-51.

基于神经网络的工厂用电状态识别方法

秦红莲^{1,2} 何玉林^{1,2} 黄哲学^{1,2}

¹(深圳大学计算机与软件学院 大数据技术与应用研究所 深圳 518060)

²(深圳大学大数据系统计算技术国家工程实验室 深圳 518060)

摘 要 智能电表的迅速普及与应用引起电力消耗数据(即智能电网数据)的激增,这不仅给数据的存储与通信带来了挑战,同时也增加了对数据进行分析的难度。另外,由于生产性质和经营方式的不同,工厂的用电状态往往比较复杂。基于传统经验的人工识别不能满足实际应用的需求,该文研究了一种基于神经网络的工厂用电状态自动识别方法。首先,对采集于真实应用场景的电网大数据进行必要的预处理,包括数据的合并、清洗、标准化、打标和抽样;然后,基于预处理的电网数据构建神经网络模型用于对工厂用电状态的自动识别;最后,对提出的基于神经网络的工厂用电状态识别方法进行实验验证,证实了该方法的合理性和有效性。通过对工厂用电状态的准确识别,能够帮助供电公司指导企业进行错峰用电,进而有效缓解电力供给不平衡的问题,以达到对电能合理开发和利用的目的。

关键词 人工智能; 神经网络; 智能电表; 电网数据; 用电状态

中图分类号 TP 391 **文献标志码** A **doi**: 10.12146/j.issn.2095-3135.20190426001

Neural Networks-Based Factory Power Consumption State Recognition

QIN Honglian^{1,2} HE Yulin^{1,2} HUANG Zhexue^{1,2}

¹(Big Data Institute, College of Computer Science & Software Engineering, Shenzhen University, Shenzhen 518060, China)

²(National Engineering Laboratory for Big Data System Computing Technology, Shenzhen 518060, China)

Abstract With the rapid application of smart electric meters, a large amount of power consumption data is generated, which challenges the data storage and communication. Meanwhile, due to the extensive production properties and business modes, the power consumption states of factories are complicated, which makes data analysis difficult. In this paper, a new neural networks-based power consumption state recognition method is proposed. The main efforts include three aspects: the pre-processing of power data, the modeling of neural network to automatically recognize the power consumption states of factories and the performance testing of

收稿日期: 2019-04-26 修回日期: 2019-05-31

基金项目: 国家重点研发计划项目(2017YFC0822604-2); 中国博士后科学基金项目(2016T90799); 深圳大学新引进教师科研启动项目(2018060)

作者简介: 秦红莲, 硕士研究生, 研究方向为大数据系统计算技术与应用; 何玉林(通讯作者), 博士, 副研究员, 研究方向为机器学习与数据挖掘, E-mail: yulinhe@szu.edu.cn; 黄哲学, 博士, 教授, 研究方向为大数据系统计算技术与应用。

the proposed method. The experimental results demonstrate the rationality and effectiveness of our method. The correct recognition of power consumption states of factories is helpful to the power supply company to arrange the sensible power supply strategy and efficient power utilization plan.

Keywords artificial intelligence; neural network; smart electric meter; electric power data; electric power consumption state

1 引言

随着智能电网技术的快速发展, 智能电表被广泛应用, 这使电力工业应用领域产生了大量结构多样、复杂多变的智能电网数据。如何有效地存储和利用这些数据是目前电网公司面临的难题。另外, 这些智能电网数据中往往蕴含着巨大的有价值的信息, 高效准确地提取这些信息不仅可以提高电网自身的管理和运营水平, 同时亦可为工业界和广大用户提供更加优质的服务, 为电网公司业务的迅速拓展提供先决条件。

在对电网数据的研究中, 由于电网数据的复杂性, 仅凭经验和专业知识很难对其进行有效的数据挖掘, 因此需要将它们进行分类归并, 获得其典型代表之后再对用户的用电状态进行深入分析。相比于其他“轻量级”数据来说, 对“重量级”电网数据进行分类的工作是复杂和困难的, 这是当下很多科研人员争相研究的热点, 同时也是本文亟待解决的问题。当前, 电网业务数据大致分为三类: 一是电力企业生产数据, 如发电量、电压以及稳定性等方面的数据; 二是电力企业运营数据, 如交易电价、售电量、用电客户等方面的数据; 三是电力企业管理数据, 如企业资源计划系统(ERP)、一体化平台、协同办公等方面的数据。如果能充分利用电网业务数据, 对其进行深入分析, 便可为相关机构和从业人员提供大量的高附加值服务。这些增值服务将有利于电网安全检测与控制(包括大灾难预警与处理、供电与电力调度决策支持和更准确的用电量预

测)、客户用电状态分析与客户细分、电力企业精细化运营管理等, 实现更科学的需求化管理。

本文所研究的电网数据属于第一类, 即电力企业生产数据。针对这一类数据, 目前主要使用大数据、人工智能、机器学习等方法对其进行数据挖掘, 相关研究主要分为基于无监督学习方法的用电模式识别^[1-6]、用电量预测^[7-13]和异常值检测^[14-16]等。其中, 大数据方法的优势体现在充分利用现有数据且只需要依靠少量经验知识就能在跨专业和跨部门数据融合基础上实现多维度的数据分析; 人工智能和机器学习方法不需要人为给定具体规则就能自动挖掘出数据之间蕴含的大量信息。因此, 有效利用这些方法挖掘智能电网大数据中的有用信息, 可有效提高电网的运行管理水平和为社会及用户的服务水平。

虽然国内外在智能电网数据挖掘方面做了一些有益的尝试, 但其主要研究的是轻量级的居民或企业用电数据。同时, 由于国情的差异, 很多国外的新技术和新思想并不适用于中国国内电网数据的分析。另外, 目前还没有对来自各种类型的工厂的重量级电网数据进行用电状态识别的研究。因此, 研究一种能对工厂用电状态进行识别的算法对于学术界以及工业界是十分必要且有意义的。基于此, 本文构建了一种基于神经网络的工厂用电状态识别模型。为了对工厂用电状态进行识别, 本文主要做了以下几个方面的工作: (1) 针对电网数据的特点, 对其进行数据预处理, 包括数据合并、数据清洗、数据标准化、数据打标和数据抽样; (2) 针对预处理后的电网数

据进行神经网络建模,实现对工厂用电状态的自动识别;(3)分析实验结果,总结工厂用电状态的特点,并对工厂的用电模式进行解释。实验结果显示,本文所使用的神经网络模型能够以高于95%的精度对工厂用电状态进行识别,这不仅表示本文所做的数据预处理是有效的,而且证实采用神经网络处理电网数据是可行的。另外,还可以利用聚类方法进一步挖掘被正确识别用电状态电网数据的用电模式,以便人们更好地理解电网数据中蕴含的知识和规律。

2 基础知识

本节将对神经网络和电网数据的基本知识进行简单介绍。

2.1 神经网络

1943年,心理学家 W. S. McCulloch 和数理逻辑学家 W. Pitts 建立了神经网络的数学模型,称为 M-P 模型^[17],如图 1 所示。M-P 模型基于网络化的构造方法展示了神经元的数学描述,证明了单个神经元能执行逻辑功能,从而奠定了神经网络研究的基础。

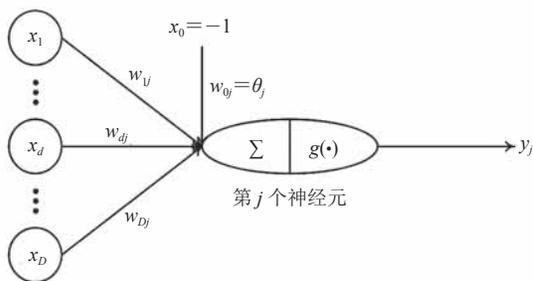


图 1 M-P 模型

Fig. 1 McCulloch-Pitts model

图 1 中, $\bar{x}=(x_1, x_2, \dots, x_D)$ 为神经元的输入; $\bar{w}_j=(w_{1j}, w_{2j}, \dots, w_{Dj})$ 表示第 j 个神经元的权重向量; θ_j 为第 j 个神经元的偏置。则神经元的输出可以被表示为:

$$y_j = g\left(\sum_{d=1}^D w_{dj} x_j - \theta_j\right) \quad (1)$$

其中, $g(\cdot)$ 为激活函数,其作用是引入非线性因素,增强神经网络的表达能力。常用的激活函数有 sigmoid 函数^[18]、tanh 函数、ReLU^[19]、Leaky ReLU^[20]以及 softmax 函数^[21]。

目前关于神经网络的定义使用最广泛的是 Kohonen^[22]于 1988 年提出的,即神经网络是由具有适应性的简单单元(即神经元)组成的广泛并行互连的网络,它的组织能够模拟生物神经系统对真实世界物体所做出的交互反应。对于神经网络的训练,关键在于神经元权重向量的确定。其中,误差反向传播算法是目前用来训练神经网络最有效的算法。

2.2 电网数据

本文所研究的电网数据是 2013 年中国广东省东莞市地区所有工厂的瞬时量表数据。针对每一家企业,每天取一条记录,每条记录按照每隔 15 min 采样一次,即每条记录包含 96 个采样点(time_point_1 到 time_point_96),然后再将其按照月份分成 12 个表。另外,每条数据里面还包括企业信息(用编号 point_id 表示)、采样日期(date_time)、实际采样数目(count_n)、采样周(week)等,具体如表 1 所示。该地区的工厂类型主要有 7 种,包括大工业、普通工业、农业和商业等,企业数量有 23 000 多家。仅按照每家工厂每天产生一条记录来计算,每个月产生的电网数据记录数就可达 700 000 条,一年的数据量超过 6 GB。同时,原始电网数据还存在大量的异常值、缺失、噪音以及数据质量较差等问题,因此对其进行一定的预处理是十分必要的。

本论文中对电网数据进行预处理的流程如下:(1)将 12 个月的数据进行合并;(2)对合并后的数据,筛选出按照每隔 15 min 进行一次采样的数据,即采样数目等于 96 的数据;(3)基于一定的先验知识与经验,对(2)中的数据进行“预处理”,该过程将在下一节中详细描述。

表 1 电网数据片段

Table 1 Electric power data segment

企业编号	采样日期	实际采样数目	修正后采样数目	采样周	采样点 1	...	采样点 96
54	2013/1/6	95	96	0	137	...	139
54	2013/1/11	95	96	5	137	...	137
54	2013/1/16	96	96	3	134	...	130
54	2013/1/21	95	96	1	137	...	135
54	2013/1/26	93	96	6	583	...	698
54	2013/1/34	96	96	4	134	...	134
56	2013/1/4	96	96	5	3 518	...	3 427

3 用电状态识别

对于工厂而言, 基于瞬时量表数据, 可以识别其两种用电状态: 用电或非用电。因此, 本文研究的是如何对瞬时量表数据的状态进行有效准确识别, 所做的工作主要是数据预处理和神经网络建模。

3.1 数据预处理

直接对原始电网数据进行分析是十分困难的, 这是因为原始电网数据存在以下问题。

(1) 存在大量缺失值: 具体表现为存在缺失的数据条数占总数据条数的 19%。

(2) 存在大量噪声数据: 噪声主要来源于智能电表内部电路中的各种器件(包括电阻、电容、运算放大器等)产生的固有噪声。

(3) 存在大量异常值: 人工录入错误导致有异常值存在; 智能电表故障(计时错误和读数不准确等)导致存在无效数据或错误数据。

(4) 数据来自多个异构数据源: 即数据字段格式不一致。因为数据是由不同人在不同时间存放在不同表中的, 因此很有可能会出现某些字段格式甚至是名称不一致的问题。

(5) 数据格式不被所使用算法接受: 本文所采用的神经网络能够消除变量间量纲关系且具有类标的 96 维数据, 而原始数据不满足这一要求等。

上述这些问题的存在, 严重降低了数据质量, 非常不利于机器学习模型的训练, 将直接导致不被实际应用所接受的数据挖掘结果。因此, 需要结合实际需求以及数据特性对原始电网数据进行预处理, 本文的数据预处理主要分为以下 5 个步骤。

(1) 数据合并

按照字段间的对应关系将 12 个月的瞬时量表数据合并起来。例如, 按照字段间的对应关系将 2013 年 12 个月的 12 个瞬时量表数据表并成一个表, 该表包含 2013 年东莞地区所有工厂的瞬时量表数据。

(2) 数据清洗

第一步: 筛选出每天的采样点数目都为 96 的工厂的瞬时量表数据。由于数据采集技术原因, 有些设备是每隔 1 h 进行一次采样, 一天只有 24 个采样点, 这造成了严重的数据特征的丢失。

第二步: 去掉异常数据以及缺失数据。由于设备质量不过关以及维修人员业务能力有限等原因, 有些设备出故障后没能及时修理, 造成数据采样异常或缺失。

第三步: 去掉数据条数不足 365 条(一年 365 天, 每天一条)的工厂数据。

(3) 数据标准化

由于工厂规模大小的不同, 具有相同用电模式(如同为三班倒模式)的不同工厂之间的用电量

大小会有很大的区别。例如,大工厂的用电量每小时上万度,而小工厂每小时的用电量只有几百度。因此,就必须对数据进行标准化,消除变量间的量纲关系,从而使数据具有可比性。本文使用的数据标准化方法是 min-max 方法,通过线性变换将数据落到[0,1]区间,因此也被称为 0-1 标准化方法。转换方法如下:

$$v'_n = \frac{v_n - \min}{\max - \min} \in [0,1] \quad (2)$$

其中, \max 和 \min 分别为数据集 $\{v_1, v_2, \dots, v_N\}$ 对应的最大值、最小值; $v_n \in \mathbb{R} (n=1, 2, \dots, N)$ 。

(4) 数据打标

如果将每个工厂的用电数据按照工作日(周一至周五)与非工作日(周六、周日)分开进行画图观察,会发现工作日里面有很多条数据明显是非用电状态。同样,非工作日里也有一些数据明显是用电状态,如图 2 所示。这说明仅仅按照工作日和非工作日对电网数据进行划分是不可行的。因此,需要重新对数据进行打标。

对数据进行打标的过程如下:对于任意一个工厂,不论它属于哪一种上班模式,如两班倒、三班倒或四班倒等,当其不开工(处于非用电状态)时只需要维持一些基础设施的用电,此时用电量曲线趋于平缓且数值较小。基于这一原理,只要为每个工厂找到用以区别用电状态与非用电状态的分界线(分界线以下为非用电状态,以上

则为用电状态),那么就能将每个工厂的数据分为“用电状态”和“非用电状态”这两类。通过对 2 000 多个工厂的原始数据进行画图观察,总结出了以下求分界线的方法。

第一步:求出每一维度(time_point_i)下的最大值。

第二步:对这 96 个局部最大值进行移动平均。

第三步:求出移动平均过后的全局最大值。

第四步:分界值。

$$Boundary = \alpha \times \max \quad (3)$$

其中, $\alpha \in (0,1)$ 为增大率,用以控制分界值与全局最大值之间的倍数关系。

得到每个工厂的分界值后,按照公式(4)对每条用电数据进行打标,即确定用电状态。

$$f(\vec{x}_{ij}) = \begin{cases} 0, & \max(\vec{x}_{ij}) < Boundary_i \\ 1, & \text{其他} \end{cases} \quad (4)$$

其中, $Boundary_i$ 表示第 i 个工厂的用电状态分界; \vec{x}_{ij} 表示第 i 个工厂的第 j 条用电数据(96 维)。

(5) 数据抽样

经过数据清洗之后的数据将近 2 000 MB,包含 6 000 多家工厂的电网数据。对于数据分析而言,这样的数据量是很大的,所以需要对其进行抽样,进而缩减数据规模。通过观察发现,这些工厂的用电模式主要分为一班倒、两班倒、三班倒及四班倒四种,图 3 展示了这四种不同的用电模式。由于整个实验过程是在单机上完成的,实

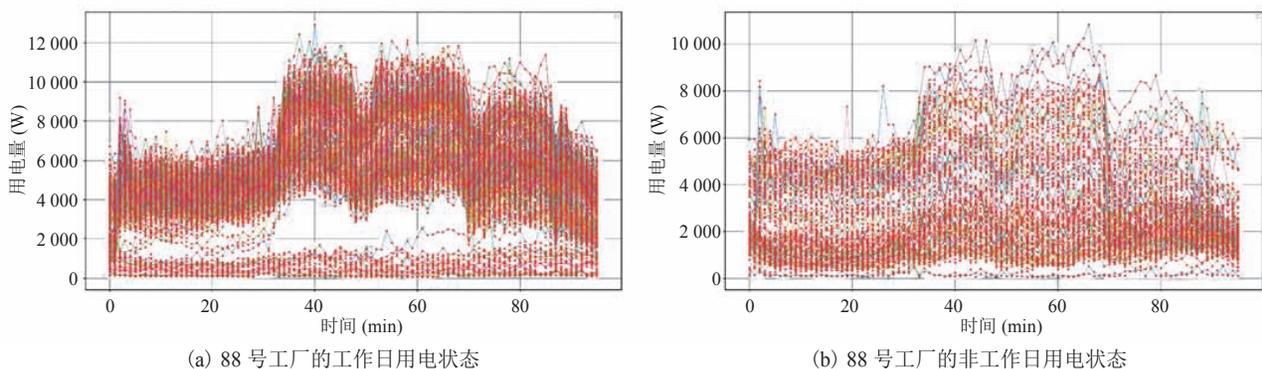


图 2 88 号工厂的用电状态

Fig. 2 The electric power consumption states of point_id 88

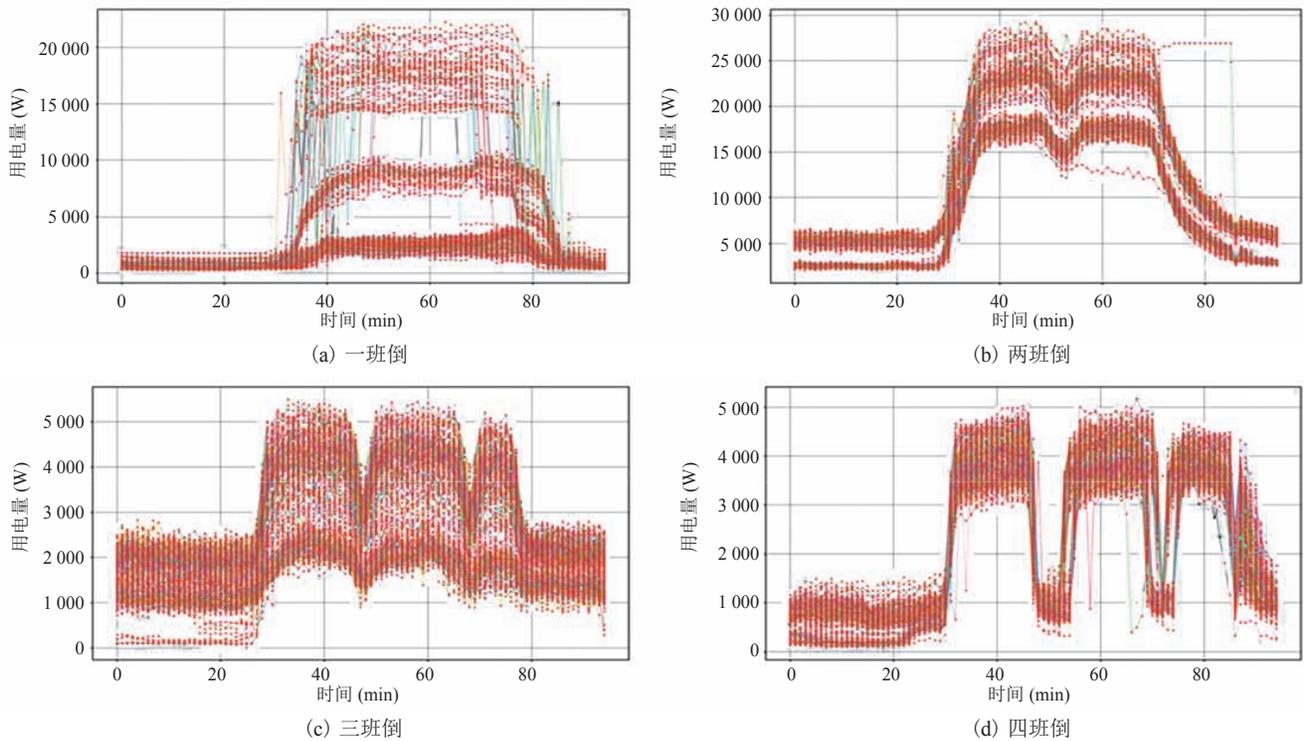


图 3 四种用电模式

Fig. 3 Four power consumption modes

验所使用的数据量大小受单机性能的限制, 因此需要对整体大数据进行抽样。本文使用分层抽样对原始数据进行抽样, 所获得的大数据随机样本划分(Random Sample Partition, RSP)数据块具有与大数据整体一致的概率分布, 能够保证在给定的显著性水平下基于抽样数据获得的分析结果逼近基于整体数据获得的分析结果^[23-24]。首先, 将这 6 000 多家工厂的用电数据看作一个整体; 其次, 将整体数据按照工厂用电模式的不同划分为四类; 接着, 在每种类别中随机选择 60 个工厂的用电数据作为子样本; 最后, 将这些被选择出来的工厂用电数据合并成一个总样本。经过这样一个抽样过程, 在保证实验数据可靠的前提下, 使得实验数据条数由原来的百万级别变到了万级别, 所占内存从 2 000 MB 降为 20 MB, 大大降低了数据量; 在保证实验结果可靠的前提下, 大大缩短了算法的运行时间, 进而提升算法的性能。在经过以上 5 个步骤对数据进行预处理后,

数据质量得到了很大的提升。

3.2 神经网络模型

对工厂用电状态的识别是一个有监督机器学习算法训练过程, 如果工厂的用电数据作为输入, 工厂用电状态类型作为输出, 那么需要研究、得到的就是输入与输出之间的关系。实验数据具有 96 个条件属性、1 个目标属性, 且目标属性的属性值分为用电状态与非用电状态两种。考虑到神经网络^[25]具有良好的非线性函数拟合能力, 本文采用其对工厂的用电状态进行识别。

由于数据条件属性有 96 个, 目标属性有 2 个属性值, 即“用电状态”和“非用电状态”, 因此神经网络的输入层节点个数为 96, 输出层节点个数为 2。下面通过对比实验来确定神经网络的隐含层个数以及各隐含层节点个数, 实验结果如表 2 所示。对于 9 种不同的网络结构, 实验结果是运行 20 次程序(70% 的数据用作训练数据、30% 的数据用作测试数据)的平均训练精度和平

均测试精度。

当神经网络的隐含层个数较少时,训练得到的模型会欠拟合;而当隐含层个数较多时,训练得到的模型会过拟合。另外,对于确定隐含层个数的神经网络,隐含层节点个数也是影响网络训练的关键:节点太少,预测结果不精确;节点太多,训练过程太缓慢。通过观察表2可以发现,隐含层个数为3、网络结构为96/64/32/16/2的神经网络的训练精度和测试精度都是最好的。

表2 9种不同结构神经网络的实验结果

Table 2 The experimental results of deep neural networks with the different structures

隐含层个数	网络结构	训练精度 (%)	预测精度 (%)
2	96/64/32/2	98.8	95.4
2	96/64/16/2	98.9	95.2
2	96/32/16/2	98.9	95.4
3	96/64/32/16/2	99.6	96.5
3	96/64/16/8/2	99.1	95.9
3	96/32/16/8/2	98.9	94.6
4	96/64/32/16/8/2	99.2	95.4
4	96/64/16/8/4/2	99.0	95.1
4	96/32/16/8/4/2	98.9	94.9

4 用电状态识别结果与分析

在该节中,首先对基于神经网络的用电状态识别模型的精度进行测试,然后与基于其他机器学习方法的识别模型进行对比,最后对实验结果进行简要的分析。

4.1 用电状态识别效果测试

本次实验基于 TensorFlow 框架进行神经网络的编程,网络参数设置如表3所示。关于 TensorFlow 框架下神经网络模型编译和训练的其他参数解释如下。

(1) 损失函数(loss):采用二值交叉熵的形式,其定义如公式5所示。

$$loss = - \sum_{n=1}^N \left(\frac{y_n}{t_n} - \frac{1-y_n}{1-t_n} \right) \quad (5)$$

其中, N 为样本个数; y_n 和 t_n 分别为神经网络对应第 n 个样本的期望输出和实际输出。

表3 神经网络参数设置

Table 3 The parameter setting of deep neural network

网络层	节点数	激活函数	Dropout 层
输入层	96	ReLU	0.2
隐含层 1	64	ReLU	无
隐含层 2	32	ReLU	0.1
隐含层 3	16	ReLU	无
输出层	2	softmax	无

(2) 优化器(optimizer):采用 RMSProp^[26] 自适应学习率方法,让 cache 每次都损失一部分,使得步长不再是单调递减,从而解决了学习率急剧下降的问题,并且让算法收敛速度更快。

(3) 评估标准(metrics):采用分类正确率(accuracy)来评估模型分类效果。

(4) 每次梯度更新的样本数(batch_size):即每次训练在训练集中取 batch_size 个样本训练。一般在合理范围之内,越大的 batch_size 使下降方向越准确,震荡越小。但是,若 batch_size 过大,则会出现局部最优的情况;若 batch_size 过小,则训练过程中引入的随机性较大,难以达到收敛,只有在极少数情况下才能获得满意的训练效果。

(5) 训练网络的迭代次数(epochs):一个 epoch 等于使用训练集中的全部样本训练一次。一般来说,epochs 太小,会使得训练不够充分;epochs 太大,则会使得训练时间很长。图4、表4分别给出了 epochs 为 1~128、1~20 时的神经网络训练结果。

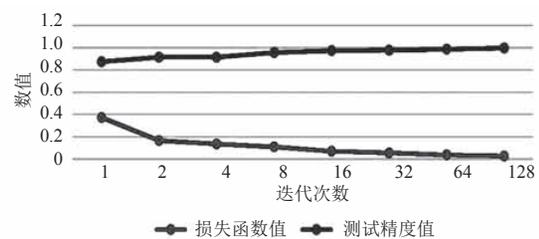


图4 迭代次数对损失函数值和测试精度的影响

Fig. 4 The effect of epochs on loss function values and test accuracy

表 4 不同迭代次数对应的神经网络训练

Table 4 The training of neural network with different

epochs		
迭代次数	损失函数	测试精度
1	0.501 5	0.830 9
2	0.244 2	0.913 8
3	0.176 6	0.913 8
4	0.157 4	0.913 8
5	0.146 7	0.913 8
6	0.138 7	0.913 8
7	0.129 2	0.919 9
8	0.123 0	0.935 6
9	0.113 8	0.951 4
10	0.111 4	0.956 0
11	0.104 5	0.961 2
12	0.099 1	0.966 3
13	0.092 9	0.711 0
14	0.091 3	0.967 6
15	0.086 0	0.970 4
16	0.082 6	0.971 6
17	0.078 9	0.970 0
18	0.079 9	0.972 9
19	0.074 8	0.973 0
20	0.070 8	0.974 1

表 5 迭代次数为 128 时对应的神经网络训练

Table 5 The training of neural network with epochs=128

梯度更新样本数	迭代次数	训练精度 (%)	测试精度 (%)
128	128	98.9	96.1
256	128	98.9	96.3
512	128	99.1	96.5
1 024	128	98.9	96.3

表 6 梯度更新样本数为 512 时对应的神经网络训练

Table 6 The training of neural network with batch_size=512

梯度更新样本数	迭代次数	训练精度 (%)	测试精度 (%)
512	8	99.0	96.4
512	16	98.9	96.3
512	32	98.9	96.4
512	64	99.1	96.4
512	128	98.9	96.5

表 5 和表 6 分别展示了在给定 epochs、batch_size 的情况下, 不同的 batch_size 和 epochs

对神经网络用电状态识别效果的影响。通过分析上述实验结果可以得出如下实验结论:

(1) 通过观察图 4 和表 4 发现, 网络训练时, 随着 epochs 的增大, loss 越来越接近 0, accuracy 越来越接近 1, 并且当 epochs 大于 20 后, 二者都逐渐趋于稳定。这说明网络在经过 20 个 epochs 的训练后就基本上收敛了。

(2) 通过观察表 5-6 发现, 对于不同的 batch_size、epochs 组合, 网络的训练精度和测试精度都高于 95%。当 batch_size 为 512、epochs 为 128 时, 神经网络的测试精度达到最高值 96.5%。

(3) 上述实验结果证实利用网络结构为 96/64/32/16/2 的神经网络对工厂用电状态的测试精度能够达到 95% 以上, 这说明本文提出的电网数据预处理方法是有效的, 并且采用神经网络进行工厂用电状态的识别是可行的。

(4) 通过对电网数据进行基于神经网络的建模, 实现了使用人工智能方法自动识别工厂用电状态的目的, 有助于有关部门和相关单位自动监测工厂的上下班情况以及进一步分析工厂的生产状况。

4.2 对比实验

为了证明本文所使用的神经网络模型的合理性和有效性, 将该方法与其他机器学习方法(包括决策树、支持向量机、线性分类器 LR 和随机梯度下降法)在测试精度方面进行了实验对比, 结果如表 7 所示。表 7 中的机器学习方法是专门面向机器学习的 Python 开源框架 Scikit-learn^[27]

表 7 文中方法测与其他几种机器学习方法测试精度的对比

Table 7 Comparison of prediction accuracy between our method and other machine learning methods	
方法	测试精度 (%)
本文方法	96.5
决策树	95.1
支持向量机	95.0
线性分类器 LR	95.3
随机梯度下降法	95.5



图5 迭代次数为128时对应的神经网络训练

Fig. 5 The training of neural network with epochs=128

中的标准算法，算法中的参数均为 Scikit-learn 的默认值。

表7结果显示，文中所使用的神经网络模型相比于 Scikit-learn 中的机器学习方法效果更好，证实了采用神经网络进行用电状态识别的合理性和有效性。

4.3 结果分析

对于以上实验结果，可以对其进行更深入的研究。例如，通过对用电状态数据进行简单聚类，就可得出用电状态数据包含的用电模式，如图5所示，其中用电状态数据包含6种用电模式。这6种用电模式按照用电量从高到低的顺序为 $2 > 0 > 5 > 4 > 1 > 3$ ，具体对应为高用电的三班倒用电模式、高用电的四班倒用电模式、高用电的两班倒用电模式、中等用电的两班倒用电模式、中等用电的夜班模式和低用电的夜班模式。

对于电网公司，由于电能是不能存储的，当所产电能过多时，会造成能源浪费；而当所产电能不能满足用户需求时，会造成用户损失。所以，电网公司可将工厂生产状况与用电量进行结合，预测客户的用电习惯与需求，为形成合适的激励机制、实施有效的需求侧管理(需求响应)提

供依据，从而实现电能的合理开发与利用。同时，电网公司资产巨大，资产的监测和运维涉及到大量复杂的数据，通过数据分析，可提高资产利用率和设备管理水平，这存在着巨大经济效益。另外，由于大多数企业都呈现出“早-中-晚”三班倒的用电趋势，这会导致高峰期供电不足的情况出现。因此，可采取相应措施将部分企业进行错峰生产，缓解电力供给不平衡的问题。对于用户，由于提前买电会比较便宜，用户可结合自身的用电习惯，推断出未来一段时间内需要购买的用电量，这一政策在有些国家(如德国)已被普遍推广。

5 结束语

本文通过对电网数据进行预处理及使用神经网络进行建模，实现了对工厂用电状态的自动高效识别，从而能够帮助相关部门更加准确高效地分析不同工厂的生产状况，对实际工作进行有针对性的指导。在未来工作中，将针对智能电网数据中的缺省值插补、异常点挖掘以及概率密度函数估计作进一步的探讨和研究。

参 考 文 献

- [1] 彭显刚, 赖家文, 陈奕. 基于聚类分析的客户用电模式智能识别方法 [J]. 电力系统保护与控制, 2014, 42(19): 68-73.
- [2] 徐良军, 张笑第, 王立军. 基于聚类分析的用户分类和用电行为分析 [J]. 山西电力, 2016, 4: 23-27.
- [3] 沈玉玲, 吕燕, 陈瑞峰. 基于大数据技术的电力用户行为分析及应用现状 [J]. 电气自动化, 2016, 38(3): 50-52.
- [4] 陈敬德, 盛戈皞, 吴继健, 等. 大数据技术在智能电网中的应用现状及展望 [J]. 高压电器, 2018, 54(1): 35-43.
- [5] Zhou K, Yang C, Shen J. Discovering residential electricity consumption patterns through smart-meter data mining: a case study from China [J]. Utilities Policy, 2017, 44: 73-84.
- [6] Singh S, Yassine A, Shirmohammadi A. Incremental mining of frequent power consumption patterns from smart meters big data [C] // Proceedings of the 2016 IEEE Electrical Power and Energy Conference, 2016: 1-6.
- [7] Antoch J, Prchal L, Rosa MRD. Functional linear regression with functional response: application to prediction of electricity consumption [M] // Functional and Operatorial Statistics. Physica-Verlag HD, 2008: 23-29.
- [8] Kaur H, Ahuja S. Time series analysis and prediction of electricity consumption of health care institution using ARIMA model [C] // Proceedings of the Sixth International Conference on Soft Computing for Problem Solving, 2017: 347-358.
- [9] Ahmad A, Anderson TN, Rehman SU. Prediction of electricity consumption for residential houses in New Zealand [C] // Proceedings of International Conference on Smart Grid Inspired Future Technologies, 2018: 165-172.
- [10] 李文彬, 张春梅. 多算法融合的电网用电量预测系统研究和实现 [J]. 现代计算机(专业版), 2017, 22: 77-80.
- [11] 曲朝阳, 张率, 刘洪涛. 基于用电影响因素回归的小区用电预测模型 [J]. 东北电力大学学报, 2015, 35(1): 73-77.
- [12] Wijaya TK, Vasirani M, Humeau S, et al. Cluster-based aggregate forecasting for residential electricity demand using smart meter data [C] // Proceedings of the 2015 IEEE International Conference on Big Data, 2015: 879-887.
- [13] 张为金. 基于机器学习的电力异常数据检测 [D]. 成都: 电子科技大学, 2018.
- [14] 刘凤魁. 基于密度峰值聚类算法的电力大数据异常值检测及用电行为分析研究 [D]. 北京: 中国电力科学研究院, 2017.
- [15] Jakkula V, Cook D. Outlier detection in smart environment structured power datasets [C] // Proceedings of the 2010 Sixth International Conference on Intelligent Environments, 2010: 29-33.
- [16] 刘凤魁, 邓春宇, 王晓蓉, 等. 基于改进快速密度峰值聚类算法的电力大数据异常值检测 [J]. 电力信息与通信技术, 2017, 15(6): 36-41.
- [17] McCulloch WS, Pitts WH. A logical calculus of ideas imminent in nervous activity [J]. The Bulletin of Mathematical Biophysics, 1942, 5: 115-133.
- [18] Han J, Moraga C. The influence of the sigmoid function parameters on the speed of backpropagation learning [C] // International Workshop on Artificial Neural Networks: from Natural to Artificial Neural Computation, 1995: 195-201.
- [19] Hahnloser RHR, Seung HS. Permitted and forbidden sets in symmetric threshold-linear networks [C] // Advances in Neural Information Processing Systems, 2001: 217-223.
- [20] 蒋昂波, 王维维. ReLU 激活函数优化研究 [J]. 传感器与微系统, 2018, 37(2): 50.
- [21] Bishop CM. Pattern Recognition and Machine Learning [M]. New York: Springer-Verlag, 2006.
- [22] Kohonen T. Self-organization and associative memory: 3rd edition [J]. Applied Optics, 1988, 8(1): 3406-3409.
- [23] Salman S, Huang JZ, He YL. Random sample partition: a distributed data model for big data analysis [J]. IEEE Transactions on Industrial Informatics, 2019, doi: 10.1109/TII.2019.2912723.
- [24] Salman S, Huang JZ, He YL, et al. An asymptotic ensemble learning framework for big data analysis [J]. IEEE Access, 2019, 7: 3675-3693.
- [25] Rivals I, Personnaz L. MLPs (mono-layer polynomials and multi-layer perceptrons) for nonlinear modeling [J]. Journal of Machine Learning Research, 2003, 3: 1383-1398.
- [26] Tieleman T, Hinton G. Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude [J]. COURSERA: Neural Networks for Machine Learning, 2012, 4(2): 26-31.
- [27] BSD License. Scikit-learn: machine learning in Python [EB/OL]. [2019-04-12]. https://scikit-learn.org/stable/supervised_learning.html#supervised-learning.