

基于LSTM的智能家庭用电预测模型研究

周 游, 徐 丹, 赵 灿, 谭宇渲

(国网江苏省电力有限公司苏州供电公司, 江苏 苏州 215004)

✉542790714@qq.com; xudan19870725@126.com; 342677492@qq.com; 852901313@qq.com



摘 要: 家庭用电是能源市场的一个重要组成部分, 预测家庭用电需求能够实现智能供电, 可以有效地提高供给率, 但目前预测方法大多效果不佳。针对此, 提出了一种基于LSTM的面向家庭智能用电预测算法, 建立了端到端的智能家庭用电预测模型。其在Boruta特征筛选的基础上设计了特征选择方法, 对多个特征进行重要性计算, 选取其中重要性高的部分进行建模, 然后利用LSTM网络与全连接层对时间序列数据进行训练, 得到预测模型。实验结果表明, 所提方法的预测效果明显优于其他三种模型, 能与真实数据较好地拟合。

关键词: 智慧能源; 用电预测; 特征选择; 时间序列; LSTM网络

中图分类号: TP399 **文献标识码:** A

Research on Smart Power Consumption Prediction Model of Smart Home based on LSTM

ZHOU You, XU Dan, ZHAO Can, TAN Yuxuan

(Suzhou Power Supply Branch, State Grid Jiangsu Electric Power Limited Company, Suzhou 215004, China)

✉542790714@qq.com; xudan19870725@126.com; 342677492@qq.com; 852901313@qq.com

Abstract: Household power consumption is an important part of the energy market. Predicting household power demand can effectively improve energy supply efficiency, but most of the current prediction methods are not effective. To address the problem, this paper proposes an LSTM (Long Short-term Memory)-based power consumption prediction algorithm for smart home, where an end-to-end smart home power prediction model is established based on LSTM. A feature selection method that utilizes Boruta feature screening is also designed. The importance of multiple features is calculated, the most important part of those features is selected for modeling, and then LSTM network and the full connection layer are used to train the time series data to obtain the prediction model. The experimental results show that the prediction effect of the proposed method is significantly better than the other three models, and it can fit the real data well.

Keywords: smart energy; power consumption prediction; feature selection; time series; LSTM network

1 引言(Introduction)

近年来, 能源产业正迎来一个新的发展, 智慧能源与能源互联网等概念越来越得到重视^[1]。智慧能源就是结合信息技术、人工智能技术、大数据技术的新能源模型形态, 是能源互联网的基础架构^[2-4]。作为能源市场的重要组成部分, 家庭用电的稳定关系着民生问题。对用电量进行全面的理解有助

于减少家庭的电费支出, 也有利于进行能源合理分配^[5]。

物联网与人工智能等技术的兴起, 以及家庭传感器的广泛采用^[6], 积累了大量的时间序列特征数据, 可以用来对家庭用电模式进行分析。长短时记忆网络(Long Short Term Memory, LSTM)在处理时间序列预测问题上有明显的优势^[7], 可以获得较佳的预测效果。

2 相关工作(Related work)

进行家庭用电预测常用的预测方法主要包括随机森林(Random Forest, RF)^[8]、极限树(ExtraTrees, ET)^[9]、差分自回归移动平均模型(Auto Regressive Integrated Moving Average Model, ARIMA)^[10]等。

2.1 常用预测方法

随机森林算法先使用 T 个弱分类器分别对 T 个由随机采样而来的训练集进行训练,然后对多个弱分类器进行组合,最后由投票或取均值得出最终结果。大量实验表明,随机森林算法通过这种方式使得模型整体的泛化能力及准确度明显优于其他Bagging算法^[11]。

极限树算法也称为极限森林,意为极其随机的森林。在随机森林算法中,为了对需要划分的节点进行选择,需要在特征子集上随机进行寻找,以找出最优划分特征^[12]。极限树算法直接使用随机的特征和阈值来进行划分,这样可以得到形状更随机、差异更大的决策树,可见极限树算法比随机森林更激进^[13]。

ARIMA是一种基于平稳时间序列来进行预测的算法,用于电力序列模型预测^[14],可以使用三元组 (p, d, q) 来建模。在模型构建过程中,主要根据PACF图和ACF图来分别确定 p 和 q 的取值,其中的 d 代表使序列数据成为平稳时间序列所做的差分次数。

2.2 长短时记忆网络

LSTM是一种特殊的循环神经网络(Recurrent Neural Network, RNN)结构,用来处理RNN面临的一个无法解决的长期依赖问题。在标准RNN中,内部是一种重复神经网络模块的简单链式结构。由于这种结构过于简单,因此在经过 t 个时刻之后会导致0时刻的信息几乎被遗忘。

整体上看,LSTM与其结构相同,但LSTM中重复的模块结构与RNN不同。对比可发现,RNN中的神经网络层只有单一的一个,而LSTM中有四个,并且以一种非常特殊的方式进行交互^[15]。通过这种复杂的结构,可以使以前时刻的重要信息得以保存,避免发生遗忘。

3 数据处理(Data processing)

3.1 数据描述

本文实验采用美国某家庭从2016年1月11日17时到2016年5月27日18时之间,每隔10分钟所记录下的共19,736条数据作为训练样本。每个样本由时间(date)、气温(T)、湿度(RH)、天气(weather)、光线(lights)、随机变量(rv1、rv2)等共28组特征数据所组成。其中,气温和湿度由该家庭所安装的9个温度传感器所获取;天气由室外温度(T_out)、露点(Tdewpoint)、室外湿度(RH_out)、大气压(Press_mm_hg)、风速(Windspeed)、可见度(Visibility)组成。本文

希望通过这28组数据组成的特征来实现对某家庭用电量的准确预测。某家庭各时刻用电量详情如图1所示。

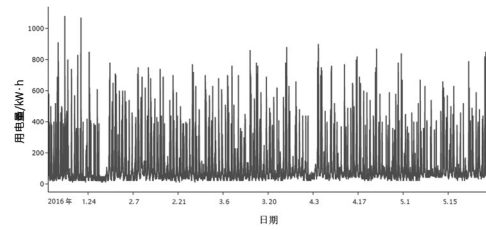


图1 某家庭用电量模式图

Fig.1 Power consumption pattern diagram of a certain household

3.2 特征筛选

除日期和需要预测的用电量之外,采集到的数据共包括27个特征数据。对于实际的模型构建来说,数据集的变量太多,而大多数变量与目标问题无关。当数据集的特征过多时会有缺点:首先,特征过多会占用过多资源,导致算法速度慢,使用起来很不方便;其次,当变量的数量显著高于最优时,会导致很多机器学习算法的准确率下降,降低模型的性能。

针对上述问题,本文采用Boruta算法来对特征进行筛选,通过对各个无偏的弱分类器—决策树的投票来进行分类,对各个特征的重要性给出数值估计,并使用Z-Score来计算特征重要度,从而选择出对因变量影响较为重要的特征^[14]。

$$Z_Score = avg_loss / s(\sigma)$$

式中, avg_loss 为平均损失, $s(\sigma)$ 为标准差。

经过100轮迭代,Boruta特征筛选模型成功地从28个特征中筛选出21个重要特征,分别为T1、T2、T3、T4、T5、T7、T8、RH_1、RH_2、RH_3、RH_4、RH_5、RH_6、RH_7、RH_8、RH_9、T_out、Tdewpoint、RH_out、Press_mm_hg、Windspeed。所选取的21个特征重要度排名如图2所示。

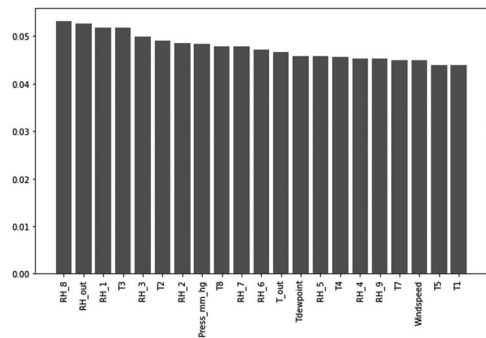


图2 特征重要度排名

Fig.2 Feature importance ranking

4 预测算法(Prediction algorithm)

4.1 处理流程

为了建立面向智能家庭用电预测模型,在筛选完特征

后, 还需将样本数据序列化, 以满足LSTM模型的训练要求。预测模型建立的主要步骤包括采集数据、选取特征、特征筛选、样本数据序列化、特征数据归一化、LSTM数据输入、模型参数调优、用电预测建模、训练并保存模型等, 具体流程如图3所示。

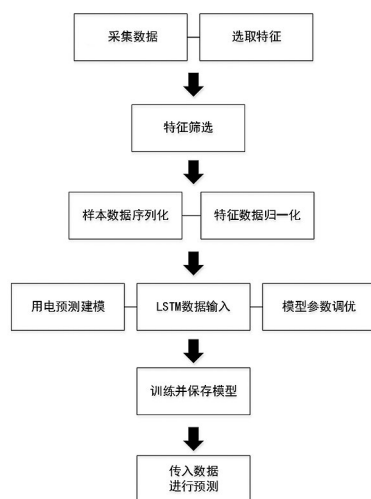


图3 面向家庭的智能用电预测流程图

Fig.3 Flow chart of home-oriented smart power consumption prediction

4.2 算法

基于LSTM的面向家庭智能用电预测算法如下。

算法1 基于LSTM的面向家庭智能用电预测算法

输入: 数据集I, 特征矩阵X

输出: 预测模型

```

1: Boruta_selector(X);
2: Repeat:
3:   shadow_features=shuffle(X)
4:   new_features=shadow_features+real_features
5:   Z_Score_real=score(real_features)
6:   Z_Score_shadow=score(shadow_features)
7:   Zmax=max( Z_Score_shadow)
8:   if Z_Score_real > Zmax then
       选择该特征
9: series_to_supervised()
10: StandardScaler()
11: for k=1 to 70 do
12:   训练得到LSTM模型
13: Return 预测模型
  
```

5 实验及分析(Experiment and analysis)

5.1 模型训练

训练前, 将样本数据集按7:3的比例划分为训练集和测试集。模型训练过程中训练集损失值和测试集损失值的变化

过程如图4所示。从图4可以看出在训练过程中, 训练集损失值和测试集损失值不断下降, 最终收敛于一个较低值, 表明本文实验模型训练效果良好。

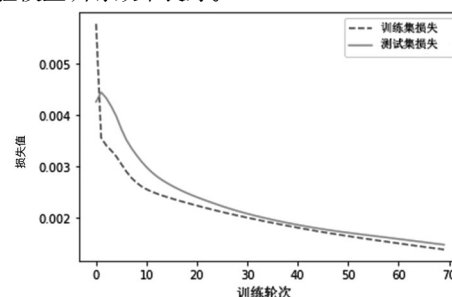


图4 训练过程损失曲线

Fig.4 Loss curve in the training process

5.2 模型训练结果分析

训练结束后, 在样本数据集中进行随机采样, 获取6,000个样本数据。将随机选取的数据传入最终预测模型, 观察其预测结果的准确度。本文所提出模型的预测结果与实际数据拟合情况如图5所示。从图5可发现, 本文所提出的预测模型可对真实值进行较好地拟合。

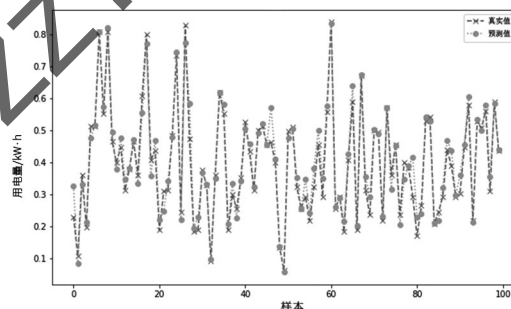


图5 本文所提出模型的预测结果与实际数据拟合情况

Fig.5 Fitting of the proposed model prediction results with the actual data

5.3 对比试验分析

为进一步对本文所提出的预测模型进行评估, 另外选取了当前常用的随机森林算法、极限树算法、ARIMA算法对同一样本数据进行实验, 并以R2_SCORE(R方值: 取值范围为[0, 1], 越接近1表示预测效果越好)和RMSE(均方根误差)的值(越小表示效果越好)作为算法的评估指标。四种方法的评估结果如表1所示。

表1 各模型评估结果

Tab.1 Evaluation results of each model

方法	Test_R2_Score	Train_R2_Score	RMSE
随机森林	0.829	0.557	0.665
极限树	0.629	0.890	0.609
ARIMA	0.629	0.240	1.864
本文方法	0.976	0.978	0.026

(下转第38页)