

一种基于迁移学习的非平稳电力运检成本预测数学方法

潘 军¹, 陈 倩¹, 金绍君^{2*}

¹国网浙江省电力有限公司金华供电公司, 浙江 金华

²国网浙江省电力有限公司, 浙江 杭州

Email: *120706793@qq.com

收稿日期: 2020年12月16日; 录用日期: 2021年1月5日; 发布日期: 2021年1月20日

摘 要

电力行业是国家发展的重要能源产业,也是国民经济的第一基础产业。随着电网规模不断扩大,运行条件日益复杂,电力数据采集范围和频率不断增加,由于电力数据具有数据样本大、类型多、价值密度低等特征,如何合理运用电力大数据,高效快速挖掘有价值的信息,提高电力数据利用率,为电网运行的可靠性提供理论依据,满足实际需求,成为了一个新的研究热点。针对电力大数据的特点,本文利用时间序列、支持向量回归等人工智能方法,通过深度迁移学习,为标准成本预测任务建立数据挖掘网络模型,提取数据的关联性特征,提高数据预测的精度和效率。实验结果表明,本文模型在小样本数据集上得到较好的预测结果,验证了深度迁移模型的可行性,相比作业成本法、传统预测方法,本文方法平均绝对误差降低10%,具有有效性与优越性。

关键词

迁移学习, 循环神经网络, 门控循环单元, 非平稳, 运检, 电网

A Mathematical Method for Operation and Maintenance Cost Prediction Based on Transfer Learning under Non-Stationary Power Data

Jun Pan¹, Qian Chen¹, Shaojun Jin^{2*}

¹Jinhua Power Supply Company, State Grid Zhejiang Electric Power Co., Ltd., Jinhua Zhejiang

²State Grid Zhejiang Electric Power Co., Ltd., Hangzhou Zhejiang

*通讯作者。

Email: *120706793@qq.com

Received: Dec. 16th, 2020; accepted: Jan. 5th, 2021; published: Jan. 20th, 2021

Abstract

The electric power enterprise is an important basic energy industry for national development, and it is also the first basic industry of the national economy. With the continuous expansion of State Grid, the progressively complex operating conditions, and the increasing scope and frequency of data collection, how to make reasonable use of electrical big data, improve utilization, and provide a theoretical basis for the reliability of State Grid operation, has become a new research hot spot. Since electrical data has the characteristics of large volume, multiple types, low value density, and fast processing speed, it is a challenge to mine and analyze it deeply, extract valuable information efficiently, and serve for actual problem. According to the features of these data, this paper uses artificial intelligence methods such as time series and support vector regression to establish a data mining network model for standard cost prediction through transfer learning. The experimental results show that the model in this paper obtains better prediction results on a small sample data set, which verifies the feasibility of the deep transfer model. Compared with activities based costing and the traditional prediction method, the average absolute error of the proposed method is reduced by 10%, which is effective and superior.

Keywords

Transfer Learning, Recurrent Neural Networks, Gate Recurrent Units, Non-Stationary, Operation and Maintenance, Smart Grid

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着计算机技术的快速发展,基于人工智能大数据分析技术愈发成熟,数据的价值越来越受到重视,籍此开启了一次重大的时代转型。我国电网规模庞大、分布地域广阔、运行条件繁琐,其产生的数据满足大数据的特征。通过当下机器学习方法,挖掘出电力数据背后隐藏的信息,对于电网和民生具有重要的研究意义。

本文研究课题是基于浙江省金华供电公司 2014~2019 年历史标准成本数据,通过历年数据的挖掘分析,掌握浙江公司的成本规律,确定每一项成本产生的动因和其影响因素,建立完善成本动因与成本之间的函数模型并预测 2021 年各模块成本下达。该数据以年为单位的数据,不涉及季度、月份数据,且存在部分地区数据难获取和丢失问题,难点在于样本关联性低,需要非线性模型探究深层特征,而非简单数学模型。针对电力数据非常规约束问题,网络训练难以收敛,效果不理想,引入迁移学习,面向跨领域非平稳数据的预测任务,并基于时间序列、支持向量回归的相关理论研究,本文提出一种基于迁移学习的非平稳电力运营检修成本预测模型。通过迁移学习使得神经网络的模型具有复用性,小样本数据集可以利用大样本训练所得的特征,这些特征相较于直接训练于小样本有着更高的区分度及鲁棒性,极大

程度地解决了由于样本数量过少导致的过拟合问题，经过实验验证，本文方法在小样本训练上得到较好的预测结果。

论文安排如下：第 1 节我们首先简要回顾了之前关于迁移学习一些工作；在第 2 节中，我们阐述了两种算法：长短期记忆网络和支持向量回归，然后提出模型体系结构。数值实验和结果在第 3 节中进行。最后，我们的工作总结在第 4 节中展示。

2. 相关工作

机器学习(Machine Learning) [1]通过训练让机器模拟人脑学习，是人工智能的技术基石。传统方法往往基于训练集与测试集满足相同数据分布的假设，但是在一些领域此假设无法满足。“大”和“小”两者是矛盾统一的，人们希望既能分析大数据，也能挖掘小数据。在机器学习发展过程中，不再局限于大样本、有监督的学习，而更希望能够通过某些方法完成无监督、小样本学习。如何挖掘分析非平稳下的大数据是现代机器学习的前沿方向之一。

2.1. 迁移学习

人类不仅能在大样本上学习，也可以借助与所学任务相近的信息，辅助在小样本或零样本情况下完成学习。人类具有知识迁移的能力，日常生活所说的举一反三、触类旁通都属于迁移。机器正是欠缺这种迁移能力，为此，将大数据“迁移”到小数据，用于解决小样本中数据少、知识少等问题，研究学者提出迁移学习(Transfer Learning) [2] [3] [4]。迁移学习尝试实现人的知识迁移能力，放宽了上述数据分布假设，该理论研究始于 20 世纪 90 年代，直到 Pan 和 Yang [2]对相关文献进行了整理，发表了一篇代表性综述，给出迁移学习的具体定义，基本原理如图 1 所示。

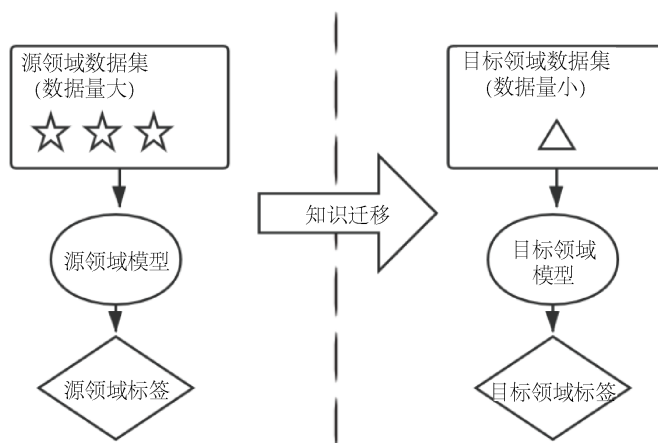


Figure 1. Demonstration for transfer learning
图 1. 迁移学习基本示例

2.2. 迁移学习的分类

迁移学习还可以根据“迁移什么”分成四类。第一种是实例迁移学习(Instance-based Transfer Learning)经过样本评估，并从源域中挑出部分匹配的样本迁移到目标域，用于完成任务[5] [6] [7]。Wu 等[8]根据支持向量机，增强数据训练，提出的 LP-SVM 方法；Dai 等[9]根据样本的相似性，提出 TrAdaBoost 算法；Quanz 等[10]根据样本均值差，提出了 LMPROJ 算法；Xu 等[11] [12]提出了多视角 Adaboost 算法；Chen 等[13]提出了多源学习的策略；Jiang 等[14]提出了知识模糊系统的算法。第二种是特征的迁移学习(Feature

Representation Transfer Learning)通过潜在的共同特征空间,完成迁移学习[15] [16]。Argyriou 等[17]提出正则化多任务学习结构; Pan 等[18] [19]提出基于流形结构的 MMDE 算法; Tu 等[20]提出域适应算法; Gao [21] [22]等提出核空间模型用于行人检测问题。模型迁移学习(Model based Transfer Learning)和关系迁移学习(Relation based Transfer Learning)受关注较少。模型迁移学习假设两个域共享算法的参数或者某些先验知识,由此完成知识的迁移[23] [24] [25]。关系迁移学习假设两个域之间知识具有相关性,例如将人们对书籍、音乐的评价与电影喜好联系起来,利用这些共性信息来完成知识的迁移[26]。

3. 基于迁移学习的非平稳电力成本预测

针对电力系统数据缺少问题,利用模型迁移学习,将知识从源域迁移到目标域,利用源域网络参数在目标域建立模型。由于电力数据呈现的时间特征,因此采用时间序列方法,同时也选取支持向量回归作为预测算法。本节阐述两种算法的基本原理,并结合迁移学习,进而建立小样本数据集下的混合预测模型,并给出模型预测效果和评价方法。预测框架如表 1:

Table 1. Transfer learning prediction framework
表 1. 基于迁移学习预测框架

基于迁移学习预测框架

第一步:在源域和目标域的数据预处理

第二步:在源域中模型预训练

第三步:模型调整

第四步:训练后的模型参数迁移

第五步:在目标域中的模型再生训练

第六步:微调

第七步:目标域模型

基于 GRU 预测模型

传统人工神经网络的基础上,RNN [27]引入了定向循环,使得隐藏节点定向连接成环,这样的内部结构有助于网络对时间信息的传递,展示动态时序行为,从而挖掘时间序列的特征,最后做出分类或预测。然而,在处理长时间序列时,会产生梯度消失问题,即随着信息在时间维度上的传递,后面的神经元对较前面的神经元的感知力会下降,信息会逐渐丢失。针对这个问题,在隐藏层中加入细胞单元和三个控制门,这个结构称为 LSTM [28] [29],在 1997 年被提出,并且逐步被改良为现在经典的版本,它适合用于处理长时间序列问题。LSTM 作为一种特殊的循环神经网络,在隐藏层增加了一个细胞状态,通过引入四个逻辑门来控制记忆的更新、遗忘和输出,从而有效地避免了梯度消失和爆炸。

在 LSTM 细胞解 t 总细胞状态 C_t 表示传输到 t 的长时记忆,而隐藏状态 h_t 表示短时记忆。 C_t 和 h_t 将在四个不同的门的控制下忘记一些旧的记忆并保存新的记忆。具体地说, f 是遗忘门,它控制 C_t 将被遗忘多少个长期记忆, i 是输入门,它决定是否保留新来的数据 x_t , g 对 x_t 的保存有影响,最后一个输出门 o 决定输出的大小。参数更新的精确公式为(1)~(3),其中 W 、 b 和 σ 分别是参数矩阵、偏差项和 sigmoid 函数。LSTM 单元的结构如图 2 所示。

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix} + b \quad (1)$$

$$c_t = f \cdot c_{t-1} + i \cdot g \tag{2}$$

$$h_t = o \cdot \tanh(c_t) \tag{3}$$

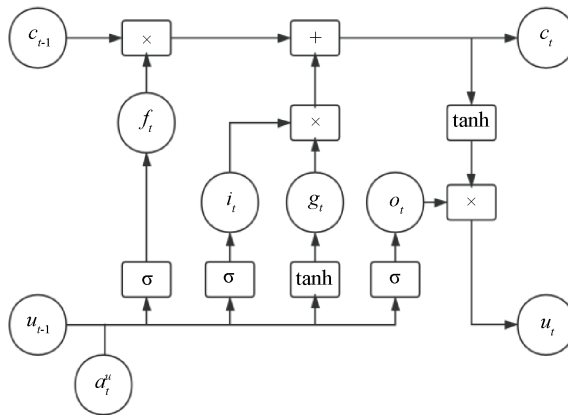


Figure 2. Network structure of LSTM
图 2. 长短时间记忆网络结构

但同时加入更多的逻辑控制导致了参数数量的增加,致使网络的时间复杂度和空间复杂度都有所提高,导致网络的训练时间变长,需要更大的训练样本并且对设备的算力和空间资源的需求也变得更大。

在 2014 年, Kyunghyun Cho [30]等人提出了 GRU (Gated Recurrent Unit)循环神经网络模型。作为一种 LSTM 神经网络的改良,其在保留其指导思想的同时,巧妙地使用一个逻辑门来同时控制遗忘和选择记忆,从而将原先总共四个逻辑门减少到两个。这种设计有效地减少了参数数量,减少了训练时间和资源需求,让整个模型更加轻量化。

图 3 刻画了 GRU 单元的内部结构。在 GRU 网络中, h_{t-1} 为上一个节点传输进来的状态, x_t 为当前节点的输入状态, \tilde{h}_t 为候选隐藏层(candidate hidden layer),表示当前时刻的新信息。GRU 单元会使用两个逻辑门对输入的状态进行处理。其中 r_t 为重置门(reset gate),用来控制需要保留多少的之前的记忆; z_t 为更新门(update gate),用来控制需要遗忘多少从前一节点输入的信息 h_{t-1} 。 h_t 为当前输出到下一节点的信息。具体的计算公式为(4)-(8):

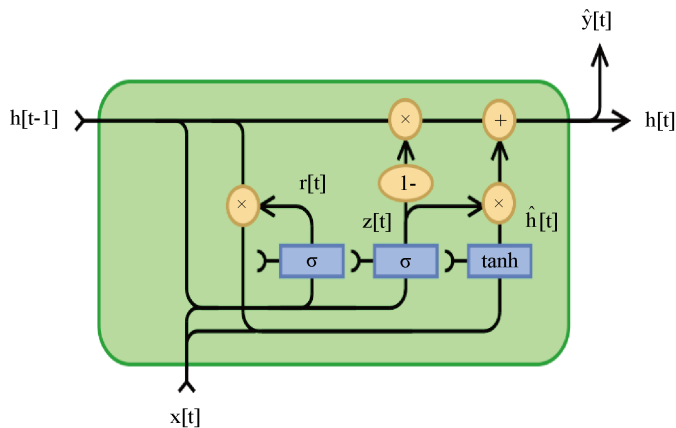


Figure 3. Network structure of GRU
图 3. Gated Recurrent Unit 结构

$$r_t = \sigma(W^r x_t + U^r h_{t-1}) \quad (4)$$

$$z_t = \sigma(W^z x_t + U^z h_{t-1}) \quad (5)$$

$$\tilde{h}_t = \tanh(W_h x_t + r^t h_{t-1}) \quad (6)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (7)$$

$$\hat{y}_t = \sigma(W^o h_t) \quad (8)$$

其中 W^r 和 U^r 为 r_t 中的待训练参数, W^z 和 U^z 为 z_t 中的待训练参数, W_h 为 \tilde{h}_t 中的待训练参数, W^o 为网络输出层 \hat{y}_t 的待训练参数。 σ 为 sigmoid 函数。

考虑到 GRU 在实现长期时间序列预测方面的显著优势, 我们采用 GRU 模型预测成本的波动。GRU 模型的训练过程如图 4 所示。

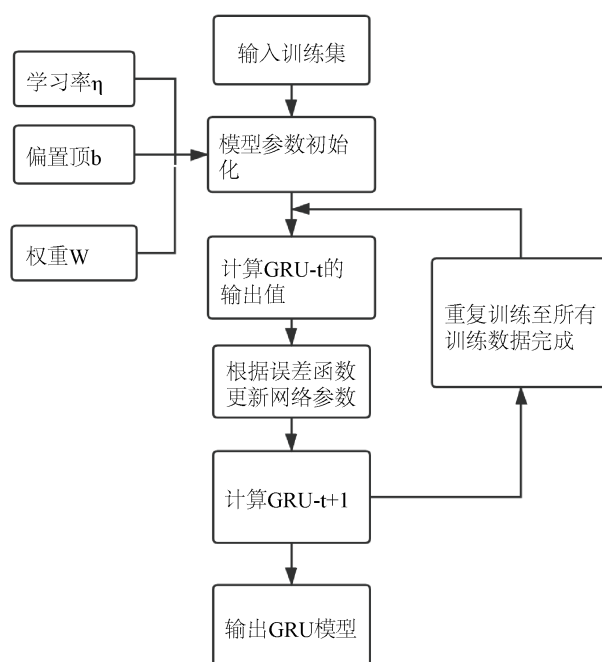


Figure 4. Training process of GRU
图 4. Gated Recurrent Unit 训练过程

整个训练过程可以归结为一个最优化求极小问题。将神经网络视为算子 f , 令网络的训练集为 (x, y) , 且其待训练参数为 W 。若选取损失函数为 L , 则有最小值问题

$$\min_W L(f(x), y) \quad (9)$$

选取适当的加速算法, 如随机梯度下降(Stochastic gradient descent), 即可求解该问题。

在训练过程中, 根据时间窗的大小, 从训练集中提取所需的数据, 得到输出值。利用损失函数计算网络的损耗值和权重梯度, 优化网络参数。重复上述计算和更新步骤, 直到达到指定的结果。

4. 实验结果与分析

本文的 GRU 模型是基 Tensorflow 深度学习框架, 训练时使用 GPU 加速。实验配置: 实验在 Ubuntu

18.04 系统下进行, 主机 CPU 为 Intel i7-7829HK, 显卡为 GTX1080, 内存为 32 GB。

4.1. 数据

实验采用的数据来源于中国浙江省金华供电局, 数据时间范围从 2014 年至 2019 年, 包含营销运检标准成本、主配网运检标准成本, 主要数据类型有电能计量、用电营业、智能用电、市场能效、供电服务, 10 kV 配电架空线路、10 kV 配电电缆线路、10 kV 配电变压器、低压台区、20 kV 配电架空线路、20 kV 配电电缆线路、20 kV 变电站等。对比各类数据集之后, 最终迁移学习源域数据采用金融领域中的某支股票价格数据, 包括 1990 年 12 月~2015 年 12 月 25 年间每天的开盘价, 收盘价以及当日最高最低价格。以股票集为训练集, 电网 2014~2018 年数据作为验证集, 2019 年为预测集。以营销运检标准成本为例, 对提出的预测模型的有效性进行评估。

选择股票数据作为迁移源域的主要原因有以下两点。它和我们要处理的电力成本营销数据有许多在金融时间序列中经常出现的杠杆效应和异方差性。杠杆效应, 是指当某一变量发生较小幅度变化时, 另一相关变量较大幅度变动。在股票等金融市场里, 它更多指信息对变差的影响的不对称性, 通常为好的消息总是没有坏的消息对市场的影响大。当股票价格下跌时, 公司的净股东权益会下降但负债的情况却没有任何变化。所以, 公司的债务与权益比率会变大并且公司会有更高的杠杆率(负债/所有者权益)。而更高的杠杆会致使更进一步的公司的股票价格下降。

异方差是一个统计学概念, 它的定义是一个随机变量的随机误差的变差不是不变的, 其中这个变差可以是方差或是其他关于离散程度的度量。在金融时间序列中, 如股票指数或某种金融产品的投资回报率, 它往往体现为这些数值不与时间独立。股票指数是同方差的, 即与时间独立, 那么我们就可以拿历史上任意一个时间的数据来预测指数未来的值, 但实际情况并非如此。某支股票未来的走势, 与最近一段时间的数值相关, 但更多受该企业的业绩, 利好消息和相关政策所影响, 这些往往是脱离数据本身的。而此时, 方差是随时间改变而改变, 甚至完全是随机不可控的。我们要处理的营销成本数据每年的变动很大程度上取决于政策的改变, 其分配方式也是随时间改变而改变, 故它有异方差性。

4.2. 训练与模型结构

1) 数据预处理, 首先将股票数据每六天分为一组, 其中将前五天的收盘价作为 x , 第六天的值作为 y , 取最后 20 组数据作为测试集, 然后将剩余数据归一化以后以 9:1 的比例分为训练集和验证集。然后分别归一化, 将股票数据测试集和电力成本数据以相同的均值和方差进行归一化。归一化的目的是便于计算, 同时防止梯度爆炸和消失的现象, 保证学习过程中排除所有训练中的信息, 以此来保证模型的真实性和可靠性。

2) 模型结构, GRU 模型模型一共四层: 输入, GRU1, GRU2 和输出(一维的全连接层), 如图 5 所示。其中, 设置 time step 为 5, 待优化的损失函数为均方误差, 优化算法为 Adam, 并采用自衰减学习率, 即当验证集的损失持续 20 个 epochs 没有减少时, 会按照一定比例减小学习率。

3) 训练, 利用预处理训练集对 GRU 模型进行训练, 并且保存所有参数。然后, 将成本数据反馈给微调网络。

4.3. 预测结果

利用股票数据预测结果见图 6 和图 7。为了更直观地体现预测模型性能, 通过计算 MAPE

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_i^*}{y_i} \right| \quad (10)$$

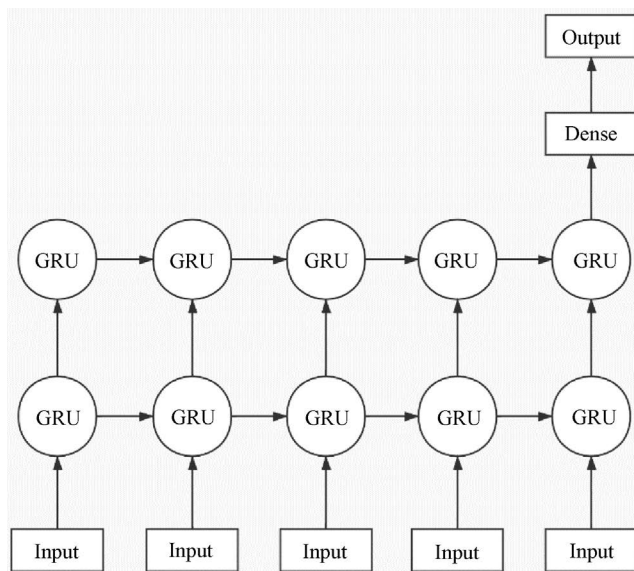


Figure 5. GRU Training network structure

图 5. GRU 网络结构

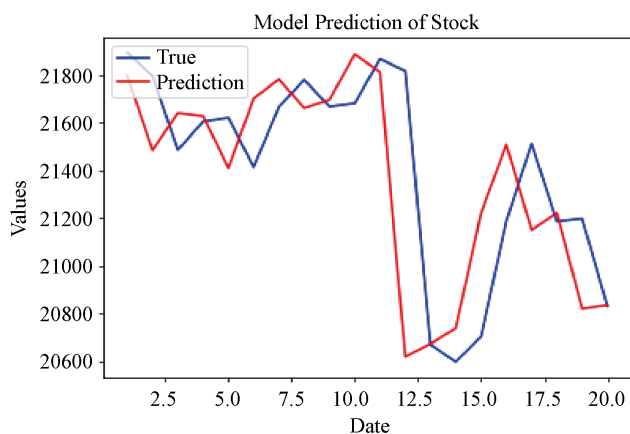


Figure 6. Source domain-stock GRU forecast results

图 6. 源域 - 股票 GRU 对股票数据的预测结果

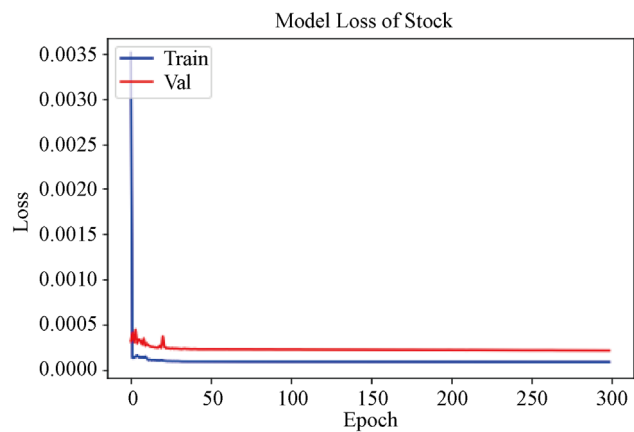


Figure 7. Stock loss function

图 7. 股票损失函数

其中 y_i 为实际值为预测值, n 为时间序列长度)来评估各种方法的性能。将作业成本法(Activity based Costing, ABC)、线性回归(Linear Regression, LR)与本文方法进行对比, 其结果见表 2。通过指标值对比可以看出, 提出的方法在各项指标中都是比较好的, 极大降低了以往凭作业成本法下达的预算误差, 本文提出的基于迁移学习的非常规电力成本预测模型为未来电力公司的财务预算规划提供了一定的参考依据。

Table 2. Comparison of algorithm prediction results

表 2. 算法预测结果比较

	Learning Rate	LSTM 单元	Reduced Factor	MAPE (%)	TotalMAPE (%)
ABC				39.52	38
LR				32.83	31.96
GRU	0.001	128	0.2	22.88	23.64
GRU	0.00088	64	0.2	24.44	18.47
GRU	0.00092	64	0.2	23.86	19.59

4.4. 实验分析

按照国网总部“四个典型”(典型资产、典型设备、典型项目、典型作业)的测算逻辑, 浙江省各地区的实际情况进行作业化, 内容细致化, 其类型数量较国网多, 浙江测算结果普遍高于国网测算结果。同时, 每年国网标准均有变化, 且随着电力系统运行过程中新技术新业务的出现, 原有的定额中一部分已不适用, 且新业务新作业的成本标准定额尚未确定, 加之政策性变化, 行业大环境等原因致使浙网电力成本波动比较大。因此成本难以预测, 以往电网凭借作业成本法下达新一年的测算, 不仅误差大, 而且缺乏理论依据。本文在传统预测方法基础上, 尝试采用神经网络完成预测任务, 虽然结果仍有待进一步提升, 但文章思路为以后研究指明了一定方向。

5. 结语

本文基于非平稳电力数据, 提出一种迁移学习的神经网络预测模型, 对源域特征空间进行有效的学习, 并将其迁移到电力数据预测模型中, 通过微调, 显著提高了该模型的预测准确性。实验结果验证了基于迁移学习的非常规约束下的电力成本预测的有效性与优越性。本文是基于迁移学习对电力成本预测的初探, 那么今后如何根据电力系统实际业务和政策因素来优化和制定预测模型, 充分全面地考虑上述各项参数, 是我们未来的研究方向。对于迁移学习, 要加强算法、有效性、可迁移性研究, 同时避免负迁移。更重要的是在当下要结合实际应用场景, 如智能电网, 以顺应大数据时代的新浪潮。

致 谢

本文由国网浙江省电力有限公司资助。

基金项目

国网浙江省电力有限公司科技项目资助, 项目名称: 基于机器学习的标准成本作业化改造研究与应用(项目编号: 5211JH1900LZ)。

参考文献

- [1] Mitchell, T.M. (2003) Machine Learning. McGraw-Hill, New York.
- [2] Pan, S.J. and Yang, Q. (2010) A Survey on Transfer Learning. *IEEE Transactions on Knowledge & Data Engineering*,

- 22, 1345-1359. <https://doi.org/10.1109/TKDE.2009.191>
- [3] Duan, L., Tsang, I.W. and Xu, D. (2012) Domain Transfer Multiple Kernel Learning. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **34**, 465-479. <https://doi.org/10.1109/TPAMI.2011.114>
- [4] Tu, W. and Sun, S. (2012) A Subject Transfer Framework for EEG Classification. *Neurocomputing*, **82**, 109-116. <https://doi.org/10.1016/j.neucom.2011.10.024>
- [5] Daume, H.C. and Marcu, D.C. (2006) Domain Adaptation for Statistical Classifiers. *Journal of Artificial Intelligence Research*, **26**, 101-126. <https://doi.org/10.1613/jair.1872>
- [6] Bickel, S., Bruckner, M. and Scheffer, T. (2007) Discriminative Learning for Differing Training and Test Distributions. In: *Proceedings of the 24th International Conference on Machine Learning*, ACM, New York, 81-88. <https://doi.org/10.1145/1273496.1273507>
- [7] Bickel, S., Sawade, C. and Scheffer, T. (2009) Transfer Learning by Distribution Matching for Targeted Advertising. In: *Proceedings of the 21st Annual Conference on Neural Information Processing Systems*, MIT Press, Cambridge, 145-152.
- [8] Wu, P.C. and Dietterich, T.G. (2004) Improving SVM Accuracy by Training on Auxiliary Data Sources. In: *Proceedings of the 21st International Conference on Machine Learning (ICML)*, ACM, New York, 110-117. <https://doi.org/10.1145/1015330.1015436>
- [9] Dai, W.Y., Yang, Q., Xue, G.R., et al. (2007) Boosting for Transfer Learning. In: *Proceedings of the 24th International Conference on Machine Learning (ICML)*, ACM, New York, 193-200. <https://doi.org/10.1145/1273496.1273521>
- [10] Quanz, B. and Huan, J. (2009) Large Margin Transductive Transfer Learning. In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM)*, ACM, New York, 1327-1336. <https://doi.org/10.1145/1645953.1646121>
- [11] Xu, Z.J. and Sun, S.L. (2011) Multi-View Transfer Learning with Adaboost. In: *Proceedings of the 23rd Conference on Tools with Artificial Intelligence*, IEEE, Boca Raton, 399-402. <https://doi.org/10.1109/ICTAI.2011.65>
- [12] Xu, Z.J. and Sun, S.L. (2012) Multi-Source Transfer Learning with Multi-View Adaboost. *Neural Information Processing*, **7665**, 332-339. https://doi.org/10.1007/978-3-642-34487-9_41
- [13] Chen, M., Weinberger, K.Q. and Blitzer, J. (2011) Co-Training for Domain Adaptation. In: *Proceedings of the 25th Conference on Neural Information Processing Systems (NIPS)*, Curran Associates, Inc., New York, 1231-1240.
- [14] Jiang, Y., Deng, Z. and Wang, S. (2012) Mamdani-Larsen Type Transfer Learning Fuzzy System. *Acta Automatica Sinica*, **38**, 1393-1409. <https://doi.org/10.3724/SP.J.1004.2012.01393>
- [15] Zhu, M.-Q., Cheng, Y.-H., Li, M., et al. (2012) A Hybrid Transfer Algorithm for Reinforcement Learning Based on Spectral Method. *Acta Automatica Sinica*, **38**, 1765-1776. <https://doi.org/10.3724/SP.J.1004.2012.01765>
- [16] Jiang, W.H. and Chung, F.L. (2012) Transfer Spectral Clustering. In: *Proceedings of the 2012 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, Springer-Verlag, Berlin, Heidelberg, 789-803. https://doi.org/10.1007/978-3-642-33486-3_50
- [17] Argyriou, A., Micchelli, C.A., Pontil, M., et al. (2007) A Spectral Regularization Frame Work for Multi-Task Structure Learning. In: *Proceedings of Advances in Neural Information Processing Systems (NIPS 2008)*, MIT Press, Cambridge, 25-32.
- [18] Pan, S.J.L., Kwok, J.T. and Yang, Q. (2008) Transfer Learning via Dimensionality Reduction. In: *Proceedings of the 23rd International Conference on Artificial Intelligence*, AAAI Press, Palo Alto, 677-682.
- [19] Pan, S.J.L., Ni, X.C., Sun, J.T., et al. (2010) Cross-Domain Sentiment Classification via Spectral Feature Alignment. In: *Proceedings of the 19th International Conference on World Wide Web (WWW10)*, ACM, New York, 751-760. <https://doi.org/10.1145/1772690.1772767>
- [20] Tu, W. and Sun, S. (2011) Transferable Discriminative Dimensionality Reduction. In: *Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence (CTAI)*, IEEE, Boca Raton, 865-868. <https://doi.org/10.1109/ICTAI.2011.134>
- [21] Gao, X., Wang, X., Li, X., et al. (2011) Transfer Latent Variable Model Based on Divergence Analysis. *Pattern Recognition*, **44**, 2358-2366. <https://doi.org/10.1016/j.patcog.2010.06.013>
- [22] Gao, X., Wang, Z., Yan, P., et al. (2013) Transfer Learning for Pedestrian Detection. *Neurocomputing*, **100**, 51-57. <https://doi.org/10.1016/j.neucom.2011.12.043>
- [23] Lawrence, N.D. and Platt, J.C. (2004) Learning to Learn with the Informative Vector Machine. *Proceedings of the 21st International Conference on Machine Learning*, 65-73. <https://doi.org/10.1145/1015330.1015382>
- [24] Evgeniou, T. and Pontil, M. (2004) Regularized Multi-Task Learning. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, 22-25 August 2004, 109-117. <https://doi.org/10.1145/1014052.1014067>

- [25] Zhang, L. and Zhang, D. (2014) Domain Adaptation Transfer Extreme Learning Machines. *Proceedings of ELM-2014 Volume 1: Algorithms and Theories*, Vol. 3, 103-119. https://doi.org/10.1007/978-3-319-14063-6_10
- [26] Davis, J. and Domingos, P. (2009) Deep Transfer via Second-Order Markov Logic. *Proceedings of the 26th Annual International Conference on Machine Learning*, 217-224. <https://doi.org/10.1145/1553374.1553402>
- [27] Schmidhuber, J. (2015) *Deep Learning in Neural Networks*. Elsevier Science Ltd., Amsterdam. <https://doi.org/10.1016/j.neunet.2014.09.003>
- [28] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [29] Wu, Y., Yuan, M., Dong, S., *et al.* (2017) Remaining Useful Life Estimation of Engineered Systems Using Vanilla LSTM Neural Networks. *Neurocomputing*, **275**, 167-179. <https://doi.org/10.1016/j.neucom.2017.05.063>
- [30] Cho, K., Van Merriënboer, B., Gulcehre, C., *et al.* (2014) Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 25-29 October 2014, 1724-1734. <https://doi.org/10.3115/v1/D14-1179>