

多尺度分析与数据互迁移相结合的短期电力负荷预测方法

刘世昌 金 敏

(湖南大学信息科学与工程学院 长沙 410006)

摘 要 为了提升短期电力负荷的预测效果,提出一种多尺度分析与数据互迁移相结合的短期电力负荷预测方法。一方面,针对多尺度分析预测法中分解得到的子序列在建模和预测的过程中没有对原序列中的隐含相关信息加以利用的问题,采用互信息特征选择法选取合适的原负荷序列历史值并将其加入到原负荷序列近似分量的特征集中,通过特征扩充为学习器提供更多的信息,进而提高近似分量的预测精度。另一方面,针对不同类型的数据在学习过程中会相互影响的问题,采用了一种基于核岭回归的数据互迁移学习方法,将其他几种类型中与待预测类型日相似的数据迁移到待预测类型日的数据中,既利用了这些数据的相似性,又兼顾了这些数据的差异性。测试案例显示,所提方法在 MAPE,MAE 和 RMSE 这 3 个误差评价指标上相对于单模型方法分别降低了 6.2%,3.4%和 5.5%。

关键词 短期电力负荷预测,多尺度分析,特征扩充,数据互迁移,核岭回归

中图分类号 TM734

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2018.07.053

Short-term Load Forecasting Method Combining Multi-scale Analysis with Data Co-transfer

LIU Shi-chang JIN Min

(College of Computer Science and Electronic Engineering, Hunan University, Changsha 410006, China)

Abstract In order to improve the performance of short-term load forecasting, this paper proposed a short-term load forecasting method combining multi-scale analysis with data co-transfer. On the one hand, aiming at the problem that the hidden information in the original series which is related to the subseries isn't fully utilized in the modeling and prediction of subseries in multi-scale analysis forecasting method, this paper used mutual information feature selection method to select appropriate past loads and introduced them into the feature set of the approximation component of original load series. By expanding feature set, more information can be provided for learning algorithm, which can further improve the forecasting accuracy of the approximate component. On the other hand, aiming at the problem that different kinds of data can influence model's performance, this paper proposed a transfer learning method based on kernel ridge regression to transfer similar data to data corresponding to days to be forecasted. By doing this, the similarity of these data was used and the difference of these data was taken into account during modeling. Case study shows that the proposed method outperforms in MAPE, MAE and RMSE which are decreased by 6.2%, 3.4% and 5.5% respectively when compared with single model forecasting method.

Keywords Short-term load forecasting, Multi-scale analysis, Feature-expanding, Data co-transfer, Kernel ridge regression

1 引言

短期电力负荷预测是智能电网建设中的重要环节,人们已经对短期电力负荷预测问题进行了许多研究,其中大多研究是对负荷序列进行整体建模,而没有针对负荷序列的内在规律进行进一步研究与利用。多尺度分析是一种能将时间序列在小波域分解为一个低频子序列与数个高频子序列的方法。它借助小波变换,使这些分解得到的子序列相较于原始序列更具规律性且更平稳,因此更有利于学习算法捕捉到序列中的特征^[1]。近年来,越来越多的研究者关注基于多尺度

分析的方法,如文献[2-4]。然而现有的多尺度分析预测方法在对分解得到的各个子序列分别建模时仅考虑了当前子序列自身的历史因素,而未将原序列中与子序列相关的因素引用到对子序列的建模中,从而导致获得的数据中的信息没有得到充分利用。

多尺度分析分解得到的子序列可以看作是时间序列的一个特例。近年来,随着人工智能领域研究的兴起,越来越多的学者关注基于人工智能的方法。不同于传统的时间序列预测法,人工智能方法能够更好地挖掘时间序列趋势与多种影响因素之间的非线性关系,因此其预测效果显著优于传统时间

到稿日期:2017-05-18 返修日期:2017-07-11 本文受国家自然科学基金项目(61374172)资助。

刘世昌(1991—),男,硕士生,主要研究方向为电力负荷预测,E-mail:ilsubject@hnu.edu.cn;金敏(1973—),女,博士,教授,主要研究方向为电力负荷预测,E-mail:jimin@hnu.edu.cn(通信作者)。

序列预测法。模糊推理系统^[5]、神经网络^[6-7]、支持向量机^[8]等多种人工智能方法近期在有关短期电力负荷预测的研究中被广泛使用,然而这些方法均存在明显的不足。模糊推理系统在建模过程中需要大量的历史数据才能取得较好的效果;对于神经网络,虽然增加神经元的数量能令其更易于训练,但这也会导致神经网络容易过拟合;而支持向量机在样本较多的情况下花费的时间较长。核岭回归是一种能同时处理高维与非线性数据的回归方法,其实现简单、处理速度快且具有高精度^[9]。借助核方法,核岭回归能够将数据映射到高维空间以使其线性可分,同时保证其运算过程仍在低维空间进行。这一特性使得核岭回归在具有非线性拟合能力的同时还能够避免维度灾难。而且借助岭准则,核岭回归能够避免过拟合问题。核岭回归的这些特性使得它在非线性预测问题中成为不错的选择^[10],这也使得其被广泛应用于金融预测^[11]、风速预测^[9]等问题。

在现有的研究中,对于一周内每日的历史负荷特征具有不同模式的问题,通常的解决方法是使用一个取值范围为1~7的整数特征或一个7维布尔特征向量来对一周内的每日数据进行区分,或者将这7日的数据分成互不相交的7个数据集并分别建模。对于前一种方法,尽管引入了用于区分一周内7日数据的特征,仍难以避免这7种不同类型数据的相互干扰。对于后一种方法,虽然将这7种类型的数据彻底分隔开,但在对某一类型的数据进行训练时,并没能使用其他类型数据中隐含的与该类型数据相关的特性。迁移学习是近年来的一个研究热点,是一种能将在其他任务中学习到的知识应用到目标任务的学习过程中的一种学习框架。迁移学习大致可以分为基于实例的迁移学习、基于特征表示的迁移学习、基于模型参数的迁移学习以及基于相关知识的迁移4种

类型。它甚至支持跨域迁移学习。而迁移学习算法在近年来也被运用到诸如自然语言处理等越来越多的领域,引起了人们的关注。在电力负荷预测领域,Zhang等^[12]提出了一种基于迁移学习的短期电力负荷预测算法,该方法将被预测城市的临近城市数据中隐含的知识迁移到被预测城市的数据中,提高了预测的精度。但该方法并没有对一周内7日的数据进行某种程度上的区分,从而使这些数据在建模的过程中不可避免地相互影响。

本文从引入原负荷序列特征以及对一周内7种不同类型的数据进行相似数据互迁移两方面入手,提出一种多尺度分析与数据互迁移相结合的短期电力负荷预测方法。一方面,针对现有的多尺度分析预测法没有充分利用原序列中隐含的与子序列相关的信息的问题,提出将原负荷序列的历史值加入到近似分量的特征中,从而为学习器提供更多的信息。另一方面,针对不同类型的数据在学习过程中会相互影响的问题,采用迁移学习的方法,将其他类型日中与待预测日相似的数据迁移到待预测类型日中,这样既利用了这些数据的相似性,又考虑了这些数据的差异性。图1给出了本文方法的流程。为了捕捉负荷序列在不同尺度上的特征与规律,负荷序列首先由多尺度分析分解为近似分量序列与细节分量序列。接着对这两个子序列分别建模。对于近似分量序列,为了充分利用原序列中与之相关的信息,通过互信息法从原负荷序列中选取合适的历史值作为其特征,从而为预测过程提供更多的信息。随后,将这两个子序列所对应的数据集送入到对应的学习算法中。其中近似分量使用的是由本文提出的基于核岭回归的数据互迁移方法。最后,将近似分量的预测结果与细节分量的预测结果进行叠加,以得到原负荷序列的预测结果。

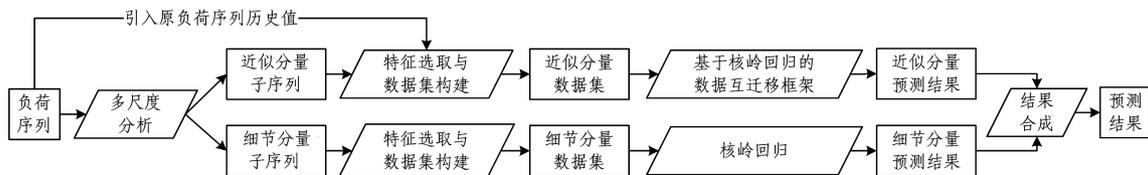


图1 方法概览

Fig. 1 Overview of proposed method

本文第2节介绍本文方法在多尺度分析预测上的改进;第3节介绍了本文提出的基于相似数据互迁移的核岭回归预测法;第4节对本文提出的方法进行评估;最后总结全文。

2 多尺度分析预测法的特征扩充

2.1 多尺度分析

多尺度分析建立在离散小波变换的基础上,它能在不同尺度上将序列分解为数个子序列。给定小波函数 $\psi(t)$ 与尺度函数 $\phi(t)$,负荷序列 $X(t)$ 的离散小波变换为:

$$\omega_{l,k} = \frac{1}{\sqrt{M}} \sum_t X(t) \psi_{l,k}(t)$$

$$\nu_{j,k} = \frac{1}{\sqrt{M}} \sum_t X(t) \phi_{j,k}(t)$$

其中, t 表示时间, j 表示分解层数。小波函数可被视为一个高通滤波器,而尺度函数则可被视为一个低通滤波器。变换

得到的 $\omega_{l,k}$ 被称为小波系数,而 $\nu_{j,k}$ 被称为尺度系数。则将 $X(t)$ 表示为:

$$X(t) = \sum_{k=-\infty}^{\infty} \nu_{j,k} \phi_{j,k}(t) + \sum_{l=1}^j \sum_{k=-\infty}^{\infty} \omega_{l,k} \psi_{l,k}(t) \quad (1)$$

式(1)被称为序列 $X(t)$ 的一个多尺度分析,它可以进一步被简化为:

$$X = \mathcal{S}_j + \sum_{l=1}^j \mathcal{D}_l \quad (2)$$

此时称 \mathcal{S}_j 为负荷序列 X 的一个近似分量,而 \mathcal{D}_l 则为 X 的第 j 层细节分量。这两类分量分别反映了序列 $X(t)$ 的低频与高频特征,使得原来直接对负荷序列 X 预测的问题转变为对相对更有规律的近似分量 \mathcal{S}_j 与数个细节分量 \mathcal{D}_l 的预测问题。

图2给出了深圳市2016年10月15日-2016年12月15日的日均负荷及其对应的近似分量与细节分量。

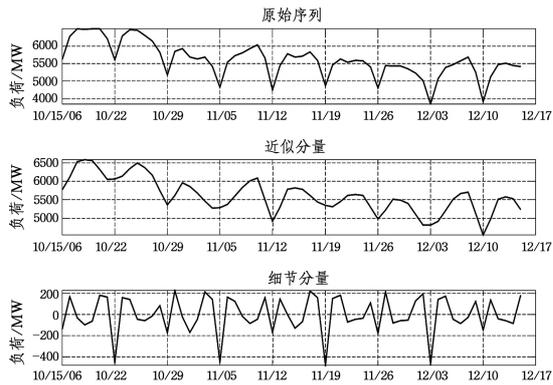


图 2 负荷序列及其对应的由 coif2 小波多尺度分析得到的各分量
Fig. 2 Load series and its components obtained by coif2

2.2 边界处理问题

由于负荷序列的多尺度分量具有周期性,因此在对这些分量建立预测模型时,其数据集中的特征必然包含其历史值。考虑近似分量序列 $\{S_j(1), S_j(2), \dots, S_j(n)\}$ 以及其对应的原始负荷序列 $\{X(1), X(2), \dots, X(n)\}$, 当预测 $S_j(n+1)$ 时, $S_j(n)$ 将作为其特征之一。由于分量序列的每一点都依赖于它所对应的原序列中相应点之前以及之后一段范围内的点, 即 $S_j(n)$ 的值依赖于 $X(n+1), X(n+2), \dots, X(n+p)$ (p 与小波函数以及分解层数有关), 而这些值当前未知, 因此需要对原序列进行延拓。本文使用镜像对称延拓法, 即令 $X(n+1) = X(n), X(n+2) = X(n-1), \dots$ 。最终 $S_j(n)$ 的值可以使用延拓后的负荷序列计算得到。由于延拓的部分通常与其对应的真实序列存在差异, 因此使用延拓后的序列分解得到的子序列的边界部分通常不同于真实序列分解得到的子序列的边界部分, 这种现象被称为边界失真。

小波函数的选择与分解层数的确定对负荷预测的最终效果有着重要的影响。本文方法使用 coif2 小波并将分解层数确定为 2 层。这是因为 coif 系列小波具有正交性, 所以分解得到的分量之间不存在冗余。其次, 由于阶数 (coif2 名称中的 2 即表示其阶数) 或分解层数越高, 分解得到的近似分量能更好地反映原负荷序列的长期趋势, 但过高的阶数或分解层数会导致 p 值过大, 扩大边界失真的影响范围, 进而使预测精度受到更大的不利影响。因此经过反复实验, 最终确定阶数为 2, 分解层数为 1。

2.3 特征选择与扩充

为了利用各个子序列自身的周期性规律, 首先将子序列的历史值作为其特征的一部分。对于细节分量, 使用待预测日前 14 日的的数据作为其候选历史值特征; 而对于近似分量, 将待预测日前 10 日的的数据加入到候选集合中。其次, 针对前文提到的现有的多尺度分析预测法中原序列中的相关信息没有得到充分利用的问题, 本文提出利用原始负荷序列与近似分量序列之间的相关性, 将待预测日前 10 日的原始负荷值也加入到上述候选集合中, 实现对特征集合的扩充, 从而利用原始历史负荷值中所包含的信息来提升近似分量的预测精度。因此, 待预测日 $n+1$ 及其对应的待预测近似分量值 $S_j(n+1)$ 的历史值特征候选集合为:

$$\{S_j'(n), S_j'(n-1), \dots, S_j'(n-9), X(n), X(n-1), \dots, X(n-9)\} \tag{3}$$

其中, $S_j'(\cdot)$ 是使用了镜像对称法的序列分解后得到的子序列中受镜像对称法影响的值, 它相对于对应的真实值 $S_j(\cdot)$ 存在失真, 且越靠近边界 (待预测日) 其失真越严重。 $X(\cdot)$ 为额外引入的原始负荷序列的历史值。本文采用互信息法来从上述候选集合中选出与 $S_j(n+1)$ 最相关的特征子集。互信息是一种基于信息论的能够度量两个变量间相关性的方法。连续随机变量 X, Y 之间的互信息定义为:

$$MI(X, Y) = \int_Y \int_X p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right) dx dy \tag{4}$$

其中, $p(x, y)$ 为 X 和 Y 的联合概率密度函数, $p(x)$ 和 $p(y)$ 分别为 X 与 Y 的边缘概率密度函数。 $MI(X, Y)$ 越大, 表明 X 与 Y 之间的相关性越强。由于通常无法得到 X 和 Y 对应的 $p(x), p(y)$ 以及 $p(x, y)$, 因此连续随机变量间的互信息一般无法直接使用式 (4) 计算, 而是退而求其次, 计算互信息的估计量。本文使用的是 Kraskov 等提出的一种基于最近邻法的互信息估计法 (简称 Kraskov 法), Korrinska 等已在其研究^[13]中展现了该方法在电力负荷预测中的有效性。 Kraskov 法对连续随机变量 X, Y 的互信息的估计如下:

$$MI(X, Y) = \psi(k) - \frac{1}{k} - \frac{1}{N} \sum_{i=1}^N [\psi(n_x(i)) + \psi(n_y(i))] + \psi(N)$$

其中, $\psi(\cdot)$ 为双伽玛函数 $\psi(x) = \Gamma(x)^{-1} \Gamma'(x)$, k 为最近邻数。 $n_x(i)$ 为满足 $\|x_i - x_j\| \leq \epsilon_x(i)/2$ 的点 x_j 的个数, $n_y(i)$ 为满足 $\|y_i - y_j\| \leq \epsilon_y(i)/2$ 的点 y_j 的个数, 其中 $\epsilon_x(i)/2$ 为 x_i 到其第 k 近邻的距离, $\epsilon_y(i)/2$ 为 y_i 到其第 k 近邻的距离。图 3 给出了在深圳市 2005 年—2006 年的数据上使用 Kraskov 法估计出的式 (3) 中每个候选特征与 $S_j(n+1)$ 之间的互信息。由图 3 可知, 一些原始序列的历史特征与待预测近似分量之间的相关性甚至要比一些近似分量的历史特征更强, 因此新引入的特征将为预测过程提供更多的信息。

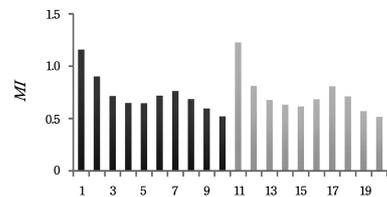


图 3 候选特征与待预测量之间的相关性

Fig. 3 Correlation between candidate features and predictor

由于不同类型的数据集, 其特征与输出量之间的互信息的大小不尽相同, 因此本文并未使用一个固定的互信息值作为相关特征的选取下限, 而是考虑到电力负荷序列中越靠近当日且与当日的间隔接近于一周的整数倍的历史值与当日值越相关的这一特性, 将最靠近当日的那一周中的候选特征与待预测量之间互信息的最小值作为选取的下限。例如, $S_j(n+1)$ 的特征选取下限为:

$$\min(MI(S_j(n+1), S_j'(n)), MI(S_j(n+1), S_j'(n-1)), \dots, MI(S_j(n+1), S_j'(n-6)))$$

3 基于相似数据迁移的核岭回归预测法

3.1 核岭回归

核岭回归是对最小二乘准则的一种改进,即将岭回归准则与核方法相结合而得到的一种非线性回归方法。对于输入 X 及其对应的输出 y ,如果构建线性回归模型:

$$y = X\beta \tag{5}$$

并使用最小二乘准则估计参数 β ,即 $RSS(\beta) = (y - X\beta)^T (y - X\beta)$ (RSS 表示极小化残差的平方和),则有 $\hat{\beta} = (X^T X)^{-1} X^T y$ 。

而岭回归准则在最小二乘准则的基础上加入了一个惩罚项,减轻了最小二乘准则容易过拟合的问题。使用岭回归准则估计参数 β 时,有 $RSS(\beta) = (y - X\beta)^T (y - X\beta) + \lambda\beta^T \beta$,其解为 $\hat{\beta}^{ridge} = (X^T X + \lambda I)^{-1} X^T y$ 。

某些问题在它们原本的特征空间中线性不可分,然而如果使用某种函数 $\phi(\cdot)$ 将它们从原始 p 维特征空间映射到一个更高的 q 维特征空间,它们就能够线性可分。在引入映射函数 $\phi(\cdot)$ 后,式(5)将变为 $y = \phi(X)\beta$ 。而它在岭回归准则下对应的解为:

$$\hat{\beta}^{ridge} = (\phi(X)^T (X) + \lambda I_q)^{-1} \phi(X)^T y \tag{6}$$

由于 q 可能很大甚至无穷,因此直接计算 $\phi(X)^T \phi(X)$ 通常是很困难甚至是不可能的。式(6)还可变换为:

$$\hat{\beta}^{ridge} = \phi(X)^T (\phi(X)\phi(X)^T + \lambda I_N)^{-1} y$$

其中, N 表示训练集中实例的个数。令 $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle = \phi(x_i) \phi(x_j)^T$, 并称 $K(\cdot, \cdot)$ 为核函数。对于输入 X' , 其输出 y' 的预测值为:

$$\begin{aligned} \hat{y}' &= \phi(X')^T \hat{\beta}^{ridge} \\ &= \phi(X') \phi(X)^T (\phi(X)\phi(X)^T + \lambda I_N)^{-1} y \\ &= K(X', X) (K(X, X) + \lambda I_N)^{-1} y \end{aligned} \tag{7}$$

式(7)即为核岭回归的预测公式。将该计算式中的 I_N 替换为 wI_N (w 为 p 维向量),就能够控制 X 中每个实例的惩罚量,从而达到加权的目的。

在核岭回归的实现中,通常使用如下高斯核(也被称为径向基函数)作为其核函数^[9]:

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$$

其中, σ 表示核宽度。 σ 与 λ 的最佳值可以使用不同的方法进行经验估计。本文采用 k 折交叉验证法来确定它们的值。

3.2 迁移学习

在迁移学习中,源任务的输入与目标任务的输入通过一些隐含关系与目标任务的输出构成关联,如图 4 所示。

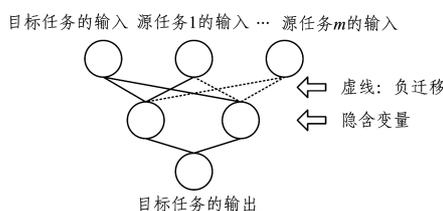


图 4 目标任务与源任务的关系图

Fig. 4 Relationship between target tasks and source tasks

通过这种关联,额外的知识和信息得以从源任务迁移到目标任务中。因此,足够数量的与目标任务相关的源任务能够给目标任务的输出带来正面的影响,提高了模型的精确度。但是与目标任务关联度低的源任务却可能给目标任务的输出带来负面影响,从而降低模型的预测精度。

3.2.1 源任务的选择

源任务的选择是迁移学习中抑制负迁移的关键步骤。对于一组任务 $\mathcal{T} = \{T_k | k = 1, \dots, M\}$, 如果 \mathcal{T} 中的子集 $\mathcal{T}_{Target} = \{T_k | T_k \in \mathcal{T}\}$ 为目标任务集合,那么源任务选择的目的是从 \mathcal{T} 余下的任务中挑选出一个合适的子集。假设 \mathcal{T} 中剩余的任务的个数为 N ,则所有可能的源任务子集的个数为 $2^N - 1$ 。由于搜索空间会随着 N 的增加而急剧增大,因此不可能对所有可能的子集进行评估。为此,使用如下方法进行源任务集合的确定:

首先,对于任务 i 和 j ,以及它们所对应的特征向量 X_i 和 X_j ,使用下式来衡量它们的相似性:

$$Sim(T_i, T_j) = \rho_{X_i, X_j}$$

其中, ρ_{X_i, X_j} 表示 X_i 和 X_j 的皮尔逊相关系数。 $Sim(i, j)$ 越大,表明任务 i 和任务 j 的相关性越强。因此,候选源任务集合 $\mathcal{T}'_{Source} = \mathcal{T} - \mathcal{T}_{Target}$ 中的第 i 个源任务与目标任务集合之间的相似性,可以表示为它与目标任务集合中每个任务的相似性的平均,即:

$$Sim(T_i, \mathcal{T}_{Target}) = \frac{1}{n} \sum_{j=1}^n Sim(T_i, T_j) \tag{8}$$

其中, n 表示目标任务集合中目标任务的个数。根据式(8)得到的相似性,可以设定一个阈值 λ ,最终选取相似性大于 λ 的任务作为该目标任务集合所对应的源任务集合。

3.2.2 数据互迁移方法

TrAdaBoost^[14]是由 Dai 等提出的一种基于 AdaBoost 的迁移学习框架。它利用 AdaBoost 能够调整数据集中不同实例的权重这一特性来达到迁移学习的目的。TrAdaBoost 最初被用于分类问题,之后 Pardoe 等在其研究^[15]中提出了适用于回归问题的 TrAdaBoost. R2 算法,如算法 1 所示。

算法 1 TrAdaBoost. R2

输入:含有 n 个实例的目标任务集合 \mathcal{T}_{Target} , 含有 m 个实例的源任务集合 \mathcal{T}_{Source} , 迭代次数 N 与基学习算法 L

输出: $h_t(x)$ 在 $\lceil N/2 \rceil \leq t \leq N$ 上的加权中值 $h_t(x)$ (其中 h_t 的权重为 $\ln(1/\beta_t)$)

初始化权重向量 $w_1^i = 1/(n+m)$, 其中 $1 \leq i \leq n+m$ 。

For $t = 1, \dots, N$:

1. 对分布为 w^i 的训练集 $\{\mathcal{T}_{Target}, \mathcal{T}_{Source}\}$ 调用 L 并得到模型 h_t ;

2. 计算每一个实例的误差;

3. 令 $D_t = \max_{j=1}^{n+m} |y_j - h_t(x_j)|$

4. 然后令 $\epsilon_t^i = |y_j - h_t(x_j)| / D_t$

5. 计算模型 h_t 的误差:

6. $\epsilon_t = \sum_{i=1}^n \epsilon_t^i w_t^i$; 如果 $\epsilon_t \geq 0.5$, 则跳出循环并将 N 设定为 $t-1$

7. 令 $\beta_t = \epsilon_t / (1 - \epsilon_t)$, $\beta = 1 / (1 + \sqrt{2 \ln(n+m) / N})$

8. 更新权重向量

9. $w_t^{i+1} = \begin{cases} w_t^i \beta_t^{-\epsilon_t^i} / Z_t, & 1 \leq i \leq n \\ w_t^i \beta_t^{\epsilon_t^i} / Z_t, & n+1 \leq i \leq n+m \end{cases}$, 其中 Z_t 为一个标准化常数

本文使用 3.1 节中的加权核岭回归法作为 TrAdaBoost. R2 的基学习器。

3.3 数据互迁移

受人类周期性活动的影响,电力负荷序列具有以星期为单位的周期性,而近似分量也继承了这种特性。这种以星期为单位的周期性使得在近似分量对应的实例中,相同星期的待预测日所对应的实例往往具有相似性,而不同星期的待预测日所对应的实例往往存在差异。因此,如果不加以区分地将全部实例用来训练模型,不同类别的实例就会因为其差异性影响模型最终的训练效果。为此,本文建立了 7 个模型以对应一周内的 7 日,既避免了不同星期数据的差异对建模效果的影响,又利用了这些有差异的数据中隐含的相似特性。我们对这 7 个模型全部使用迁移学习的框架。首先,以星期为标准分割实例,得到周一至周日的 7 种不同类型的实例集合。随后,对于某个特定的星期,将该星期的实例作为目标任务,而将剩余 6 种星期实例作为候选源任务集合,并使用 3.2.1 节中的方法从候选源任务集合中筛选出源任务集合。最后将目标任务与源任务集合应用到以核岭回归为基学习器的 TrAdaBoost.R2 算法中,以建立该星期对应的模型。如图 5 所示,每建立一个特定星期的模型,都会借助这个星期之外的数据,并将其中与该星期具有相似性的数据迁移到该星期所对应的模型的建立过程中。通过不同星期类型之间数据的互迁移来提升最终集成模型的整体预测效果。

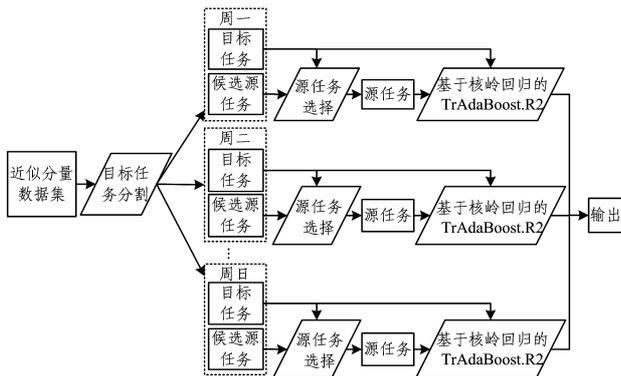


图 5 基于核岭回归的数据互迁移方法

Fig. 5 Data co-transfer method based on kernel ridge regression

4 案例研究

本节使用深圳市 2005 年—2007 年的数据来测试所提方法。将前两年的数据作为训练集,将 2007 年的数据作为测试集,并使用历史值、星期以及节假日与特殊日这 3 类特征。其中,星期特征使用取值范围为 1~7 的整数变量表示,如 1 表示周一,7 表示周日。节假日与特殊日特征包含 3 个布尔特征以分别标记当前实例是否为周六、周日以及周一,以及一个用以标记当前实例是否为春节的布尔特征。

为了评估本文方法的性能,本文采用平均百分比误差(MAPE)、平均误差(MAE)以及均方根误差(RMSE)3 个评价指标,它们能够对序列 L 的预测量 \hat{L} 的预测精度进行评估,其定义如下:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{L_i - \hat{L}_i}{L_i} \right|$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |L_i - \hat{L}_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (L_i - \hat{L}_i)^2}$$

其中, n 是序列 L 中点的个数, L_i 是序列 L 中第 i 个点的值,而 \hat{L}_i 则是序列 \hat{L} 中第 i 个点的值。

4.1 案例 1

案例 1 考查了近似分量实例的特征扩充对近似分量预测效果的提升情况,为此本文对比了使用镜像对称延拓法分解得到的近似分量在使用了特征扩充和未使用特征扩充两种情况下的预测效果。当没有使用特征扩充时,近似分量的历史特征值为 $\{\mathcal{J}'_j(n), \mathcal{J}'_j(n-1), \mathcal{J}'_j(n-2), \mathcal{J}'_j(n-3), \mathcal{J}'_j(n-5), \mathcal{J}'_j(n-6), \mathcal{J}'_j(n-7)\}$ 。使用了特征扩充后,近似分量的历史特征值在之前的基础上又增加了 $\{X(n), X(n-1), X(n-2), X(n-5), X(n-6), X(n-7)\}$ 。

表 1 列出了使用 MAPE, MAE 和 RMSE 3 个指标对这两种情况下的预测结果的评价。可以看出,3 种不同的评测指标均显示在使用特征扩充后,近似分量的预测效果有了较明显的提升,这是因为特征扩充使得对近似分量进行建模和预测时能够利用其对应的原始序列中与近似分量相关的信息。

表 1 案例 1 中不同方法预测结果的对比

Table 1 Comparison of all methods in case 1

是否使用特征扩充	MAPE	MAE	RMSE
否	2.93	165.15	238.53
是	2.50	141.45	205.47

4.2 案例 2

案例 2 对比了 3 种不同的多尺度分析预测法以及核岭回归法对原负荷序列的预测效果,具体使用以下 4 个方案。

M1:采用对称延拓法进行序列延拓的多尺度分析预测法,其中每个子序列均使用核岭回归法进行建模。

M2:使用文献[16]中提出的预测延拓法进行序列延拓的多尺度分析预测法,其中每个子序列均使用核岭回归法进行建模。

M3:使用核岭回归法直接对原负荷序列进行建模。

M4:本文提出的方法,即在 M1 的基础上对近似分子序列进行特征扩充。

表 2 列出了上述 4 种方案在 2007 年全年预测的情况。可以看出在这 4 种方法中,本文方法的预测效果最好。这是因为本文提出的特征扩充方法提升了近似分量的预测精度,从而为最终的原负荷序列的预测打下了良好的基础。M1 的预测效果优于 M2,这是因为 M2 所使用的预测延拓法的效果很大程度上受其所使用的预测方法预测得到的结果精度的影响,因此不够稳定。

表 2 案例 2 中 4 种方案的预测结果

Table 2 Forecasting results of four methods in case 2

方法	MAPE	MAE	RMSE
M1	3.15	179.95	252.40
M2	3.56	198.43	281.12
M3	2.89	162.61	232.93
M4	2.80	160.25	223.28

4.3 案例 3

案例 3 考查了本文提出的数据互迁移方法,为此使用以

下 3 个方案进行对比。

M1:使用本文提出的特征扩充多尺度分析法。其中对于近似分量,对一周内的 7 日所对应的 7 种不同类型的数据分别进行建模。

M2:使用本文提出的特征扩充多尺度分析法。近似分量与细节分量均使用单模型方法。

M3:使用本文提出的特征扩充多尺度分析法。其中近似分量使用本文提出的相似数据互迁移方法。由于本案例使用的数据量有限,为了避免损失过多的信息,将源任务选择过程中的 λ 设为 0.4。

图 6 给出了这 3 种方法在 2007 年全年以一周内的 7 种不同类型日计算得到的 MAPE 的对比。可以看到,除周四和周五,在其他时间上本文方法均取得了最好的效果。M1 的效果除周五以外均是最差的,这是因为单纯地将数据集分成 7 份进行单独建模,无法利用其余数据中隐含的有用信息,从而使预测结果不理想。M2 虽然将所有的数据一起建模,却没有对一周内的 7 种不同类型日进行适当地区分,从而使不相关的数据在训练的过程中相互影响,最终导致预测结果不理想。表 3 列出了采用这 3 种方案预测的全年近似分量情况的对比,可以看到本文方法依然是最优的,这再次验证了本文方法的有效性。最后,表 4 列出了 M2 和 M3 在 2007 年全年预测情况的对比,得益于近似分量预测精度的提升,本文提出的方法在最终原负荷序列的结果上也表现出了最好的效果。

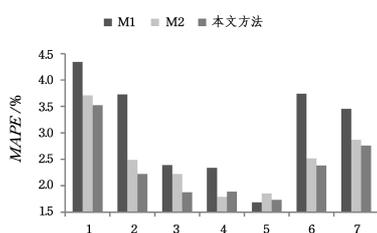


图 6 案例 3 中 3 种方法的近似分量在一周内 7 日数据的对比

Fig. 6 Comparison of approximation component forecasting for three methods in case 3 during a week

表 3 案例 3 中 3 种方案全年近似分量的对比

Table 3 Comparison of one year approximation component forecasting by three methods in case 3

方法	MAPE	MAE	RMSE
M1	3.10	177.61	290.83
M2	2.50	141.45	205.47
M3	2.38	136.46	199.89

表 4 案例 3 中 2 种方法全年预测情况的对比

Table 4 Comparison of one year forecasting by two methods in case 3

方法	MAPE	MAE	RMSE
M2	2.80	160.25	223.28
M3	2.71	157.04	220.14

4.4 案例 4

案例 4 考查了源任务选择参数 λ 的选取对本文方法的影响。本文在案例 3 中 M3 方案的基础上,将源任务选择过程中的 λ 分别设置为 0.3,0.4,0.5。表 5 列出了使用这 3 种不同参数时本文方法在 2007 年全年预测情况的对比。 λ 为

0.4 时预测的结果优于 λ 为 0.3 或 0.5 时,这是因为 λ 过小会导致一些与目标任务相关性过低的任务被选作源任务,从而降低预测精度;而 λ 过大时,又会导致源任务的数量过少,同样会导致预测精度的降低。

表 5 λ 取不同值时全年预测情况的对比

Table 5 Comparison of one year forecasting with different value of λ

λ	MAPE	MAE	RMSE
0.3	2.77	159.68	222.74
0.4	2.71	157.04	220.14
0.5	2.73	157.75	220.91

结束语 本文提出一种多尺度分析与数据互迁移相结合的负荷预测方法。为了提高预测性能,首先,针对多尺度分析预测法中原序列中的隐含信息没有在子序列的建模与预测过程中得到充分利用的问题,本文采用互信息特征选择法选取合适的原负荷序列历史值并将其加入到原负荷序列近似分量的特征集合中,通过引入原负荷序列历史值实现了特征的扩充,从而为学习器提供更多的信息,最终达到提升近似分量预测精度的目的。其次,针对一周内的 7 日所对应的 7 种不同类型数据在学习过程中会相互影响的问题,采用了一种基于核岭回归的数据互迁移方法,该方法将其他几种类型日对应的数据中与待预测类型日数据相似的数据迁移到待预测类型日中,对一周内的 7 日所对应的 7 种不同类型的数据进行数据互迁移,这样既利用了这些数据的相似性,又考虑了数据的差异性。将本文方法在真实的数据集上进行了测试。实验结果表明,相较于改进前的方法,本文方法的性能有所提升,从而证明了本文方法的有效性。由于本文方法能处理具有强非线性且受随机因素影响的序列,因此它同样适用于非电力能源消耗、交通流量等预测场合。

参 考 文 献

- [1] MALLAT S G. A theory for multiresolution signal decomposition: the wavelet representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(7): 674-693.
- [2] MOAZZAMI M, KHODABAKHSHIAN A, HOOSMAND R. A new hybrid day-ahead peak load forecasting method for Iran's National Grid[J]. Applied Energy, 2013, 101: 489-501.
- [3] GHOFRANI M, GHAYEKHLOO M, ARABALI A, et al. A hybrid short-term load forecasting with a new input selection framework[J]. Energy, 2015, 81: 777-786.
- [4] LI S, GOEL L, WANG P. An ensemble approach for short-term load forecasting by extreme learning machine[J]. Applied Energy, 2016, 170: 22-29.
- [5] MASTOROCOSTAS P A, THEOCHARIS J B, BAKIRTZIS A G. Fuzzy modeling for short term load forecasting using the orthogonal least squares method[J]. IEEE Transactions on Power Systems, 1999, 14(1): 29-36.
- [6] BAKIRTZIS A G, PETRIDIS V, KLARTZIS S J, et al. A neural network short term load forecasting model for the Greek power system[J]. IEEE Transactions on Power Systems, 1996, 11(2): 858-863.
- [7] MANDAL P, SENJYU T, FUNABASHI T. Neural networks

- approach to forecast several hour ahead electricity prices and loads in deregulated market[J]. *Energy Conversion and Management*, 2006, 47(15-16): 2128-2142.
- [8] WU C H, TZENG G H, LIN R H. A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression[J]. *Expert Systems with Applications*, 2009, 36(3): 4725-4735.
- [9] DOUAK F, MELGANI F, BENOUDJIT N. Kernel ridge regression with active learning for wind speed prediction[J]. *Applied Energy*, 2013, 103: 328-340.
- [10] EXTERKATE P, GROENEN P J F, HEIJ C, et al. Nonlinear forecasting with many predictors using kernel ridge regression [J]. *International Journal of Forecasting*, 2016, 32(3): 736-753.
- [11] KIM H H, SWANSON N R. Forecasting financial and macro-economic variables using data reduction methods; New empirical evidence[J]. *Journal of Econometrics*, 2014, 178: 352-367.
- [12] ZHANG Y L, LUO G M. Short term power load prediction with knowledge transfer[J]. *Information Systems*, 2015, 53: 161-169.
- [13] KOPRINSKA I, RANA M, AGELIDIS V G. Correlation and instance based feature selection for electricity load forecasting[J]. *Knowledge-Based Systems*, 2015, 82: 29-40.
- [14] DAI W, YANG Q, XUE G R, et al. Boosting for transfer learning[C]// *Proceedings of the 24th International Conference on Machine Learning*. 2007: 193-200.
- [15] PARDOE D, STONE P. Boosting for Regression Transfer[C]// *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 2010: 863-870.
- [16] REIS A J R, DA SILVA A P A. Feature extraction via multi-resolution analysis for short-term load forecasting[J]. *IEEE Transactions on Power Systems*, 2005, 20(1): 189-198.
- (上接第 285 页)
- [9] WILSON D H, LONG A C, ATKESON C. A context-aware recognition survey for data collection using ubiquitous sensors in the home [C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Portland, USA, 2005: 1865-1868.
- [10] NEBELING M, SPEICHER M, NORRIE M. W3Touch: metrics-based web content adaptation for touch [C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Paris, France, 2013: 2311-2320.
- [11] JUN-KI M, AFSANEH D, JASON W, et al. 'N' turn: smartphone as sleep and sleep quality detector [C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Toronto, Canada, 2014: 477-486.
- [12] INTILLE S S, RONDONI J, KUKLA C, et al. A context-aware experience sampling tool [C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Fort Lauderdale, USA, 2003: 972-973.
- [13] FIGUEIREDO I N, LEAL C, PINTO L, et al. A lemos exploring smartphone sensors for fall detection [J]. *The Journal of Mobile User Experience*, 2016, 5(1): 1-17.
- [14] HONG J H, RAMOS J, DEY A K. Toward personalized activity recognition systems with a semipopulation approach [J]. *IEEE Transactions on Human-Machine Systems*, 2016, 46(1): 101-112.
- [15] BRICON-SOUF N, NEWMAN C R. Context awareness in health care: a review [J]. *International Journal of Medical Informatics*, 2007, 76(1): 2-12.
- [16] BAE N J, KWAK K H, et al. Context-aware control service model based on ontology for greenhouse environment [J]. *Advances in Computer Science and Its Applications*, 2014, 279: 321-326.
- [17] CAPRA L, EMMERICH W, MASCOLO C. CARISMA: context-aware reflective middleware system for mobile applications [J]. *IEEE Transactions on Software Engineering*, 2003, 29(10): 929-945.
- [18] FETTER M, GROSS T. CAESSA: visual authoring of context-aware experience sampling studies [C]// *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2011: 2341-2346.
- [19] HAN L, LIU Z J, LI H, et al. A method based on context-aware for remote user experience data capturing [J]. *Chinese Journal of Computer*, 2015, 38(11): 2234-2246. (in Chinese)
韩立, 刘正捷, 李晖, 等. 基于情境感知的远程用户体验数据采集方法[J]. *计算机学报*, 2015, 38(11): 2234-2246.
- [20] CARD S, MORAN T P, NJWELL A. *The psychology of human-computer interaction* [M]. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.
- [21] ANDERSON J R, LEBIERE C. *Atomic components of thought* [M]. Hillsdale, NJ: Lawrence Erlbaum Associates, 1998.
- [22] LAIRD J E, NEWELL A. SOAR: An architecture for general intelligence [J]. *Artificial Intelligence*, 1987, 33(1): 1-64.
- [23] JUST M A, CARPENTER P N. A capacity theory of comprehension: individual differences in working memory [J]. *Psychological Review*, 1992, 99(99): 122-49.
- [24] MEYER D E, KIERAS D E. A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms [J]. *Psychological Review*, 1997, 104(1): 3-65.
- [25] MEYER D E, KIERAS D E. A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of Psychological Refractory Period Phenomena [J]. *Psychological Review*, 1997, 104(4): 749-791.
- [26] LIU Y, ROBERT F, TSIMHONI O. Queuing network-model human processor (QN-MHP): A computational architecture for multitask performance in human-machine systems [J]. *ACM Transactions on Computer-Human Interaction*, 2006, 13(1): 37-70.
- [27] MO T, LI W P, WU Z H, et al. Framework of context-aware based service system [J]. *Chinese Journal of Computer*, 2010, 33(11): 2084-2092. (in Chinese)
莫同, 李伟平, 吴中海, 等. 一种情境感知服务系统框架 [J]. *计算机学报*, 2010, 33(11): 2084-2092.