

实现影响因素多源异构融合的短期负荷预测支持向量机算法

吴倩红¹, 高 军², 侯广松², 韩 蓓¹, 汪可友¹, 李国杰¹

(1. 电力传输与功率变换控制教育部重点实验室(上海交通大学), 上海市 200240;

2. 国网山东省电力公司菏泽供电公司, 山东省菏泽市 274000)

摘要: 针对智能电网大数据环境下, 导致电力系统负荷波动的诸多因素存在多源异构性的问题, 利用多核函数来对其多源异构特性进行差异化处理和融合, 能够描述影响因素的内在分布特性并应对其变化, 提高负荷预测精度。选取历史负荷、气温、气压、相对湿度、降雨量、风向、风速、节假日及电价 9 个属性作为多源异构影响因素, 利用样本特征分布法、单变量法及核矩阵秩空间差异法来选择多核函数的构成, 采用双层多核学习算法, 建立了并行化多核支持向量机(SVM)负荷预测算法流程, 并在 Hadoop 集群上进行了仿真验证。仿真结果表明, 多核 SVM 比单核 SVM 预测平均相对误差小, 双层多核学习、基于 l_p 范数的多核 SVM 模型预测精度最高。因此, 多核 SVM 能有效处理负荷预测中的多源异构数据, 经并行化处理后, 能提高负荷预测的速度与精度。

关键词: 大数据; 多源异构特性; 支持向量机(SVM); 负荷预测; 并行化

0 引言

随着智能电网研究的不断深入^[1], 高级管理体系逐步建立, 智能用电设备逐渐推广^[2], 在这一过程中积累了大量的基础用电数据^[3], 同时分布式电源、储能、电动汽车等多种新型用电形式出现, 各类先进理念和技术不断出现, 社会对供电服务的需求日趋多样化^[4], 因此影响负荷预测的因素数量、种类、随机性以及分布特点等日益多而复杂, 传统的负荷预测方法的速度与精度已经不能满足大数据环境下的要求。

智能电网大数据具有分散性、多样性、复杂性以及高价值的特点^[5], 前三个特点使得大数据具有多源异构特性。所谓多源异构性是指在数据管理和访问中存在异构性问题, 如系统异构、数据源异构等。影响负荷预测的因素包括历史负荷数据、气象数据(气温、降雨量、气压、相对湿度、风向、风速等)、地理数据、节假日数据、电价等。这些因素各自系统的建设时间、研发单位、采用的技术和具体业务的特定要求等, 导致了数据的存储方式、数据类型以及更新频次的不同^[6], 因而具有多源异构的特点。如何基于负荷影响因素的多源异构特性进行负荷预测, 对

提高预测精度具有重要意义。

大数据环境下, 现有的负荷预测方法如时间序列^[7-8]、回归分析^[9]、神经网络^[10-12]、支持向量机(SVM)^[13-15]等在运算速度和考虑负荷影响因素的多源异构性方面存在局限性。在运算速度方面, 当数据量非常大时, 现有负荷预测方法速度很慢, 比如神经网络和模糊理论均需要较长的训练时间。在考虑负荷影响因素多源异构特性方面, 经典时间序列法对负荷影响因素考虑较少, 大多只利用历史负荷进行预测; 回归分析方法虽然考虑了部分影响因素, 但并没有考虑影响因素的异构特性, 未能对这些特征进行区分对待; 现有 SVM 大多是基于单核进行负荷预测, 其将所有的特征都并列堆叠成一个向量, 而后一起输入到一个具有固定形式与固定参数的单一核函数中进行建模, 无法处理负荷影响因素来自不同数据源的不同种类特征。

目前已有一些文献对大数据环境下负荷预测速度的提高做了研究。文献^[16]提出基于随机森林算法的并行负荷预测方法, 缩短了负荷预测时间, 提高了随机森林算法对大数据的处理能力; 文献^[17]基于局部加权线性回归和云计算平台, 建立并行局部加权线性回归模型, 提高了预测速度。以上文献都是从运算速度方面对负荷预测算法进行改进, 但从考虑负荷影响因素多源异构特性方面来提升负荷预测精度的研究还很少, 文献^[13]采用两个不同尺度的高斯核函数进行负荷预测, 但多分辨率核函数是

收稿日期: 2016-02-29; 修回日期: 2016-05-03。

上网日期: 2016-07-12。

国家自然科学基金资助项目(51407116); 国家科技支撑计划资助项目(2015BAA01B02)。

靠经验构建,且两个核函数是简单相加。

目前文献中关于多核函数构成选择的方法,主要有两种:第一种是直接选取一个全局核函数 K_{global} 和一个局部核函数 K_{local} , 构成混合核函数 $K_3 = \lambda K_{\text{global}} + (1-\lambda) K_{\text{local}}$, 其中权重 $\lambda \in [0, 1]$, 这种方法没有考虑已有数据集的特征,且常用的全局核函数有两种,该方法选择全局核函数时具有盲目性;第二种是首先利用单核 SVM 进行预测或分类,选取精度最高的两种核函数组合形成多核函数,这种方法忽略了来自不同数据源的不同种类特征,没有充分利用多核函数对多源异构数据进行有效处理。因此如何根据负荷多源异构影响因素特点而构建多核函数,需要深入探究。

本文将多核函数 SVM 算法应用于负荷预测中,对负荷影响因素存在的多源异构特性进行差异化处理和融合,研究最优多核函数的选择方法以及并行化多核 SVM 负荷预测算法流程,并在 Hadoop 集群上进行了仿真验证。

1 多核学习 SVM 预测模型

1.1 多核函数

多核函数学习是通过针对每一种类型的数据创建相应的核函数,再将这多个核函数组合后进行机器学习,能有效处理多源异构数据。多核函数的构造有两种:线性多核函数和非线性多核函数^[18]。采用非线性多核函数学习时,相关的优化问题不具备凸性或很难保证获得局部的最优解^[19],因此在本文中采用线性多核函数。线性多核函数为:

$$K_{\text{MKL}}(x, y) = \sum_{h=1}^m \lambda_h K_h(x, y) \quad (1)$$

式中: $K_h(x, y)$ 为单核函数; $\lambda_h \in [0, 1]$, 为核函数权重; m 为单核函数个数。

常用的单核函数有线性核函数、多项式核函数、径向基核函数以及 Sigmoid 核函数。单核函数 $K_h(x, y)$ 有两种选择:一是针对数据异构性选择参数不同的核函数,如根据不同的特性可以选择不同宽度的径向基核函数;二是针对数据异构性选择不同种类的核函数,如根据不同的特性可以在学习过程中同时选择高斯核函数和多项式核函数。图 1 为各核函数曲线图。图中, σ, k, d 分别为径向基核函数、Sigmoid 核函数、多项式核函数的参数。径向基核函数是局部核函数,距离样本点很近的数据点对核函数的值影响较大,学习能力强、泛化性能较弱; Sigmoid 核函数、线性核函数及多项式核函数是全局核函数,距离样本点很远的的数据点都可以对核函数的值产生影响,泛化性能强、学习能力较弱。由于

局部核函数和全局核函数的内推与外推能力存在差异,若将两者组合成新的多核函数,则能够扬长避短,兼顾其构成中的普通核函数的优势,得到性能更加优良的 SVM 预测模型^[19]。

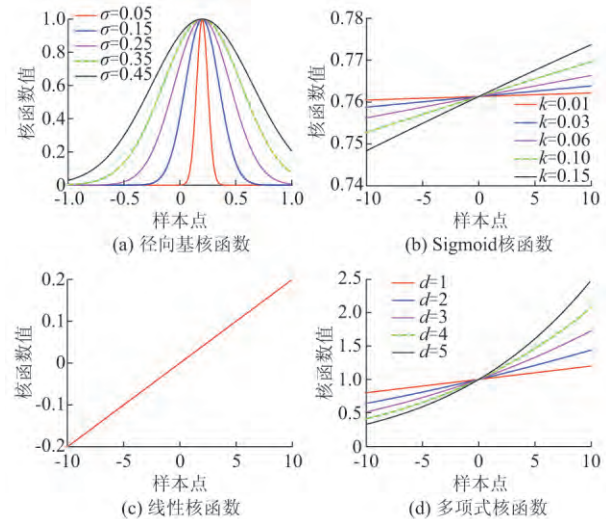


图 1 各核函数曲线

Fig.1 Curves of each kernel function

1.2 基于 l_p 范数的多核 SVM 预测模型

基于范数的多核 SVM 有 l_1 与 l_p ($1 < p < +\infty$) 范数,由于基于 l_1 范数易得到稀疏解,因此本文采用基于 l_p 范数,其目标函数及约束条件^[18]为:

$$\left\{ \begin{array}{l} \min_{\lambda} \max_{\alpha, \hat{\alpha}} \left(\sum_{i=1}^N y_i (\alpha_i - \hat{\alpha}_i) - \epsilon \sum_{i=1}^N (\alpha_i + \hat{\alpha}_i) - \left\| \frac{1}{2} \sum_{h=1}^m \lambda_h \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \hat{\alpha}_i) (\alpha_j - \hat{\alpha}_j) \cdot K_{\text{MKL}p}(x_i, x_j) \right\|_{p^*} \right) \\ \text{s.t.} \quad \sum_{i=1}^N (\alpha_i - \hat{\alpha}_i) = 0 \\ 0 \leq \alpha_i \leq C \quad i = 1, 2, \dots, N \\ 0 \leq \hat{\alpha}_i \leq C \quad i = 1, 2, \dots, N \\ \lambda_h \geq 0 \quad h = 1, 2, \dots, m \end{array} \right. \quad (2)$$

式中: $K_{\text{MKL}p}(x_i, x_j) = \sum_{h=1}^m \lambda_h K_h(x_i, x_j)$; $p^* = p/(p-1)$; x_i 为训练样本输入值; y_i 为训练样本目标值; N 为训练样本总数; λ 为核函数权重向量; ϵ 为精度; α_i 和 $\hat{\alpha}_i$ 分别为拉格朗日乘子向量 α 和 $\hat{\alpha}$ 中的元素; C 为惩罚因子。

本文采用双层多核学习来求解参数,一层优化核函数权重 λ , 一层优化拉格朗日乘子 α 和 $\hat{\alpha}$, 两层交替优化,直至满足停时准则,得到最优核函数权重

λ^* 与最优拉格朗日乘子 α^* 和 $\hat{\alpha}^*$ 。具体算法步骤见附录 A。最终回归估计函数为:

$$f(x) = \sum_{i=1}^N (\alpha_i^* - \hat{\alpha}_i^*) K_{\text{MKL},p}(x_i, x) + b \quad (3)$$

式中: $b \in \mathbf{R}$ 为阈值参数。

2 基于多核学习的 SVM 并行负荷预测

2.1 负荷多源异构影响因素的多核函数构成

本文采用多源异构影响因素的样本特征分布法、单变量法以及核矩阵秩空间差异(RSD)法,提出最优多核函数构建方法,如图 2 所示。

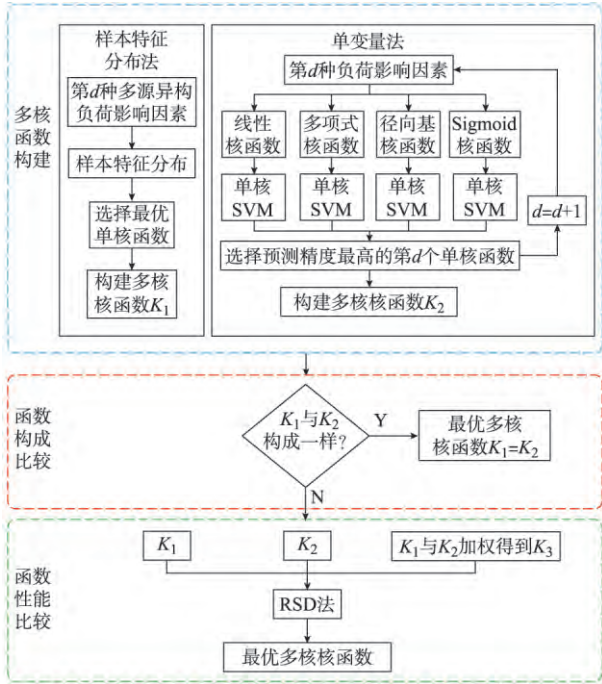


图2 最优多核函数构建方法
Fig.2 Construction method of optimal multiple kernel function

设影响负荷预测的 Q 个多源异构因素的集合为 $\{X_d, d=1, 2, \dots, Q\}$, $X_d = [x_1, x_2, \dots, x_T] \in \mathbf{R}_{1 \times T}$, 目标负荷值为 $Y_i = [y_1, y_2, \dots, y_T] \in \mathbf{R}_{1 \times T}$ 。将每种影响因素样本 X_d 的特征在相应坐标系下表示,根据图中样本的分布,判断该影响因素的分布特性;根据样本分布特性,选择其对应最优核函数 $K_{d1}(x_i, x_j)$,构建多核函数:

$$K_1(x_i, x_j) = \sum_{d=1}^Q K_{d1}(x_i, x_j) \quad (4)$$

采用单变量法,比较 4 种核函数的预测结果,选择预测精度最高的单核函数作为该影响因素的核函数 $K_{d2}(x_i, x_j)$,构建多核函数:

$$K_2(x_i, x_j) = \sum_{d=1}^Q K_{d2}(x_i, x_j) \quad (5)$$

若 $K_1(x_i, x_j)$ 与 $K_2(x_i, x_j)$ 的构成相同,则此多核函数为负荷多源异构影响因素融合的最优多核函数 $K_{\text{MKL},p}(x_i, x_j)$;若两者的构成不同,则利用加权法形成多核核函数:

$$K_3(x_i, x_j) = aK_1(x_i, x_j) + (1-a)K_2(x_i, x_j) = a \sum_{d=1}^Q K_{d1}(x_i, x_j) + (1-a) \sum_{d=1}^Q K_{d2}(x_i, x_j) = \sum_{d=1}^Q K_{d3}(x_i, x_j) \quad (6)$$

式中: a 为权系数, $0 < a < 1$, 在文献[20]的基础上进行改进,得到 a 的计算方法如式(7)所示。

$$a = - \frac{\sum_{i=1}^T \sum_{j=1}^{i-1} B y_i y_j}{2 \sum_{i=1}^T \sum_{j=1}^{i-1} A y_i y_j} = \frac{\sum_{i=1}^T \sum_{j=1}^{i-1} E y_i y_j}{\sum_{i=1}^T \sum_{j=1}^{i-1} (D + E) y_i y_j} \quad (7)$$

$$\begin{cases} D = K_1(x_i, x_i) - 2K_1(x_i, x_j) + K_1(x_j, x_j) \\ E = K_2(x_i, x_i) - 2K_2(x_i, x_j) + K_2(x_j, x_j) \\ A = D + E \\ B = -2E \end{cases} \quad (8)$$

分别计算在 $K_{d1}(x_i, x_j)$, $K_{d2}(x_i, x_j)$, $K_{d3}(x_i, x_j)$ 核函数映射下的核矩阵:

$$A_{ds} = \begin{bmatrix} K_{ds}(x_1, x_1) & K_{ds}(x_1, x_2) & \cdots & K_{ds}(x_1, x_T) \\ K_{ds}(x_2, x_1) & K_{ds}(x_2, x_2) & \cdots & K_{ds}(x_2, x_T) \\ \vdots & \vdots & \ddots & \vdots \\ K_{ds}(x_T, x_1) & K_{ds}(x_T, x_2) & \cdots & K_{ds}(x_T, x_T) \end{bmatrix} \quad (9)$$

式中: $d=1, 2, \dots, Q; s=1, 2, 3$ 。

采用 RSD 法,计算式(9)得到的 $A_{1,s}, A_{2,s}, \dots, A_{Q,s}, s=1, 2, 3$ 的秩空间差异性度量^[21]为:

$$R_s(A_{1,s}, A_{2,s}, \dots, A_{Q,s}) = \frac{1}{Q} \cdot$$

$$\sum_{d=1}^Q (2\text{rank}([A_{d,s}^T, A_{1,s}^T, \dots, A_{d-1,s}^T, A_{d+1,s}^T, \dots, A_{Q,s}^T]^T) - \text{rank}(A_{d,s}) - \text{rank}([A_{1,s}^T, \dots, A_{d-1,s}^T, A_{d+1,s}^T, \dots, A_{Q,s}^T]^T)) \quad (10)$$

RSD 值越大,核矩阵的秩越大,核矩阵的正定性越好,样本集在特征空间中的线性可分程度越高^[21],此时多核函数性能越好。因此,比较 R_1 和 R_2 与 R_3 值的大小,选取最大值对应的函数作为负荷多源异构影响因素融合的最优多核函数 $K_{\text{MKL},p}$ 。

2.2 基于多核学习的 SVM 并行化负荷预测

实验表明,在同样样本数量、单机运行的前提下,多核学习的 SVM 预测算法比单核 SVM 预测算

法速度慢很多,为利用多核 SVM 处理多源异构数据的优势同时提升其预测速度,本文提出了基于多核学习的 SVM 的 MapReduce 并行化负荷预测实现方法,以满足智能电网大数据环境下对负荷预测速度和精度的要求,整个流程如图 3 所示。图中,SVR 表示支持向量回归。具体步骤如下。

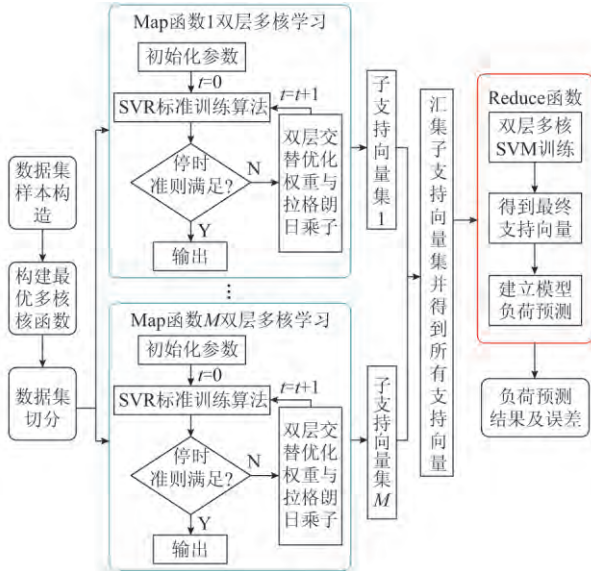


图 3 多核 SVM 并行化负荷预测流程
Fig.3 Process of parallel multiple kernel SVM load forecasting

步骤 1:采集原始数据,经数据预处理后构造出训练样本集(含 N 个训练样本点)与待预测集(含 S 个待预测点)。

步骤 2:按照图 2 所示流程选取多核函数的组成核函数。

步骤 3:数据集切分,将训练样本集的 N 个样本点划分为 M 个子训练集,每个子训练集含有约 N/M 个训练样本点,并上传至 Hadoop 分布式文件系统。

步骤 4: M 个 Map 函数并行训练,每个 Map 函数获得步骤 3 中的子训练集后,双层多核学习算法进行 SVM 训练,将子支持向量集作为 Value 值输出。

步骤 5:一个 Reduce 函数汇集 M 个子支持向量集,得到全部的支持向量,对所有支持向量再次利用双层多核学习算法进行最终的 SVM 训练,得到最终支持向量;建立最终多核 SVM 负荷预测模型,对待预测集进行预测,并计算预测误差。

3 实证分析

3.1 仿真系统介绍

根据大数据环境下负荷预测速度和精度的高要

求,本文在虚拟机中建立了一个含 3 个节点的 Hadoop 集群,一个节点作为主节点,其余两个节点作为从节点,三个节点的操作系统均为 Ubuntu,仿真基于该平台进行。实验数据来自山东省菏泽市供电公司、气象局,所有数据的采集频率均为 15 min。

3.2 预测效果评测指标

本文选取平均相对误差作为负荷预测结果评测指标,其计算公式为:

$$e = \frac{1}{96} \sum_{i=1}^{96} \left| \frac{\hat{L}_i - L_i}{L_i} \right| \times 100\% \quad (11)$$

式中: L_i 和 \hat{L}_i 分别为负荷真实值和预测值。

3.3 实验多源异构特征值及多核函数选取

本文选取的影响负荷预测的多源异构数据特征共 9 种:历史负荷、气温、气压、相对湿度、风向、风速、降雨量、电价、节假日。利用样本特征分布法与单变量法,对周一至周日共 7 d 的影响因素进行分析,得到结果见附录 B 图 B1。可知:历史负荷、气温、气压、相对湿度、降雨量以及电价的分布呈局部特性,因此适于选择局部核函数;风向、风速以及节假日的分布呈全局特性,适于选择全局核函数。

采用单变量法,对 2014 年 8 月 1 日至 9 月 30 日共两个月进行负荷预测,统计这 61 d 中,单变量法与特征分布法所得多核函数组成相同与不同时的比例:8 月份组成相同的一共有 21 d,占一个月的 67%,组成不同的一共有 10 d,占一个月的 33%;9 月份组成相同的一共有 18 d,占一个月的 60%,组成不同的一共有 12 d,占一个月的 40%。组成不同时结果有很多种,总结其中两种于附录 B 表 B1。

当与样本特征分布法确定的多核函数组成相同时,呈现局部特性的影响因素在局部径向基核函数下的预测误差最小,呈现全局特性的影响因素在全局 Sigmoid 核函数下的预测误差最小。当与样本特征分布法确定的多核函数组成不同时,附录 B 表 B1 不同组成 1 中相对湿度和降雨量与特征分布法对应的核函数性质不同,附录 B 表 B1 不同组成 2 中气温、气压、风速以及降雨量与特征分布法对应的核函数性质不同。

3.4 仿真算例及仿真结果

算例 1:以 2014 年 6 月 16 日至 8 月 2 日为训练数据,预测 2014 年 8 月 3 日至 8 月 9 日的负荷,样本特征分布法与单变量法确定的多核函数构成相同。分别采用单核径向基核函数(简称函数 1)、基于 l_1 范数多核函数(简称函数 2)、基于 l_p (p 取 8/7, 6/5, 4/3, 3/2, 2)范数多核函数(简称函数 3)进行预测,结果如表 1 所示。图 4 是 8 月 3 日的日负荷

预测曲线。其中,多核函数采用径向基核函数与 Sigmoid 核函数构成。

表 1 平均相对误差预测结果
Table 1 Prediction results of average relative error

时间	平均相对误差/%						
	函数 1	函数 2	函数 3				
			8/7	6/5	4/3	3/2	2
8月3日	2.79	1.69	1.63	1.621	1.629	1.62	1.61
8月4日	4.11	3.73	3.68	3.682	3.67	3.67	3.65
8月5日	3.95	3.02	2.99	2.98	2.97	2.89	2.89
8月6日	4.36	3.88	3.82	3.81	3.812	3.80	3.80
8月7日	2.98	1.81	1.74	1.71	1.72	1.72	1.73
8月8日	3.64	2.87	2.77	2.78	2.74	2.75	2.75
8月9日	3.10	2.31	2.27	2.27	2.24	2.27	2.26

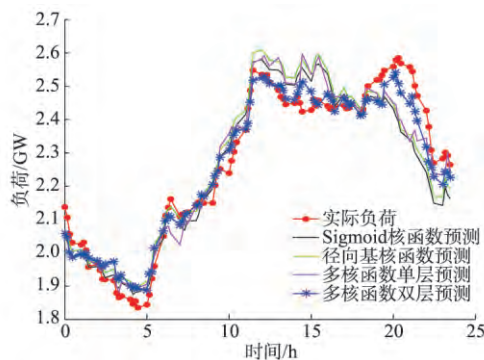


图 4 日负荷预测曲线
Fig.4 Prediction curves of daily load

从表 1、图 4 可以看出:利用多核 SVM 预测模型的平均相对误差预测值低于单核 SVM 预测模型,基于 l_p 范数多核函数的预测误差低于基于 l_1 范数多核函数,在两个高峰处、负荷曲线末尾的预测结果得到显著改善。此仿真结果说明多核 SVM 预测模型比单核 SVM 预测模型更加优良,将局部径向基核函数与全局 Sigmoid 核函数组合成多核函数,并通过算法求解最优权重,使得多核函数实现优势互补,能够有效处理影响负荷预测的多源异构因素,提高了负荷预测精度。

算例 2:以 2014 年 6 月 16 日至 8 月 2 日为训练数据,预测 2014 年 8 月 23 日的负荷,该日的样本分布法与单变量法确定的多核函数组成不同,且单变量法确定的组成情况为附录 B 表 B1 中的不同组成 1。与组成情况相同、不同组成 2 作对比。当采用不同组成 1 时,加权法得到的 RSD 大于样本特征分布法与单变量法得到的 RSD,对应的预测平均相对误差最小;不同组成 1 得到的 RSD 均大于组成情况相同、不同组成 2 两种情况下的 RSD,是因为当天的单变量法确定的为不同组成 1,说明 2.1 节提出的选取多核函数的方案可行。具体预测结果见附录 B

表 B2。

算例 3:以一天为例,利用 2014 年 6 月 16 日至 8 月 2 日数据预测 2014 年 8 月 3 日的负荷,比较多核函数预测在核函数个数不同时单机与 Hadoop 集群运行的运算时间,结果如图 5 所示。可以看出,当核函数个数较少时,单机与 Hadoop 集群运算时间差别较小,但当核函数个数增多时,单机运算时间急剧增加,Hadoop 集群运算时间远小于单机运算时间。因此当在大数据环境下考虑更多负荷影响因素、训练样本数量剧增时,并行化多核 SVM 负荷预测能有效提高负荷预测精度与速度。

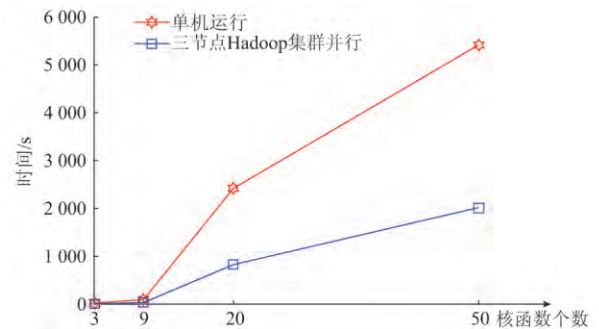


图 5 单机与 Hadoop 集群运算时间对比
Fig.5 Consume time contrast between one machine and Hadoop cluster

4 结语

本文针对现有负荷预测算法无法解决大数据环境下影响负荷预测的因素具有多源异构特性的问题,利用多核函数处理多源异构影响因素,提出最优多核函数选择方法,并在 Hadoop 集群上进行并行化多核 SVM 负荷预测,提高了负荷预测的速度与精度。仿真结果表明,多核 SVM 比单核 SVM 预测精度高,采用双层学习法求解、基于 l_p 范数的 SVM 预测模型精度最高。

相比于传统单核 SVM 核函数参数选取困难的问题,多核函数可扩大参数选择范围,例如多核函数由 100 个径向基核函数构成,则可任选 100 个值作为多核函数的参数,减小了参数选择的难度。本文采用虚拟机中三节点的 Hadoop 集群进行并行运算,当核函数个数较多时耗时较长,实际中可采用更大规模的 Hadoop 集群,缩短运算时间。

附录见本刊网络版 (<http://www.aeps-info.com/aeps/ch/index.aspx>)。

参考文献

- [1] 祝恩国, 龚健. 用电信息采集系统双向互动功能设计及关键技术[J]. 电力系统自动化, 2015, 39 (17): 62-67. DOI: 10.7500/AEPS20150414002.

- ZHU Enguo, DOU Jian. Design of two-way interaction function based on electricity information acquisition system and its key technology[J]. Automation of Electric Power Systems, 2015, 39(17): 62-67. DOI: 10.7500/AEPS20150414002.
- [2] 王守相, 孙智卿, 刘喆. 面向智能用电的家庭能量协同调度策略[J]. 电力系统自动化, 2015, 39(17): 108-113. DOI: 10.7500/AEPS20140625016.
- WANG Shouxiang, SUN Zhiqing, LIU Zhe. Co-scheduling strategy of home energy for smart power utilization[J]. Automation of Electric Power Systems, 2015, 39(17): 108-113. DOI: 10.7500/AEPS20140625016.
- [3] 郭晓利, 于阳. 基于云计算的家庭智能用电策略[J]. 电力系统自动化, 2015, 39(17): 114-119. DOI: 10.7500/AEPS20150310017.
- GUO Xiaoli, YU Yang. A residential smart power utilization strategy based on cloud computing[J]. Automation of Electric Power Systems, 2015, 39(17): 114-119. DOI: 10.7500/AEPS20150310017.
- [4] 李扬, 王蓓蓓, 李方兴. 灵活互动的智能用电展望与思考[J]. 电力系统自动化, 2015, 39(17): 2-9. DOI: 10.7500/AEPS20150730004.
- LI Yang, WANG Beibei, LI Fangxing. Outlook and thinking of flexible and interactive utilization of intelligent power[J]. Automation of Electric Power Systems, 2015, 39(17): 2-9. DOI: 10.7500/AEPS20150730004.
- [5] 张东霞, 苗新, 刘丽平, 等. 智能电网大数据技术发展研究[J]. 中国电机工程学报, 2015, 35(1): 2-11.
- ZHANG Dongxia, MIAO Xin, LIU Liping, et al. Research on development strategy for smart grid big data[J]. Proceedings of the CSEE, 2015, 35(1): 2-11.
- [6] 耿焕同, 张明哲, 张勇. 多源异构数据的多维决策分析与可视化方法[J]. 信息技术, 2013, 37(11): 49-53.
- GEN Huantong, ZHANG Mingzhe, ZHANG Yong. A multi-dimensional decision-making analysis and visualization method for solving multi-source and heterogeneous datasets[J]. Information Technology, 2013, 37(11): 49-53.
- [7] 牛东晓, 魏亚楠. 基于 FHNN 相似日聚类自适应权重的短期电力负荷组合预测[J]. 电力系统自动化, 2013, 37(3): 54-57.
- NIU Dongxiao, WEI Yanan. Short-term power load combinatorial forecast adaptively weighted by FHNN similar-day clustering[J]. Automation of Electric Power Systems, 2013, 37(3): 54-57.
- [8] 王鹏, 邵能灵, 王波, 等. 针对气象因素的短期负荷预测修正方法[J]. 电力系统自动化, 2008, 32(13): 92-96.
- WANG Peng, TAI Nengling, WANG Bo, et al. A short-term load forecasting correction method considering climatic factors[J]. Automation of Electric Power Systems, 2008, 32(13): 92-96.
- [9] 钟光科. 偏最小二乘回归分析在短期负荷预测中的应用[D]. 邯郸: 河北工程大学, 2011.
- [10] 赵登福, 张涛, 杨增辉, 等. 基于 GN-BFGS 算法的 RBF 神经网络短期负荷预测[J]. 电力系统自动化, 2003, 27(4): 1-4.
- ZHAO Dengfu, ZHANG Tao, YANG Zenghui, et al. Short-term load forecasting using radial basis function (RBF) neural networks based on GN-BFGS algorithm[J]. Automation of Electric Power Systems, 2003, 27(4): 1-4.
- [11] 舒菲. 基于 RBF 神经网络与模糊理论的电力系统短期负荷预测[D]. 西安: 西安理工大学, 2008.
- [12] BAKIRTZIS A G, THEOCHARIS J B. Short term load forecasting using fuzzy neural networks[J]. IEEE Trans on Power Systems, 1995, 10(3): 1518-1524.
- [13] 畅广辉, 刘涤尘, 熊浩. 基于多分辨率 SVM 回归估计的短期负荷预测[J]. 电力系统自动化, 2007, 31(9): 37-41.
- CHANG Guanghui, LIU Dichen, XIONG Hao. Short term load forecasting based on multi-resolution SVM regression[J]. Automation of Electric Power Systems, 2007, 31(9): 37-41.
- [14] 黄帅栋, 卫志农, 高宗和, 等. 基于非负矩阵分解的相关向量机短期负荷预测模型[J]. 电力系统自动化, 2012, 36(11): 62-66.
- HUANG Shuaidong, WEI Zhinong, GAO Zonghe, et al. A short-term load forecasting model based on relevance vector machine with nonnegative matrix factorization[J]. Automation of Electric Power Systems, 2012, 36(11): 62-66.
- [15] 王奔, 冷北雪, 张喜海, 等. 支持向量机在短期负荷预测中的应用概况[J]. 电力系统及其自动化学报, 2011, 23(4): 115-121.
- WANG Ben, LENG Beixue, ZHANG Xihai, et al. Application profiles of support vector machine in short-term load forecasting[J]. Proceedings of the CSU-EPSA, 2011, 23(4): 115-121.
- [16] 王德文, 孙志伟. 电力用户侧大数据分析与并行负荷预测[J]. 中国电机工程学报, 2015, 35(3): 527-537.
- WANG Dewen, SUN Zhiwei. Big data analysis and parallel load forecasting of electric power user side[J]. Proceedings of the CSEE, 2015, 35(3): 527-537.
- [17] 张素香, 赵丙镇, 王风雨, 等. 海量数据下的电力负荷短期预测[J]. 中国电机工程学报, 2015, 35(1): 37-42.
- ZHANG Suxiang, ZHAO Bingzhen, WANG Fengyu, et al. Short-term power load forecasting based on big data[J]. Proceedings of the CSEE, 2015, 35(1): 37-42.
- [18] 邵喜高. 基于统计学习理论的多核预测模型研究及应用[D]. 长沙: 中南大学, 2013.
- [19] 曹懿. 混合核函数支持向量机在信用评估中的应用[D]. 哈尔滨: 哈尔滨工业大学, 2007.
- [20] 王行甫, 俞璐. 混合核函数中权重求解方法[J]. 计算机系统应用, 2015, 24(4): 129-133.
- WANG Xingfu, YU Lu. Weight solving method in hybrid kernel function[J]. Computer Systems & Application, 2015, 24(4): 129-133.
- [21] 罗林开. 支持向量机的核选择[D]. 厦门: 厦门大学, 2007.
- 吴倩红(1991—), 女, 博士研究生, 主要研究方向: 大数据在智能电网中的应用. E-mail: mailtowqh@163.com
- 高 军(1972—), 女, 高级工程师, 主要研究方向: 电网调度. E-mail: hzldkzx@sina.com
- 侯广松(1974—), 男, 硕士, 高级工程师, 主要研究方向: 电网稳定与经济运行. E-mail: hezehougong@sina.com
- 韩 蓓(1984—), 女, 通信作者, 博士, 讲师, 主要研究方向: 含新能源接入的电力系统分析、微电网分析与控制. E-mail: han_bei@sjtu.edu.cn

(编辑 万志超)

(下转第 92 页 continued on page 92)

photovoltaic power stations[J]. Proceedings of the CSEE, 2013, 33(36): 10-17.

- [17] CAMM E H, BEHNKE M R, BOLADO O, et al. Wind power plant collector system design considerations: IEEE PES wind plant collector system design working group[C]// IEEE Power & Energy Society General Meeting, July 26-30, 2009, Calgary, Canada.
- [18] KOCEWIAK L H. Harmonics in large offshore wind farms [D]. Aalborg: Aalborg University, 2012.
- [19] XIAOMING Y, MERK W, STEMLER H, et al. Stationary-frame generalized integrators for current control of active power filters with zero steady-state error for current harmonics of concern under unbalanced and distorted operating conditions [J]. IEEE Trans on Industry Applications, 2002, 38(2):

523-532.

严亚兵(1989—),男,通信作者,博士研究生,主要研究方向:含大规模变流器接入的系统稳定分析与控制。E-mail: yb_yan@hust.edu.cn

苗 淼(1982—),男,博士,主要研究方向:电力系统稳定与控制。E-mail: miaomiao_b116@sina.cn

袁小明(1966—),男,博士,教授,博士生导师,主要研究方向:可再生能源发电设备及其控制、含大规模可再生能源电力系统的控制与保护、大容量变换器工业应用。E-mail: yuanxm@hust.edu.cn

(编辑 章黎)

Impact of Renewable Power Plant Collector System on Stability of Grid-connected Multi-converters

YAN Yabing¹, MIAO Miao², YUAN Xiaoming¹

(1. State Key Laboratory of Advanced Electromagnetic Engineering and Technology, Huazhong University of Science and Technology, Wuhan 430074, China; 2. State Grid Qinghai Electric Power Company, Xining 810000, China)

Abstract: The distributed parameters of collector feeders on the photovoltaic and wind farm have an important influence on the resonance of converters based on LCL filters, presenting a challenge to converter control system design. In order to investigate the impact of the renewable power plant collector system on the stability of grid-connected multi-converters, a mathematical model for the multi-converter system is developed using the network admittance matrix in the complex frequency domain, and frequency characteristics are analyzed. It is shown that the influence on the stability of converters varies with topological form and feeder length of different feeder types in the collector system. The validity of the relevant conclusions is verified by simulation results based on MATLAB/Simulink.

This work is supported by National Natural Science Foundation of China (No. 51190104), National Basic Research Program of China (973 Program) (No. 2012CB215100) and State Grid Corporation of China (No. 52283014003).

Key words: grid-connected multi-converter system; current control loop; collector system; topology; renewable energy

(上接第 72 页 continued from page 72)

Short-term Load Forecasting Support Vector Machine Algorithm Based on Multi-source Heterogeneous Fusion of Load Factors

WU Qianhong¹, GAO Jun², HOU Guangsong², HAN Bei¹, WANG Keyou¹, LI Guojie¹

(1. Key Laboratory of Control of Power Transmission and Transformation (Shanghai Jiao Tong University), Ministry of Education, Shanghai 200240, China;

2. Heze Power Supply Company of State Grid Shandong Electric Power Company, Heze 274000, China)

Abstract: A method to select optimal multiple kernels developed from multiple kernel function is proposed for short-term load forecasting in the big data environment of smart grid, multi-source heterogeneous load factors taken into account. The multiple kernel function is able to describe the distribution characteristics of the factors, cope with their variations and improve the accuracy of load forecasting. Load factors such as historical load, air temperature, air pressure, relative humidity, rainfall, wind direction, wind speed, holidays and electricity price are selected as multi-source heterogeneous factors. Three methods (the sample distribution method, single variable method and rank space diversity method) are adopted to establish optimal multiple kernels, and parallel multiple kernel support vector machine (SVM) load forecasting algorithm is based on double layer multi kernel learning algorithm. A Hadoop cluster is built for conducting experiments of short-term load forecasting. Experimental results show that the average relative error of multiple kernel SVM is smaller than single kernel SVM's, and the accuracy of multiple kernel SVM model based on double layer multiple kernel learning algorithm and norm is the highest. Therefore, multiple kernel SVM can tackle the multi-source heterogeneous data in the load forecasting effectively, and the speed and accuracy of load forecasting can be improved by parallel processing.

This work is supported by National Natural Science Foundation of China (No. 51407116) and National Key Technologies R&D Program (No. 2015BAA01B02).

Key words: big data; multi-source heterogeneous characteristics; support vector machine (SVM); load forecasting; paralleling