

## 面向物联网隐私数据分析的分布式弹性网络回归学习算法

方维维<sup>\*①</sup> 刘梦然<sup>①</sup> 王云鹏<sup>①</sup> 李阳阳<sup>②</sup> 安竹林<sup>③</sup>

<sup>①</sup>(北京交通大学计算机与信息技术学院 北京 100044)

<sup>②</sup>(社会安全风险感知与防控大数据应用国家工程实验室 北京 100041)

<sup>③</sup>(中国科学院计算技术研究所 北京 100190)

**摘要:** 为了解决基于集中式算法的传统物联网数据分析处理方式易引发网络带宽压力过大、延迟过高以及数据隐私安全等问题, 该文针对弹性网络回归这一典型的线性回归模型, 提出一种面向物联网(IoT)的分布式学习算法。该算法基于交替方向乘子法(ADMM), 将弹性网络回归目标优化问题分解为多个能够由物联网节点利用本地数据进行独立求解的子问题。不同于传统的集中式算法, 该算法并不要求物联网节点将隐私数据上传至服务器进行训练, 而仅仅传递本地训练的中间参数, 再由服务器进行简单整合, 以这样的协作方式经过多轮迭代获得最终结果。基于两个典型数据集的实验结果表明: 该算法能够在几十轮迭代内快速收敛到最优解。相比于由单个节点独立训练模型的本地化算法, 该算法提高了模型结果的有效性和准确性; 相比于集中式算法, 该算法在确保计算准确性和可扩展性的同时, 可有效地保护个体隐私数据的安全性。

**关键词:** 物联网; 隐私保护; 弹性网络回归; 分布式算法; 交替方向乘子法

中图分类号: TP391.4

文献标识码: A

文章编号: 1009-5896(2020)00-0001-09

DOI: [10.11999/JEIT190739](https://doi.org/10.11999/JEIT190739)

## A Distributed Elastic Net Regression Algorithm for Private Data Analytics in Internet of Things

FANG Weiwei<sup>①</sup> LIU Mengran<sup>①</sup> WANG Yunpeng<sup>①</sup> LI Yangyang<sup>②</sup> AN Zhulin<sup>③</sup>

<sup>①</sup>(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

<sup>②</sup>(National Engineering Laboratory for Public Safety Risk Perception and Control, Beijing 100041, China)

<sup>③</sup>(Institute of Computing Technology, Chinese Academy of Science, Beijing 100190, China)

**Abstract:** In order to solve the problems caused by the traditional data analysis based on the centralized algorithm in the IoT, such as excessive bandwidth occupation, high communication latency and data privacy leakage, considering the typical linear regression model of elastic net regression, a distributed learning algorithm for Internet of Things (IoT) is proposed in this paper. This algorithm is based on the the Alternating Direction Method of Multipliers (ADMM) framework. It decomposes the objective problem of elastic net regression into several sub-problems that can be solved independently by each IoT node using its local data. Different from traditional centralized algorithms, the proposed algorithm does not require the IoT node to upload its private data to the server for training, but rather the locally trained intermediate parameters to the server for aggregation. In such a collaborative manner, the server can finally obtain the objective model after several iterations. The experimental results on two typical datasets indicate that the proposed algorithm can quickly converge to the optimal solution within dozens of iterations. As compared to the localized algorithm in which each node trains the model solely based on its own local data, the proposed algorithm improves the validity and the accuracy of training models; as compared to the centralized algorithm, the proposed algorithm can guarantee the accuracy and the scalability of model training, and well protect the individual private data from leakage.

**Key words:** Internet of Things(IoT); Privacy protection; Elastic net regression; Distributed algorithm; Alternating Direction Method of Multipliers (ADMM)

收稿日期: 2019-09-25; 改回日期: 2020-xx-xx; 网络出版: 2020-05-17

\*通信作者: 方维维 wwfang@bjtu.edu.cn

基金项目: 北京市自然科学基金(L191019), 赛尔网络下一代互联网创新项目(NGII20190308)

Foundation Items: The Beijing Municipal Natural Science Foundation (L191019), The CERNET Innovation Project (NGII20190308)

## 1 引言

随着物联网(Internet of Things, IoT)技术的广泛应用,物联网节点数目不断增加,然而由于资源所限,节点无法完成计算密集型任务。因此面向物联网数据分析的传统机器学习通常基于集中式算法,由具备超强存储和计算能力的专用服务器对数据进行集中式地处理<sup>[1]</sup>。然而进入万物互联和大数据时代<sup>[2]</sup>以来,节点数量急剧增加的同时产生海量实时数据,基于集中式算法的传统机器学习对如此大规模的物联网数据进行分析处理时,主要面临3方面挑战:首先,直接将节点的海量数据上传至服务器,会造成网络带宽压力过大和计算资源的浪费<sup>[1]</sup>;其次,传统的集中式算法在应用于求解大规模机器学习问题时可扩展性较差<sup>[3]</sup>;最后,物联网数据若直接上传至服务器或其他设备进行训练,会面临数据隐私泄露的风险<sup>[4]</sup>。事实上,为了解决大数据下传统机器学习面临的问题,各分布式数据并行平台纷纷研发分布式机器学习库,然而由于资源限制<sup>[5]</sup>和隐私问题<sup>[6]</sup>,它们并不适用于资源受限且异构的物联网节点。因此分布式优化算法逐渐涌现出来,将服务器难以完成的计算任务拆分成多个小计算任务分布式地部署到多个物联网节点上执行,然后将各节点的执行结果整合成最终结果并返回。相比于传统的集中式算法,分布式优化算法能够减轻网络带宽压力、降低通信成本,保护数据的隐私性。

本文主要研究如何利用分布式优化算法对物联网数据进行回归分析,重点研究目标是弹性网络回归<sup>[7]</sup>这一典型的线性回归技术。本文提出一种基于多个物联网节点的协同弹性网络回归问题模型。针对该模型,引入交替方向乘子法(Alternating Direction Method of Multipliers, ADMM)算法<sup>[1]</sup>,提出一种基于ADMM的分布式弹性网络回归学习算法,将需要由服务器集中式求解的目标优化问题分解成多个可以由物联网节点进行分布式独立求解的子问题。该算法并不要求节点将原始数据上传至服务器,而是由节点独立处理数据,仅仅向服务器传递中间结果,再由服务器整合并返回最终结果。服务器与节点之间以这种协作方式进行多次迭代直至模型收敛。为了验证所提算法的有效性以及评估该算法的性能,本文在两个典型数据集上进行大量的仿真实验,结果表明:所提算法可在几十轮内快速收敛到最优解;相比于本地化算法,提高了结果的有效性和准确性;相比于集中式算法,目标函数值和预测精度可逼近集中式算法的最优值,而且可降低网络传输带宽压力,提高计算的可扩展性,保护隐私数据的安全性。

## 2 问题模型

### 2.1 弹性网络回归

弹性网络回归作为一种典型的线性回归技术,所解决的优化问题基本形式为

$$\min \frac{1}{2} \sum_{i=1}^N (\mathbf{w}^T \mathbf{x}_i + b - y_i)^2 + u_1 \|\mathbf{w}\|_1 + \frac{u_2}{2} \|\mathbf{w}\|_2^2 \quad (1)$$

其中,  $\{\mathbf{x}_i, y_i\}$  是数据样本,  $\mathbf{x}_i \in R^n$  是特征向量,  $y_i$  是相应的因变量, 特征权重向量  $\mathbf{w} \in R^n$ , 截距  $b \in R$ , 正则化参数  $u_1, u_2 > 0$ 。当  $u_1 = 0$  时, 弹性网络回归退化为岭回归; 当  $u_2 = 0$  时, 则退化为 Lasso 回归

$$y = \mathbf{w}^T \mathbf{x} + b \quad (2)$$

弹性网络回归模型建立的目标是通过训练求得得到的  $(\mathbf{w}, b)$  的值, 根据式(2), 对于给定的特征向量  $\mathbf{x} \in R^n$  能准确地预测出因变量  $y$  的值。虽然弹性网络回归具备很好的预测性能<sup>[8]</sup>, 但集中式算法增大网络带宽压力和隐私数据泄露的风险。即使采用网络安全机制, 服务器端仍然存在泄露用户隐私数据的可能。

### 2.2 问题模型

基于弹性网络回归问题模型的基本形式(1), 本文进一步研究基于多个物联网节点的协同弹性网络回归学习问题。考虑由一个中心服务器与  $N$  个物联网节点组成的物联网系统, 其中每个节点拥有相同的传感器组。首先, 物联网节点  $i \in \{1, 2, \dots, N\}$  在一定时间内将通过板载传感器生成的原始数据转换为特征向量, 每个特征向量包含  $n$  个预测变量, 即  $\mathbf{x}_{ij} \in R^n$ , 并对应于一个因变量  $y_{ij} \in R$ , 其中  $j \in \{1, 2, \dots, M_i\}$ ,  $M_i$  是节点  $i$  提供的数据样本个数; 其次, 物联网节点  $i$  将数据样本  $D_i = \{(\mathbf{x}_{ij}, y_{ij}), j = 1, 2, \dots, M_i\}$  由本地上传至服务器, 由服务器对收集到的数据进行回归分析; 最后, 通过建立特征向量与因变量之间的回归模型, 可以通过特征向量准确地预测出因变量的值。此时弹性网络回归解决的优化问题形式为

$$\min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^{M_i} (\mathbf{w}^T \mathbf{x}_{ij} + b - y_{ij})^2 + u_1 \|\mathbf{w}\|_1 + \frac{u_2}{2} \|\mathbf{w}\|_2^2 \quad (3)$$

其中, 特征向量  $\mathbf{x}_{ij} \in R^n$ , 对应的因变量  $y_{ij} \in R$ , 特征权重向量  $\mathbf{w} \in R^n$ , 截距  $b \in R$ , 正则化参数  $u_1, u_2 > 0$ 。训练所得回归模型可以根据给定的新的特征向量  $\mathbf{x}_{ij}$  预测出因变量  $y_{ij}$  的值。

### 3 基于ADMM的分布式弹性网络回归学习算法

#### 3.1 ADMM算法概述

ADMM算法是一个简洁高效的分布式优化算法，它独特的分布式架构而非常适用于分布式环境下的并行求解<sup>[1]</sup>。它解决的优化问题形式为

$$\left. \begin{aligned} \min \quad & F(x) + G(y) \\ \text{s.t.} \quad & Ax + By = C, x \in X, y \in Y \end{aligned} \right\} \quad (4)$$

其中， $F$ 和 $G$ 是凸函数， $X$ 和 $Y$ 是非空凸集， $x$ 和 $y$ 是两个需要优化的原始变量， $A, B, C$ 是等式约束的参数。目标函数关于原始变量 $x, y$ 可以分解成两个函数之和的形式，因此可使用ADMM算法求得最优解。

利用增广拉格朗日法<sup>[9]</sup>对原问题式(4)进行求解，可得到增广拉格朗日函数形式为

$$\begin{aligned} L_\rho(x, y, \lambda) = & F(x) + G(y) + \lambda^T(Ax + By - C) \\ & + \frac{\rho}{2} \|Ax + By - C\|_2^2 \end{aligned} \quad (5)$$

其中， $\lambda \geq 0$ 是对偶变量， $\rho > 0$ 是惩罚参数。

原始的最小化问题式(4)转化为拉格朗日函数式(5)中原始变量 $x, y$ 最小化，对偶变量 $\lambda$ 最小化问题。在每轮迭代过程中，首先原始变量 $x, y$ 进行交替优化，然后再更新对偶变量 $\lambda$ 。求解过程总结为

$$x^{k+1} = \arg \min_x L_\rho(x, y^k, \lambda^k) \quad (6)$$

$$y^{k+1} = \arg \min_y L_\rho(x^k, y, \lambda^k) \quad (7)$$

$$\lambda^{k+1} = \lambda^k + \rho(Ax^{k+1} + By^{k+1} - C) \quad (8)$$

文献<sup>[10]</sup>给出了ADMM算法收敛性的证明，在中等精度要求下，经过几十次迭代后即可收敛<sup>[1]</sup>。

#### 3.2 基于ADMM的分布式弹性网络回归学习算法

由于问题模型式(3)无法根据两个变量分解成两个函数之和的形式，不能使用ADMM算法进行求解。本文引入一组辅助变量 $\{(\mathbf{w}_i, b_i), i = 1, 2, \dots, N\}$ ，由此可转化为与原问题等价的新的优化问题，具体形式为

$$\left. \begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^{M_i} (\mathbf{w}_i^T \mathbf{x}_{ij} + b_i - y_{ij})^2 \\ & + u_1 \|\mathbf{w}\|_1 + \frac{u_2}{2} \|\mathbf{w}\|_2^2 \\ \text{s.t.} \quad & \mathbf{w}_i = \mathbf{w}, b_i = b, i = 1, \dots, N. \end{aligned} \right\} \quad (9)$$

其中， $\{(\mathbf{w}_i, b_i), i = 1, 2, \dots, N\}$ 是物联网节点 $i$ 的中间参数， $\{(\mathbf{w}, b)\}$ 是服务器整合节点的中间参数得到的全局参数。目标函数可以根据两个参数 $\{(\mathbf{w}, b)\}$ 和 $\{(\mathbf{w}_i, b_i), i = 1, 2, \dots, N\}$ 分解成两个函数

之和的形式，因此可使用ADMM算法进行求解。利用增广拉格朗日法对问题式(9)进行求解，可得增广拉格朗日函数形式为

$$\begin{aligned} L_\rho(\alpha, \beta, \gamma) = & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^{M_i} (\mathbf{w}_i^T \mathbf{x}_{ij} + b_i - y_{ij})^2 \\ & + u_1 \|\mathbf{w}\|_1 + \frac{u_2}{2} \|\mathbf{w}\|_2^2 \\ & + \sum_{i=1}^N ((\mathbf{w}_i - \mathbf{w})^T \gamma_{i,w} + (b_i - b) \gamma_{i,b}) \\ & + \sum_{i=1}^N \frac{\rho}{2} ((\mathbf{w}_i - \mathbf{w})^T (\mathbf{w}_i - \mathbf{w}) + (b_i - b)^2) \end{aligned} \quad (10)$$

其中， $\alpha = \{(\mathbf{w}, b)\}$ ， $\beta = \{(\mathbf{w}_i, b_i), i = 1, 2, \dots, N\}$ ， $\gamma = \{(\gamma_{i,w}, \gamma_{i,b}), i = 1, 2, \dots, N\}$ 。通过对拉格朗日函数式(10)中的参数 $\alpha, \beta$ 和 $\gamma$ 的迭代更新，可最终求得原问题的最优解，接下来将分别介绍各参数的更新过程。

(1)  $\alpha$ -更新部分： $\alpha$ 更新时需要解决的优化问题具有如式(11)的形式

$$\begin{aligned} \min_{\alpha^{k+1}} \quad & u_1 \|\mathbf{w}\|_1 + \frac{u_2}{2} \|\mathbf{w}\|_2^2 + \frac{\rho N}{2} \mathbf{w}^T \left( \mathbf{w} - 2\bar{\mathbf{w}}^k - \frac{2\bar{\gamma}_{\mathbf{w}}^k}{\rho} \right) \\ & + \frac{\rho N}{2} b \left( b - 2\bar{b}^k - \frac{2\bar{\gamma}_b^k}{\rho} \right) \end{aligned} \quad (11)$$

其中， $\bar{\theta}^k$ 为在第 $k$ 次迭代时，向量 $\theta_i (i = 1, 2, \dots, N)$ 的均值。该优化问题由于包含L1范数而不可微，因此本文采用次梯度演算法(subgradient algorithm)对其进行求解。

令  $\eta = \frac{\rho N}{\rho N + u_2} \left( \bar{\mathbf{w}}^k + \frac{\bar{\gamma}_{\mathbf{w}}^k}{\rho} \right)$ ， $\varphi = \frac{u_1}{\rho N + u_2}$ ，求解结果具有如式(12)的形式

$$\mathbf{w}^{k+1} = \begin{cases} \eta - \varphi, & \eta > \varphi \\ 0, & \eta \in [-\varphi, \varphi] \\ \eta + \varphi, & \eta < -\varphi \end{cases} \quad (12)$$

$$b^{k+1} = \bar{b}^k + \frac{\bar{\gamma}_b^k}{\rho} \quad (13)$$

(2)  $\beta$ -更新部分：当 $\alpha$ 通过更新得到 $\alpha^{k+1}$ 即 $\{(\mathbf{w}^{k+1}, b^{k+1})\}$ 后， $\beta$ 更新时需要解决的优化问题具有如式(14)的形式

$$\begin{aligned} \min_{\beta^{k+1}} \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^{M_i} (\mathbf{w}_i^T \mathbf{x}_{ij} + b_i - y_{ij})^2 \\ & + \sum_{i=1}^N \frac{\rho}{2} \mathbf{w}_i^T \left( \mathbf{w}_i - 2\mathbf{w}^{k+1} + \frac{2\gamma_{i,w}^k}{\rho} \right) \\ & + \sum_{i=1}^N \frac{\rho}{2} b_i \left( b_i - 2b^{k+1} + \frac{2\gamma_{i,b}^k}{\rho} \right) \end{aligned} \quad (14)$$

问题式(14)可以分解成 $N$ 个独立的子问题,并部署到多个物联网节点求解,节点 $i$

$$\begin{aligned} \min_{\beta_i} & \frac{1}{2} \sum_{j=1}^{M_i} (\mathbf{w}_i^T \mathbf{x}_{ij} + b_i - y_{ij})^2 \\ & + \frac{\rho}{2} \mathbf{w}_i^T \left( \mathbf{w}_i - 2\mathbf{w}^{k+1} + \frac{2\gamma_{i,w}^k}{\rho} \right) \\ & + \frac{\rho}{2} b_i \left( b_i - 2b^{k+1} + \frac{2\gamma_{i,b}^k}{\rho} \right) \end{aligned} \quad (15)$$

问题式(15)是一个典型的非线性规划问题,本文引入PRP(Polak, Ribiere and Polyak, PRP)共轭梯度法<sup>[11]</sup>对其进行求解,具体流程详见表1。首先,将问题式(15)看作关于 $\mathbf{w}_i$ 的函数 $F(\mathbf{w}_i)$ ,令 $b_i = b_i^*$ 求得使函数 $F(\mathbf{w}_i)$ 最小化的最优解 $\mathbf{w}_i^*$ ;其次,固定 $\mathbf{w}_i = \mathbf{w}_i^*$ ,将问题式(15)看作关于 $b_i$ 的函数 $F(b_i)$ ,求得使函数 $F(b_i)$ 最小化的最优解 $b_i^*$ ;最后,两组变量以这种方式多次交替更新可以求得最优解 $\mathbf{w}_i^{t+1} =$

$\mathbf{w}_i^*, b_i^{t+1} = b_i^*$ 。篇幅所限,本文仅介绍 $\mathbf{w}_i$ 的求解方法,对 $b_i$ 的求解同理。物联网节点 $i$ 关于 $\mathbf{w}_i$ 的目标优化问题为

$$\begin{aligned} F(\mathbf{w}_i) = & \frac{1}{2} \sum_{j=1}^{M_i} (\mathbf{w}_i^T \mathbf{x}_{ij} + b_i - y_{ij})^2 + \\ & \frac{\rho}{2} \mathbf{w}_i^T \left( \mathbf{w}_i - 2\mathbf{w}^{k+1} + \frac{2\gamma_{i,w}^k}{\rho} \right) \\ & + \frac{\rho}{2} b_i \left( b_i - 2b^{k+1} + \frac{2\gamma_{i,b}^k}{\rho} \right) \end{aligned} \quad (16)$$

(3)  $\gamma$ -更新部分:  $\alpha$ 和 $\beta$ 通过更新得到 $\alpha^{k+1} = \{(\mathbf{w}^{k+1}, b^{k+1})\}$ ,  $\beta^{k+1} = \{(\gamma_{i,w}^{k+1}, \gamma_{i,b}^{k+1})\}$ 后,  $\gamma$ 更新为

$$\gamma_{i,w}^{k+1} = \gamma_{i,w}^k + \rho(\mathbf{w}_i^{k+1} - \mathbf{w}^{k+1}) \quad (17)$$

$$\gamma_{i,b}^{k+1} = \gamma_{i,b}^k + \rho(b_i^{k+1} - b^{k+1}) \quad (18)$$

图1说明了分布式弹性网络回归学习算法的计

表1 PRP共轭梯度算法流程

算法1 PRP共轭梯度算法流程

输入: 特征向量 $\mathbf{x}_{ij}$ ; 相应变量 $y_{ij}$ ; 服务器提供的参数 $\alpha = \{(\mathbf{w}^{k+1}, b^{k+1})\}$ ; 对偶变量 $\gamma^k = \{(\gamma_{i,w}^k, \gamma_{i,b}^k)\}$ ;  $b_i^*$ ;

输出: 物联网节点 $i$ 的局部最优解 $\mathbf{w}_i^*$ ;

(1) 初始迭代次数 $t = 0$ , 初始向量 $\mathbf{w}_i^0 = 0$ , 收敛精度 $\varepsilon = 1e - 5$ , 初始搜索方向 $\mathbf{p}^0 = -g(\mathbf{w}_i^0)$ ;

(2) repeat /\*算法进行迭代\*/

(3) for  $j = -1:2:1$

(4) if  $F(\mathbf{w}_i^t + \lambda^t \mathbf{p}^t) > F(\mathbf{w}_i^t + j * \mathbf{p}^t)$  then

(5)  $\lambda^t \leftarrow j$ ;

(6) end if

(7) end for

(8)  $\mathbf{w}_i^{t+1} \leftarrow \mathbf{w}_i^t + \lambda^t \mathbf{p}^t$ ;

(9)  $\beta^t \leftarrow \frac{g(\mathbf{w}_i^{t+1})^T (g(\mathbf{w}_i^{t+1}) - g(\mathbf{w}_i^t))}{g(\mathbf{w}_i^t)^T g(\mathbf{w}_i^t)}$ ;

(10)  $\mathbf{p}^{t+1} = -g(\mathbf{w}_i^{t+1}) + \beta^t \mathbf{p}^t$ ;

(11)  $t \leftarrow t + 1$ ;

(12) until  $\|g(\mathbf{w}_i^t)\| \leq \varepsilon$ ; /\*算法达到收敛准则, 停止迭代\*/

(13)  $\mathbf{w}_i^* \leftarrow \mathbf{w}_i^t$ ;

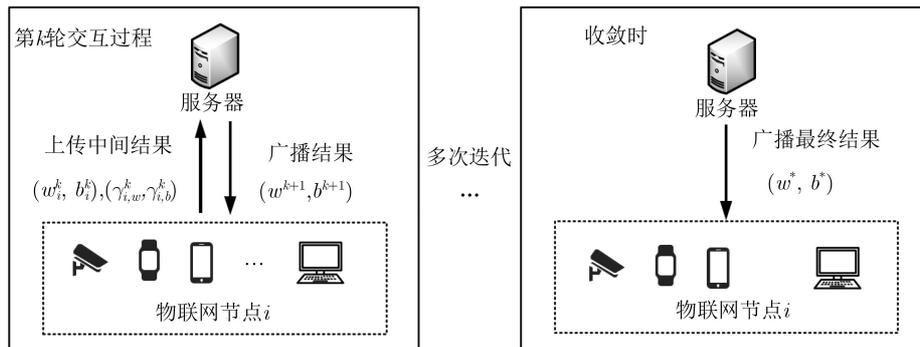


图1 分布式弹性网络回归学习算法计算流程

算流程。本文采用原始残差 $r^k$ 和对偶残差 $s^k$ 共同作为算法的收敛标准<sup>[1]</sup>，记 $\varepsilon^{\text{rel}}$ 和 $\varepsilon^{\text{abs}}$ 分别是原始残差和对偶残差的偏差阈值，取经验值为 $\varepsilon^{\text{rel}} = 1e - 2$ ， $\varepsilon^{\text{abs}} = 1e - 4$ 。当 $\|r^k\|_2 \leq \varepsilon^{\text{rel}}$ ， $\|s^k\|_2 \leq \varepsilon^{\text{abs}}$ 时，则视为算法达到收敛准则<sup>[1]</sup>。分布式弹性网络回归学习算法的具体流程详见表2。该算法的收敛性及收敛速度证明可参考文献[12,13]，考虑到其复杂性和版面限制，本文不再赘述。

## 4 实验与分析

### 4.1 实验设置

为了验证分布式弹性网络回归学习算法的有效性及其性能，本文在两个典型数据集上进行了仿真实验，其中拟合数据集根据文献[1]的描述生成，包括1500个数据样本，每个样本包含9维特征向量和1个相应的因变量。通过使用数据集，不仅可以验证所提算法的有效性，而且能够在不考虑数据质量的前提下评估各参数对算法性能的影响以及与其它方法进行性能比较。然而由于该拟合数据集数据质量高且分布均匀，缺乏真实性。因此为了进一步评估该算法在实际应用中的性能，本文在真实数据集上进行了仿真实验。该真实数据集则为文献[14]中提到的公开疾病数据集，包含442个患者的数据样本，每个样本包含10个生理特征以及一年以后疾病级数指标。在本文实验中，将数据集按照7:3的比例划分为训练集和测试集。除特殊说明外，实验参数均设置为： $\rho = 1.0$ ， $u_1 = 0.01$ ， $u_2 = 0.01$ ， $\varepsilon^{\text{rel}} = 1e - 2$ ， $\varepsilon^{\text{abs}} = 1e - 4$ <sup>[1]</sup>。为了进一步评估所提算法的性能，本文设计了相关实验将其与传统的集中式算法以及本地化算法两种方法进行比较。

## 4.2 实验结果

### 4.2.1 算法的有效性

首先，为了验证所提算法的收敛性，本文采用拟合数据集，在物联网节点数量 $N$ 的不同取值下进行了多组实验，并观察到对于不同的 $N$ 值，算法均具有良好的收敛性。采用集中式算法计算的目标函数值作为最优值，并以此为基准与分布式算法所得目标函数值进行对比。当 $N = 20$ 时算法的收敛性如下图所示。图2说明随着迭代次数的增加，目标函数值在前50次迭代过程中快速下降，迭代次数为100次时接近集中式算法求得的目标函数值。图3则表示 $\|r\|_2$ 和 $\|s\|_2$ 随迭代次数的变化，最终当迭代次数为227次时，算法达到收敛准则。实验结果表明，本文所提分布式算法可以在有限的迭代次数内收敛并接近集中式算法的目标函数值。

其次，为了进一步评估所提算法的性能，本文采用校正复相关系数( $R_a^2$ )<sup>[15]</sup>和均方根误差(root mean square error, RMSE)分别评估算法模型的拟合效果和预测精度。图4和图5说明在大约前100次迭代过程中，所提算法RMSE值持续降低， $R_a^2$ 值不断增加。当算法达到收敛准则时，RMSE的值为0.03987，逼近集中式算法的预测精度； $R_a^2$ 的值为0.998307趋近于1，表示算法模型具备较好的拟合效果。实验结果表明本文所提算法可以在有限迭代次数内收敛得到接近集中式算法的拟合效果和预测精度。

### 4.2.2 参数对算法性能的影响

为研究各参数的设置对算法性能的影响，本文在拟合数据集上进行了8组实验，以目标函数值和

表 2 分布式弹性网络回归学习算法流程

算法2 分布式弹性网络回归学习算法流程
<b>输入：</b> 物联网节点的样本数据，包括特征向量 $\mathbf{x}_{ij}$ ；相应因变量 $y_{ij}$ ；
<b>输出：</b> 最终结果 $\alpha = \{(\mathbf{w}^*, b^*)\}$ ；
(1) 服务器初始参数设置： $k = 0$ ， $\bar{\mathbf{w}} = 0$ ， $\bar{b}^0 = 0$ ， $\varepsilon^{\text{rel}} = 1e - 2$ ， $\varepsilon^{\text{abs}} = 1e - 4$ ；
(2) 物联网节点 $i$ 参数设置： $k = 0$ ， $\gamma_{i,w}^0 = 0$ ， $\gamma_{i,b}^0 = 0$ ；
(3) Repeat /*算法进行迭代*/
(4) 服务器整合物联网节点上传的中间参数 $(\mathbf{w}_i^k, b_i^k)$ 及 $(\gamma_{i,w}^k, \gamma_{i,b}^k)$ ，求得各变量均值 $\bar{\mathbf{w}}^k, \bar{b}^k, \bar{\gamma}_w^k, \bar{\gamma}_b^k$ ，根据式(12)和式(13)更新参数 $(\mathbf{w}^{k+1}, b^{k+1})$ ，并将结果广播给物联网节点；
(5) 物联网节点 $i$ 根据服务器提供的参数 $(\mathbf{w}^{k+1}, b^{k+1})$ 对问题式(14)进行求解得到参数 $(\mathbf{w}_i^{k+1}, b_i^{k+1})$ ；
(6) 物联网节点 $i$ 根据式(17)和式(18)更新对偶变量 $(\gamma_{i,w}^{k+1}, \gamma_{i,b}^{k+1})$ ；
(7) 物联网节点 $i$ 向服务器发送新的中间参数 $(\mathbf{w}_i^{k+1}, b_i^{k+1})$ 及 $(\gamma_{i,w}^{k+1}, \gamma_{i,b}^{k+1})$ ；
(8) $k \leftarrow k + 1$ ；
(9) until $\ r^k\ _2 \leq \varepsilon^{\text{rel}}$ ， $\ s^k\ _2 \leq \varepsilon^{\text{abs}}$ ；/*算法达到收敛准则，停止迭代*/

RMSE值作为算法性能的评价指标,在保证其他参数固定不变的条件下,分别评估参数 $N, \rho, u_1, u_2$ 对算法性能的影响。其中参数 $N$ 对算法性能的影响如图6所示,可以发现当 $N$ 取不同值时,算法均能收敛,而且 $N$ 值越小时,算法的收敛速度越快,但最终均能得到相同的目标函数值和RMSE值,表明算法具备较好的可扩展性。参数 $\rho$ 对算法性能的影响如图7所示,可以发现 $\rho$ 值较小时,算法可以更快地收敛,目标函数值和RMSE值较大。参数 $u_1$ 和 $u_2$ 对算法性能的影响分别如图8和图9所示,说明 $u_1$ 对算法的收敛速度影响较小,但对目标函数值和RMSE

值影响较大, $u_1$ 值越小,目标函数值和RMSE值就越小;与 $u_1$ 相似, $u_2$ 值越小,目标函数值和RMSE值也越小,不同的是, $u_2$ 对算法的收敛速度影响较大, $u_2$ 值越大,算法的收敛速度越快。

#### 4.2.3 与其它方法性能比较

相比于分布式算法,本地化算法的性能很大程度上与单个物联网节点处理的本地数据集规模大小相关,因此将数据集随机均匀地划分为 $N$ 个训练子集,即由 $N$ 个节点独立训练,并以 $R_a^2$ 和RMSE值作为评价指标,比较在 $N$ 的不同取值下两种算法的性能。值得注意的是,当 $N=1$ 时本地化算法等同于集中式算法。

图10(a)、图10(b)分别表示不同 $N$ 值下两种算法的RMSE值和 $R_a^2$ 值比较。由图10(a)可以发现,

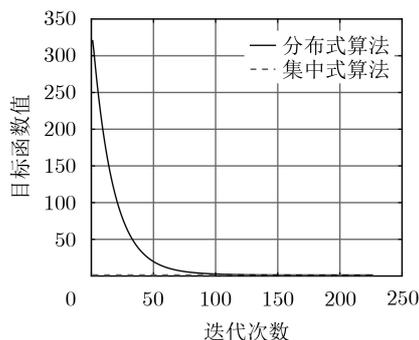
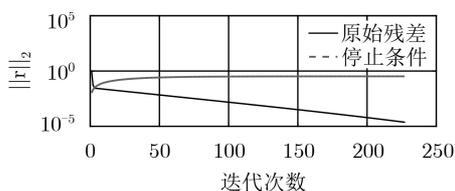
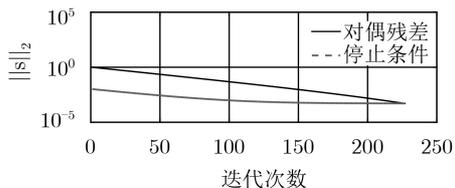


图2 目标函数值随迭代次数变化

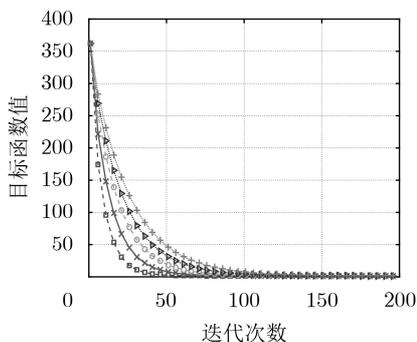


(a) 原始残差随迭代次数变化



(b) 对偶残差随迭代次数变化

图3 原始残差和对偶残差随迭代次数变化



(a) 目标函数值



(b) RMSE

图6 参数 $N$ 对算法性能的影响

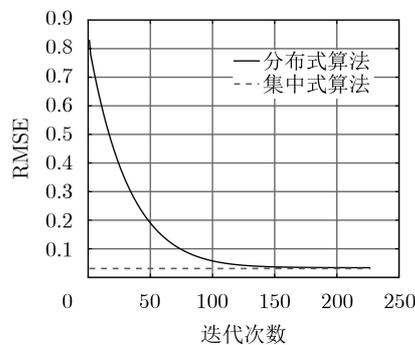


图4 RMSE值随迭代次数变化

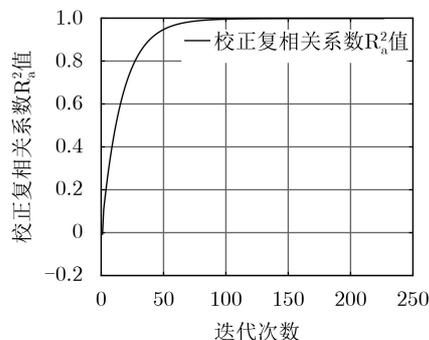


图5 调整复相关系数 $R_a^2$ 值随迭代次数变化

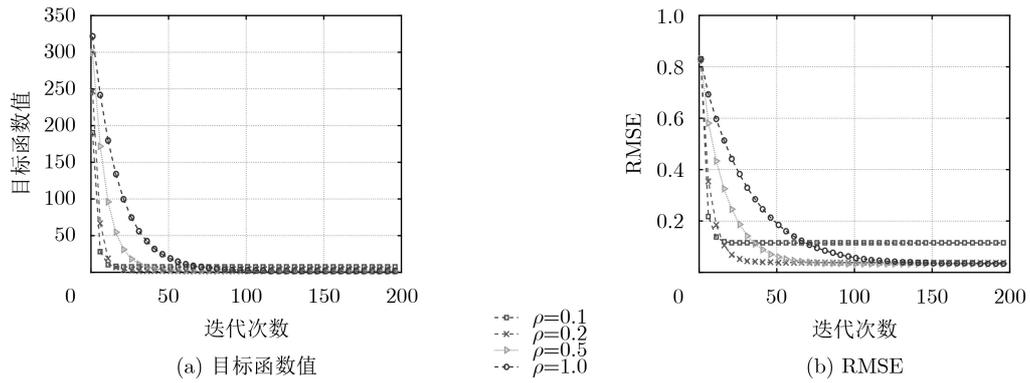


图 7 参数  $\rho$  对算法性能的影响

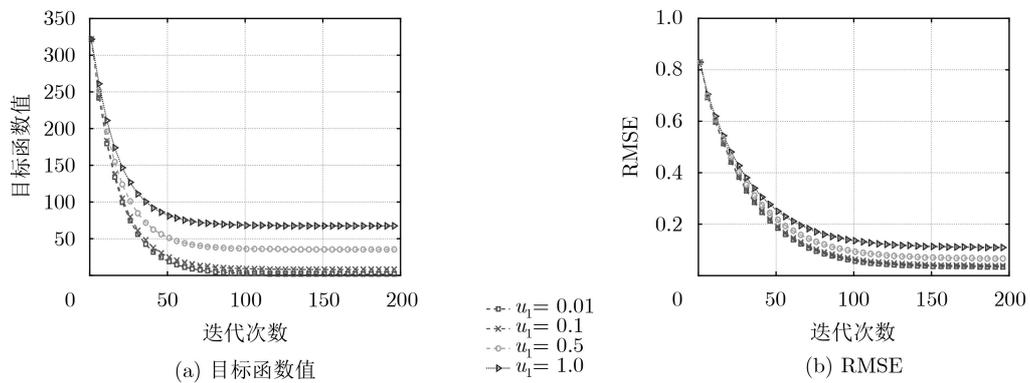


图 8 参数  $u_1$  对算法性能的影响

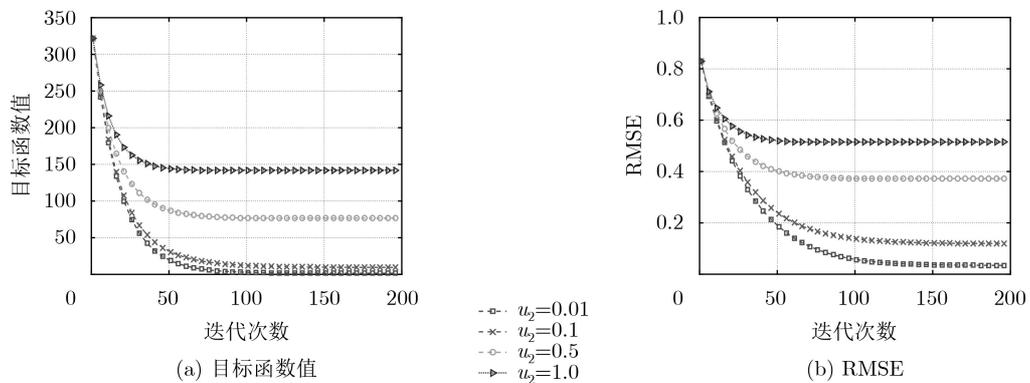


图 9 参数  $u_2$  对算法性能的影响

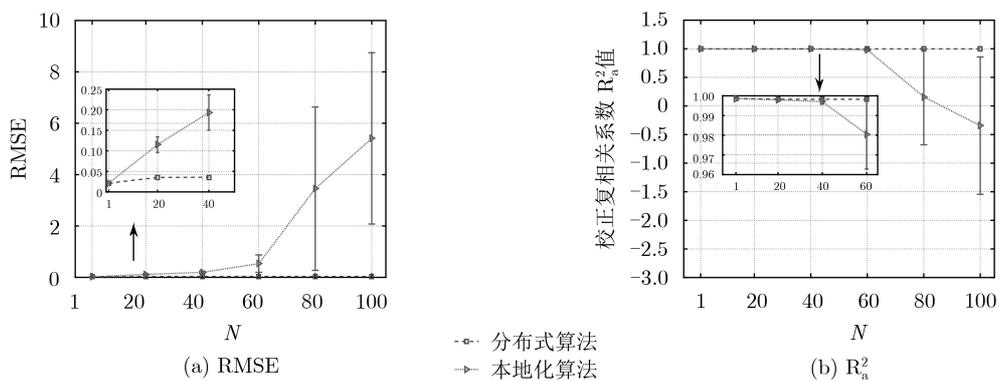


图 10 分布式算法与本地化算法之间的性能比较

所提算法RMSE均值基本不随 $N$ 值变化,始终维持在0.03左右,而本地化算法随着 $N$ 值的增大, RMSE均值不断增大。同时, RMSE值之间的差值也在增加,说明由于单个物联网节点本地数据集规模较小,不同节点之间的训练结果差异较大。由图10(b)可以发现,所提算法 $R_a^2$ 值始终稳定在0.98左右,基本不随 $N$ 值变化。相比之下,本地化算法 $R_a^2$ 值受 $N$ 值的影响较大。在 $N < 60$ 时 $R_a^2$ 值非常接近本文所提算法,然而当 $N > 60$ 时, $R_a^2$ 值显著下降,当 $N = 80$ 时,其值已经降为负数,表明此时该模型对数据集并没有拟合效果。实验结果表明,在单个物联网节点的本地数据集规模较小且缺

乏多样性的情况下,本地化算法训练得到的弹性网络回归模型性能较差,不仅预测精度降低而且拟合效果也显著下降。相比之下,本文所提算法可以利用多个物联网节点提供的样本数据,因而始终可以收敛到接近最优的状态并获得接近集中式算法性能

#### 4.2.4 应用于真实数据集的实验结果

为了进一步评估分布式算法在真实场景下的性能表现,本文使用相同的评价指标在疾病数据集<sup>[14]</sup>上进行了相关实验。由于篇幅所限,本文只列出相比于拟合数据集表现出显著不同的实验结果如图11—图13所示。

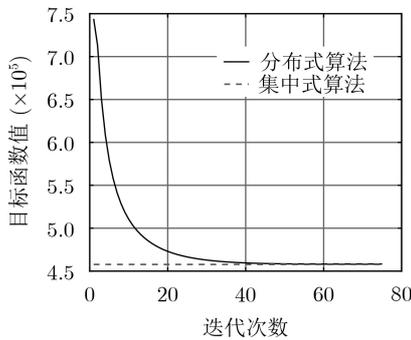


图 11 目标函数值随迭代次数变化

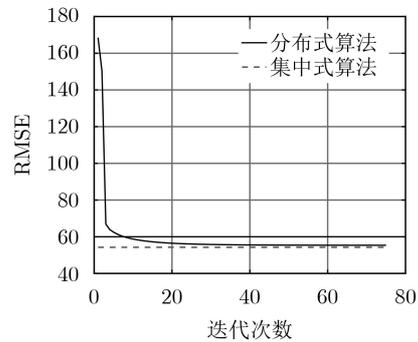


图 12 RMSE随迭代次数变化

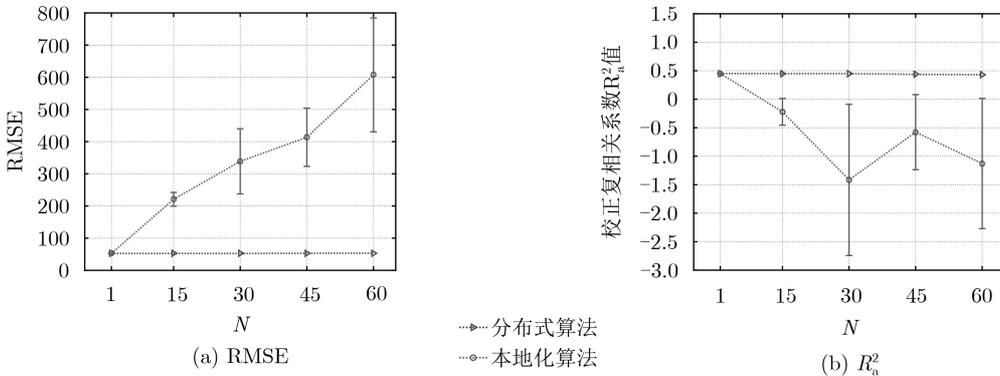


图 13 分布式算法与本地化算法之间的性能比较

首先,如图11和图12所示,当 $N = 20$ 时,本文所提分布式算法的目标函数值和RMSE值在大约前5次迭代中快速下降,大约40次迭代后就已经接近集中式方法的最优解,这种快速收敛的现象与真实数据集的规模较小有关。其次,图13说明本地化算法是3种算法中性能表现最差的。当 $N = 15$ 时,本地化算法的RMSE值远远大于集中式算法和本文所提分布式算法,而且其 $R_a^2$ 值已经降为负数,由此可以发现,当样本数据规模较小且数据缺乏多样性时,单个物联网节点很难通过对本地数据进行独立训练得到一个好的模型。最后,实验结果说明,在

应用于真实数据集时,所提算法仍然能够得到接近集中式算法性能

## 5 结束语

本文面向物联网数据提出一种分布式弹性网络回归学习算法,该算法基于ADMM算法,将需要集中式求解的弹性网络回归目标优化问题分解为多个可以由物联网节点利用本地数据进行独立求解的子问题。该算法不要求节点向服务器上传原始数据,仅需上传中间结果,由服务器进行简单整合得到最终结果并返回。本文在两个典型数据集上的实

验结果表明: 该算法能够在几十轮迭代内快速收敛到最优解; 所得目标函数值和预测精度接近集中式算法, 相比于集中式算法, 既减轻带宽压力又保护数据隐私性; 相比于本地化算法, 提高了计算结果的有效性和准确性。接下来的研究工作中, 我们将进一步研究分布式优化算法在其他机器学习问题中的应用, 以及在物联网实验中采用该算法解决实际问题, 进一步评估它在实际应用中的性能表现。

### 参考文献

- [1] BOYD S, PARIKH N, CHU E, *et al.* Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. *Foundations and Trends® in Machine Learning*, 2011, 3(1): 1–122. doi: [10.1561/22000000016](https://doi.org/10.1561/22000000016).
- [2] 邵振峰, 蔡家骏, 王中元, 等. 面向智能监控摄像头的监控视频大数据分析处理[J]. *电子与信息学报*, 2017, 39(5): 1116–1122. doi: [10.11999/JEIT160712](https://doi.org/10.11999/JEIT160712).  
SHAO Zhenfeng, CAI Jiajun, WANG Zhongyuan, *et al.* Analytical processing method of big surveillance video data based on smart monitoring cameras[J]. *Journal of Electronics & Information Technology*, 2017, 39(5): 1116–1122. doi: [10.11999/JEIT160712](https://doi.org/10.11999/JEIT160712).
- [3] DHAR S, YI C R, RAMAKRISHNAN N, *et al.* ADMM based scalable machine learning on spark[C]. Proceedings of the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, 2015: 1174–1182. doi: [10.1109/BigData.2015.7363871](https://doi.org/10.1109/BigData.2015.7363871).
- [4] 赵海涛, 朱银阳, 丁仪, 等. 车联网中基于移动边缘计算的内容感知分类卸载算法研究[J]. *电子与信息学报*, 2020, 42(1): 20–27. doi: [10.11999/JEIT190594](https://doi.org/10.11999/JEIT190594).  
ZHAO Haitao, ZHU Yinyang, DING Yi, *et al.* Research on content-aware classification offloading algorithm based on mobile edge calculation in the internet of vehicles[J]. *Journal of Electronics & Information Technology*, 2020, 42(1): 20–27. doi: [10.11999/JEIT190594](https://doi.org/10.11999/JEIT190594).
- [5] SHI Weisong, ZHANG Xingzhou, WANG Yifan, *et al.* Edge computing: State-of-the-art and future directions[J]. *Journal of Computer Research and Development*, 2019, 56(1): 69–89. doi: [10.7544/issn1000-1239.2019.20180760](https://doi.org/10.7544/issn1000-1239.2019.20180760). (本条文献为中文文献, 请联系作者确认)
- [6] CHOUDHURY T, GUPTA A, PRADHAN S, *et al.* Privacy and security of cloud-based Internet of Things (IoT)[C]. Proceedings of 2017 3rd International Conference on Computational Intelligence and Networks (CINE), Odisha, 2017: 40–45. doi: [10.1109/CINE.2017.28](https://doi.org/10.1109/CINE.2017.28).
- [7] ZOU Hui and HASTIE T. Regularization and variable selection via the elastic net[J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005, 67(2): 301–320. doi: [10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x).
- [8] SHEN Yi, HAN Bing, and BRAVERMAN E. Stability of the elastic net estimator[J]. *Journal of Complexity*, 2016, 32(1): 20–39. doi: [10.1016/j.jco.2015.07.002](https://doi.org/10.1016/j.jco.2015.07.002).
- [9] LUO Hezhi, SUN Xiaoling, and LI Duan. On the convergence of augmented lagrangian methods for constrained global optimization[J]. *SIAM Journal on Optimization*, 2007, 18(4): 1209–1230. doi: [10.1137/060667086](https://doi.org/10.1137/060667086).
- [10] BERTSEKAS D P and TSITSIKLIS J N. Parallel and Distributed Computation: Numerical Methods[M]. London: Prentice Hall, 1989. (请联系作者补充页码信息)
- [11] ZHANG Yueqin, ZHENG Hao, and ZHANG Chuanlin. Global convergence of a modified PRP conjugate gradient method[J]. *Procedia Engineering*, 2012, 31: 986–995. doi: [10.1016/j.proeng.2012.01.1131](https://doi.org/10.1016/j.proeng.2012.01.1131).
- [12] He Bingsheng and Yuan Xiaoming. On the  $O(1/n)$  convergence rate of the Douglas-Rachford alternating direction method[J]. *SIAM Journal on Numerical Analysis*, 2012, 50(2): 700–709. doi: [10.1137/110836936](https://doi.org/10.1137/110836936).
- [13] NISHIHARA R, LESSARD L, RECHT B, *et al.* A general analysis of the convergence of ADMM[C]. Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille, France, 2015: 343–352.
- [14] EFRON B, HASTIE T, JOHNSTONE I, *et al.* Least angle regression[J]. *Annals of Statistics*, 2004, 32(2): 407–451. doi: [10.1214/009053604000000067](https://doi.org/10.1214/009053604000000067).
- [15] SPENCER B, ALFANDI O, and AL-OBEIDAT F. A refinement of lasso regression applied to temperature forecasting[J]. *Procedia Computer Science*, 2018, 130: 728–735. doi: [10.1016/j.procs.2018.04.127](https://doi.org/10.1016/j.procs.2018.04.127).

方维维: 男, 1981年生, 博士, 副教授, 研究方向为物联网、边缘计算、分布式机器学习。

刘梦然: 女, 1994年生, 硕士生, 研究方向为物联网、边缘计算、ADMM。

王云鹏: 男, 1996年生, 硕士生, 研究方向为物联网、边缘计算、ADMM。

李阳阳: 男, 1987年生, 博士, 高级工程师, 研究方向为移动网络、边缘计算、系统安全。

安竹林: 男, 1980年生, 博士, 高级工程师, 研究方向为物联网、边缘计算、分布式系统。