

图正则非线性岭回归模型的异常用电行为识别

张小斐¹, 耿俊成¹, 孙玉宝²

(1. 国网河南省电力公司 电力科学研究院, 郑州 450052; 2. 南京信息工程大学 信息与控制学院, 南京 210044)

摘 要: 对于用户异常用电行为的检测, 电力企业通常采用人工检查的方法, 但该方法需要消耗大量的人力、物力, 且容易受主观因素的影响。为此, 提出一种基于岭回归模型的异常用电行为识别算法。通过收集用户用电数据, 对岭回归模型进行训练, 并将训练好的模型用于异常用电行为的自动检测。为捕获未知的用户用电行为类别样本信息, 在岭回归模型的基础上引入图正则项。考虑到用电数据的非线性分布特性, 通过核函数的方式, 将原始数据映射到高维希尔伯特空间, 得到基于图正则的非线性岭回归模型。实验结果表明, 与最小二乘、岭回归、图正则岭回归模型相比, 该算法具有更高的识别准确率。

关键词: 用电行为分析; 岭回归; 图正则; 非线性分布; 半监督学习

中文引用格式: 张小斐, 耿俊成, 孙玉宝. 图正则非线性岭回归模型的异常用电行为识别[J]. 计算机工程, 2018, 44(6): 8-12.

英文引用格式: ZHANG Xiaofei, GENG Juncheng, SUN Yubao. Abnormal electricity behavior recognition of graph regularization nonlinear ridge regression model[J]. Computer Engineering, 2018, 44(6): 8-12.

Abnormal Electricity Behavior Recognition of Graph Regularization Nonlinear Ridge Regression Model

ZHANG Xiaofei¹, GENG Juncheng¹, SUN Yubao²

(1. Electric Power Research Institute, State Grid Henan Electric Power Company, Zhengzhou 450052, China;

2. School of Information and Control, Nanjing University of Information Science and Technology, Nanjing 210044, China)

【Abstract】 For the detection of abnormal electricity behavior by users, power companies usually adopt manual inspection methods, however, this method requires a lot of manpower and material resources, and is influenced by subjective factors. Therefore, an abnormal electricity behavior recognition algorithm based on ridge regression model is proposed. By collecting user electric data, the model is trained and the trained model is used for automatic detection of abnormal electricity behavior. In order to capture the sample information of unknown users' electricity behavior categories, the graph regularization term is introduced on the basis of the ridge regression model. Taking into account the non-linear distribution characteristics of electricity data, the original data is mapped to high-dimensional Hilbert spaces through kernel functions, and a nonlinear ridge regression model based on graph regularity is gained. Experimental results show that compared with least squares, ridge regression, and graph regularization ridge regression models, this algorithm has higher recognition rate.

【Key words】 electricity behavior analysis; ridge regression; graph regularization; nonlinear distribution; semi-supervised learning

DOI: 10.19678/j.issn.1000-3428.0048320

0 概述

电力行业是我国经济飞速增长的重要推手, 而用电搜检部门是整个电力行业的主要构成部分^[1]。目前, 在电量自动化管理方面, 我国电力企业相对落后, 窃电情况比较严重, 不仅给国家造成了经济损失, 而且严重干扰了市场经济秩序。对于窃电等异

常用电行为, 早期多采用现场检查的方式, 比如检查线损率、用电设备等。该方法消耗大量的人力、物力, 不利于电力企业的管理。

近年来, 随着智能电网的快速发展以及用电信息采集系统的逐步完善^[2], 配用电数据呈现出明显的大数据特性, 即数据量大、数据类型多、增长速度快等。与此同时, 电力大数据也吸引了越来越多人的关注。

基金项目: 国家自然科学基金(61300162); 国家电网公司 2016 年科技项目。

作者简介: 张小斐(1976—), 男, 高级工程师, 主研方向为电网大数据分析与管理; 耿俊成, 高级工程师、硕士; 孙玉宝, 副教授、博士。

收稿日期: 2017-08-11 **修回日期:** 2017-09-21 **E-mail:** 397536335@qq.com

因此,在此背景下,充分合理地利用数据挖掘技术,对用户的用电数据进行分析,便能对用户的用电行为进行自动识别,从而提高用电检查的管理效率。

当前的数据挖掘方法可分为3种不同类别,即无监督学习方法、监督学习方法以及半监督学习方法,无监督学习方法不需要对训练集的样本进行预先标记,直接挖掘数据内在的结构关系,常见的有模糊C均值聚类法和K-Means聚类法^[3-5]等。与无监督学习方法不同,监督学习方法利用训练集的标记信息,学习用户的用电信息与用户行为之间的关系,常见的有支持向量机、决策树、极限学习机^[6-10]等。监督学习方法通常需要标记大量的训练样本,而在许多现实任务中,很难获取足够多的训练样本。半监督学习可以看作是监督学习和无监督学习的折中。通过挖掘大量未标记样本的信息和少量标记样本的信息,共同对模型进行训练。其中,基于图模型的半监督学习方法^[11-12]的研究较多,如文献[13]的图拉普拉斯正则的半监督学习模型、文献[14]的拉普拉斯特征映射模型。此类方法将所有样本(不管标记与否)作为图模型的顶点,不同的顶点之间通过边进行连接,以此表示它们之间的关系。然后根据连接关系将类别信息从少量含有标记信息的样本传递到其邻近样本。

本文提出一种基于图正则非线性岭回归的异常用电行为识别算法。首先通过用户的用电数据构造图模型,图模型的顶点由所有样本共同构成,顶点间的关系通过边进行连接,边的权重表示相关性的强弱,通过高斯核函数进行计算。其次利用构建好的图模型计算拉普拉斯矩阵,并由该矩阵和原始的用户用电数据建立图正则的岭回归模型。由于岭回归模型是线性模型,而用户的用电数据是非线性分布的,因此需要将用户用电数据通过核函数的方法映射到高维希尔伯特空间,从而在高维空间中进行用户异常用电行为的识别。

1 岭回归模型

岭回归将二范数正则化融入到最小二乘模型,是一种多元线性回归模型。得益于模型的简单性和灵活性,岭回归模型已经被广泛地应用于各个领域,例如人脸识别^[15]、生物信息学^[16]、化学信息学^[17]等。

假设收集到 l 个标记数据 $\{x_i, y_i\}_{i=1}^l$ 和 u 个未标记数据 $\{x_i\}_{i=l+1}^u$ 。其中, $x_i \in \mathcal{R}^d$ 表示第 i 个用户的用电特性。不同的学者通常采用不同的变量作为特征来描述用电特性。比如,文献[18]不仅使用了日用电量、日最大负荷、日最小负荷等表征负荷特性的特征,还使用了房屋面积、家庭成员数等关联特征;文献[19]从时间特征量的角度选取年度用电数据、季节用电数据、日用电数据等组成聚类特征向量。受这些文献的启发,本文采用用户每天的用电量作

为模型的特征。

为便于计算,通常可以将数据表示成矩阵的形式。例如用 $\mathbf{X} = [x_1, x_2, \dots, x_{l+u}]^T$ 表示整个数据集, $\mathbf{X}_l = [x_1, x_2, \dots, x_l]^T$ 表示训练集。线性回归的核心思想是拟合函数 $f(\mathbf{x}; \mathbf{w}) = \mathbf{x}^T \mathbf{w} + b$,使得真实值和预测值之间的误差最小,即:

$$(\mathbf{w}^*, b^*) = \arg \min_{\mathbf{w}, b} \sum_{i=1}^l \|f(x_i) - y_i\|^2 \quad (1)$$

为了方便推导,可以在每个变量 x_i 中添加一个元素1,从而将参数 b 放入向量 \mathbf{w} 中。假设 $\mathbf{Y}_l = [y_1, y_2, \dots, y_l]^T$,则目标函数可改为矩阵的形式:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \sum_{i=1}^l \|\mathbf{X}_l \mathbf{w} - \mathbf{Y}_l\|_F^2 \quad (2)$$

该函数的解为 $\mathbf{w}^* = (\mathbf{X}_l^T \mathbf{X}_l)^{-1} \mathbf{X}_l^T \mathbf{Y}_l$ 。然而,由于相邻时间段用户的用电量具有较强的相关性,使得协方差矩阵 $\mathbf{X}_l^T \mathbf{X}_l$ 往往是奇异矩阵。常见的解决方案是在 \mathbf{w} 上添加一个惩罚项,即:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \sum_{i=1}^l \|\mathbf{X}_l \mathbf{w} - \mathbf{Y}_l\|_F^2 + \alpha \|\mathbf{w}\|^2 \quad (3)$$

其中, $\alpha \|\mathbf{w}\|^2$ 即为Tikhonov正则项,统计学上称作岭回归, $\mathbf{w}^* = (\mathbf{X}_l^T \mathbf{X}_l + \alpha \mathbf{I})^{-1} \mathbf{X}_l^T \mathbf{Y}_l$ 为式(3)的解, \mathbf{I} 表示 $d \times d$ 维的单位矩阵。

2 本文算法

针对用户用电数据存在的标定样本少以及数据的非线性分布问题,本文提出一种基于图正则非线性岭回归模型的异常用电行为识别算法。该算法首先建立图正则的岭回归模型,然后联合利用标定样本和大量的未标定样本进行学习,进一步引入核函数,形成图正则的非线性岭回归模型,实现对异常用电行为的识别分类。

2.1 图正则岭回归模型

对于电力数据分析而言,可用的训练样本通常较少,因为样本的标签信息需要通过人工实地检查的方式获取。另一方面,电力部门通常能收集到大量没有标记信息的数据。为解决小样本时的模型训练问题,很多模型试图挖掘未标注样本的丰富信息。图模型是其中一种比较有效的方法,它通过“顶点”和“边”将所有样本(不管标记与否)联系起来。因此,本文以图模型与岭回归模型为基础,提出了图正则岭回归模型。

构建一个图 G ,其顶点数为 $l+u$,每个顶点对应一个样本。假设相似度矩阵为 \mathbf{S} ,其元素 S_{ij} 表示任意2个顶点 i 和 j 的相似度,通过下式进行计算:

$$S_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}}, & x_i \in N_k(x_j), x_j \in N_k(x_i) \\ 0, & \text{其他} \end{cases} \quad (4)$$

其中, $N_k(x_j)$ 表示样本 x_j 的 k 个最近邻样本的集合, σ 是核函数的参数。图正则项假设图 G 中邻近的样本具有相似的标注值,基于该假设,可推导出如下约

束项:

$$\mathbf{f}^* = \arg \min_{\mathbf{f}} \sum_{i,j=1}^{l+u} \|f(x_i) - f(x_j)\|^2 S_{ij} = \mathbf{f}^T \mathbf{L} \mathbf{f} \quad (5)$$

其中, $\mathbf{f} = [f(x_1), f(x_2), \dots, f(x_l), \dots, f(x_{l+u})]$, \mathbf{L} 表示图拉普拉斯矩阵。经过一些简单的推导, 可得 $\mathbf{L} = \mathbf{D} - \mathbf{S}$, 其中, \mathbf{D} 为对角矩阵, 矩阵中每个对角元素等于 \mathbf{S} 矩阵中相应行向量或列向量所有元素的和, 即 $D_{ii} = \sum_j S_{ji}$ 。将式(5)代入式(3), 可构造出图正则岭回归模型的最终目标函数为:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \sum_{i=1}^l \|\mathbf{X}_i \mathbf{w} - \mathbf{Y}_i\|_F^2 + \alpha \|\mathbf{w}\|^2 + \beta \mathbf{w}^T \mathbf{X}^T \mathbf{L} \mathbf{X} \mathbf{w} \quad (6)$$

显然, 式(6)为凸函数, 其全局最优解为:

$$\mathbf{w}^* = (\mathbf{X}_l^T \mathbf{X}_l + \alpha \mathbf{I} + \beta \mathbf{X}^T \mathbf{L} \mathbf{X})^{-1} \mathbf{X}_l^T \mathbf{Y}_l \quad (7)$$

当 $\beta = 0$ 时, 式(7)与式(3)相同。因此, 岭回归可看作图正则岭回归的特殊情况。

2.2 图正则非线性岭回归模型

图正则岭回归模型假设数据是线性可分的, 而实际的用户用电数据通常是非线性可分的。如图1所示, 将部分用户用电数据从高维空间通过主成分分析 PCA 投影到二维空间, 在该空间数据明显是线性不可分的。因此, 本文进一步利用核函数方法, 建立图正则非线性岭回归模型, 有效地将用电数据从非线性空间投影到线性空间。核方法的目的是将数据投影到高维、甚至无穷维的希尔伯特空间 H , 随后在 H 空间寻找一个线性超平面, 将数据分成不同类型。该线性超平面在原空间的投影是一个非线性超平面。

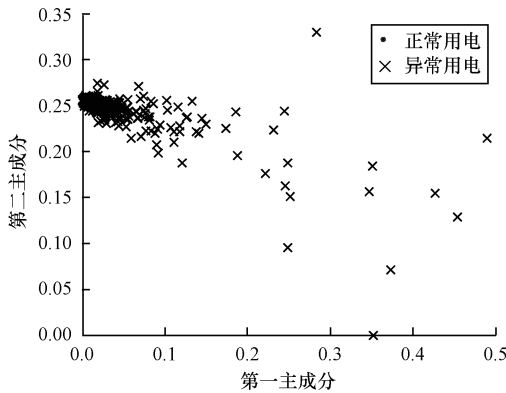


图1 部分用户用电数据的非线性分布

给定非线性映射函数 $\phi(x): \mathcal{R}^d \rightarrow H$, 投影之后的数据集可以分别表示为 $\phi_l = [\phi(x_1), \phi(x_2), \dots, \phi(x_l)]^T$ 和 $\phi = [\phi(x_1), \phi(x_2), \dots, \phi(x_l), \dots, \phi(x_{l+u})]^T$ 。此外, 根据表示定理^[20]可知: $\mathbf{w} = \sum_{i=1}^l \phi(x_i) a_i = \phi_l^T \mathbf{A}$ 。其中, a_i 表示权重系数, $\mathbf{A} = [a_1, a_2, \dots, a_l]^T$ 。代入式(6), 图正则化非线性岭回归模型表示为:

$$\mathbf{A}^* = \arg \min_{\mathbf{A}} \|\phi_l \phi_l^T \mathbf{A} - \mathbf{Y}_l\|_F^2 + \alpha \mathbf{A}^T \phi_l \phi_l^T \mathbf{A} + \beta \mathbf{A}^T \phi_l \phi^T \mathbf{L} \phi \phi_l^T \mathbf{A} \quad (8)$$

从式(8)可以看出, 直接计算映射函数 $\phi(x)$ 几乎是不现实的。核技巧可以很好地避开这个问题, 间接求取任意2个样本之间的内积值, 其表达式为 $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$, 其中, 操作符 $\langle \cdot \rangle$ 表示内积, $K(\cdot)$ 表示核函数。因此, 式(8)可以重新写成如下形式:

$$\mathbf{A}^* = \arg \min_{\mathbf{A}} \|\mathbf{K}_l \mathbf{A} - \mathbf{Y}_l\|_F^2 + \alpha \mathbf{A}^T \mathbf{K}_l \mathbf{A} + \beta \mathbf{A}^T \mathbf{K} \mathbf{L} \mathbf{K}^T \mathbf{A} \quad (9)$$

式(9)对 \mathbf{A} 求导并令导数为0, 可求得解为:

$$\mathbf{A}^* = (\mathbf{K}_l^T \mathbf{K}_l + \alpha \mathbf{K}_l + \beta \mathbf{K} \mathbf{L} \mathbf{K}^T)^{-1} \mathbf{K}_l^T \mathbf{Y}_l \quad (10)$$

对于任意一个用户 x_j , 模型预测值为:

$$f(x_j) = \sum_{i=1}^l \langle \phi(x_i), \phi(x_j) \rangle a_i = \sum_{i=1}^l K(x_i, x_j) a_i$$

3 实验结果与分析

3.1 实验数据

本节通过实验验证本文图正则非线性岭回归算法(KGRR)的性能。所使用的用户用电数据采集自某市电力部门的实际业务系统, 包含8 956个用户300余天的用电量数据。该数据中有2 176名用户被标识为违规用电用户, 由电力执法人员现场调查得到, 其余为正常用电数据。图2为部分用户的用电量曲线。

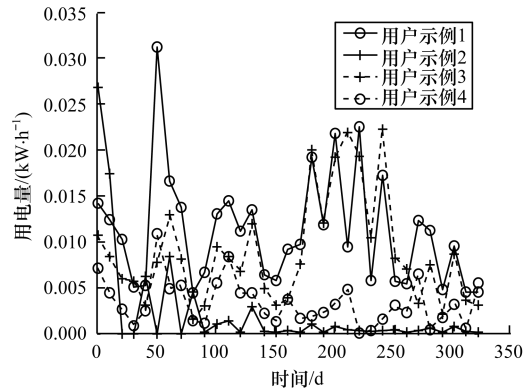


图2 部分用户用电负荷曲线

3.2 数据预处理

采集的用电数据通常会存在一些错漏或缺失的问题, 例如某些用户的用电量数据明显异常(如为负数)或缺失(如为空值), 需要对这些数据进行处理, 后续算法才能正常执行。常见的处理方式包括: 用预先设定的默认值进行填充, 用用电数据的平均值、中位数值、最大值、最小值等进行填充, 或者直接将该用户数据删除等^[10]。本文采用平均值的方法对异常数据进行处理。

除对数据进行填充外, 由于用电数据的特征量包含不同的量纲和单位, 还需要通过归一化消除指标之间的量纲影响。归一化处理是各种数据挖掘以及机器学习算法的基础, 因为对于不同的评价标准, 通常存在不同的单位和量纲, 从而影响了数据分析最后的结果以及不同标准之间的可比性。经过归一化处理后, 不同的评价标准可以处于同一个数量级,

使得评价结果更客观真实。在各种归一化方法中,Min-max 归一化方法是其中较为简单有效的一种。该方法通过对原始数据进行一些线性变换,便能将结果值映射到 $[0,1]$ 范围之间。具体而言,对于第 i 个用户 x_i ,其归一化后的结果可表示为:

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

其中, x_{\max} 表示所有用户在当天用电量的最大值, x_{\min} 表示所有用户在当天用电量的最小值。

3.3 结果分析

为了验证 KGRR 算法的优越性,将其与 4 种不同的方法进行了对比。包括图正则岭回归算法(GRR)、岭回归算法(RR)、最小二乘回归算法(LS)以及支持向量机模型(SVM)。由于异常用电行为识别是一个分类问题,而 KGRR 及其对比算法 GRR、RR 和 LS 均为回归模型,需要将标签信息进行变换。具体操作如下:将异常用户的标签记为向量 $[1,0]$,正常用户的标签记为向量 $[0,1]$,则回归算法的输出结果为一个二维的向量。当向量的第 1 个元素数值大于第 2 个元素时,该用户即为异常用户,反之则为正常用户。对于 KGRR 和 GRR 算法中的正则化参数 α 和 β ,采用交叉验证的方法从候选值 $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ 中进行筛选。所有算法的分类性能通过整体分类准确率来度量,即测试集中正确识别的用户数量与所有用户数量的比值。

表 1 为 5 种算法在使用不同比例的训练样本时获得的分类结果。考虑到选择标定样本的随机因素,所有算法都被重复执行 5 次,所列结果为 5 次结果的平均值。表 1 中所有加粗的字体表示同等条件下获得的最好结果。

表 1 5 种算法在不同标定比例下的分类准确率 %

标定样本比例	LS	RR	GRR	SVM	KGRR
1	50.38	60.98	69.05	70.71	72.20
2	52.18	62.93	70.89	72.14	74.28
5	52.09	63.21	72.33	75.66	77.10
10	52.96	65.61	75.05	78.39	80.40
20	54.10	66.03	75.95	78.97	80.73
50	59.06	69.48	76.84	80.12	82.16

从表 1 可以得到以下结论:

1) 当标定样本的数量增加时,所有模型的分类准确率均呈现上升的趋势。说明算法的性能容易受到训练样本数量的影响。

2) 与 LS 算法相比,RR 能够明显地提升分类准确率。因为用户每天的用电量具有一定的相关性,导致 LS 算法中的协方差矩阵 $X^T X$ 是奇异的。通过 Tikhonov 正则项,RR 算法能够很好地避开该问题,并能够自适应地从所有特征中筛选出对算法而言最优的特征。

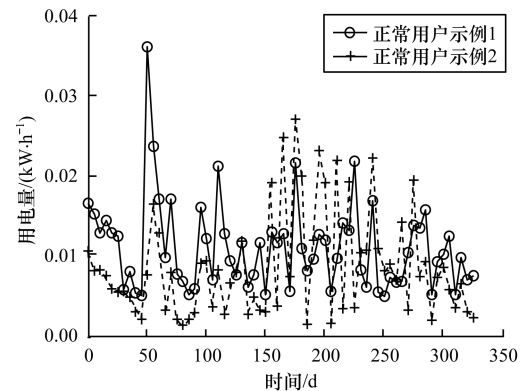
3) 由于训练样本的数量通常较少(50% 以下),

使得 RR 算法的性能受到了一定的限制。不同于 RR 算法,GRR 算法不仅能够利用训练样本的信息,还能够充分挖掘未标注样本的信息,从而进一步提高了分类准确率。

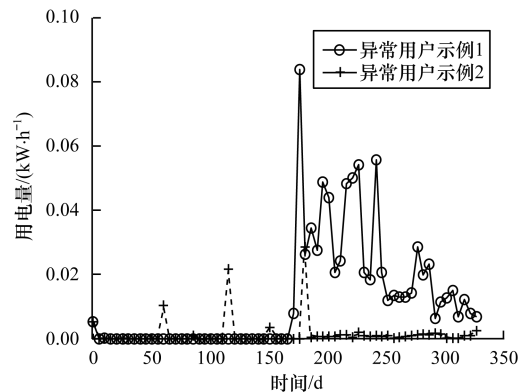
4) LS、RR 和 GRR 算法均为线性模型,较难捕获用户数据的非线性分布特征。通过核函数的方式将低维数据映射到高维空间,KGRR 算法能够有效地解决以上问题,从而获得最好的分类结果。

5) 尽管 SVM 同样可以通过核函数的方法解决非线性分类问题,但由于其仅利用了标注样本的信息,并没有充分利用未标注样本的信息,因此其性能不如 KGRR 算法。

图 3 为本文 KGRR 算法分类结果的用户用电曲线示例。其中,图 3(a)为正常用户用电曲线,图 3(b)为异常用电曲线。从图 3 可以看出:正常用户的用电数据尽管也有部分波动,但整体表现出一致的趋势;而异常用户的用电数据大多数时间为较小的值,只有少数时间涌现为较大的用电量。



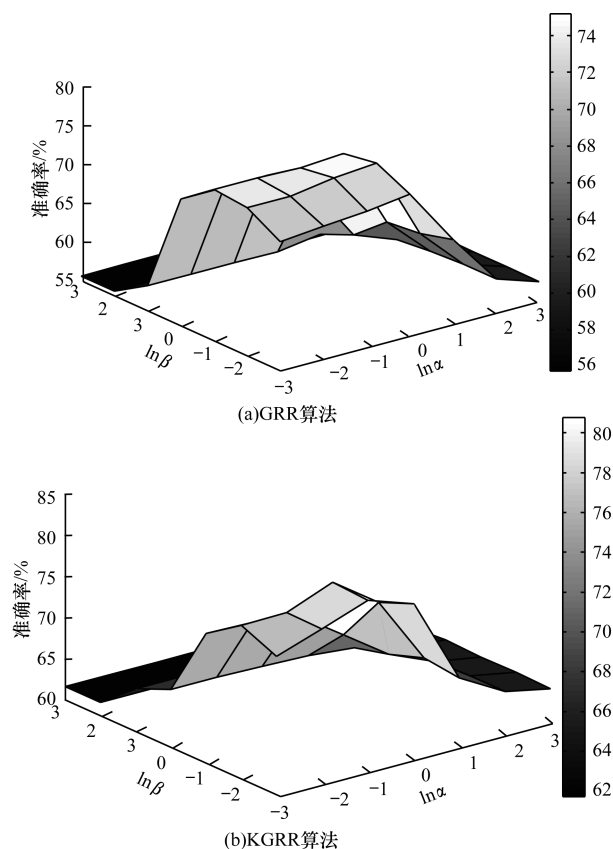
(a)正常用电曲线示例



(b)异常用电曲线示例

图 3 本文算法分类结果的用户用电曲线

图 4 分析了 GRR 和 KGRR 算法中正则化参数 α 和 β 变化时对分类结果的影响。实验中所用标定样本的比例为 10%。图 4 中 x 轴表示参数 α 的对数, y 轴表示参数 β 的对数, z 轴表示分类准确率。显然,随着 α 值和 β 值的增加,分类准确率先增大后减小,并在 $\alpha=0.01, \beta=0.1$ 时取得最大值。

图 4 参数 α 和 β 对分类结果的影响

4 结束语

本文提出一种图正则非线性岭回归的异常用电行为识别算法。通过图正则项利用未标注样本的信息,并根据核函数的方法将用户数据投影到高维空间,从而解决实际数据中存在的线性不可分问题。实验结果表明,与最小二乘、岭回归、图正则岭回归模型相比,本文算法获得了更高的识别准确率。

参考文献

- [1] 魏 瑶,朱伟义,龚桃荣,等.基于数据挖掘技术的用电异常分析系统设计[J].电力信息与通信技术,2014,12(5):70-73.
- [2] 胡江溢,祝恩国,杜新纲,等.用电信息采集系统应用现状及发展趋势[J].电力系统自动化,2014,38(2):131-135.
- [3] 王星华,陈卓优,彭显刚.一种基于双层聚类分析的负荷形态组合识别方法[J].电网技术,2016,40(5):1495-1501.
- [4] 赵 莉,候兴哲,胡 君,等.基于改进 k-means 算法的海量智能用电数据分析[J].电网技术,2014,38(10):2715-2720.
- [5] ANGELOS E W S, SAAVEDRA O R, CORTES O A C, et al. Detection and identification of abnormalities in customer consumptions in power distribution systems[J]. IEEE Transactions on Power Delivery, 2011, 26(4): 2436-2442.
- [6] 简富俊,曹 敏,王 磊,等.基于 SVM 的 AMI 环境下用电异常检测研究[J].电测与仪表,2014(6):64-69.
- [7] NAGI J, YAP K S, TIONG S K, et al. Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system[J]. IEEE Transactions on Power Delivery, 2011, 26(2): 1284-1285.
- [8] MUNIZ C, FIGUEIREDO K, VELLASCO M, et al. Irregularity detection on low tension electric installations by neural network ensembles[C]//Proceedings of International Joint Conference on Neural Networks. Washington D. C., USA: IEE Press, 2009: 2176-2182.
- [9] MONEDERO I, BISCARRI F, LEON C, et al. Detection of frauds and other non-technical losses in a power utility using pearson coefficient, Bayesian networks and decision trees[J]. International Journal of Electrical Power & Energy Systems, 2012, 34(1): 90-98.
- [10] 胡殿刚,李韶瑜,楼 俏,等.ELM 算法在用户用电行为分析中的应用[J].计算机系统应用,2016(8):155-161.
- [11] ZHOU D, HUANG J, KOPF S B. Learning with hypergraphs: clustering, classification, and embedding[C]//Proceedings of International Conference on Neural Information Processing Systems. Washington D. C., USA: IEE Press, 2006: 1601-1608.
- [12] HUANG Y, LIU Q, LV F, et al. Unsupervised image categorization by hypergraph partition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 33(6): 1266-1273.
- [13] ZHU X, GHAMRANI Z, LAFFERTY J. Semi-supervised learning using gaussian fields and harmonic functions[C]//Proceedings of International Conference on Machine Learning. Washington D. C., USA: IEEE Press, 2003: 213-222.
- [14] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computing, 2003, 15(6): 1373-396.
- [15] AN S, LIU W, VENKATESH S. Face recognition using kernel ridge regression[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEE Press, 2007: 1-7.
- [16] GIGUERE S, MARCHAND M, LAVIOLETTE F, et al. Learning a peptide-protein binding affinity predictor with kernel ridge regression[J]. BMC bioinformatics, 2013, 14(1): 82-97.
- [17] HINKLEY T. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase[J]. Nature Genetics, 2011, 43(5): 487-489.
- [18] 赵 莉,候兴哲,胡 君,等.基于改进 K-means 算法的海量智能用电数据分析[J].电网技术,2014,38(10):2715-2720.
- [19] 薛承荣,顾 洁,赵建平,等.基于用户用电特性及供电成本分摊的销售侧电价机制研究[J].华东电力,2014,42(1):168-173.
- [20] SHAW-TAYLOR J, CRISTIANINI N. Kernel methods for pattern analysis[M]. Cambridge, UK: Cambridge University Press, 2004.