DOI:10.12158/j.2096-3203.2019.06.016

基于随机森林的电动汽车充电行为聚类技术研究

刘亚丽[1]，李国栋[1]，刘云[1]，洪奕[2]，刘瑜俊[2]
（1. 国网天津市电力有限公司电力科学研究院，天津 300392；
2. 东南大学电气工程学院，江苏 南京 210096）

摘　要：随着国家对新能源汽车的持续推进，成千上万的电动汽车（EV）接入电力系统，在充电过程中形成了关于EV充电行为的海量数据，因此有必要对EV充电行为特征展开研究。文中提出了一种基于随机森林的EV充电行为聚类技术，从充电行为的大量数据中辨识和分析不同类型的充电行为。采用英国Dundee市2018年1月的充电数据进行实验，分别得到该月工作日、双休日和节假日的充电行为分类。聚类分析获得的各个类别有着较为明确的特性，并以此推断出用户的充电方式、出行行为特点等。最后将该算法和欧式距离法进行对比，对比结果表明随机森林算法在EV聚类问题中的优越性。

关键词：电动汽车；充电行为；随机森林；聚类分析
中图分类号：TM910.6　　**文献标志码**：A　　**文章编号**：2096-3203（2019）06-0115-07

0　引言

电动汽车（electrical vehicle，EV）由于其较高的能量效率和较低的污染气体排放量而受到关注[1—2]。EV充电行为的随机性很大[3]，且不同的EV用户有着不同的充电行为模式[4]。随着充电设施功能的不断更新，更多样的充电数据将被记录，从而形成关于某一具体地区EV充电行为的大数据[5]。基于这种实测大数据对EV用户充电行为进行聚类分析，可提高对充电行为描述的准确性。

目前，EV的分类主要基于EV的车型种类，而其运行规则也多是参考日常生活中燃油汽车的常见情形。文献[7]按照EV充电行为的时空差异对其进行分类，但EV的特性主要基于仿真模拟，缺少足够的实际数据支撑。文献[8]根据EV的驱动构造对EV进行分类，但没有涵盖所有种类的EV。

k-means算法是最常见的EV充电行为聚类算法[9]。文献[10]提出基于引力模型评价的改进*k*-means算法，但是所选取的负荷范围仅限于某个工业园区，且*k*-means算法是直接基于欧式距离的算法，在运算前需要人为设置较多参数，所以在未知类别数目的情况下可能只得到次优解。而随机森林算法仅需要人为设定两个与聚类无直接关联的参数就可以进行聚类工作，降低了人为影响。

文中提出了一种基于随机森林的EV充电行为聚类技术。首先介绍了随机森林算法和其应用于聚类的基本原理，其次建立了基于主成分分析改进的随机森林聚类算法，最后在算例中利用随机森林自带的数据相似度辨识，从英国Dundee市公开的大量EV充电数据中辨识和分析出不同类型的充电行为特征，并对聚类结果进行了详细分析，形成多种典型的EV充电行为模型。

收稿日期：2019-05-08；修回日期：2019-07-17
基金项目：国家重点研发计划资助项目（2017YFA0700300）

1　EV行为聚类问题和随机森林

在无人干预的情况下，收集到的EV充电数据样本通常包含充电起始时间、结束时间、充电电量和充电地点等信息，而不会标记有明确的类别。所以EV充电数据样本也就是无标签数据。由于样本的标签未知，无法训练样本对应的类别，因而只能从原始样本集开始学习得到随机森林，这一问题属于非监督学习。

随机森林算法是指利用多棵决策树对样本进行训练并预测的一种机械学习方法[11]，是目前可用于处理大数据的最成功方法之一[12]，已在生态学、医学、电气等[13—15]多个领域得到应用。该算法可以处理包含多个变量的大数据，包括离散型、连续型数据、非规范化、高维度和存在许多未知特征的数据，并且能够进行高度并行化处理，计算个体之间的相关性。通过快速学习后输出高准确度的分类，聚类或者回归结果[16]。同时，随机森林算法不会产生过拟合问题，也可以评估输入变量的重要性[17]。根据其特点和优点，随机森林可以被用于基于大数据的EV行为聚类分析。

2　EV充电行为的聚类算法

2.1　基于随机森林的EV行为聚类算法主要步骤

聚类思路主要分为4个部分：

（1）使用主成分分析法（principle component analysis，PCA）进行数据降维；

（2）构造用于非监督学习的 EV 充电行为混合数据样本；训练随机森林，且统计原始数据样本间的相关性，得到相似矩阵；

（3）使用多维尺度分析法将相似矩阵的维数降至 2，在直角坐标系中绘出所有原始数据样本的对应点，并表示出相对集中点集所对应的现实意义；

（4）采用随机数检验聚类结果。

2.2 基于主成分分析法的随机森林聚类功能优化

随机森林算法具有优越的聚类功能，可以很好的实现预期的聚类目标。在实际运算过程中，由于计算机性能的限制，需要降低输入的样本特征，来减少运算量。采用主成分分析法来提取重要的样本特征，并降低样本特征的数量，但依旧保留原始样本的大部分特征。

EV 充电数据集的主成分可以通过以下的计算步骤得到，具体包括：

（1）将原始 EV 充电数据集进行标准化处理。

（2）基于原始数据集，建立原始数据间的系数矩阵：

$$\boldsymbol{R}=(r_{ij})_{m\times m} \tag{1}$$

其中：

$$r_{ij}=\frac{\sum_{k=1}^{n}(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ki}-\bar{x}_i)^2(x_{kj}-\bar{x}_j)^2}} \tag{2}$$

（3）求 $\boldsymbol{R}$ 的特征根，特征根的排序满足：$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_m\geqslant 0$。同时得到每个特征根对应的特征向量：

$$\boldsymbol{a}_l=[a_{1l}\quad a_{2l}\quad\cdots\quad a_{ml}]^{\mathrm{T}}\quad l=1,2,3,\cdots,m \tag{3}$$

（4）由此，各个主成分为：

$$F_i=a_{1i}x_1+a_{2i}x_2+\cdots+a_{mi}x_m \tag{4}$$

式中：$i=1,2,3,\cdots,m$。

（5）前 N 个主成分的累计贡献率为：

$$C_N=\frac{\sum_{k=1}^{N}\lambda_k}{\sum_{k=1}^{m}\lambda_k} \tag{5}$$

在通常情况下，当前 N 个主成分的累计贡献率高达 85%～95%时，则仅使用前 N 个主成分就可以代表原本特征的绝大部分信息。

2.3 随机森林训练流程

在非监督学习问题下，获得可用于训练随机森林的数据样本的思想。首先，设定原始数据集为数据集Ⅰ。然后，将同一数据样本中每个参数的数值进行独立置换，形成数据集Ⅱ。数据集Ⅰ和数据集Ⅱ形成混合数据集。独立置换的方法多样，文中采取按该参数的经验边界分布进行随机采用的方法。获得的数据集Ⅱ拥有独立的随机参数分布特征，但数据集Ⅱ中的所有参数都拥有和初始数据集中相应的参数相同的单变量参数分布特征。这一过程能够由随机森林算法自行模拟得到[11]。

随机森林需要将数据集Ⅱ从混合数据集中提取出来。因此，在决策树在生长过程中，它们的分枝将依赖于样本的各个参数，所得到的样本相似性也与各个参数有关[13]。

随机森林训练的主要算法基础是装袋法（bootstrap aggregating，Bagging）算法和分类与回归树（classification and regression tree，CART）算法[18]。Bagging 算法是一种基于自助法（Bootstrap）抽样的技术，其核心思想是利用 Bootstrap 采样[19]的结果来构造众多相互独立的决策树。每个决策树都不会被修剪，且在生成决策树的过程中，每个节点的判定都是基于随机选出的部分参数，随后将这种方式生成的不同的树进行组合，由此得到随机森林。森林中的每一棵树都有一个与之对应的随机且独立同分布的数据向量。随机森林的训练流程为：

（1）假设输入的 EV 混合数据集是一个有 M 个特征的训练数据集 S，包含 n 个不同的样本 $\{x_1,x_2,\cdots,x_n\}$。使用 Bootstrap 抽样法进行重复采样，得到随机生成的训练数据集 $S_1,S_2,\cdots,S_n$。该新数据集不含样本 $x_i(i=1,2,\cdots,n)$ 的概率约为 36.8%。这些样本会被排除在 Bootstrap 抽样的样本以外，部分数据被称为袋外数据（out of bag，OOB）。这些训练数据集有 $m(m<<M)$ 个特征。

（2）对于每一个训练数据集 $S_1,S_2,\cdots,S_n$，在处理决策树的每个节点时，基于训练数据集的 m 个特征计算所有可能出现的分裂方式。选择最佳的分裂方式（如最大的 Gini 度量）用于该结点的分裂。重复以上分裂过程，直到满足某个预设的条件。每棵树都会被剪枝，由此训练得到与之一一对应的决策树 $C_1,C_2,\cdots,C_n$。

（3）提取已被分离出的原始数据集 X，用步骤（2）得到的不剪枝树进行判别，并得到每个决策树末端上的样本情况 $C_1(X),C_2(X),\cdots,C_n(X)$。

（4）统计每个决策树的末端上的样本，如果 2 个样本出现在同一个末端，则两者的相关性将增加。

生成随机森林的算法流程如图 1 所示。

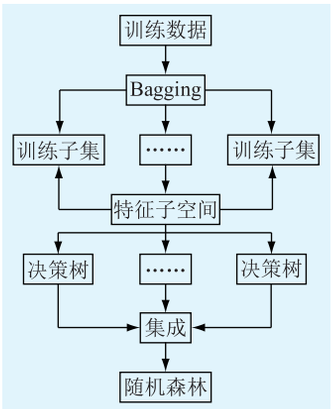


图 1 生成随机森林的算法流程

Fig.1 The flow chart for generating random forests

2.4 基于 EV 样本相关性的聚类过程

在数据集中的任意 2 个 EV 充电样本数据之间的相关性被定义为这两个样本数据在同一个决策树的末节点上出现次数的比值(小于 1)。假设数据样本的个数为 n,由随机森林产生反映数据样本关系的一个 $n\times n$ 维相似矩阵,该矩阵内所有元素的值属于[0,1],其大小表示数据样本之间的距离。研究表明,相似的数据样本在决策树的分支末端的聚集概率要大于不相似数据样本的聚集概率[16]。

数据样本的相关性能够被作为传统的聚类运算中的输入量,但并不是所有的多元结构的数据集都可以展示成聚类的形式[20]。在 EV 充电行为的聚类问题中,随机森林算法可以通过非监督机械学习来检测常规的多元结构数据集,且可以不事先假设数据集的聚类特征[21]。

多维尺度分析法是分析数据集特性的方法之一。该方法可以用来对 EV 充电行为进行聚类,其输入为 EV 充电数据样本之间的相似程度。多维尺度分析法通过对输入的初始数据集进行适当的降维处理,可以将数据样本在低维度坐标系中用坐标表示出来。在坐标系中,点与点之间的距离和聚集方式反映了对应的数据样本间的相似程度。

2.5 聚类结果检验

随机森林的特点之一是会在运算过程中自动生成随机数组,过程中产生的随机数是可控的,即让每次运行随机森林后产生的随机数组相同。因此,需要检验得到的聚类结果是基于实际的充电数据而不是自动生成的随机数组。具体做法是:

(1) 设定随机森林产生与进行原聚类时一样的随机数。

(2) 根据充电行为各个参数的概率分布,随机生成与实际充电数据数量相同的随机数组。

(3) 将步骤(2)随机生成的数组输入随机森林,并进行聚类分析。

如果实验后无法得到原聚类结果,则证明得到的聚类结果是基于实际数据而不是由随机森林自动产生的数据。

3 随机森林在 EV 充电行为分析聚类中的应用

3.1 数据来源

文中使用的 EV 充电数据是从在英国 Dundee 市城市评议会公布的 28 个充电桩或充电站采集到的。采集数据的时段是 2018 年 1 月 1 日至 1 月 31 日。用于观测变量包括充电起始时刻、充电结束时刻、充电持续时间、充电电量和充电地点。在剔除参数不全或有明显参数错误的无效充电行为后,剩下的有效充电行为共计 5 654 次,其中工作日占有 4 220次,双休日占有 1 180 次,节假日占有 254 次。

3.2 主成分分析的结果

基于 PCA 理论,对原始的 5 个观测变量进行分析,得到 5 个新参数,即主成分。图 2 给出了 5 个主成分表示整个充电数据集的贡献度比例。从图中可以看出,前 4 个主成分就可以表示原始数据 93% 的信息和特征,有助于减少随机森林的计算量。

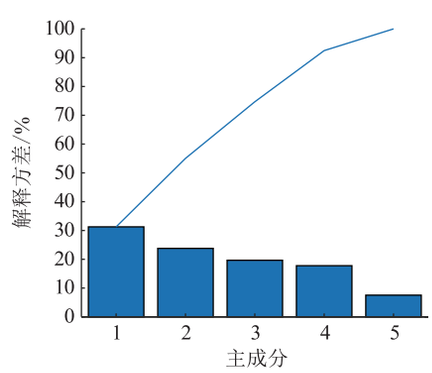


图 2 各个主成分表示原始工作日数据集的比例

Fig.2 Cumulative contribution of principal components to the entire charging data set

3.3 基于 R 语言的聚类实施过程

采用 R 语言来实现随机森林算法的聚类功能,并使用经典多维尺度(classical multidimensional scale,CMS)下的坐标表征各个充电行为之间的联系。具体过程为:

(1) 设置森林数量和每个森林中决策树的数量;

(2) 设置分类树上每个节点用来分叉的参数个数,使其等于一个整数,该整数接近参数总个数的平方根;

(3) 将充电行为数据输入到随机森林算法中,

使用 RStudio 软件中的 RFdist 函数实现随机森林算法，得到各个参数的重要程度（Gini 指数）；

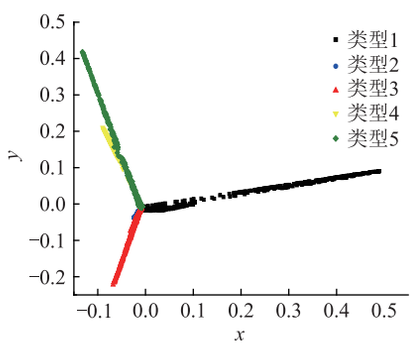
（4）使用 RStudio 软件中的 cmdscale 函数生成各个充电行为在 CMS 的二维坐标；

（5）基于步骤（4）得到的充电行为二维坐标，绘出所有充电行为的 CMS，两点间的距离表示两种充电行为的不相似度，每个点以（x，y）表示；

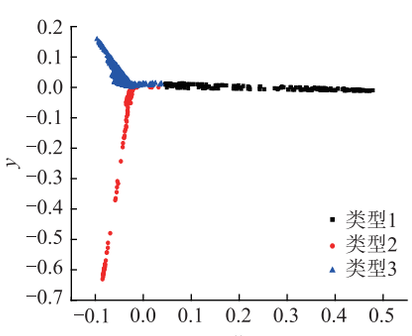
（6）根据步骤（5）得到的 CMS 图的形状特征，对图像进行划分，将密集的点归为同一类，得到 EV 用户充电行为的聚类结果。

3.4 聚类实施结果

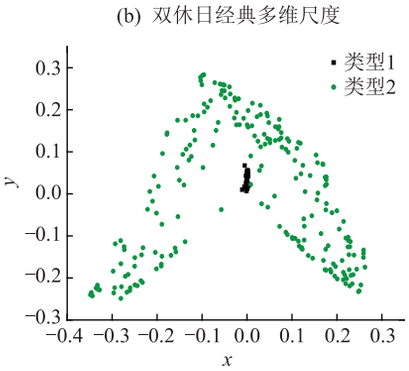
对工作日、双休日和节假日的充电行为分别进行聚类，分别得到对应的 CMS 图，如图 3 所示。同时，直接使用欧式距离法体现工作日中所有充电行为间的相关性，结果如图 4 所示。



(a) 工作日经典多维尺度



(b) 双休日经典多维尺度



(c) 节假日经典多维尺度

图 3 基于随机森林的 3 个时段经典多维尺度

Fig.3 Classical multidimensional scale map of three periods based on random forest

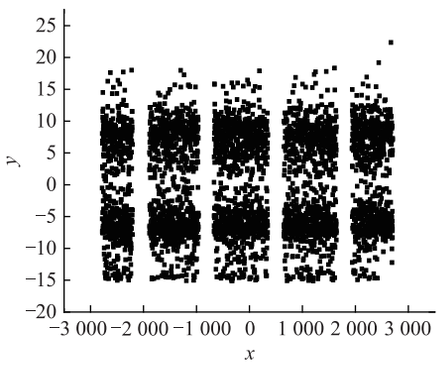


图 4 基于欧式距离法的工作日充电行为尺度

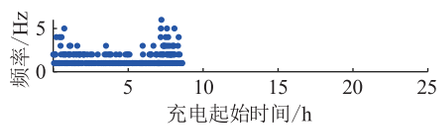
Fig.4 CMS of EV charging behavior on workdays by euclidean Distance

对比图 3（a）和图 4 可看出，由随机森林聚类得到的散点图的形状更加规整和密集，易于对图像进行分割和区别。由传统的欧式距离法得到的聚类结果过于复杂，根据 R 语言中的特征参数重要性评估，由欧式距离法得到的散点分布只取决于一个特征参数，与其余特征参数无关。此外，从坐标轴的数值可以看出，每个点之间的距离要远大于由随机森林得到的结果。所以，基于欧式距离法得到的聚类结果虽然多样，但不足以反应行为的一般特征，说明在进行 EV 充电行为聚类分析时，采用随机森林算法有着明显的优势。

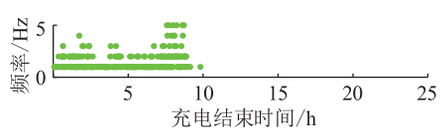
3.5 聚类结果分析

根据 3 个时段的 CMS 图的形状特征，分别进行划分，得到 3 个时段的聚类结果。以工作日的充电行为分类为例，图 5—图 9 展示了工作日中不同类别的参数分布特点，图（a）和图（b）的纵坐标都表示在某一时刻出现对应行为的次数。图（c）的纵坐标表示在某一种充电时长的出现次数。图（d）的纵坐标表示在某一种充电能量的出现次数。表 1 对 3 个时段的所有类别进行数据方面的解释。

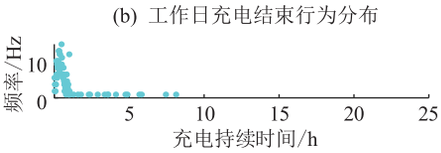
从整体结果来看，随机森林算法有效实现了对



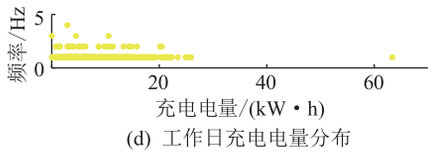
(a) 工作日充电起始行为分布



(b) 工作日充电结束行为分布



(c) 工作日充电持续时长分布



(d) 工作日充电电量分布

图 5 第一类 EV 充电行为

Fig.5 Class 1 EV charaing behavior

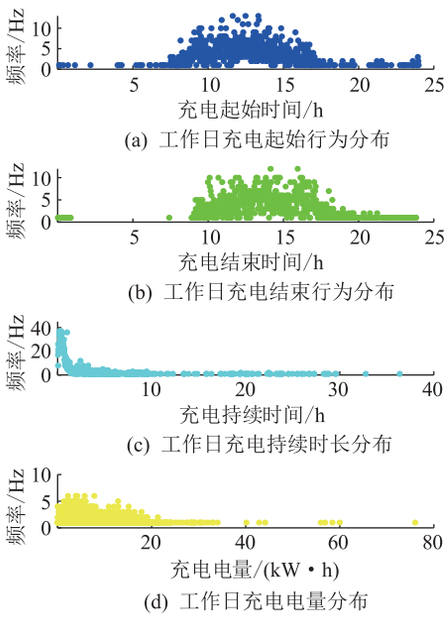


图 6 第二类 EV 充电行为

Fig.6 Class 2 EV charaing behavior

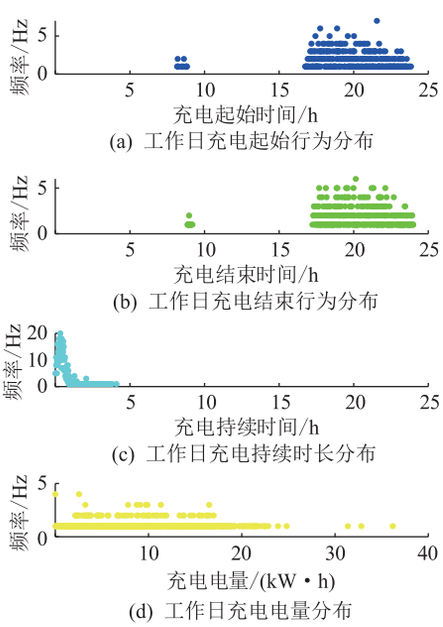


图 7 第三类 EV 充电行为

Fig.7 Class 3 EV charaing behavior

充电行为进行聚类的功能,效果优良。节假日和双休日的类别较少,其特征不如工作日的类别明显。因为节假日和双休日的数据样本较少,所以客观上存在的类别较少,EV 用户在节假日和双休日的行为模式确实会与工作日有所不同。此外,在使用 CMS 图进行类别划分时,由于节假日的数据量较少,图像较为分散和无规则,不利于进行类别划分。

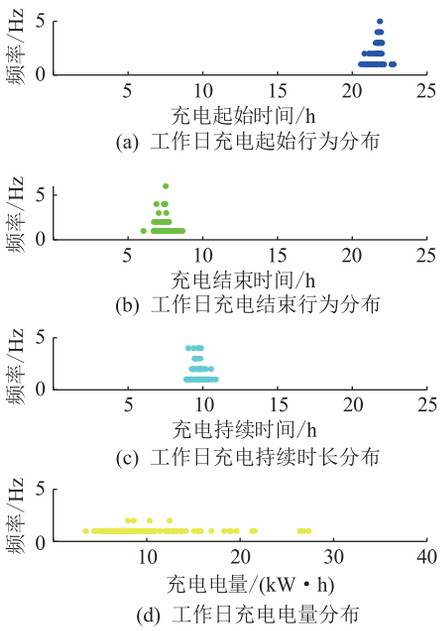


图 8 第四类 EV 充电行为

Fig.8 Class 4 EV charaing behavior

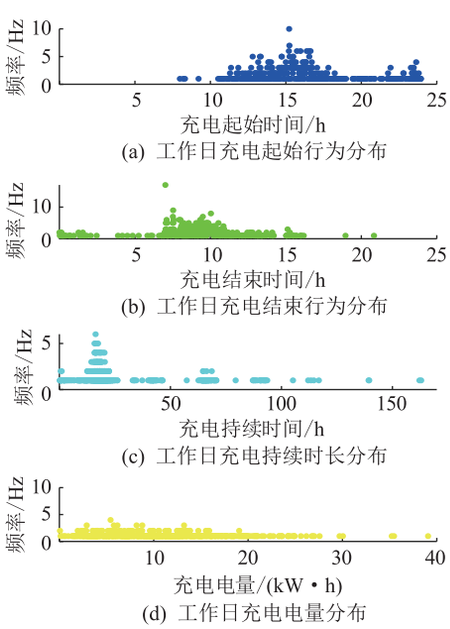


图 9 第五类 EV 充电行为

Fig.9 Class 5 EV charaing behavior

从图 5—图 9 可以看出,EV 符合总体与电网负荷高峰时段基本一致。工作日类别 2 所占比例最高,其主要充电时段是上午和下午,且消耗的电量最大,在一天中创造了第 1 个充电高峰期。工作日类别 3 和类别 4 的充电行为集中在上半夜,且以快充为主,在一天中创造了第 2 个充电高峰期。此外,双休日和节假日的充电行为趋势大致与工作日的

表 1 3 个时段的所有充电行为类别参数

Table 1 The charging behavior characteristics of each cluster

时段	类别	占比/%	起始时刻范围	结束时刻范围	平均充电时长/h	平均充电电量/(kW·h)
工作日	1	9.905 2	00:03~08:35	00:04~09:49	0.753 6	8.696 3
	2	56.729 9	08:21~18:51	08:53~23:59	3.494 4	7.413 0
	3	14.976 3	16:45~23:50	17:14~23:59	0.600 0	9.282 5
	4	2.251 2	20:35~22:46	06:00~08:38	9.696 0	10.524 5
	5	16.208 5	10:30~23:54	00:02~20:50	26.229 6	9.360 3
双休日	1	14.491 5	00:00~05:33	00:01~05:48	0.412 7	10.869 8
	2	13.644 1	00:07~23:58	00:06~23:51	18.986 4	13.089 1
	3	71.864 4	03:18~23:51	05:35~23:58	0.961 2	8.561 2
节假日	1	13.385 8	06:30~23:56	00:18~20:18	20.600 0	13.280 9
	2	86.614 2	00:00~23:18	00:07~23:49	0.690 1	8.552 2

行为接近,但单日充电电量要低于工作日,说明在工作日中 EV 用户的活动更加频繁。故需要区别考虑工作日、双休日和节假日中的 EV 充电特性。

4 结语

文中提出了一种基于随机森林和主成分分析的 EV 充电行为分类技术,用来辨识和分析不同类型的充电行为特征。采用 Dundee 市 2018 年 1 月的充电数据进行实验,分别得到该月工作日、双休日和节假日的充电行为分类。考虑充电可调度时段,可以将工作日中的 EV 充电行为分为凌晨至上午、中午至傍晚、下午至傍晚、整夜和全天 5 种,据此可以推断出用户的充电方式,同时将欧式距离法和随机森林算法进行对比,表明随机森林算法在 EV 充电行为聚类问题中具有优越性。但仍需指出的是,文中只使用了 5 个参数来表示充电行为,且部分时段的数据样本数量较少,可以研究不同类型 EV 的调度策略。

本文得到国网天津市电力公司科技项目(kj17-1-02)资助,谨此致谢!

参考文献:

[1] 文福拴,吴复立,倪以信. 电力市场环境下的发电容量充裕性问题[J]. 电力系统自动化, 2002, 19(26): 16-22.
WEN Fushuan, WU Fuli, NI Yixin. Generation capacity adequacy in the deregulated electricity market environment[J]. Automation of Electric Power Systems, 2002, 19(26): 16-22.

[2] CLEMENT K, HAESEN E, DRIESEN J. The impact of charging plug-in hybrid electric vehicles on the distribution grid[J]. IEEE Transactions on Power Systems, 2010, 25(1):371-380.

[3] OLIVELLA R P, VILLAFAFI I R, SUMPER A, et al. Probabilistic agent-based model of electric vehicle charging demand to analyse the impact on distribution networks [J]. Energies, 2015,8(5):4160-4187.

[4] 罗卓伟,胡泽春,宋永华,等. 电动汽车充电负荷计算方法[J]. 电力系统自动化, 2011, 35(14): 36-42.
LUO Zhuowei, HU Zechun, SONG Yonghua, et al. Study on plug-in electric vehicles charging load calculating[J]. Automation of Electric Power Systems, 2011, 35(14):36-42.

[5] 赵俊华,文福拴,杨爱民, 等. 电动汽车对电力系统的影响及其调度与控制问题[J]. 电力系统自动化, 2011, 35(14): 2-10.
ZHAO Junhua, WEN Fushuan, YANG Aimin, et al. Impacts of electric vehicles on power systems as well as the associated dispatching and control problem[J]. Automation of Electric Power Systems, 2011, 35(14): 2-10.

[6] STEEN D, TUAN L A, CARLSON O, et al. Assessment of electric vehicle charging scenarios based on demographical data[J]. IEEE Transactions on Smart Grid, 2012, 3(3):1457-1468.

[7] GALUS M D, WARAICH R A, NOEMBRINI F, et al. Integrating power systems, transport systems and vehicle technology for electric mobility impact assessment and efficient control [J]. IEEE Transactions on Smart Grid, 2012, 3(2):934-949.

[8] SALMASI, RAJAEI F. Control strategies for hybrid electric vehicles: evolution, classification, comparison, and future trends [J]. IEEE Transactions on Vehicular Technology, 2007, 56(5):2393-2404.

[9] 李永攀,黄兵,解大. 基于 k-means 聚类的电动汽车用户行为特征可视化分析[J]. 电气自动化, 2019, 41 (1): 12-15.
LI Yongpan, HUANG Bing, XIE Da. Visual analysis on the behavior characteristics of electric vehicle users based on k-means clustering [J]. Electrical Automation, 2019, 41(1):12-15.

[10] 郭世枭,罗晶晶,高亚静. 基于改进 *k*-means 聚类计及分布式光伏和电动汽车的园区负荷聚合体的最优构建[J]. 电力科学与工程, 2018, 34(3): 14-21.
GUO Shixiao, LUO Jingjing, GAO Yajing. Optimal construction of industrial park load polymers concerning the distributed photovoltaic and electric vehicles based on improved k-means clustering[J]. Electric Power Science and Engineering, 2018, 34(3): 14-21.

[11] WANG Y, XIA S T, TANG Q, et al. A novel consistent random forest framework: bernoulli random foests [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(8):3510-3523.

[12] DONG Y, DU B, ZHANG L. Target detection based on random forest metric learning[J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2017, 8

(4):1830-1838.

[13] 张雷,王琳琳,张旭东,等. 随机森林算法基本思想及其在生态学中的应用-以云南松分布模拟为例[J]. 生态学报, 2014, 34(3): 650-659.
ZHANG Lei, WANG Linlin, ZHANG Xudong, et al. The basic principle of random forest and its applications in ecology: a case study of Pinus yunnanensis[J]. Acta Ecologica Sinica, 2014, 34(3): 650-659.

[14] 曾青霞,杜建强,聂斌,等. 融合随机森林的偏最小二乘法及其中医药数据分析[J]. 计算机应用研究, 2018, 35(10): 1-5.
ZENG Qingxia, DU Jianqiang, NIE Bin, et al. PLS method and analysis of TCM data fusing random forest [J]. Application Research of Computers, 2018, 35(10): 1-5.

[15] 黄晗,孙堃,刘达. 基于随机森林的电力系统小时负荷预测研究[J]. 智慧电力, 2018, 46(5): 8-14.
HUANG Han, SUN Kun, LIU Da. Hourly load forecasting of power system based on Random Forest [J]. Smart Power, 2018, 46(5): 8-14.

[16] XIN Ma, JING Guo, KE Xiao, et al. PRBP: prediction of RNA-binding proteins using a random forest algorithm combined with an RNA-binding residue predictor[J]. IEEE/ACM Transactions on Computational Biology and Bio-informatics, 2015, 12(6): 1385 - 1393.

[17] 吴潇雨,和敬涵,张沛, 等. 基于灰色投影改进随机森林算法的电力系统短期负荷预测[J]. 电力系统自动化, 2015, 39(12): 50-55.
WU Xiaoyu, HE Jinghan, ZHANG Pei, et al. Power system short-term load fore casting based on improved random forest with grey relation projection [J]. Automation of Electric Power System, 2015, 39(12): 50-55.

[18] LI Y, GUO Z, YANG J, et al. Prediction of ship collision risk based on CART[J]. IET Intelligent Transport Systems, 2018, 12(10): 1345-1350.

[19] WEBBER J R, GUPTA Y P. A sieve bootstrap method for correlation analysis[J]. IEEE Transactions on Automatic Control, 2007, 52(6): 1079-1081.

[20] YU Zhiwen, CHEN Hantao, YOU Jane, et al. Adaptive fuzzy consensus clustering framework for clustering analysis of cancer data[J]. IEEE/ACM Transactions on Computational Biology and Bio-informatics, 2015, 12(14): 887 - 901.

[21] POUYAN M B, NOURANI M. Clustering single-cell expression data using random forest graphs [J]. IEEE Journal of Biomedical and Health Informatics, 2016, 21(4): 1-1.

作者简介:


刘亚丽

刘亚丽(1984),女,硕士,高级工程师,从事电能质量评估治理、新能源与电动汽车相关工作(E-mail:liuyali_sdu@163.com);

李国栋(1978),男,硕士,高级工程师,从事智能配用电技术、电能质量、电力系统分析工作;

刘云(1983),男,硕士,高级工程师,从事新能源技术、电动汽车接入技术相关工作。

Clustering technology of electric vehicle charging behavior based on Random Forest

LIU Yali[1], LI Guodong[1], LIU Yun[1], HONG Yi[2], LIU Yujun[2]

(1. State Gird Tianjin Electric Power Co., Ltd. Research Institute, Tianjin 300392, China;
2. School of Electrical Engineering, Southeast University, Nanjing 210096, China)

Abstract: Since the Chinese government continuously support the development of new energy vehicles (EVs), the charging process of EVs will generate big data regarding the EVs charging behavior. This paper proposes a big data mining technique based on Random Forest (RF) and Principle Component Analysis (PCA) for EV charging behavior to identify and analyze clusters with different charging characteristics. Then, Dundee's EV charging data in the January of 2018 is applied to conduct experiments, and respectively obtains the charging behavior clusters of the workdays, weekends and holidays. Finally, the RF algorithm in the EV clustering problem is compared to the Euclidean distance method and the clusters obtained by RF get more convinced characteristics.

Keywords: electric vehicle; charging behavior; Random Forest; cluster analysis

(编辑 杨卫星)