

模糊 c 均值在非技术损失检测中的应用

张玉振,冯晓蒲,丁巧林

(华北电力大学 电气与电子工程学院,河北 保定 071003)

摘要: 使用电力公司用户实际负荷曲线数据和模糊聚类中的经典算法模糊 c 均值(FCM)进行聚类分析,得到负荷簇和负荷代表曲线,然后通过比较偏差用户负荷曲线与所属行业负荷代表曲线的距离,以及与偏差行业代表负荷曲线的距离,分析可能发生异常电能消费行为的用户,电力公司据此可进一步采取针对性措施减少损失。

关键词: 电力系统;非技术损失;负荷曲线;模糊 c 均值

中图分类号: TP39

文献标识码: A

文章编号: 1674-7720(2010)20-0097-03

Application of fuzzy c-means in nontechnical losses detection

ZHANG Yu Zhen, FENG Xiao Pu, DING Qiao Lin

(School of Electric and Electronic Engineering, North China Electric Power University, Baoding 071003, China)

Abstract: Fuzzy c-mean (FCM) clustering algorithm is used to cluster actual load curves of electricity customers and obtain load clusters and representative load profiles, then by comparing the distance between load curves of outlier customers and the representative load profiles of their belonging industry with the distance between load curves of outlier customers and the representative load profiles of their outlier belonging industry. Thus, the customers with potential abnormal behaviors are identified, which provides a basis for power utilities to take pertinent measures to reduce losses.

Key words: power system; nontechnical losses; load profile; fuzzy c-mean

无论在发达国家还是发展中国家,非技术损失在电能损失中都占有很大比例,因此对电力用户的非技术损失进行检测识别已经成为一个很活跃的课题。非技术损失相关技术^[1]的研究和使用能够将可能的窃电问题筛选出来,为企业针对性处理提供依据。实践证明在预防窃电和计量错误方面可为企业追回上千万元损失,同时可为供电公司节省人力物力,具有重大意义。

负荷曲线是用户电能消费行为的直观表示,是负荷分析的数据来源。通过负荷曲线可以更清楚地观察到用户的异常和不规律性。目前,基于负荷特性的用电异常检测的方法有很多,例如粗糙集、决策树、人工神经网络、小波分析、聚类分析等。其中,聚类分析方法^[2]在异常检测中占有重要的地位。本文首先介绍了非技术损失的概念,然后选用聚类技术中的模糊 c 均值 FCM (fuzzy c-mean)对某供电公司的 100 个电力用户负荷曲线进行了聚类分析,以找出那些可能存在违规用电行为的客户。

1 非技术损失

非技术损失^[3]主要涉及电力盗窃和用户管理过程中

大量存在的各式各样欺骗公共资源的手段。大多数国家中,非技术损失占有输配电损失很大比例,这意味着比起技术损失电力部门在减少非技术损失方面更需加大力度,采取有效措施。

非技术损失包括如下活动:(1)篡改表计以便使表计少记录消费的电能;(2)绕过表计窃取或以其他方式非法连接;(3)通过贿赂读表人员提供虚假读数;(4)在内部人员的帮助下通过开出较低账单或改变电价(赋予较低电价类属或享受政策性电价优惠)等方式进行账单违规,改变账单小数点的位置或忽略未付账单。

默认情况下,电能生产总量应等于电能消费总量。然而现实中因输配过程中存在电能损失而有所不同。有一些电能损失是不可避免的,但可以使其最小化。所应用的措施中一些基于技术,另一些则依赖人工努力和创造。

减少非技术损失对供电公司至关重要。由于这些损失集中在中低压网络,遍布整个电力系统,而现行处理非技术损失的方法多基于现场检查,这样不但增加了运

应用奇葩

Example of Application

营成本并且要求大量使用人力资源,因此本研究旨在减少检测非技术损失活动的成本。

2 模糊 c 均值

目前聚类技术在许多领域都得到了广泛的研究和成功的应用,这些领域包括模式识别、数据分析、图像处理、市场研究、客户分割,Web 文档分类等^[4]。

对负荷曲线进行聚类进行负荷模式分析是获取电力用户典型负荷曲线和按照负荷特性进行用户分类的基础,对于负荷预测、负荷控制、用电异常检测甚至电价目录制定和开发营销策略等都有重要意义^[5]。

已有研究表明,通过负荷模式分析获取用户正常负荷模式可用于用电异常检测(窃电或计量错误)分析^[6],能够减少人工检测的工作量和提高命中率。

本文应用模糊聚类中经典的模糊 c 均值算法通过负荷曲线聚类获得负荷模式以检测用户的用电违规行为。

模糊 c 均值^[7]算法是一种以隶属度来确定每个数据点属于某个聚类程度的算法,该聚类算法是传统硬聚类算法的一种改进。

FCM 的算法原理如下:设数据集 $X=\{x_1, x_2, \dots, x_n\}$, 它的模糊 c 划分可用模糊矩阵 $U=[u_{ij}]$ 表示,矩阵 U 的元素 u_{ij} 表示第 $j(j=1, 2, \dots, n)$ 个数据点属于第 $i(i=1, 2, \dots, c)$ 类的隶属度, u_{ij} 满足如下条件:

$$\forall j, \sum_{i=1}^c u_{ij}=1; \forall i, j, u_{ij} \in [0, 1]; \forall i, \sum_{j=1}^n u_{ij} > 0$$

FCM 算法就是求使聚类目标函数最小化的划分矩阵 U 和聚类中心 V 。即:

$$\min J_m(U, V) = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m d_{ij}^2(x_j, v_i)$$

其中 n 是样本数据集的个数; c 是聚类中心数; m 为加权指数; d_{ij} 是样本点和聚类中心的欧氏距离,即: $d_{ij}(x_j, v_i) = \|v_i - x_j\|$ 。

算法流程如下:

- (1)调入用户负荷数据,对数据进行标准化;
- (2)输入要聚类的数目,进行聚类;
- (3)得到聚类中心矩阵、隶属度矩阵、目标函数值;
- (4)根据隶属度矩阵找到数据点所属的类别。

3 案例分析

3.1 数据来源

本文对某供电公司 100 个电力用户的实际电能消费数据进行分析,识别用电异常用户,以降低企业减少非技术损失的成本。这些数据均通过自动抄表系统收集,采集间隔为 30 min。其中包含的行业有公用事业、机械加工、建材、轻工加工、商业、行政事业。

3.2 基于 FCM 聚类分析

各行业包含多个用户,按用户行业将负荷曲线聚为 6 类,得到的聚类结果如图 1 所示(横坐标为 48 个时刻;

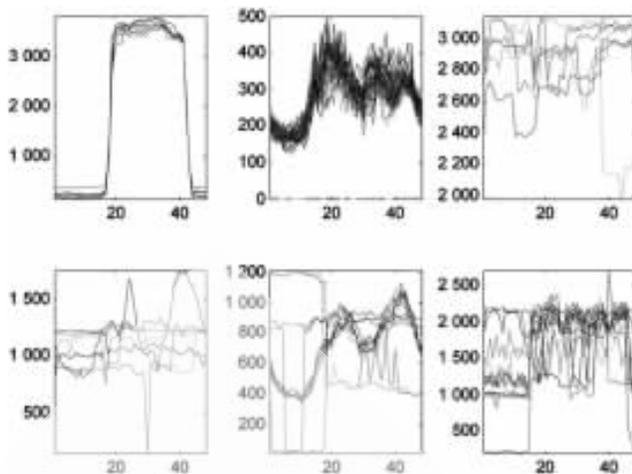


图 1 聚类后各行业的负荷曲线

纵坐标为时段内电力用户用电量,单位:kWh)。图中曲线簇代表的行业依次为商业、公用事业、轻工加工、建材、行政事业和机械加工。

聚类后各行业包含电力用户如下:

商业:85、86、87、88、89、90、91、92、93。

公用事业:1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、82。

轻工加工:71、72、73、74、75、76。

建材:48、64、65、67、68、80、81。

行政事业:62、63、66、69、70、94、95、96、97、98、99、100。

机械加工:49、50、51、52、53、54、55、56、57、58、59、60、61、77、78、79、83、84。

基于以上结果分析可知:各行业的负荷曲线大部分可以聚为一类,商业和公用事业用户负荷曲线可以准确无误地聚在一起。出现偏差的用户有 13 个,分别为 62、63、66、69、70、77、78、79、80、81、82、83 和 84。

图 2 为根据图 1 得到的各行业中心代表曲线。

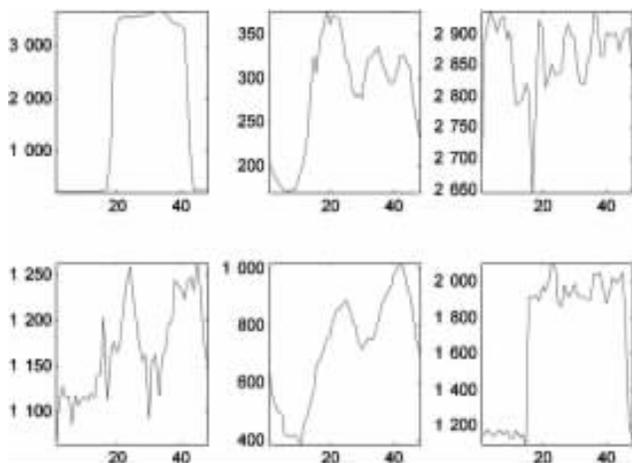


图 2 各行业的中心代表曲线

应用奇葩

Example of Application

依据负荷分布,为了进一步确定出现异常的电力用户,对偏差用户负荷曲线与原属行业和偏差行业中心代表曲线的距离进行计算。定义两条曲线的距离 d ,其中 i 表示时刻点(48个), $p_{中心,i}$ 、 $p_{用户,i}$ 分别表示行业中心代表曲线与各用户负荷曲线对应时刻的负荷值。

$$d = \sqrt{\frac{1}{48} \sum_{i=1}^{48} (p_{中心,i} - p_{用户,i})^2}$$

d 的值越小表明两条曲线间的距离越小,两条曲线的相似度就越高。计算得到偏差用户与原属行业和偏差行业的距离如表1所示。

表1 用户负荷曲线和行业中心代表曲线距离表

负荷曲线聚类 出现偏差用户	与原属行业中心 代表曲线的距离	与偏差行业中心 代表曲线的距离
62	302.448 9	235.962 0
63	58.756 0	202.965 6
66	368.639 8	339.625 0
69	401.603 7	459.054 5
70	539.542 8	288.200 2
77	128.752 5	504.961 7
78	154.011 3	561.831 0
79	381.672 2	569.710 4
80	903.578 4	617.432 5
81	1 038.800 0	729.108 7
82	2 048.300 0	307.332 3
83	448.749 0	342.178 1
84	498.589 6	309.855 7

进一步分析得知:出现偏差的用户负荷曲线与原属行业中心代表曲线的距离明显大于与偏差行业中心代表曲线距离的用户有62、66、69、70、80、81、82、83和84。这些用户曲线形状与原属行业中的其他用户的负荷曲线存在显著差异,表明该用户的用电负荷情况可能出现了异常。而对于63、77、78、79这几个用户与原属行业中心负荷曲线距离小但发生了偏差,是因为FCM聚类算法存在固有的缺陷,对初始聚类中心比较敏感,导致聚类结果不是百分之百正确,关于如何改进聚类算法,提高其精度,有待进一步研究。

对于筛选出的用户,可从以下方面排查:该用户被划入不恰当的行业类型或者该用户发生了窃电或计量出现异常。前者更正即可,后者进行现场检查排查或后续作为重点检查对象。这样可缩小用电稽查范围,避免盲目检查造成的运营成本增加和人力浪费,也可减少扰民,改善企业的社会形象。

本文研究通过模糊 c 均值对负荷曲线聚类实现用电异常用户识别,以减少供电企业非技术损失。

论文首先介绍了非技术损失和模糊 c 均值聚类方法,然后对实际企业的100个用户负荷典型曲线进行聚类,得到行业中心曲线和偏差用户,进一步通过计算偏差用户和原属行业与偏差行业的距离确认那些有用电异常行为的用户,从而有效减少用电稽查的范围,最终降低企业营运成本和节约人力资源。

参考文献

- [1] NIZAR A H, DONG Z Y. Identification and detection of electricity customer behavior irregularities[C]. Power Systems Conference and Exposition, PSCE 2009.
- [2] 刘燕, 梁云娟. 基于聚类分析的非监督式异常检测研究[J]. 河南科技学院学报, 2006, 34(2): 1673-1676.
- [3] NAGI J, YAP K S, TIONG S K. Nontechnical loss detection for metered customers in power utility using support vector machines[J]. IEEE Transactions on Power Delivery, 2010, 25(2): 1162-1171.
- [4] HAN Jia Wei, KAMBER M. 数据挖掘概念与技术[M]. 范明, 孟晓峰, 译. 北京: 机械工业出版社, 2001.
- [5] 仲伟宽. 模糊聚类方法在用户负荷曲线分析中的应用[J]. 华东电力, 2007, 32(8): 97-100.
- [6] NIZAR A H, DONG Z Y. Load profiling method in detecting non-technical loss activities in a power utility. Power and Energy Conference, 2006. PECon'06. IEEE International.
- [7] 吕晓燕, 罗立民, 李祥生. FCM算法的改进及仿真实验研究[J]. 计算机工程与应用, 2009, 45(20): 144-147.

(收稿日期: 2010-05-24)

作者简介:

张玉振,男,1973年生,工程师,硕士研究生,主要研究方向:电力信息处理。