

基于负荷数据挖掘的公变用途分类方法研究

方超, 仲春林, 季聪

(江苏方天电力技术有限公司, 江苏 南京 211102)

摘要:为了从公变行业特性角度对配电网进行有效的投资经济效益分析,需要对公变进行用途划分。文中先利用大数据技术对行业负荷特性进行分析,再通过基于 Canopy 的改进 K-means 聚类方法分析公变负荷特性,进而根据余弦相似度算法分别从四季负荷特性及节假日负荷特性等角度分析公变所属行业,最后利用行业拟合方法从不同维度匹配出公变的所属行业。采用该方法对江苏省公变设备进行用途划分取得了较高的准确率,实现了较为准确的公变设备用途划分。基于负荷数据挖掘的公变用途划分方法为配电网投资效益评价提供了有力的支撑。

关键词:负荷特性;负荷聚类;行业划分

中图分类号: TM714

文献标志码: A

文章编号: 2096-3203(2018)05-0115-06

0 引言

随着我国电力市场化改革的不断深入,供电企业作为市场经济中的主体,秉承自主经营、自负盈亏和追求效益最大化的市场原则。因此,电网企业必须在充分考虑社会效益的同时,追求投资经济效益的最大化^[1]。电网项目的投资效益分析是电网建设项目决策科学化、减少和避免决策失误以及提高项目建设经济效益的重要手段^[2]。配电网作为智能电网的重要组成部分,其安全、可靠、经济运行对于保障主电网的平稳运行和为电力用户提供优质的电力产品具有重要意义^[3-5]。其中,依据负荷特性对配电网中的公变进行有效的用途划分,进而分析公变用电量需求及用电趋势是一种评价配电网投资效益的有效方式。

目前负荷特性分析的方法众多,文献[6—8]提出由于数据匮乏,配电网的负荷研究需要采用对典型用户进行采样测量的方式进行,从概率分布函数、置信度、均值和方差等多种角度对测量数据进行统计分析,得到典型负荷模式。此种分析方法比较合理且全面,但需要以大量的人力物力为代价,在我国运用这种方法不太实际。文献[9]提出了有序用电用户可中断负荷分析方法,通过对用户的历史负荷进行聚类,得到用户典型日负荷曲线,在此基础上制定用户有序用电策略。文献[10]通过问卷调查和相关文献分析居民用电负荷特性,总结居民峰谷电价用电特性及影响因素。文献[11]运用模糊均值聚类(fuzzy C-means, FCM)方法对工业用户进行分类并对负荷特性进行分析。文献[12]采用先确定聚类数目和聚类中心后的改进型 FCM 聚

类方法对典型用户日负荷进行分类和分析。文献[13—16]以某一地区作为研究对象分析了负荷特性。行业的归类划分可以从负荷特性分析的角度进行,但是公变的负荷特性分析则需要采用聚类分析方法。基于 Canopy 的改进 K-means 聚类算法^[17-19]是一个被广泛使用的聚类算法,可以用于公变的负荷特性分析。

1 总体框架

本文从负荷特性分析的角度对公变用途进行划分,首先对多种负荷数据进行清洗,将可能影响行业划分的各种异常数据做相应的处理。在高质量数据的基础上再进行各行业负荷的归类,归类出行业负荷特性曲线。然后,对需要划分用途的公变负荷数据利用基于 Canopy 的改进 K-means 聚类算法进行聚类分析,得到公变负荷特性曲线。最后,利用余弦相似性算法计算不同维度公变与各行业的相似度,最终拟合出相似度最高的行业即为该公变的所属行业,完成公变的用途划分。基于公变的用途划分结果,相关部门可以依据相关特点分析配电网的投资经济效益,总体框架流程如图 1 所示。

2 数据清洗

高质量的配电网负荷数据是实现公变用途划分的重要保障,也是电力部门分析配电网投资效益的主要事实依据。数据库的不良数据可能会对系统的安全运行造成重大影响,因此在公变用途划分前对配电网负荷数据进行清洗是十分必要的。

负荷数据一般为高精度数且数值敏感性强,加之公变用途分析方法较为复杂,即使对少量数据处理也需要较高的存储和计算成本。整个江苏省有

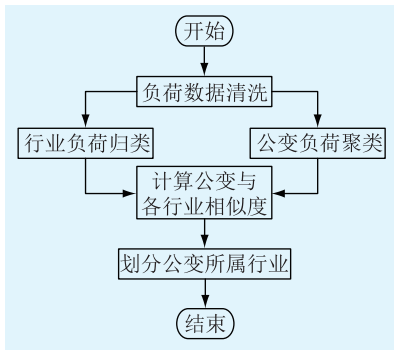


图1 总体框架流程

Fig.1 Overall framework flow chart

四十几万公变设备,每天会产生大量的负荷数据,面对如此庞大的数据,数据清洗的开销可想而知。因此,数据清洗对公变用途划分是十分重要的。

负荷数据在采集过程中存在如下的现实问题:

(1) 一天内采集的数据点个数不同,根据设备类型分为24点(每1h采一个数据),48点(每0.5h采一个数据)和96点(每15min采一个数据)。

(2) 部分数据由于各种原因没有及时获取,导致数据点的缺失。

(3) 部分数据出现明显的偏差,导致异常数据入库。

针对以上问题,需要对负荷数据进行清洗操作,具体的清洗方法如下。

2.1 数据采集点数不同的处理

针对24点,48点原始数据作规范化处理,统一换算到96点负荷。

24点负荷数据处理方法如下:由于数据处理过程中需要用到次日0点的数据,但采集系统还没有采集到次日数据,因此将当日0点数据复制到24点负荷数据序列的末尾,形成一个25点数据序列。假定25点数据序列为 $\{A_1, A_5, \dots, A_{93}, A_{97}\}$,中间空缺的数据序列即需要补全的数据,计算公式如下:

$$\begin{cases} A_{4i-2} = \frac{3A_{4i-3} + A_{4i+1}}{4} \\ A_{4i-1} = \frac{A_{4i-3} + A_{4i+1}}{2} \\ A_{4i} = \frac{A_{4i-3} + 3A_{4i+1}}{4} \end{cases} \quad (i = 1, \dots, 24) \quad (1)$$

48点负荷数据处理方法如下:在48点负荷数据序列后添加当日0点负荷数据,形成49点数据序列 $\{A_1, A_3, \dots, A_{95}, A_{97}\}$,中间空缺的数据序列即需要补全的数据,计算公式如下:

$$A_{2i} = \frac{A_{2i-1} + A_{2i+1}}{2} \quad (i = 1, \dots, 48) \quad (2)$$

2.2 缺失数据处理

针对缺失的负荷数据,例如当负荷 A_i, A_j (其中 $i > j$)已成功采集入库,而 A_i, A_j 之间 $j-i-1$ 个点没有成功采集,这时需要将缺失的数据补全。补全方法采用构建 A_i, A_j 两点确定的线性方程,补算中间缺失的负荷值:

$$A_{i+k} = A_i + k \frac{A_j - A_i}{j - i} \quad k = 1, \dots, j - i - 1 \quad (3)$$

2.3 异常数据处理

针对超大或超小数据处理,先确定比例阈值 N ,在一连续的负荷数据序列 $A_n = \{A_1, A_2, \dots, A_n\}$ 中,数据是否异常的判定方法如下:

$$A_i \begin{cases} \text{超大数据} A_i > A_{i+1} \times N \text{ 且 } A_i > A_{i-1} \times N \\ \text{超小数据} A_i < A_{i+1}/N \text{ 且 } A_i < A_{i-1}/N \\ \text{正常数据} \end{cases} \quad (4)$$

超大数据和超小数据的修正公式如下:

$$A_i = \frac{A_{i-1} + A_{i+1}}{2} \quad (5)$$

根据用户容量参数,校验用户负荷数据,排查不符合容量的异常数据,同时应用超大或超小校验,排查并换算此类异常数据,最终得到经清洗后的用户96点的日负荷数据。

3 分行业负荷特性提取

目前国内外对负荷特性的分析大多采用典型日负荷特性分析方法,即先确定行业的典型日或者典型时间段,然后根据选定的典型日或典型时间段采用相应的聚类方法获得行业的负荷特性分析结果。采用这种方法分析负荷特性存在一定问题,例如典型日选取一旦出现偏差会对行业负荷特性分析结果产生较大影响。本文对于行业负荷特性的分析采用对大量行业负荷数据直接归类的方法。在大数据技术的支持下,直接产生行业负荷特性分析结果,使大数据量的同行业负荷数据的累计效应得以展现。

首先划分四季时间段,3、4、5月为春季,6、7、8为夏季,9、10、11月为秋季,12、1、2月为冬季。根据大数据量负荷曲线展现出的负荷特性,剔除离散程度较大的曲线或者负荷点,将特性一致的负荷曲线进行归类统一,形成汽车、摩托车、燃料及零配件专门零售业四季负荷特性曲线,如图2所示。

按照同样的方法,可以得到纺织、服装及日用品批发业四季负荷特性曲线,如图3所示。

从两个行业的四季负荷特性曲线可以发现汽车、摩托车、燃料及零配件专门零售业冬、夏负荷

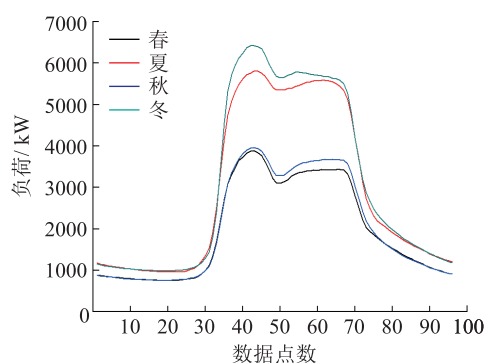


图2 汽车、摩托车、燃料及零配件专门零售业四季负荷特性

Fig.2 Four seasons load characteristics of automobile, motorcycle, fuel and spare parts specialized retail industry

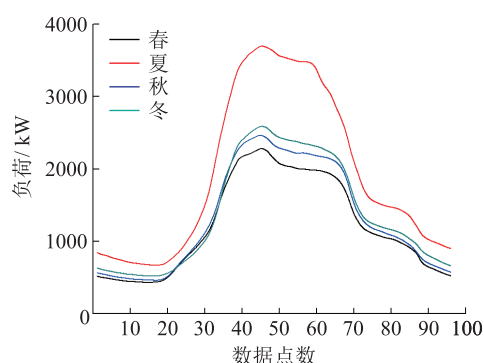


图3 纺织、服装及日用品批发业四季负荷特性

Fig.3 Four seasons load characteristics of textile, clothing and daily necessities wholesale industry

高,纺织、服装及日用品批发业夏季负荷高,体现了不同行业在不同季节的需求量的差异。

为了突出负荷特性曲线的行业特点,减少因数据绝对值差异导致的行业负荷特点分散的现象,需要将行业负荷特性数据做归一化处理,便于后续的行业匹配。

通过大数据汇集产生的行业本身的负荷特性展现方法可以很好地归类出行业间的负荷特性区别,此方法不仅可以归类出行业四季负荷特性,也可以获取行业节假日负荷特性等信息,为后续公变的负荷特性匹配提供多维度的参考角度。

4 公变负荷特性分析

公变负荷特性分析是根据公变的历史负荷数据,采用聚类算法进行多次迭代聚类,聚合出公变的四季、节假日等负荷特性分析结果,用于对公变的用途进行划分。

为了提高普通 K -means 聚类算法的运算效率,对于公变的负荷特性聚类采用基于 Canopy 的改进 K -means 聚类算法。分两个阶段执行:第一阶段把

处理后的每一个数据看做一个中心点,按 Canopy 算法产生一些可覆盖的划分,初始划分以随机选的第一个数据作为本集合标识。对于后续每个数据都须判断它是否落入先前产生的集合中,如果没有落入集合中则产生一个新的集合,并以此数据为集合标识。然后对每个集合内的点采用不同的距离度量方法,形成重叠的集合。第二阶段用根据 Canopy 算法计算产生的众多集合代替 K -means 算法初始随机选择的 K 个聚类中心点,由于已经对数据进行可覆盖的划分,在计算数据间距离时不必计算所有的距离,只需要计算本集合内的数据距离即可,然后对子集内的点进行迭代聚类,可以得到公变的负荷特性聚类分析结果。这种改进的聚类方法进一步提高了聚类算法的计算效率,减少了传统聚类算法中对所有数据点进行精确计算的计算量,另外允许有重叠的子集合也起到了增加容错性和消除孤立点作用。

5 公变用途分析

通过前面介绍的公变负荷特性分析方法对公变的四季、节假日等不同维度负荷特性进行聚类分析,得到公变各维度的聚类结果。在负荷特性曲线上取合适的比较点(一般为 96 点),根据余弦相似度算法计算各维度中行业负荷特性与公变负荷聚类结果的相似度,得到公变四季和节假日的最相似行业,由于不同维度计算出的相似行业不尽相同,需要通过各维度分析结果拟合出最相似的行业,拟合过程如下。

四季负荷分析:计算春、夏、秋、冬四季的公变负荷与各行业负荷相似度,为了扩大分析范围,提高行业划分准确度,取每个季节中相似度最高的 3 个行业,分行业计算所有季节的相似度之和。

节假日负荷分析:与四季负荷类似,计算清明节、劳动节、端午节和国庆节的公变负荷与各行业负荷的相似度,取每个节假日中相似度最高的 3 个行业。分行业计算所有节假日的相似度之和。

最相似行业拟合:取四季和节假日维度中相似度之和最高的两个行业进行比较,记相似度之和较高的行业为公变的所属行业。根据负荷历史数据比较发现,四季的行业特性优于节假日的行业特性,因此,如果两者的相似度之和相等,则取四季负荷中相似度之和最高的行业为公变所属行业,完成公变用途划分。

6 算例分析

以江苏省为例,从四季负荷特性中的春季负荷

角度出发对公变用途划分进行说明,利用全省春季行业数据归类提取行业负荷特性曲线,为了便于比较,对数据进行归一化处理,得到全省各行业春季负荷特性曲线,其中部分行业春季负荷特性曲线如图4所示。

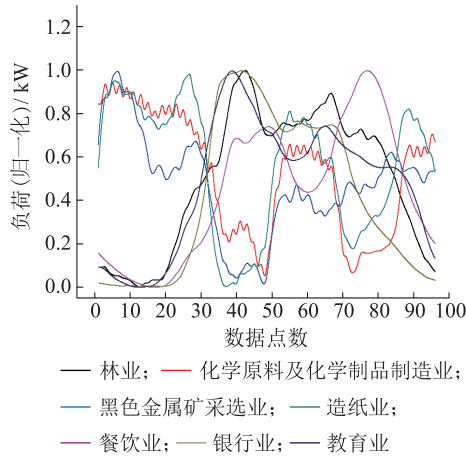


图4 部分行业春季负荷曲线

Fig.4 Part of the industry spring load curve

对全省四十几万公变的历史春季负荷数据利用基于Canopy的改进K-means聚类算法进行迭代聚类,得到12类公变春季负荷聚类数据。再进行归一化处理,得到公变春季负荷聚类曲线,如图5所示。

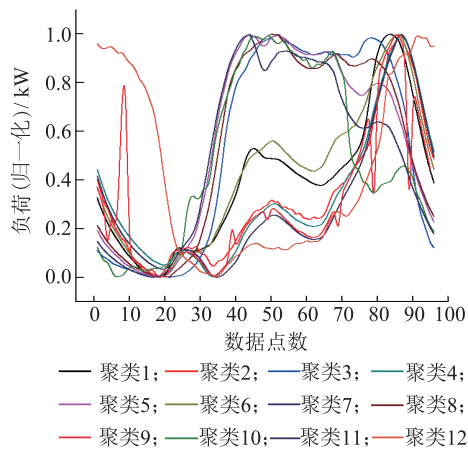


图5 公变春季负荷聚类曲线

Fig.5 Spring load clustering curve of public transformers

利用余弦相似度算法,将公变春季负荷各聚类结果与行业春季负荷特性进行匹配。例如,有公变设备A存在于聚类8集合中,而与聚类8最相似的前3个行业分别是房地产业、水的生产和供应业以及农业,相似度分别为98.21%、91.23%和85.34%。图6为与聚类8最相似的前3个行业春季负荷特性匹配图。

由春季负荷分析可知,公变A最有可能属于房

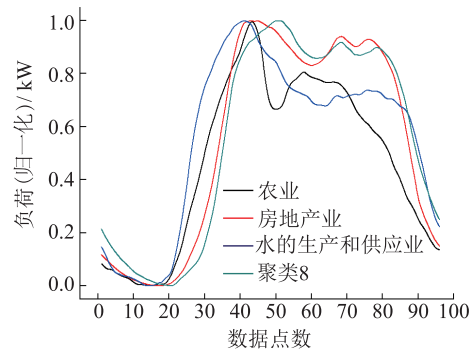


图6 聚类8春季行业负荷匹配

Fig.6 Clustering 8 spring industry load matching chart

地产业。继续按照同样方法计算公变A所处聚类结果集夏、秋、冬三季以及各节假日负荷与各行业的相似度,每个维度中取相似度最高的3个行业纳入比较范围。分行业计算四季和节假日中相似度之和,取四季维度内相似度之和最高的行业与节假日维度内相似度之和最高的行业比较。相似度之和较高的行业即为公变A的所属行业;如果相似度之和相等,则取四季负荷中相似度之和最高的行业为公变A的所属行业,完成公变用途划分。

对江苏省公变设备采用该方法进行公变用途划分并对划分结果进行统计分析,将分析结果与公变档案信息进行对比分析后发现用途划分平均准确率达到91%,其中,工业的用途划分准确率最高,达到95%。

7 结语

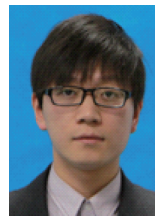
本文以一个新的角度分析配电网投资效益,即根据公变的行业特性划分用途,进而根据公变所属行业分析配电网投资的经济效益。公变用途划分过程中先对初始数据进行数据清洗,保证数据质量。再根据大数据汇集方法分析行业的负荷特性,从四季和节假日的角度对公变负荷特性采用基于Canopy的改进K-means聚类算法进行负荷特性分析,利用余弦相似度算法计算行业相似度,最后根据相似度比较分析出公变的所属行业,完成公变用途划分。实践证明,采用该方法对公变进行用途划分是行之有效的。

参考文献:

- [1] 柴文,白建勇. 电网项目投资决策问题研究[J]. 商情, 2013(23):169.
CHAI Wen, BAI Jianyong. Research on investment decision of power grid project[J]. Shang Qing, 2013(23):169.
- [2] 顾韧,陈俊,沈炯,等. 配电网投资收益评价模型及应用[J]. 智能电网, 2017, 5(5):480-485.
GU Ren, CHEN Jun, SHEN Jiong, et al. Evaluation model for

- investment income of distribution network and its application [J]. Smart Grid, 2017, 5(5):480-485.
- [3] 刘福炎. 基于投资关联性分析的配电网投资效益评价体系研究[J]. 浙江电力, 2016, 35(3):68-71.
LIU Fuyan. Study on investment efficiency evaluation system for distribution networks based on Investment correlation analysis [J]. Zhejiang Electric Power, 2016, 35(3):68-71.
- [4] 吴哲慧, 王光增, 钟伟, 等. 配电网发展与投资决策动态评估方法[J]. 广东电力, 2017, 30(4):11-14, 49.
WU Zhehui, WANG Guangzeng, ZHONG Wei, et al. Dynamic assessment method for power distribution network development and investment decision-making [J]. Guangdong Electric Power, 2017, 30(4):11-14, 49.
- [5] 林女贵. 电力市场下多投资方微电网竞价均衡研究[J]. 广东电力, 2018, 31(6):40-45.
LIN Nyugui. Research on bidding equilibrium of microgrid with multiple investors under electric power market [J]. Guangdong Electric Power, 2018, 31(6):40-45.
- [6] JARNIDT J A, TAHAN C M V, GOCVEA M R, et al. Daily load profiles for residential, commercial and industrial low voltage consumers [J]. IEEE Transactions on Power Delivery, 2000, 15(1):375-380.
- [7] SEPPALA. Statistical distribution of customer load profiles [A]. International Conference on Energy Management and Power Delivery, 1995, 2:696-701.
- [8] CHARYTIONIUK W, CHEN M S, KOTIS P, et al. Forecasting in power distribution system using nonparametric probability density estimation [J]. IEEE Transactions on Power System, 1999, 14(4):1200-1206.
- [9] 颜庆国, 薛溟枫, 范洁, 等. 有序用电用户负荷特性分析方法研究[J]. 电力工程技术, 2014, 33(6):48-50.
YAN Qingguo, XUE Mingfeng, FAN Jie, et al. Load property analysis method for demanders participating orderly power utilization [J]. Jiangsu Electrical Engineering, 2014, 33(6):48-50.
- [10] 林启开, 王珂, 余昆, 等. 峰谷电价下居民用电聚合响应特性分析[J]. 电力工程技术, 2017, 36(3):88-93.
LIN Qikai, WANG Ke, YU Kun, et al. Analysis on the polymeric response characteristics of residents under the peak and valley electricity price [J]. Electric Power Engineering Technology, 2017, 36(3):88-93.
- [11] 蔡敏, 沈培锋, 朱红, 等. 基于模糊聚类的配电网负荷特性分析[J]. 陕西电力, 2015, 43(4):24-27.
CAI Min, SHEN Peifeng, ZHU Hong, et al. Load characteristic analysis of distribution grid based on fuzzy clustering [J]. Shaanxi Electric Power, 2015, 43(4):24-27.
- [12] 刘永光, 孙超亮, 牛贞贞, 等. 改进型模糊C均值聚类算法的电力负荷特性分类技术研究[J]. 电测与仪表, 2014, 51(18):5-9.
LIU Yongguang, SUN Chaoliang, NIU Zhenzhen, et al. Research on the improved fuzzy C-means clustering algorithm based power load characteristic classification technology [J]. Electrical Measurement & Instrumentation, 2014, 51(18):5-9.
- [13] 杨建萍. 南京市居民用电负荷特性分析与思考[J]. 电力需求侧管理, 2002, 4(2):24-27.
YANG Jianping. Analysis and study of Nanjing residential load profile [J]. Power Demand Side Management, 2002, 4(2):24-27.
- [14] 姜勇. 南京市居民夏季用电负荷特性分析[J]. 继电器, 2003, 31(4):24-26.
JIANG Yong. Analysis of residents' summer load characteristics in Nanjing [J]. Relay, 2003, 31(4):24-26.
- [15] 薛少华. 银川地区配电网负荷特性分析及预测[J]. 宁夏电力, 2013(5):1-4.
XUE Shaohua. Analysis and forecast of load characteristics of distribution network in Yinchuan area [J]. Ningxia Electric Power, 2013(5):1-4.
- [16] 高蓉, 徐讯, 周江昕, 等. 上海市安居小区电力负荷特性分析[J]. 华东电力, 2012(6):1064-1066.
GAO Rong, XU Xun, ZHOU Jiangxin, et al. Power load characteristics of affordable housing project quarters [J]. East China Electric Power, 2012(6):1064-1066.
- [17] 刘莉, 王刚, 翟登辉. K-means 聚类算法在负荷曲线分类中的应用[J]. 电力系统保护与控制, 2011, 39(23):65-68.
LIU Li, WANG Gang, ZHAI Denghui. Application of K-means clustering algorithmic load curve classification [J]. Power System Protection and Control, 2011, 39(23):65-68.
- [18] 毛典辉. 基于 Mapreduce 的 Canopy-Kmeans 改进算法[J]. 计算机工程与应用, 2012, 48(27):22-26.
MAO Dianhui. Improved Canopy-Kmeans algorithm of based on mapreduce [J]. Computer Engineering and Applications, 2012, 48(27):22-26.
- [19] 赵庆. 基于 Hadoop 平台下的 Canopy-Kmeans 高效算法[J]. 电子科技, 2014, 27(2):29-31.
ZHAO Qing. Efficient algorithm of Canopy-Kmeans based on hadoop Platform [J]. Electronic Sci. & Tech., 2014, 27(2):29-31.

作者简介:



方超

方超(1985—),男,本科,工程师,从事电力软件开发与设计工作(E-mail: 13770328566@163.com);

仲春林(1981—),男,本科,工程师,从事电力软件开发管理工作(E-mail: 15905166657@139.com);

季聪(1988—),男,硕士,工程师,从事电力软件开发与设计工作。

Analytical Method of the Industry Which Public Transformers Belong Based on Load Data Mining

FANG Chao, ZHONG Chunlin, JI Cong

(Jiangsu Frontier Electric Technology Co.,Ltd., Nanjing 211102, China)

Abstract: In order to effectively analyze the economic benefits of investment of the electricity distribution network from the perspective of the characteristics of the public transformer industry, it is necessary to classify the industry of public transformers. In this paper, the use of big data technology to analyze the industry load characteristics, and then the improved K -means algorithm based on canopy is used to analyze the characteristics of the public transformer load. After that, according to the cosine similarity algorithm, from the perspective of the characteristics of the four seasons load and the characteristics of holiday load, the analysis of public transformers belong to which industry. Finally, the industry fitting method from different dimensions is used to match the industry which public transformers belong. For public transformers in Jiangsu province, the use of the method has been achieved after the high accuracy rate, to achieve a more accurate analytical method of the public transformers industry. The analytical method of the industry which public transformers belong based on load data mining provide a strong support for the evaluation of economic benefits of investment of the electricity distribution network.

Key words: load characteristic; load clustering; industry partition

(编辑 钱悦)

(上接第 91 页)

The Control Method of HVDC Commutation Failure Based on UPFC

LIU Jiankun¹, WANG Meng², ZHANG Ningyu¹, YANG Zhichao²

(1. State Grid Offshore Wind Power Grid Joint Laboratory(State Grid Jiangsu Electric Power Co.,Ltd. Research Institute), Nanjing 211103,China; 2. School of Electrical Engineering, Nanjing Institute of Technology,Nanjing 211167,China)

Abstract: Aiming at the problem of commutation failure in high voltage direct current (HVDC) transmission system, the unified power flow controller(UPFC) is applied to reduce the probability of commutation failure of HVDC system, and improve the original control strategy of the unified power flow controller. When the AC side faults, UPFC shunt control strategy using AC bus voltage reference value adjustment method to suppress commutation failure. The severity of fault is introduced into the fault level FL to characterize the HVDC system, through theoretical analysis and simulation verification of PSCAD/EMTDC; When the fault level of $F_L = 18.7\%$, which can control the original control strategy and improvement strategy of UPFC inhibition after commutation failure; fault level $F_L = 23.1\%$, only the improved control strategy can effectively inhibit the occurrence of commutation failure; failure is more serious, when the fault level reached $F_L = 93.56\%$, the improved control strategy can effectively reduce the probability of occurrence of continuous commutation failure. Reactive power compensation can be affected by AC system strength enhanced HVDC system using UPFC dynamic, improve the commutation failure resistance, thereby effectively improving the operation performance of HVDC system.

Key words: HVDC; commutation failure; UPFC; PSCAD/EMTDC; fault level

(编辑 钱悦)