

一种变电站负荷行业构成比例估算方法

施佳君¹, 秦川¹, 鞠平¹, 徐春雷², 赵静波³, 陈彦翔¹

(1. 河海大学 能源与电气学院, 江苏 南京 211100; 2. 国网江苏省电力有限公司, 江苏 南京 210029;
3. 国网江苏省电力有限公司电力科学研究院, 江苏 南京 211103)

摘要:为解决负荷模型的时变性问题,提高电网运行的安全性和经济性,提出基于能量管理系统底层负荷出线 and 专线用户的日负荷数据估算变电站负荷行业构成比例的方法。以实际电网数据为例,首先采用因子分析法对高维度的日采样数据进行降维;然后根据降维结果进行K-means聚类,并基于聚类结果进行负荷特性分析,从而获取典型行业日负荷曲线;最后根据所有底层出线和专线用户的行业归属情况及功率,自下而上聚合得到220 kV变电站的负荷行业构成比例。与用电信息采集系统用户数据的比对结果表明,所提方法估算得到的上层变电站负荷行业构成比例与实际总体相符。

关键词:因子分析;K-means聚类;负荷行业构成比例;时变性;负荷建模

中图分类号:TM 714

文献标志码:A

DOI:10.16081/j.epae.201909012

0 引言

负荷模型准确与否对电网有着重要影响,过于悲观的负荷模型会导致资金投入过多,浪费严重;过于乐观的负荷模型会给电网的安全稳定运行造成隐患^[1]。在电力系统仿真计算中,综合负荷模型SLM(Synthesis Load Model)应用最为广泛。其中,电动机比例是SLM的一个重要参数,该参数的获取主要有2种方法:一是总体测辨法^[2-3],通过实测动态响应辨识模型参数获取电动机比例,但该方法仅针对故障时刻有效,且扰动数据不易获取;二是统计综合法,通过离线调查变电站负荷行业构成比例及各行业动静负荷构成以获取总的电动机比例^[4],但实际负荷构成复杂,统计调查耗时费力,且难以反映负荷构成的时变性。

负荷构成的时变性是负荷模型具有时变性的主要原因^[5],而用户和变电站的日负荷曲线含有反映负荷构成变化的丰富信息。由于220 kV变电站下负荷构成复杂,直接分解其日负荷曲线存在困难,如果能在线分析海量底层用户的行业类型,从而自下而上估算220 kV变电站的负荷行业构成比例,则有助于在线建立最终的时变负荷模型。

目前,已有不少文献利用负控系统或信息采集系统AMI(Advanced Metering Infrastructure)的用户

日负荷曲线挖掘用户行业特征。由于受测量设备故障等因素的影响,用户的日负荷曲线并非一成不变,因此首先需要提取典型日负荷曲线:一种思路是求平均值获取典型特性,如文献[6]选取一周工作日,通过对日负荷曲线求平均值获取典型日负荷曲线,但该方法没有考虑异常日负荷曲线的影响;另一种是基于聚类的思想^[7-9],如文献[7]采用基于密度的聚类算法DBSCAN(Density-Based Spatial Clustering of Applications with Noise)剔除异常日负荷曲线,但该方法没有考虑线路运行方式变化的影响。然后基于提取的典型日负荷曲线挖掘用户行业特征进行后续分析。文献[10-11]基于负控系统的部分大用户日负荷曲线得到行业典型特征矩阵,并对离线统计调查获得的变电站负荷行业构成比例进行在线修正。文献[12-14]基于AMI的海量底层用户日负荷曲线,通过聚类提取用户用电模式,从而获取用户行业情况。但目前国家电网调度部门主要基于能量管理系统EMS(Energy Management System)展开日常工作,而EMS与AMI的网络拓扑命名规则不一致,导致两者数据自动匹配率较低,实际应用时需要人工校核,耗时耗力。

因此,本文提出直接利用EMS底层出线,即最小数据颗粒密度为10 kV出线的日负荷曲线,自下而上估算220 kV变电站的负荷行业构成比例。本文以某地调3座220 kV变电站为例,采集其下所有底层出线和专线用户的功率数据,其中,专线用户接入单个大用户,行业属性已从调度部门获知,因此只需对底层出线进行分析。首先将各底层出线全年的日负荷曲线按照季节进行划分,聚类求平均值,分别提取每条底层出线各季的典型日负荷曲线(下文中称为“纵向”聚类);其次,按季节对所有底层出线的典型日负荷曲线再次进行聚类分析(下文中称为“横

收稿日期:2018-11-22;修回日期:2019-07-11

基金项目:国家自然科学基金资助项目(51837004);111引智计划“新能源发电与智能电网学科创新引智基地”项目(B14022);国网江苏省电力有限公司科技项目(J2018042)

Project supported by the National Natural Science Foundation of China(51837004), 111 Project of “Discipline Innovation and Research based on Renewable Energy Generation and Smart Grid”(B14022) and the Science and Technology Program of State Grid Jiangsu Electric Power Co.,Ltd.(J2018042)

向”聚类),获得220 kV变电站下所有底层出线的聚类情况;然后,提取各个类型的典型日负荷曲线,根据其用电行为判断行业属性;最后,根据所有出线和专线的行业归属情况及功率,自下而上聚合,获取最终220 kV变电站的负荷行业构成比例。在聚类过程中,针对因输入数据维度较高导致聚类结果不理想的问题,首先采用因子分析法对高维日负荷数据进行降维,再基于降维结果采用K-means算法聚类。

1 基于因子分析降维的K-means聚类

目前EMS采样间隔为5 min,维数较高,对于电网中海量的日负荷数据,若直接聚类将导致算法效率低、耗费存储空间大,而且聚类有效性不够理想。因此,首先采用因子分析法对高维度数据进行降维,然后基于降维结果采用K-means算法进行聚类。K-means算法理论可靠,设计思想比较简单;虽然有迭代过程,但相比层次聚类等其他聚类算法,收敛速度较快,在处理大量数据时具有一定优势^[15]。

1.1 聚类有效性指标

聚类分析以类内的紧促度最高、类间的差异度最大为原则,将数据分为不同类别。由于没有先验信息来衡量聚类效果,因此,需要引入有效性指标对聚类结果进行评价^[14]。本文引入以下2个指标。

(1) MIA (Mean Index Adequacy) 指标。用于衡量类内的紧促度,表示类内距离的平均值,计算方法为:

$$d(U_k, C_k) = \sqrt{\frac{1}{n_k} \sum_{n=1}^{n_k} d^2(U_{k,n}, C_k)} \quad (1)$$

$$MIA = \sqrt{\frac{1}{K} \sum_{k=1}^K d^2(U_k, C_k)} \quad (2)$$

其中, K 为聚类数; U_k 为第 k 个类簇; $U_{k,n}$ 为第 k 个类簇中第 n 个样本; n_k 为样本数; C_k 为聚类中心, $C_k = \frac{1}{n_k} \sum_{n=1}^{n_k} U_{k,n}$; $d(U_{k,n}, C_k)$ 为样本成员与聚类中心间的欧氏距离。

MIA 值越小表示类内越紧促,聚类效果越好。

(2) DBI (Davies-Bouldin) 指标。又称为分类适确性指标,用于衡量类内紧促度和类间的差异度,计算公式为:

$$DBI = \frac{1}{K} \sum_{k=1}^K \max_{j \neq k} \frac{d(U_k) + d(U_j)}{d(C_k, C_j)} \quad (3)$$

其中, $d(C_k, C_j)$ 为聚类中心间的欧氏距离; $d(U_k)$ 和 $d(U_j)$ 为矩阵内部距离的均值。

DBI 越小表示类内紧促度越高,类间差异度越高,聚类结果越理想。

1.2 基于因子分析法降维的K-means聚类

因子分析通过研究多个变量间相关系数矩阵的

内部依赖关系,将具有错综复杂关系的变量综合为少数几个相互独立的公共因子,以探讨高维度的原始变量与公共因子之间的关系^[16]。即因子分析是对原始变量的重新组合,保留了变量原始信息,从而使得降维结果更合理。

公共因子是不可直接观测但又客观存在的共同影响因素,每一个原始变量都可以表示成公共因子的线性函数与特殊因子之和。设样本原始变量为 $X = (X_1, X_2, \dots, X_p)^T$, 公共因子为 $F = (F_1, F_2, \dots, F_m)^T$, 因子模型可表示为:

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + \varepsilon_2 \\ \vdots \\ X_p = a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + \varepsilon_p \end{cases} \quad (4)$$

其中, a_{ij} 为初始因子载荷,即公共因子与原始变量间的相关系数; $\varepsilon_1 \sim \varepsilon_p$ 为特殊因子,表示公共因子以外的影响因子,当提取的公共因子包含大部分原始变量信息时,特殊因子对原始变量的影响极小,可忽略不计。

根据公共因子的提取方法,因子分析可分为不同类别。根据文献^[17]的研究结果,本文采用基于主成分分析法的思路提取公共因子,因子分析的基本步骤如下。

步骤1 判断样本是否适用于因子分析。因子分析一般需要满足前提条件:样本数与原始变量个数之比至少大于5^[18]。当确定了输入样本后,提取KMO (Kaiser-Meyer-Olkin) 统计量。KMO 统计量是用于比较变量间简单相关系数和偏相关系数的指标,反映变量间的相关程度。

简单相关系数的计算公式为:

$$\bar{X} = \frac{1}{G} \sum_{g=1}^G X_{(g)} = (\bar{X}_1, \dots, \bar{X}_p)^T \quad (5)$$

$$S = \frac{1}{G-1} \sum_{g=1}^G (X_{(g)} - \bar{X})(X_{(g)} - \bar{X})^T = (s_{ij})_{p \times p} \quad (6)$$

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} \quad i, j = 1, 2, \dots, p \quad (7)$$

其中, $\bar{X}$ 为原始变量的均值; $X_{(g)}$ 为第 g 个样本; G 为样本总数; S 为协方差矩阵, s_{ij} 为协方差矩阵中的元素; r_{ij} 为变量 X_i 和 X_j 间的简单相关系数。

偏相关系数的计算公式为:

$$p_{ij} = \frac{-r_{ij}}{\sqrt{r_{ii}r_{jj}}} \quad i \neq j \quad (8)$$

其中, p_{ij} 为变量 X_i 和 X_j 间的偏相关系数。

因此, KMO 统计量可表示为:

$$KMO = \sum_{j=1}^p \sum_{i=1, i \neq j}^p \frac{r_{ij}^2}{p_{ij}^2 + r_{ij}^2} \quad (9)$$

KMO值接近于1时,变量间相关性越强,越适合作因子分析;KMO值趋向于0时,则相反。实际分析中,KMO值大于0.7时,即可认为相关性较高^[19]。

步骤2 确定初始公共因子数。根据原始变量间的相关系数矩阵 $R=(r_{ij})$,求出 R 的特征根以及对应的特征向量,分别记为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ 和 $u_1, u_2, \dots, u_p$ 。特征根的值用来衡量公共因子的影响力,通常以特征根大于1为原则确定公共因子。计算公共因子特征值之和与所有因子特征值之和的比值,得到特征值累计贡献率,实际分析中,当该值大于等于85%时,即可较好地反映原始变量^[20]。

步骤3 按照式(4)所示的因子模型计算初始因子载荷矩阵 $A=(a_{ij})$,计算公式为:

$$A = (A_1, A_2, \dots, A_p)^T = \left(\sqrt{\lambda_1} u_1, \sqrt{\lambda_2} u_2, \dots, \sqrt{\lambda_p} u_p \right)^T \quad (10)$$

步骤4 对初始因子载荷矩阵进行旋转,即对初始公共因子进行线性组合,以利于数据信息的解释。旋转方法采用方差最大正交旋转,对初始因子载荷矩阵右乘一个正交矩阵,使载荷矩阵中各元素向0和1两端分化,从而突出每个公共因子和载荷较大的变量间的联系,而其余公共因子的载荷趋向于0,即用少量公共因子表达原始变量大部分的信息。

步骤5 计算因子得分系数矩阵。由于公共因子不可直接观测,因此需要获取原始变量与公共因子之间的关系,估计出每个原始变量对各公共因子的得分。本文基于回归的思想建立回归方程估计因子得分,将公共因子表示成变量的线性表达式 $F=XB$,如式(11)所示。

$$F_j = b_{j1}X_1 + b_{j2}X_2 + \dots + b_{jp}X_p \quad j = 1, 2, \dots, m \quad (11)$$

该式又称为因子得分函数。根据最小二乘估计可得 $B=(X^T X)^{-1} X^T F$,又因为 $A \approx F^T X$,则:

$$F = XB = X(X^T X)^{-1} A^T \quad (12)$$

保存计算所得的因子得分系数矩阵 F ,作为后续K-means算法聚类分析的基础。

K-means算法的基本思想是将各个样本数据与聚类中心的距离作为分类标准,距离最近的归为一类。关于K-means算法已有较多文献进行了详述,在此不再列出具体分析步骤。

2 变电站负荷行业构成比例估算

本文利用基于因子分析法降维的K-means聚类,采用“先纵向,后横向”聚类的方法获得底层负荷出线行业归属情况及其实际功率;再结合专线用户情况,自下而上聚合得到220 kV变电站的负荷行业构成比例。具体流程如图1所示。

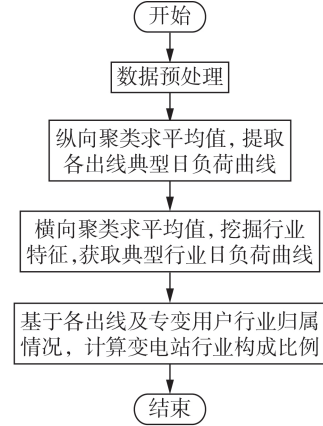


图1 变电站负荷行业构成比例估算流程

Fig.1 Estimation flowchart of industry composition for substation loads

主要包括以下几个步骤。

(1)数据预处理。在数据量测与采集的过程中,量测设备故障、环境干扰等因素可能会导致数据异常,因此首先需要进行异常数据的识别、剔除与修正^[21]。本文中数据清洗方法参考文献[10]。

然后,对清洗后的数据进行归一化,一方面消除量纲对聚类结果的影响,另一方面提高算法的收敛速度以缩短聚类过程的时间^[12]。为保证归一化后数据间的联系,同时为保证输出数据的区间为 $[0, 1]$,本文采用极差归一化方法,计算公式为:

$$z_i = \frac{z_i - \min(Z)}{\max(Z) - \min(Z)} \quad i = 1, 2, \dots, 288 \quad (13)$$

其中, z_i 为第 i 个采样点数据; Z 为日负荷曲线的所有采样点数据。

(2)纵向聚类求平均值,提取各出线典型日负荷曲线。首先选取该出线一段时间内的日负荷数据,采用基于因子分析降维的K-means算法进行纵向聚类。纵向聚类数目的确定采用文献[22]所述的GS(Gap Statistic)方法,利用间隙统计量评估出最佳聚类数目,以保证聚类数为1时仍具有良好的适应性。

根据聚类结果,剔除异常日负荷曲线;对其余各类日负荷曲线分别求平均值,如式(14)所示,从而获得各条出线的典型日负荷曲线。

$$v_i = \frac{1}{Y} \sum_{y=1}^Y v_{y,i} \quad i = 1, 2, \dots, 288 \quad (14)$$

其中, $v_{y,i}$ 为第 y 天第 i 个采样点的负荷值; Y 为出线典型类中的样本数; v_i 为出线典型日负荷曲线第 i 个采样点的负荷值。

(3)横向聚类求平均值,获取典型行业日负荷曲线。仍采用基于因子分析降维的K-means算法,以各出线的典型日负荷曲线为数据集,进行横向聚类。横向聚类数目以有效性指标最小为依据,同时应综合考虑有效性与简洁性。聚类数不宜过多,应控制

在7类以内。

根据聚类结果对同类出线的原始日负荷曲线,即降维前的日负荷曲线求平均值,然后分析其用电行为,获取其行业类型,从而得到典型行业日负荷曲线。

(4)变电站负荷行业构成比例计算。根据上述分析结果,可确定各底层负荷出线典型日负荷曲线的行业归属情况;再结合纵向聚类结果,可获取各底层负荷出线不同日期下的行业属性;最后结合出线相应日期的功率,自下而上逐层聚合,便可获得各层级变电站时变的负荷行业构成比例。其中,由于配电网损耗对最终负荷行业构成比例的计算影响较小,因此忽略不计。最终220 kV变电站的负荷行业构成比例计算方法如式(15)所示。

$$\rho_i = \frac{\sum_{j=1}^{N_i} P_{ij}}{\sum_{i=1}^D \sum_{j=1}^{N_i} P_{ij}} \quad i=1,2,\dots,D \quad (15)$$

其中, ρ_i 为第*i*类行业的占比; P_{ij} 为第*i*类行业的第*j*条出线或专线的负荷值; N_i 为第*i*类行业的出线及专线总数; D 为行业类型总数。

3 算例分析

本文从某地调选取3个典型220 kV站点,采集EMS中其下所有底层负荷出线和专线用户2017全年日负荷数据,其中底层负荷出线共216条,110 kV专线用户2个,均为重工业用户。采样间隔为5 min。3座220 kV变电站的出线及专线的分布情况如表1所示。

表1 220 kV变电站基本情况

Table 1 Basic situation of 220 kV substations

变电站	出线数/条	专线数/条	变电站性质
A	97	1	非工业为主
B	75	0	工业非工业混合
C	44	1	工业为主

由于典型行业日负荷曲线可能受季节、节假日等因素的影响而发生变化,应对不同情况分别讨论。但受篇幅所限,本文仅以用电行为相似的春秋季节工作日为例展开分析。

3.1 因子分析降维对纵向聚类的影响

取春秋季节工作日对各出线进行纵向聚类。对比降维前后各出线的平均运行时间,如表2所示。本文的实验环境为MATLAB R2015a。

从表中可以看出,基于因子分析降维的K-means算法聚类平均运行时间有所缩短。对于本算例中共216条出线而言,总运行时间可以大幅缩短。因此,面对海量负荷数据,采用先降维后聚类的方式

表2 降维前后平均运行时间

Table 2 Average running time before and after dimension reduction

聚类方法	平均运行时间/s
K-means	1.164
因子分析+K-means	0.859

能有效提高运行效率。

纵向聚类结果表明,剔除异常日负荷曲线后,大部分出线均对应同一种典型日负荷曲线;7条出线由于运行方式或用户发生变化,而对应2种典型日负荷曲线,即共获得223条典型日负荷曲线。

3.2 横向聚类数目的确定

首先采用K-means算法对纵向聚类获得的样本进行横向聚类,聚类数目与有效性指标之间的关系如图2中实线所示。

然后采用基于因子分析法降维的K-means算法,由于因子分析一般要求样本量至少是变量数的5倍,其中变量为288个日采样点功率,而由各出线典型日负荷曲线构成的样本数为223,因此对各条出线增加4月和10月的10个工作日数据,以满足此要求。由于输入样本均来自春秋季节的工作日,因此因子得分系数矩阵仍能较好地反映原始变量。输入样本数据,KMO统计量为0.913,数据间相关性较强,适用于因子分析。对原始日负荷数据降维,保留公共因子个数为15,特征值累计贡献率为92.21%,可较好地反映原始变量。取各出线典型日负荷曲线的降维结果进行K-means聚类,聚类数目与有效性指标之间的关系如图2中虚线所示。

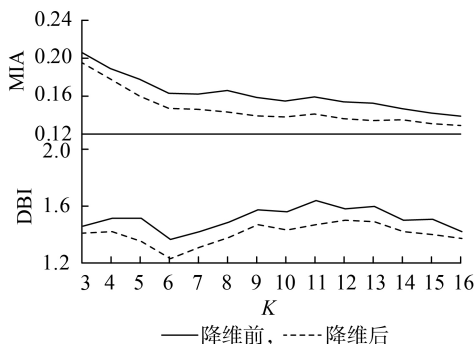


图2 聚类数与有效性指标的关系

Fig.2 Relationship between number of clusters and evaluation indexes

从图中可得,基于因子分析法降维的K-means聚类有效性指标值均更小,聚类效果更理想。

从降维后的有效性指标可以看出,MIA指标总体上逐渐减小,而DBI指标在6类时较小。综合考虑简洁性与有效性,最终选定聚类数目为6类。

3.3 典型行业日负荷曲线

基于上述分析,最终得到6类典型行业日负荷

曲线如图 3 所示,图中负荷均为标么值。

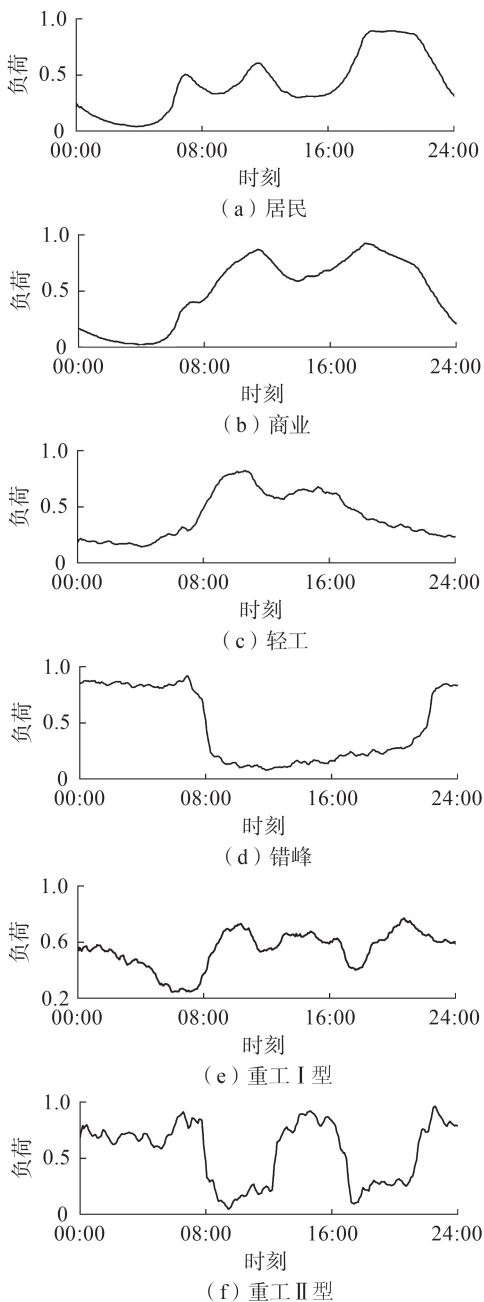


图 3 典型行业日负荷曲线

Fig.3 Typical daily load profiles of different industries

第一类曲线夜间负荷水平低,白天负荷水平高,功率波动较大;用电高峰出现在 07:00—08:00、11:00—12:00 之间及 20:00 左右,与居民的生活作息时间相似,为典型居民日负荷曲线。

第二类曲线昼夜用电差异明显;09:00—22:00 负荷水平相对较高,13:00—16:00 负荷出现小低谷,曲线整体比较平稳。这类曲线与商业用电模式相似,为典型商业日负荷曲线。

第三类曲线用电高峰为 09:00—18:00,夜间负荷水平较低;12:00 左右负荷出现小低谷是由于工人

休息;上午负荷水平高于下午是由于工厂机组开启瞬间有功需求较大。这类曲线与工业负荷中的劳动密集型产业用电行为相似,为典型轻工行业日负荷曲线。

第四类曲线呈现昼低夜高的情况,一般为工业负荷。由于电网实行峰谷电价,对用电价格较高且用电量较大的工业负荷而言,错开用电高峰可带来巨大的经济效益。这类曲线记为典型错峰工业日负荷曲线。

第五类曲线昼夜负荷水平相差较小;用电低谷为 08:00、12:00、18:00 左右,与工人换班时间基本一致;这类曲线与钢厂、化工厂等重工行业的用电行为相似,为典型重工 I 型日负荷曲线。

第六类曲线负荷水平呈规律性突变,在夜间、13:00—16:00 用电低谷时期负荷水平陡增,体现一定的错峰特性。这类出线较少,仅有 2 条,与用采数据人工比对可知,其下主要用户为钢压延、金属用品制造等重工企业,因此记为典型重工 II 型日负荷曲线。

各典型行业对应的出线数如表 3 所示。

表 3 春秋季 220 kV 变电站典型行业出线数

Table 3 Feeder number of typical industries for 220 kV substations in spring and autumn

变电站	出线数 / 条					
	居民	商业	轻工	错峰	重工 I 型	重工 II 型
A	61	28	8	1	1	—
B	24	25	16	5	8	—
C	5	—	14	5	20	2
合计	90	53	38	11	29	2

3.4 变电站负荷行业构成比例

由上述分析即可获得各出线在春秋季工作日的行业标签。以 2017 年 4 月 28 日为春季典型日,根据所有出线和专线的行业归属情况及其日平均负荷,可得到各 220 kV 变电站日平均负荷行业构成比例。其中,由于第五、六类同属于重工行业,且第六类出线只有 2 条,平均负荷仅占总负荷的 1.76%,对最终电动机比例的计算影响较小,因此合并为重工型负荷。最终结果如表 4 所示。

表 4 春季典型日 220 kV 变电站负荷行业构成比例

Table 4 Industry composition of 220 kV substation loads in spring

变电站	比例 / %				
	居民	商业	轻工	错峰	重工
A	31.16	36.21	4.09	2.36	26.18
B	19.65	30.81	21.11	5.03	23.40
C	1.62	0	22.71	8.89	66.78

同理,以 2017 年 7 月 28 日、10 月 19 日、12 月 18 日分别为夏、秋、冬季典型日,也可获得相应的平均负荷行业构成比例,如图 4 所示。

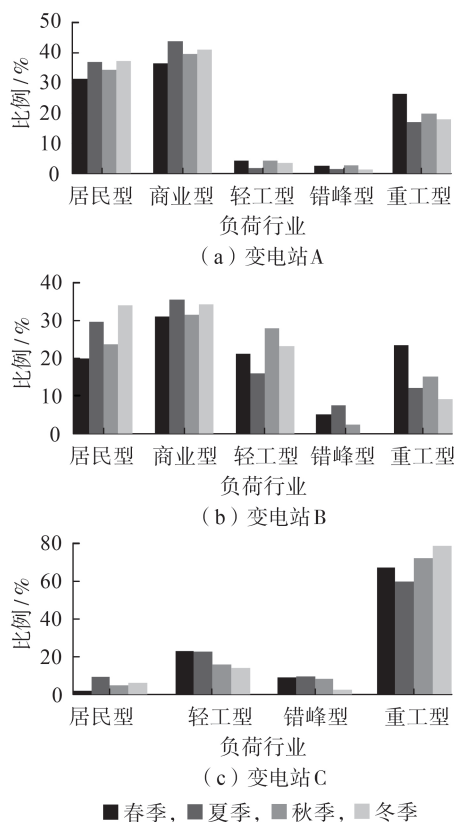


图4 不同季节典型日变电站负荷行业构成比例

Fig.4 Industry composition of 220 kV substation loads for typical days in different seasons

3.5 负荷行业构成比例对比校核

结合表1可以看出,变电站负荷行业构成比例整体情况与变电站性质基本一致。从图4可以看出,冬夏两季受空调负荷的影响,非工业负荷的占比总体增加,其中,夏季又高于冬季。但不同变电站负荷具体行业构成比例变化情况存在差异,与AMI的用户功率进行人工匹配及比对发现,是受工厂搬迁、新增线路、线路运行方式变化等多种因素的影响,不同季节负荷行业构成比例总体合理,但在线应用时,需根据底层负荷出线和专线用户的实际变化情况及其实时功率定期更新负荷行业构成比例。

比对的结果及分析如下。

(1)由于110 kV专线为重工业大用户,负荷水平较高,因此变电站负荷行业构成比例受其影响较大。其中变电站A下某110 kV专线用户在该年下半年搬迁,因此秋季重工行业比例较春季低,而夏季受空调负荷水平较高的影响,重工行业占比低于春季。

(2)由于不同类型的用户区域分布存在一定的规律性,因此,大部分底层负荷出线下虽然用户较多,但行业构成相对单一。以出线甲为例,聚类结果为商业型,其下日平均负荷行业构成比例如表5所示,可看出公变和商业专变占比较大。各行业典型日负荷曲线见图5,从用电行为可以看出,公变用户也以商业

用电为主,因此该出线下主要为商业用户。

表5 10 kV出线下日平均负荷行业构成比例
Table 5 Industry composition of average daily load under 10 kV feeders

出线	比例/%			
	公变	商业专变	轻工专变	重工专变
甲	65.74	32.86	1.40	—
乙	18.73	—	—	81.27

注:公变用户通常以商业或居民类非工业用户为主。

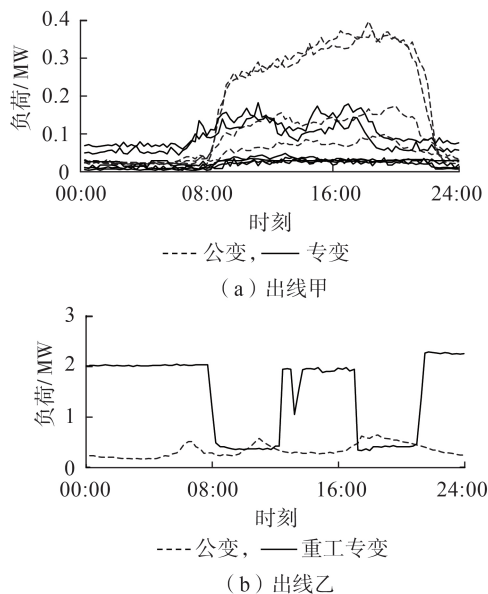


图5 10 kV出线下各行业典型日负荷曲线

Fig.5 Typical daily load profiles of different industries under 10 kV feeders

(3)部分10 kV出线下虽然用户构成较为复杂,但是以某一行业用户为主,以出线乙为例,聚类结果为重工Ⅱ型,从表5和图5可以看出其下重工用户所占比例较高,这类出线的行业归属情况取决于其所占比例较大的用户行业类型。

(4)由于受用户用电行为的随机性、线路运行方式变化等多方面因素的影响,底层负荷出线下主要用户及其用电行为会发生变化,导致其典型日负荷曲线发生变化。如变电站B,冬季时由于原错峰型出线退出运行和运行方式变化,因而错峰型负荷冬季占比为0。因此后期在线应用时,需要定期更新底层负荷出线行业归属情况。

4 结论

本文提出利用EMS底层出线和专线用户的日负荷数据,基于因子分析法降维和K-means聚类获取典型行业日负荷曲线,根据各条出线和专线用户的行业归属情况,自下而上聚合得到上层变电站时变的负荷行业构成比例。本文所提方法综合考虑了可行性与准确性,以某地调实测数据为例进行分析并进行校核,结果表明所提方法是合理且有效的。

但受篇幅所限,本文仅考虑季节这一典型因素对负荷行业构成比例的影响。后期在线应用时,应根据线路和专线用户的实际变化情况,定期更新底层出线的行业归属情况,同时结合各出线 and 专变的实时功率,得到各层级变电站时变的负荷行业构成比例。

参考文献:

- [1] 鞠平,马大强. 电力系统负荷建模[M]. 2版. 北京:中国电力出版社,2008:2-3.
- [2] GE Y, FLUECK A J, KIM D K, et al. An event-oriented method for online load modeling based on synchrophasor data[J]. IEEE Transactions on Smart Grid, 2015, 6(4): 2060-2068.
- [3] ZHANG Kaiqing, ZHU Hao, GUO Siming. Dependency analysis and improved parameter estimation for dynamic composite load modeling[J]. IEEE Transactions on Power Systems, 2017, 32(4): 3287-3297.
- [4] 褚壮壮,梁军,张旭,等. 基于仿射传播聚类算法的广义负荷稳态特性建模及其应用[J]. 电力自动化设备, 2016, 36(3): 115-123.
CHU Zhuangzhuang, LIANG Jun, ZHANG Xu, et al. Modeling of generalized load steady-state characteristics based on affinity propagation clustering algorithm and its application[J]. Electric Power Automation Equipment, 2016, 36(3): 115-123.
- [5] ARIF A, WANG Zhaoyu, WANG Jinahui, et al. Load modeling—a review[J]. IEEE Transactions on Smart Grid, 2018, 9(6): 5986-5999.
- [6] 林顺富,田二伟,符杨,等. 基于信息熵分段聚合近似和谱聚类的负荷分类方法[J]. 中国电机工程学报, 2017, 37(8): 2242-2253.
LIN Shunfu, TIAN Erwei, FU Yang, et al. Power load classification method based on information entropy piecewise aggregate approximation and spectral clustering[J]. Proceedings of the CSEE, 2017, 37(8): 2242-2253.
- [7] 苏适,李康平,严玉廷,等. 基于密度空间聚类和引力搜索算法的居民负荷用电模式分类模型[J]. 电力自动化设备, 2018, 38(1): 129-136.
SU Shi, LI Kangping, YAN Yuting, et al. Classification model of residential power consumption mode based on DBSCAN and gravitational search algorithm[J]. Electric Power Automation Equipment, 2018, 38(1): 129-136.
- [8] KEVIN M, FREDERICK D, CHRIS D. Two-stage load pattern clustering using fast wavelet transformation[J]. IEEE Transactions on Smart Grid, 2016, 7(5): 2250-2259.
- [9] 刘思,李林芝,吴浩,等. 基于特性指标降维的日负荷曲线聚类分析[J]. 电网技术, 2016, 40(3): 797-803.
LIU Si, LI Linzhi, WU Hao, et al. Cluster analysis of daily load curves using load pattern indexes to reduce dimensions[J]. Power System Technology, 2016, 40(3): 797-803.
- [10] 李欣然,姜学校,钱军,等. 基于用户日负荷曲线的用电行业分类与综合方法[J]. 电力系统自动化, 2010, 34(10): 56-61.
LI Xinran, JIANG Xuejiao, QIAN Jun, et al. A classifying and synthesizing method of power consumer industry based on the daily load profile[J]. Automation of Electric Power Systems, 2010, 34(10): 56-61.
- [11] 徐振华,李欣然,钱军,等. 变电站用电行业负荷构成比例的在线修正方法[J]. 电网技术, 2010, 34(7): 52-57.
XU Zhenhua, LI Xinran, QIAN Jun, et al. An online method to modify substation's structural proportion of synthetic load for power consuming-industries[J]. Power System Technology, 2010, 34(7): 52-57.
- [12] 赵岩,李磊,刘俊勇,等. 上海电网需求侧负荷模式的组合识别模型[J]. 电网技术, 2010, 34(1): 145-151.
ZHAO Yan, LI Lei, LIU Junyong, et al. Combinational recognition model for demand side load profile in Shanghai power grid[J]. Power System Technology, 2010, 34(1): 145-151.
- [13] 方超,仲春林,季聪. 基于负荷数据挖掘的公变用途分类方法研究[J]. 电力工程技术, 2018, 37(5): 115-120.
FANG Chao, ZHONG Chunlin, JI Cong. Analytical method of the industry which public transformers belong based on load data mining[J]. Electric Power Engineering Technology, 2018, 37(5): 115-120.
- [14] 赵文清,龚亚强. 基于Kernel K-means的负荷曲线聚类[J]. 电力自动化设备, 2016, 36(6): 203-207.
ZHAO Wenqing, GONG Yaqiang. Load curve clustering based on Kernel K-means[J]. Electric Power Automation Equipment, 2016, 36(6): 203-207.
- [15] 吴霜,季聪,孙国强. 基于CUDA技术海量电力负荷曲线聚类算法[J]. 电力工程技术, 2018, 37(4): 65-70.
WU Shuang, JI Cong, SUN Guoqiang. A clustering algorithm based on CUDA technology for massive electric power load curves[J]. Electric Power Engineering Technology, 2018, 37(4): 65-70.
- [16] 程序,关永刚,张文鹏,等. 基于因子分析和支持向量机算法的高压断路器机械故障诊断方法[J]. 电工技术学报, 2014, 29(7): 209-215.
CHENG Xu, GUAN Yonggang, ZHANG Wenpeng, et al. Diagnosis method on the mechanical failure of high voltage circuit breakers based on factor analysis and SVM[J]. Transactions of China Electrotechnical Society, 2014, 29(7): 209-215.
- [17] 张斌,庄池杰,胡军,等. 结合降维技术的电力负荷曲线集成聚类算法[J]. 中国电机工程学报, 2015, 35(15): 3741-3749.
ZHANG Bin, ZHUANG Chijie, HU Jun, et al. Ensemble clustering algorithm combined with dimension reduction techniques for power load profiles[J]. Proceedings of the CSEE, 2015, 35(15): 3741-3749.
- [18] 张文彤. SPSS统计分析高级教程[M]. 北京:高等教育出版社, 2004: 220.
- [19] 林海明. 因子分析应用中一些常见问题的解析[J]. 统计与决策, 2012(15): 65-69.
LIN Haiming. Analysis of some common problems in factor analysis[J]. Statistics & Decision, 2012(15): 65-69.
- [20] 李卫东. 应用多元统计分析[M]. 2版. 北京:北京大学出版社, 2015: 205-206.
- [21] 丁晓,孙虹,郑海雁,等. 基于配用电大数据的短期负荷预测[J]. 电力工程技术, 2018, 37(3): 21-27.
DING Xiao, SUN Hong, ZHENG Haiyan, et al. Distribution and consumption big data based short-term load forecasting[J]. Electric Power Engineering Technology, 2018, 37(3): 21-27.
- [22] TIBSHIRANI R, WALTHER G, HASTIE T. Estimating the number of clusters in a dataset via the Gap Statistic[R]. Palo Alto, USA: Stanford University, 2000.

作者简介:



施佳君

施佳君(1993—),女,江苏南通人,硕士研究生,主要研究方向为电力系统负荷建模(E-mail: jjs321q@sina.cn);

秦川(1981—),男,江苏靖江人,副教授,博士研究生导师,博士,主要研究方向为电力系统负荷建模、可再生能源发电系统建模与控制(E-mail: cqin@hhu.edu.cn)

鞠平(1962—),男,江苏靖江人,教授,博士研究生导师,博士,主要研究方向为电力系统建模与控制(E-mail: pju@hhu.edu.cn)。

(下转第223页 continued on page 223)

- [12] 宗鹏鹏. 含分布式光伏的主动配电网状态估计研究[D]. 南京:东南大学,2017.
ZONG Pengpeng. Research on state estimation of active distribution network with distributed photovoltaic system[D]. Nanjing:Southeast University,2017.
- [13] 候小虎. 含分布式电源的中低压配电网状态估计研究[D]. 秦皇岛:燕山大学,2017.
HOU Xiaohu. Research on a state estimation of medium and low voltage distribution network with distributed generation [D]. Qinhuangdao:Yanshan University,2017.
- [14] 孙一冰. 分布式动态状态估计算法及其在电力系统中的应用[D]. 济南:山东大学,2017.
SUN Yibing. Distributed dynamic state estimation algorithms and applications in power systems [D]. Jinan:Shandong University,2017.

- [15] ZUO Shan, DAVOUDI A, SONG Y, et al. Distributed finite-time voltage and frequency restoration in islanded AC microgrids [J]. IEEE Transactions on Industrial Electronics, 2016,63(10):5988-5997.

作者简介:



马 建

马 建(1967—),男,河南濮阳人,高级工程师,从事电能计量、电能质量方面的研究工作(**E-mail**:jian67@sina.com);

窦晓波(1979—),男,江苏南京人,教授,博士研究生导师,博士,主要研究方向为分布式电源高渗透配电网(**E-mail**:dxb_2001@sina.com)。

Design of network information observer based on Kalman filtering algorithm

MA Jian¹, DOU Xiaobo², CHEN Kexu¹, JIAO Yang², GE Pudong²

(1. State Grid Jiangxi Electric Power Research Institute, Nanchang 330096, China;

2. School of Electrical Engineering, Southeast University, Nanjing 210096, China)

Abstract: The traditional power system mostly adopts unified control and management, and has a strong dependence on measurement equipment and communication facilities. Once communication delay, packet loss and other failures occur, the system status monitoring and system control will be affected. Therefore, a non-communication network information observer based on Kalman filtering algorithm for distributed generation connected power system is proposed, which can realize the observation of system node voltage and distributed generation local information without communication information. The effectiveness of the proposed observer is verified by the results of simulation analysis.

Key words: electric power systems; communication; monitoring; Kalman filtering; information measurement; network information observer

(上接第 207 页 continued from page 207)

Estimation method of industry composition of substation loads

SHI Jiajun¹, QIN Chuan¹, JU Ping¹, XU Chunlei², ZHAO Jingbo³, CHEN Yanxiang¹

(1. College of Energy and Electrical Engineering, Hohai University, Nanjing 211100, China;

2. State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210029, China;

3. Electric Power Research Institute of State Grid Jiangsu Electric Power Co., Ltd., Nanjing 211103, China)

Abstract: In order to solve the time variation problem of load model and improve the operation security and economy of power grid, an estimation method of industry composition of substation loads is proposed based on the daily load data of bottom load feeders and special line users acquired by energy management system. The data of an actual power grid is taken as an example. Firstly, the factor analysis method is used to reduce the dimension of high-dimension daily sampling data. Then, K-means clustering is carried out according to the dimension reduction results, and the load characteristics are analyzed based on the clustering results to obtain the daily load curves of typical industries. Finally, according to the categories and powers of all the bottom load feeders and special line users, the industry composition of 220 kV substation loads is obtained by bottom-up aggregation. The comparison results with the user data from AMI (Advanced Metering Infrastructure) show that the industry composition of upper substation loads estimated by the proposed method is correspond to the actual situation.

Key words: factor analysis; K-means clustering; industry composition of loads; time variation; load modeling