

# Classification of Electric Power Load in Xinjiang Based on Daily Load Curve

Honghao Guan, Jiayu Bian, Jin Yu, Rui Luo

Institute of Economic and Technological Research, State Grid Xinjiang Electric Power Co. Ltd., Urumqi Xinjiang  
Email: 13730161728@163.com

Received: Nov. 27<sup>th</sup>, 2019; accepted: Dec. 12<sup>th</sup>, 2019; published: Dec. 19<sup>th</sup>, 2019

---

## Abstract

Power load classification is the basic data and key indicators of power system planning and operation control. The location characteristics and load characteristics of Xinjiang region obviously determine the classification of power load is particularly important. Therefore, it is of great engineering value to carry out accurate classification research of power load in Xinjiang. This paper proposes a power load classification method based on daily load curve and K-means clustering algorithm in Xinjiang. Firstly, the load characteristics of Xinjiang region are analyzed, and the load classification ideas and classification indicators of Xinjiang region using daily load curve data processing are proposed. Then the load classification in Xinjiang is realized based on K-means clustering. The results show the correctness and effectiveness of the proposed method.

## Keywords

Daily Load Curve, K-Means Clustering Algorithm, Xinjiang, Power Load Classification

---

# 基于日负荷曲线的新疆地区电力负荷分类

关洪浩, 边家瑜, 余金, 罗锐

国网新疆电力有限公司经济技术研究院, 新疆 乌鲁木齐  
Email: 13730161728@163.com

收稿日期: 2019年11月27日; 录用日期: 2019年12月12日; 发布日期: 2019年12月19日

---

## 摘要

电力负荷分类是负荷建模的基础依据, 负荷分类的准确性直接影响负荷模型的实用性, 进而影响电力系

统仿真结果的可靠性。新疆地区区域特点与负荷特性明显决定了电力负荷分类工作尤其重要,因此开展新疆地区电力负荷准确分类研究具有重要工程价值。论文提出了一种基于日负荷曲线和K-means聚类算法的新疆地区电力负荷分类方法。首先分析了新疆地区的负荷特点,提出采用日负荷曲线数据进行新疆地区负荷分类的方法,然后基于K-means聚类算法实现了新疆地区负荷分类。结果表明了文中方法的正确性与有效性。

## 关键词

日负荷曲线, K-Means聚类算法, 新疆, 电力负荷分类

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

将电力负荷正确分类,可以从中看出负荷结构情况,指导电力系统的运行控制,有利于电力系统的安全稳定运行,意义重大。与其它地区相比,新疆地区电力负荷有其特殊性,比如,高耗能产业占比大,冬季电采暖负荷非常普遍。此外,需要特别注意的是,新疆地区和其它地区有两个多小时的时差,很多研究人员很容易忽略这一点。因此,开展新疆地区的负荷分类更具重要性。

目前,在工程实践中,电力系统负荷分类大多采用统计学的方法,通过大量的调研得到结果,调研过程中需要大量的人力、物力,而且得到的结果只是一个时间或者一段时间的负荷情况,不具备时变性[1] [2] [3]。上述这种比较传统的方法,虽然思路清晰,但是随着电网结构的扩大,调研的工作量太大,很难实际操作。除了上述传统方法外,比较新颖的就是基于日负荷曲线的电力负荷分类,这种方法的原理是根据负荷特性的差异会体现在日负荷曲线上,所以先对日负荷曲线进行预处理,选取合适的聚类指标,之后进行聚类分析,完成对电力负荷的分类。关于电力负荷聚类的算法非常多,比如近邻传播算法、主成分分析法、K-means 算法、模糊 C 均值算法、自组织映射神经网络算法(SOM)等等[4] [5] [6]。

本文将新疆地区典型区域变电站的日负荷曲线进行了分离,借助分离的结果求出了相应的聚类指标,之后运用 K-means 聚类算法将新疆地区典型区域变电站进行了聚类,最后实现了新疆地区电力负荷的分类。

## 2. 新疆地区负荷曲线的分离

### 2.1. 新疆负荷特点

受地理位置、资源结构、人们生活方式的影响,新疆地区负荷特点和中东部地区有着很大差异。① 负荷构成,新疆地区高耗能产业所占比重较大,导致大部分日负荷曲线都比较平缓,负荷特性比较稳定。② 高耗能产业的发展与国家政策的指向密切相关,一旦发生经济危机,新疆地区月平均负荷率会大幅度下降。③ 新疆地区与中东部地区的气候差异很大,这种差异体现在日负荷曲线上,就是高峰的差异。和中东部地区相比,新疆地区冬季高峰高于夏季高峰,这是因为新疆地区冬季普遍采用电采暖。④ 虽然新疆地区负荷特性有其独特性,但是毕竟地处西北,可以并入西北主网。⑤ 受中央新疆工作座谈会的影响,

新疆地区负荷将快速发展，第三产业和居民用电将占取很大比重，高耗能产业所占比重会有所下降。此外，新疆地区和一般的地区存在两个小时左右的时差，在进行日负荷曲线分时段处理的时候一定要注意[7][8]。

## 2.2. 新疆负荷聚类指标的确定

目前，电力系统都装有用于数据采集与监视控制的 SCADA 系统，系统每隔 5 分钟或者 1 分钟采集一个数据点，形成日负荷曲线。经过观察对比不同站点的日负荷数据发现，有的站点的日负荷曲线是 5 分钟采集一个点，而有的站点则是 1 分钟采集一个点，为了研究方便，必须进行统一化处理。统一化处理后，日负荷曲线呈现为 5 分钟采集一个点的数据文件，每条日负荷曲线有 288 组数据，如果直接进行聚类分析，数据维数巨大，不仅聚类速度慢，而且操作困难。因此，有必要进行降维处理。降维处理就意味着要选取合适的聚类指标，要求选取的聚类指标能够可靠地反映不同类型的负荷。考虑到上述因素，选用 $[\rho_g, \rho_n, \rho_s, \rho_j, \gamma, \beta]$ 这六个指标，这六个指标的定义如表 1 所示：

Table 1. The connotation of clustering indicators

表 1. 聚类指标的定义

聚类指标	全称	定义
$\rho_g$	工业负荷比例	工业的日平均负荷/整体的日平均负荷的
$\rho_n$	农业负荷比例	农业的日平均负荷/整体的日平均负荷的
$\rho_s$	商业负荷比例	商业的日平均负荷/整体的日平均负荷的
$\rho_j$	市政及居民生活负荷比例	市政居民的日平均负荷/整体的日平均负荷的
$\gamma$	日负荷率	$\gamma = \frac{P_{d,av}}{P_{d,max}}$
$\beta$	日最小负荷率	$\beta = \frac{P_{d,min}}{P_{d,max}}$

## 2.3. 新疆地区日负荷曲线分离

日负荷曲线是各行业负荷的叠加，需要对日负荷曲线进行分离，得到 $[\rho_g, \rho_n, \rho_s, \rho_j]$ 这四个聚类指标。日负荷曲线的分离步骤：① 确定典型日负荷曲线的四个时间区间；② 求解各特征时刻的负荷功率平均值；③ 构造分行业负荷构成方程式；④ 求解各分行业负荷在各特征时刻的负荷比例；⑤ 日负荷曲线的分离；⑥ 求解各分行业的用电负荷比例。

根据不同类型负荷的用电特点，可以进行分时段处理。不论从统计资料还是运行经验，可以得出：工业负荷全天几乎没有波动，农业负荷在夜间比较突出，商业负荷主要集中在白天，市政居民负荷则和人们的生活习惯息息相关。考虑到新疆地区的时差，新疆地区日负荷曲线分时段如下：

低谷时段：2:00~9:00，低谷特征时刻：5:00~7:00；

早高峰时段：9:00~14:00，早高峰特征时刻：12:00~13:00；

平缓时段：14:00~19:00；

晚高峰时段：19:00~2:00，晚高峰特征时刻 22:00~0:00。

选取了新疆地区典型变电站对日负荷曲线进行了分离，日负荷曲线分离结果如表 2 所示。

**Table 2.** Daily load curve separation results  
**表 2.** 日负荷曲线分离结果

	阿拉尔	八户梁	巴彦岱变	白碱滩变	拜城变	彩虹变
工业在低谷时刻所占比例	1	1	1	1	1	1
工业早高	0.977008006	0.496885151	0.817770416	0.876200184	0.822586463	1.036241135
工业平缓	0.972178153	0.53583321	0.789127048	0.89819742	0.830042573	1.001036219
工业晚高	0.824917826	0.53283285	0.747649016	0.949458677	0.834512254	0.988696443
农业在低谷时刻所占比例	0	0	0	0	0	0
农业早高	0	0.072686906	0	0.024490425	0.008982804	0
农业平缓	0	0	0	0	0	0
农业晚高	0	0	0	0	0	0
商业在低谷时刻所占比例	0	0	0	0	0	0
商业早高	0.027960068	0.430427942	0.218527121	0.099309391	0.168430733	0
商业平缓	0.027821847	0.46416679	0.210872952	0.10180258	0.169957427	0
商业晚高	0.023607543	0.461567722	0.199789065	0.107612581	0.170872628	0
居民在低谷时刻所占比例	0	0	0	0	0	0
居民早高	0	0	0	0	0	0
居民平缓	0	0	0	0	0	0
居民晚高	0.151474631	0.005599428	0.052561919	0	0	0.012327002

### 3. 基于 K-Means 的新疆负荷聚类

#### 3.1. K-Means 聚类算法

K-means 聚类算法,也被称为 K-均值聚类算法,与其它聚类算法相比,该算法聚类效果良好。K-means 聚类算法的过程如图 1 所示。

#### 3.2. 基于 K-Means 聚类算法的新疆地区典型区域负荷分类的实现

从上述 K-means 聚类算法过程可以看出,聚类数和初始聚类中心对聚类结果的影响比较大,若是出现初始聚类中心的两个点本来就处于同一类,极有可能陷入局部最优解。

K-means 聚类算法聚类数的确定有手肘法和轮廓系数法。手肘法主要借助  $SSE$  指标,  $SSE$  计算的是所有样本点的聚类误差,通过  $SSE$  可以从某种程度上反应聚类效果。 $SSE$  的计算如式(1)所示。

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (1)$$

式中,  $C_i$  表示聚类结果的第  $i$  类,  $p$  是  $C_i$  中的样本点,  $m_i$  是  $C_i$  的中心点。

随着聚类数的增大,  $SSE$  的值会变小。当聚类数小于真实聚类数时,  $SSE$  值的下降幅度比较大;当聚类数大于真实聚类数时,  $SSE$  值的下降幅度明显变小。根据手肘法,下降幅度转折点就是真实聚类数。 $SSE$  值计算结果如图 2 所示。

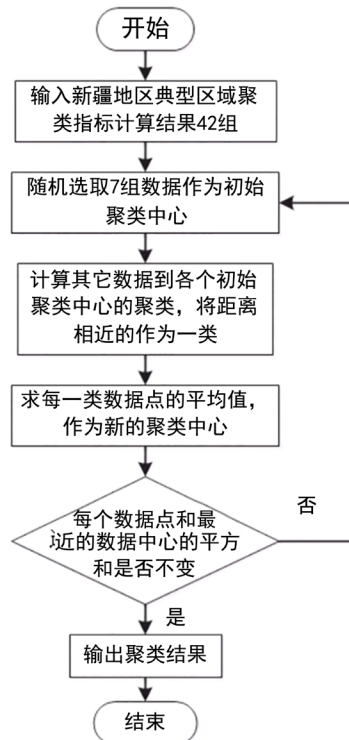


Figure 1. K-means clustering algorithm flow chart  
图 1. K-means 聚类算法流程图

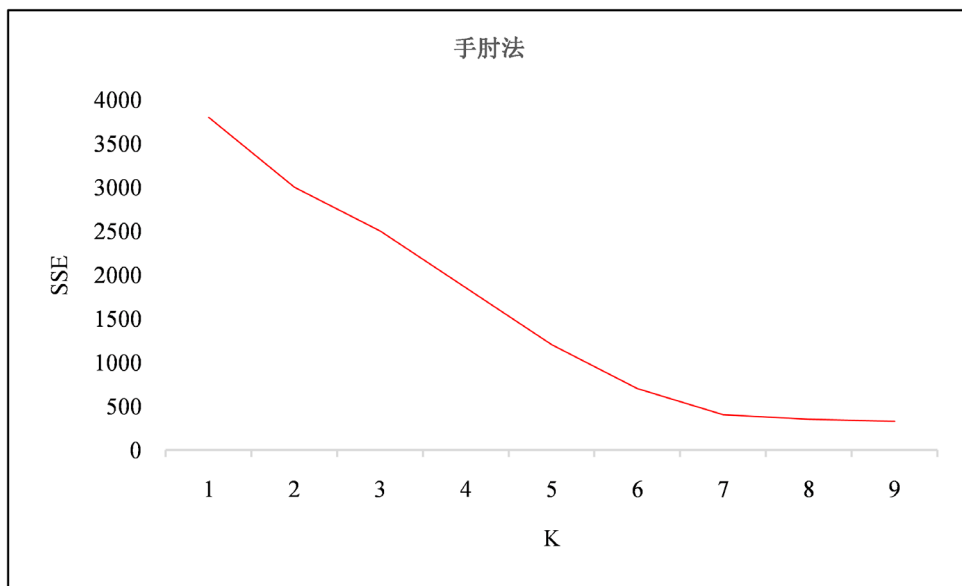


Figure 2. SSE calculation result  
图 2. SSE 值计算结果

根据手肘法，可以确定聚类数为7。结合新疆电网负荷构成特点，确定初始聚类中心。

表 3 为部分站点夏季最大运行方式下的聚类指标结果。变电站聚类结果如表 4 所示，行业聚类结果如图 3 所示。考虑到气候差异，同一变电站在冬季和夏季的供电类型可能不同，将冬季最大运行方式下的日负荷曲线进行聚类，变电站聚类结果如表 5 所示，行业聚类结果如图 4 所示。

**Table 3.** Partial site clustering indicator (Xia Da)  
**表 3.** 部分站点聚类指标(夏大)

变电站	$\rho_s$	$\rho_n$	$\rho_s$	$\rho_j$	$\gamma$	$\beta$
T1	0.933472542	0	0.018774214	0.048758633	0.870026692	0.775682694
T2	0.617535287	0.017644639	0.36304624	0.001773833	0.78753517	0.475842759
T3	0.835457107	0	0.155568415	0.016576699	0.861062909	0.677417167
T4	0.936509937	0.005296025	0.074847576	0	0.920576908	0.856519444
T5	0.874000187	0.001952919	0.125673056	0	0.890487979	0.756192498
T6	1.004193656	0	0	0.003670176	0.942834051	0.898659978
T7	0.808657896	0.002175037	0.205726788	0	0.876273226	0.70657552
T8	0.924761497	0.003626641	0.063381477	0.008230385	0.923288899	0.803837955

**Table 4.** Clustering results (summer)  
**表 4.** 聚类结果(夏季)

类别	变电站
第 1 类	拜城变、丁香变、纺织工业园变、龟兹变、海楼变、夹河子变、祥云变、向阳变、长宁变
第 2 类	八户梁、老满城、三宫变
第 3 类	大墩子变、奇台变、于田变
第 4 类	巩留变、宁远变
第 5 类	沙湾变、玉龙变
第 6 类	阿拉尔、白碱滩变、彩虹变、崇光变、额敏变、阜东变、克信变、乐土驿、棉城变、棉城变、台远变、兴盛变、瑶池变
第 7 类	巴彦岱变、昌吉变、锦华变、克拉玛依变、奎屯变、米泉变、深喀变

**Table 5.** Clustering results (winter)  
**表 5.** 聚类结果(冬季)

类别	变电站
第 1 类	阿拉尔、拜城变、彩虹变、崇光变、丁香变、额敏变、纺织工业园变、海楼变、夹河子变、锦华变、奎屯变、棉城变、沙湾变、石河子西、台远变
第 2 类	八户梁、老满城、宁远变、三宫变
第 3 类	巴彦岱变、巩留变、龟兹变、梨城变、米泉变、宁州户、祥云变、向阳变、长宁变
第 4 类	白碱滩变、阜东变、克信变、乐土驿、青年变、兴盛变
第 5 类	瑶池变
第 6 类	昌吉变、克拉玛依变、奇台变、深喀变、玉龙变、于田变
第 7 类	大墩子变

图 3 是夏季最大运行方式下日负荷曲线的聚类情况,从图中可以看出:第一类负荷低谷集中在凌晨,白天负荷相对较高,晚间达到最高,但变化幅度偏小,与人们的生活作息有一定关系;第二类负荷低谷也出现在凌晨,晚间负荷相对较高,峰值出现在白天,与商业用电特点类似;第三类负荷低谷出现在白天,凌晨负荷相对较高,晚间达到最高,说明农业用电的比例较高;第四类与第一类趋势类似,但明显

变化幅度比第一类大很多,说明与人们的生活作息关系密切;与第四类负荷相对比,第五类负荷在晚间向凌晨过渡期间,有一个明显的下降趋势;第六类负荷整体比较平稳,可见工业用电占比非常大,可将其归为工业负荷,此外这一类的变电站数目也是最多的,与新疆地区高耗能产业较高的事实也是吻合的,说明了聚类结果的正确性;与其它六类负荷相比,第七类负荷波动比较频繁,呈现三峰三谷的特点,说明这类变电站供电类型多样,可称之为混合负荷供电。

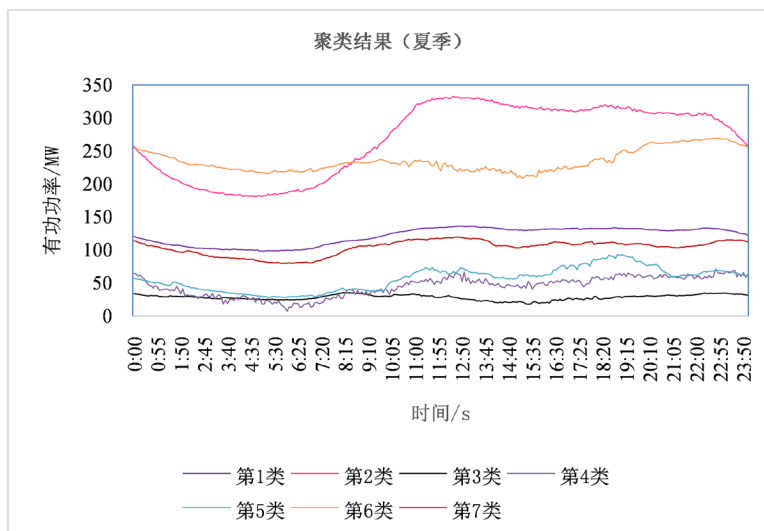


Figure 3. Clustering results (summer)

图 3. 聚类结果(夏季)

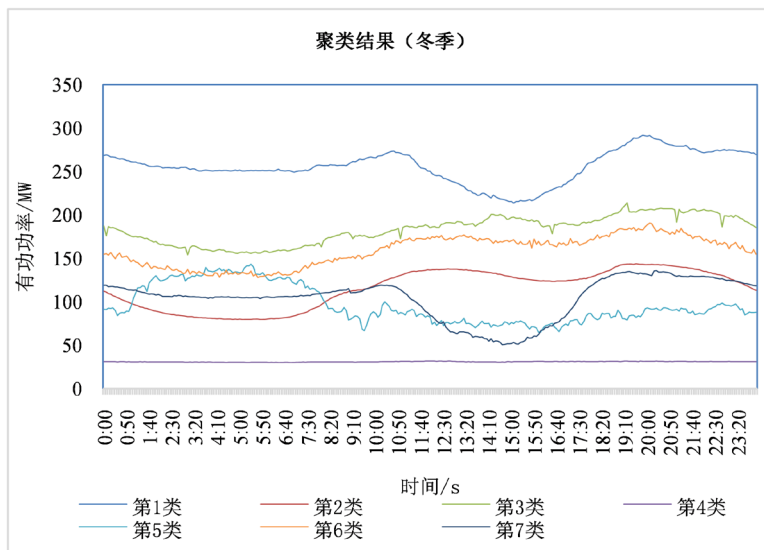


Figure 4. Clustering results (winter)

图 4. 聚类结果(冬季)

对比表 4 和表 5,图 3 和图 4,可以看出,某些变电站的供电结构会随气候的变化而发生变化。需要注意的是,冬季第 5 类变电站和第 7 类变电站的日负荷曲线的峰值出现在夜间,而且夜间普遍比白天高,在夏季中找不到类似的,结合实际,这主要是由于在这些地区冬季电采暖比较普遍。新疆地区的这种差异,也印证了在进行新疆地区电力负荷分类时,考虑气候差异的必要性。

### 3.3. 聚类效果评价

MIA、CDI、SMI、DBI、SI 这 5 个指标在聚类效果评价中较为常用，本文取 MIA 和 CDI 这两个指标进行分析。Mean Index Adequacy (MIA) 表示的是分配给聚类的每个输入向量与其中心之间的平均距离，MIA 指标越小，聚类结果内部越紧凑，其值与聚类效果成负相关。Clustering Dispersion Indicator (CDI)，表示的是同一集群中输入向量之间的平均基础距离与类代表负荷曲线之间的基础距离的比率，其值与聚类效果也成负相关。MIA 和 CDI 的计算结果如表 6 所示。

**Table 6.** MIA and CDI calculation results

**表 6.** MIA 和 CDI 的计算结果

聚类算法	K-means 算法	FCM 算法	Ward 算法	SOM 算法
平均 MIA 指标	0.1456	0.1521	0.1488	0.1719
平均 CDI 指标	0.0805	0.0955	0.0804	0.1048

与 FCM、Ward、SOM 算法相比，K-means 算法的平均 MIA 和 CDI 的值相对更小，聚类效果更好。此外，本文聚类分析的结果已在新疆电网典型区域负荷建模及参数辨识中得到应用。对新疆电网负荷模型进行了细化，结果表明，本文提出的聚类方法是有效的。

## 4. 结论

本文提出了一种基于日负荷曲线和 K-means 聚类算法的新疆地区电力负荷分类方法，成功将新疆地区的气候差异、时差差异、负荷特点考虑在内，以  $[\rho_g, \rho_n, \rho_s, \rho_j, \gamma, \beta]$  这六个指标为切入点，对典型区域变电站的日负荷曲线进行分离，得到聚类指标结果，按照 K-means 聚类算法的流程进行聚类。实际应用表明，该方法是有效的。

## 参考文献

- [1] 贺仁睦, 周文. 电力系统负荷模型的综合与分类[J]. 电力系统自动化, 1999, 23(19): 12-16.
- [2] 张红斌. 电力系统负荷模型结构与参数辨识的研究[D]: [博士学位论文]. 北京: 华北电力大学, 2003.
- [3] 鞠平, 陈谦, 熊传平, 等. 基于日负荷曲线的负荷分类和综合建模[J]. 电力系统自动化, 2006, 30(16): 6-9.
- [4] 陈焯, 吴浩, 史俊祎, 等. 奇异值分解方法在日负荷曲线降维聚类分析中的应用[J]. 电力系统自动化, 2018, 42(3): 105-111.
- [5] Jain, A.K. (2010) Data Clustering: 50 Years beyond K-Means. *Pattern Recognition Letters*, **31**, 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- [6] Rahman, M.A. and Islam, M.Z. (2014) A Hybrid Clustering Technique Combining a Novel Genetic Algorithm with K-Means. *Knowledge-Based Systems*, **71**, 345-365. <https://doi.org/10.1016/j.knsys.2014.08.011>
- [7] 徐邦恩, 蔺红. 基于改进模糊聚类的典型日负荷曲线选取方法[J]. 电测与仪表, 2019, 56(4): 21-26.
- [8] Zhou, L., Peng, Z.W., Deng, C.R., Qi, X.Z. and Li, P.Q. (2018) A Generalized Synthesis Load Model Considering Network Parameters and Allvanadium Redox Flow Battery. *Protection and Control of Modern Power Systems*, **3**, 315-327. <https://doi.org/10.1186/s41601-018-0105-1>