基于概率改进K均值聚类算法的地区电网调度分析

任 冲[1]，卫 琳[1]，樊国伟[2]，王胜利[2]，徐 志[2]，程 林[1]

（1. 国家电网公司西北分部，陕西 西安 710048；2. 新疆电力公司，新疆 乌鲁木齐 830002）

摘要：为了解决电网风险验证方法不足的缺点，提出了一种基于概率改进K均值聚类算法，对地区电网风险进行分析。首先利用K均值聚类的算法确定数据的最佳分类数和类内数据，根据峰谷差的大小，在弃光校验时，选择光伏、水电峰谷差大和负荷峰谷差小的类数据，通过拟合的负荷数据校验产生弃光的负荷概率；在限负荷校验时，选择峰谷差大的负荷和峰谷差小的光伏、水电数据，通过拟合的水电数据进行限负荷校验。最后对比拟合数据和实际最相似数据的调度偏差以及未概率改进的地区电网风险，结果表明所提出的方法可以发现地区电网风险。

关键词：K均值聚类；概率；数据拟合

Regional power network dispatching analysis based on probability improved K-means clustering algorithm

REN Chong[1], WEI Lin[1], FAN Guowei[2], WANG Shengli[2], XU Zhi[2], CHENG Lin[1]

(1. Northwest Branch of State Grid Corporation of China, Xi'an 71048, China; 2. State Grid Xinjiang Electric Power Company, Urumqi 830002, China)

Abstract: In order to solve the shortcomings of insufficient power grid risk, a K-means clustering algorithm based on probability improvement is proposed to analyze the regional power grid risk. Firstly, the K-means clustering algorithm is used to determine the best classification number and intra-class data of the data. The size of the peak-to valley difference, when the light-abandonment check, the data of the class of photovoltaic and hydropower peaks and valleys is selected, and then the data of the fitted load is verified to generate the probability of the light-abandonment load; When the load-limit check the load data of the peak-to-valley difference is selected, and then the fitted hydropower data is verfied, The scheduling deviation of the fitted data with the actual most similar data, and the regional power grid risk without the probability improvement are compared . The proposed method can identify regional grid risks .

Key words: K-means clustering; probability; data fitting

0 引言

“扩展克强指数”中的“全社会用电量、货运总量和工商业贷款额”可以更加准确地预测经济发展[1]，其中电能作为目前社会的主要能量来源，在经济和社会发展中有着更为重要的作用[2]。由于电力建设周期和经济发展周期不同步，导致某些地区的调峰能力不足而产生限负荷和弃光的现象[3-5]，考

虑到光伏的反调峰特性增加，光伏建设不能缓解限负荷的情况且会增加弃光[6-7]，火电机组建设周期较长且缺煤的现实[8]，网架补强工程不能满足当前调峰的需要，为了提高电网的安全裕度和促进高新能源消纳，新疆某地区计划通过建设储能的方法来提高电网的调峰能力，进而解决该地区限负荷和弃光的问题。

目前的电力规划主要是通过基于随机生产模拟[9-11]和基于大数据的聚类[12-14]的方法。随机生产模拟的方法是将一段时间的数据转化为持续性负荷，按照概率方法进行计算，由于概率只能反映出地区电网静态的风险概率，并不能对实际生产进行动态的电网调度运行；而基于大数据聚类的方法是根据一定的聚类算法，按照所占据的比例，在比例最多的数据中按照需要选取数据后进行校验，通过剔除畸数据得到误差更小的典型日数据，基于分层数据的典型日选取方法克服了聚类数据选取的盲目性。

上述分析方法只是在数据的分类上提出了更加准确的聚类方法，在进一步的分析中，由于占比例最高的数据和调峰情况最严重的数据并不一定在同一类，导致不能包括最严重的调度情况；本文在K均值聚类的基础上，按照峰谷差选出最严重的调峰情况，然后根据不同调峰情况的需要，选取最严重的情况或者占比最多的情况，再按照数据拟合的方法，校验电网风险和分析实际的调度过程。

1 数据的选取

1.1 K均值聚类算法

采集到历史的负荷、水电和光伏第i日第t时刻的功率用x_{it}表示，任意两天功率的距离d_{ij}表达式为

$$d^2_{ij}=\sqrt{|x_{i1}-x_{j1}|^2+|x_{i2}-x_{j2}|^2+\cdots+|x_{it}-x_{jt}|^2} \quad (1)$$

K均值聚类的方法是每类数据到数据中心点距离二次方和的大小，按照拐点处的中心点数k作为最佳的数据分类。

首先将数据随机分为k个数据中心，计算出数据中心和距离二次方和，按照新的数据中心得到新的距离二次方和，直到距离二次方和大小不再改变为止；距离二次方和J按照式（2）计算为

$$J=\sum_{k=1}^{k}\sum_{n=1}^{n}d(x_n,z_k) \quad (2)$$

式中，z_k为第k个数据中心；x_n为第n日数据点。

根据K均值聚类的算法，将数据分类后，按照分类中最多的类中的数据进行检验。在弃光检验中，选择光伏峰谷最大的光伏发电曲线做调峰检验，水电和负荷的选择按照概率密度最大的曲线考虑；在进行限负荷检验的过程中，负荷选择峰谷差最大曲线做调峰检验，光伏和水电按概率密度最大的方法考虑。

1.2 概率改进的K均值聚类算法模型

由于K均值聚类算法在校验时不能校验风险的大小，因此本文利用概率改进的算法进行校验。

在类中选择数据时，数据最多的类和峰谷差最大类不一定相同，因此在K均值聚类后，需要判断该类内是否包含峰谷差最大的数据。然后根据弃光确定光伏水电的最大峰谷差，负荷选择数据多的校验；而在限负荷时，负荷选择峰谷差最大的数据，光伏选择概率密度最多的数据，水电选择数据多的校验。

$$\sigma_{k,i}=P^{\max}_{k,i}-P^{\min}_{k,i} \quad (3)$$

式中，$\sigma_{k,i}$为第k类数据内第i天的峰谷差；$P^{\max}_{k,i}$为第k类数据内第i天最大的功率；$P^{\min}_{k,i}$为第k类数据内第i天最小的功率。

$$\sigma^{\max}=\max|\sigma_{k,i}| \quad (4)$$

式中，$\sigma^{\max}$为峰谷差最大的数据日。

在弃光校验中，首先根据K均值聚类算法

确定数据的类，光伏和水电选择峰谷差最大的类数据，负荷为占比例最高的类。其中负荷的拟合，根据条件约束确定出储能正好满足临界的数据日，然后利用式（1）计算出和拟合数据的距离，选择出和该距离最小的数据日，根据该类负荷平均值的概率分布确定弃光的概率。

$$P_{it}=(P_{it}^{\max}-P_{it}^{\min})\lambda+P_{it}^{\min} \tag{5}$$

式中，$P_{it}^{\max}$为第i天第t时刻最大功率；$P_{it}^{\min}$为第i天第t时刻最小功率；λ为最大功率和最小功率之间的系数。

在限负荷校验中，首先根据K均值聚类算法确定数据的类，负荷选择峰谷差最大的类数据，光伏和水电为占比例最高的类。其中光伏为概率密度最高的数据，水电的拟合根据条件约束，确定出储能正好满足临界的数据日，然后利用式（1）计算出和拟合数据的距离，选择出和该距离最小的数据日，根据该类水电的平均值的概率分布，确定限负荷的概率。

2 地区电力模型

（1）系统功率平衡约束

$$P_{\mathrm{f}}+P_{\mathrm{ph}}+P_{\mathrm{hdr}}+P_{\mathrm{TR}}=P_{\mathrm{s}}+(P_{\mathrm{re,c}}-P_{\mathrm{re,f}}) \tag{6}$$

式中，P_{f}为火电机组功率；P_{ph}为光伏功率；P_{hdr}为水电功率；P_{TR}为联络线功率；P_{s}为系统功率；$P_{\mathrm{re,c}}$为储能存储功率；$P_{\mathrm{re,f}}$为储能放电功率。

（2）火电机组约束

1）出力约束

$$P_{\mathrm{f}}^{\min}\leqslant P_{\mathrm{f}}\leqslant P_{\mathrm{f}}^{\max} \tag{7}$$

式中，$P_{\mathrm{f}}^{\min}$为火电最小功率；$P_{\mathrm{f}}^{\mathrm{man}}$为火电最大功率。

2）爬坡约束

$$\frac{\mathrm{d}T}{\mathrm{d}t}P_{\mathrm{f}}\leqslant R_{\mathrm{up}} \tag{8}$$

$$\frac{\mathrm{d}T}{\mathrm{d}t}P_{\mathrm{f}}\geqslant R_{\mathrm{down}} \tag{9}$$

式中，R_{up}为火电上行爬坡速率约束；R_{down}为火电下行爬坡速率约束。

（3）联络线约束

$$P_{\mathrm{TR}}\leqslant P_{\mathrm{TR,S}}^{\max} \tag{10}$$

$$P_{\mathrm{TR}}\leqslant P_{\mathrm{TR,X}}^{\max} \tag{11}$$

式中，$P_{\mathrm{TR,S}}^{\max}$为联络线最大外送功率；$P_{\mathrm{TR,X}}^{\max}$为联络线最大下网功率。

（4）储能调用约束

可调用容量约束

$$Q_{\mathrm{re}}\leqslant Q_{\mathrm{re}}^{\max} \tag{12}$$

式中，$Q_{\mathrm{re}}^{\max}$为储能最大储存容量。

（5）水电约束

本地主要是径流来水，水电调节能力已经达到极限，直接采用实际的发电曲线。

3 仿真验证

3.1 地区情况分析

新疆某地区火电机组装机容量为250 MW，但是由于该地区煤资源短缺，实际上单机最大的功率为80 MW/h，光伏总装机容量为594 MW，水电装机容量为362 MW，该地区建设的储能容量为100 MW。

该地区2018年的光伏曲线如图1所示，采样间隔为1 h。

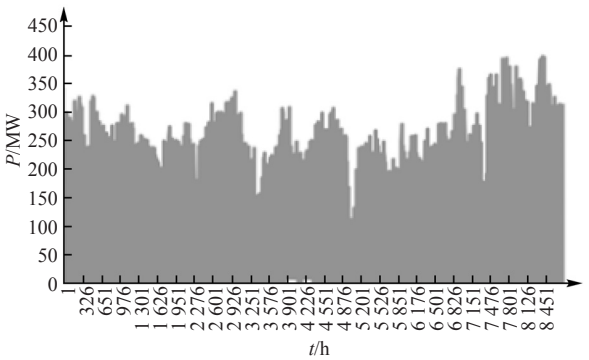


图1 光伏曲线

Fig. 1 Curve of PH

基于K均值聚类的分类数k和距离二次方和的关系如图2所示。

由图2可知，负荷和水电在k=2、光伏在k=3时，距离二次方和在拐点处，继续增大聚类数k，对距离二次方和减少已经很小，因此负荷、水电分为两类，光伏分为三类。其中

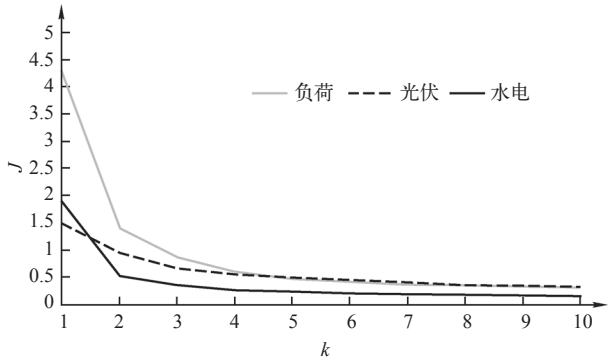


图2 距离二次方和J与聚类个数k的关系

Fig. 2 Relationship between distance of squared sum J and cluster number k

的数据中每个类别所占的比例如下表所示。

表 每类数据所占比例

Tab. Proportion of each type of data

(%)

	类别Ⅰ	类别Ⅱ	类别Ⅲ
负荷	47.2	52.8	—
光伏	21.6	29.6	48.8
水电	29.6	70.4	—

图3为光伏所占比例最高的类别Ⅲ的出力特征，蓝色为峰谷差最大的数据日，红色为概率密度最高的数据日。

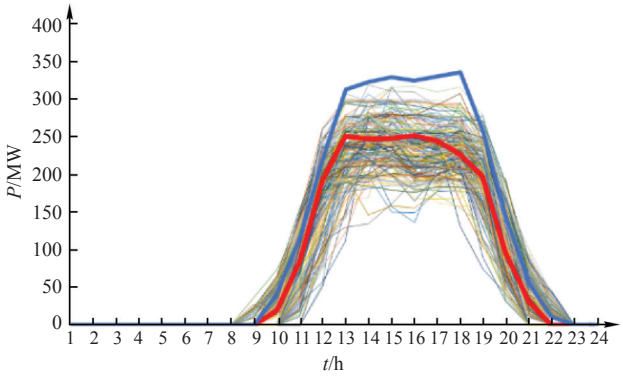


图3 类Ⅲ光伏出力样本特征

Fig. 3 Characteristics of PH samples output in type Ⅲ

负荷和水电的选择方法相同。

3.2 调度方法研究

（1）原调度方法

弃光校验时，水电选择类别Ⅱ，光伏选择类别Ⅲ，然后选择峰谷差最大的曲线，此时火电机组选择停机，其满足约束的调度曲线如图4所示。由图4可知，按照该方法，没有弃光。

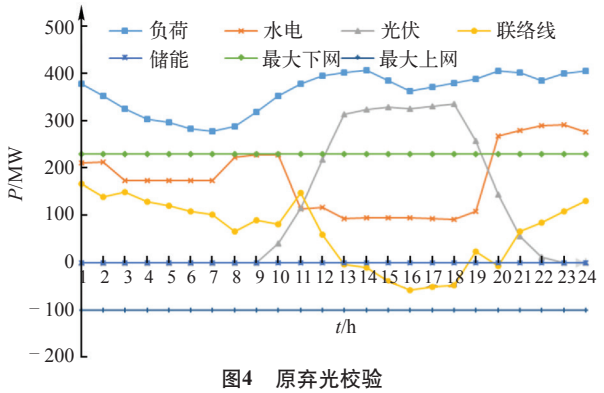
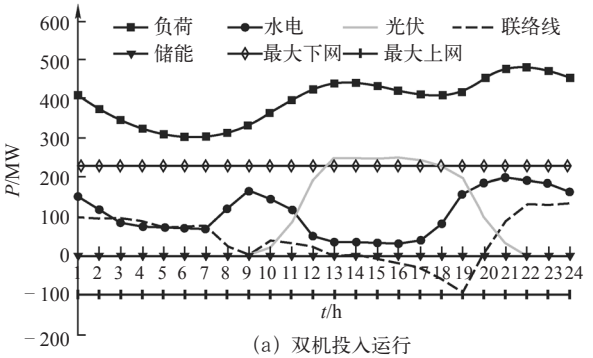


图4 原弃光校验

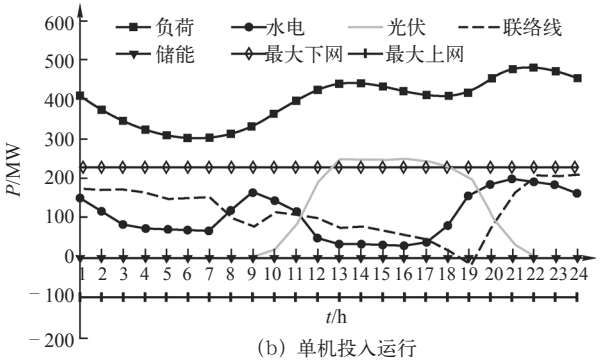
Fig. 4 Original light-abandonment check

在限负荷校验时，负荷为类别Ⅱ中的数据，水电选择类别Ⅱ，光伏选择类别Ⅲ，然后选择峰谷差最小的数据；分为火电双机、单机和无火电机组投入运行，在火电机组投入时，其功率选择为最大功率。双机投入运行，其调度曲线如图5a所示；单机投入运行，其调度曲线如图5b所示；无火电机组投入运行，其满足约束的调度曲线如图5c所示。

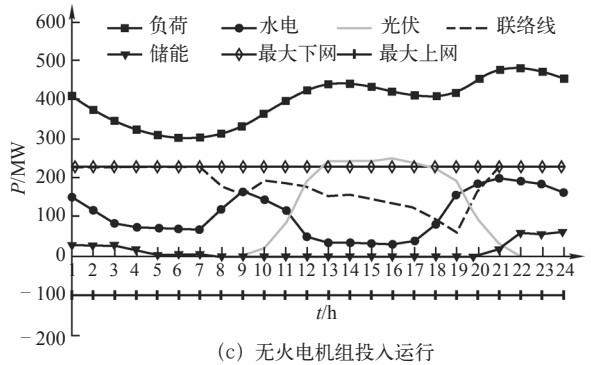
由图5a可知，该地区双机投入运行时，没有限负荷风险。由图5b可知，该地区单机投入运行时，没有限负荷风险。由图5c可知，该地区电网有限负荷风险，其中限负荷的天数概率为按照基于概率改进的K均值聚类。



(a) 双机投入运行



(b) 单机投入运行



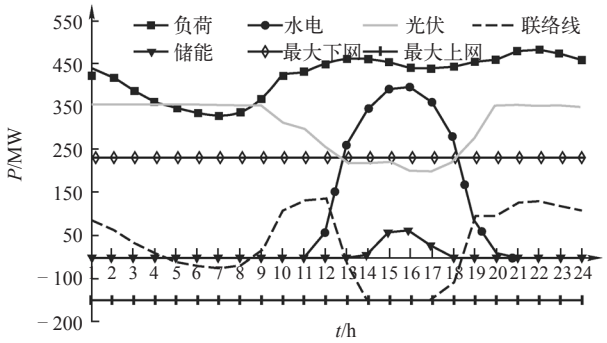
（c）无火电机组投入运行

图5　原限负荷校验

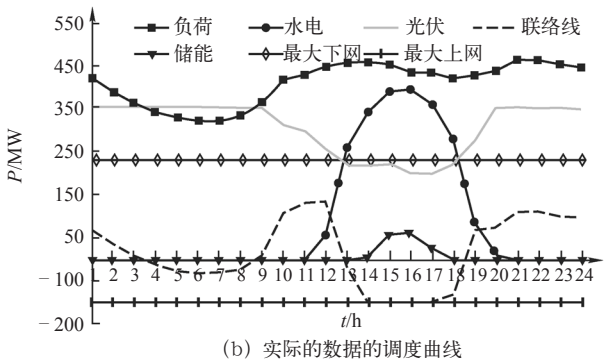
Fig. 5 Original load limit check

（2）改进的调度方法

在弃光校验时，如果按照K均值聚类的算法，光伏选取类别Ⅲ，水电选择为峰谷差最大的类别Ⅰ，而此时是没有弃光风险的，光伏和水电选择该类峰谷差最大的曲线，拟合数据负荷满足条件的调度曲线如图6a所示,实际的数据调度曲线如图6b所示。



（a）拟合数据负荷满足条件的调度曲线



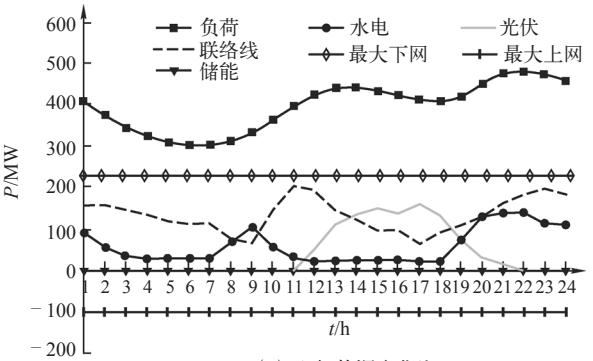
（b）实际的数据的调度曲线

图6　改进的弃光校验

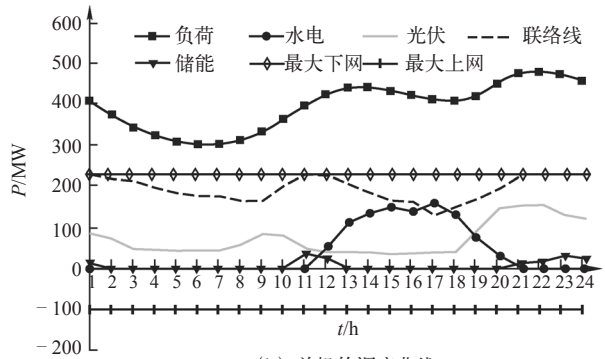
Fig. 6 Improved light-abandonment check

所有的拟合数据和实际数据均不满足条件，拟合数据需要配备的储能容量为140 MW，实际数据需要配备的储能容量为150 MW，拟合负荷和实际负荷的储能装置配置和实际数据的偏差为9.3%，平均相对偏差为2.9%。

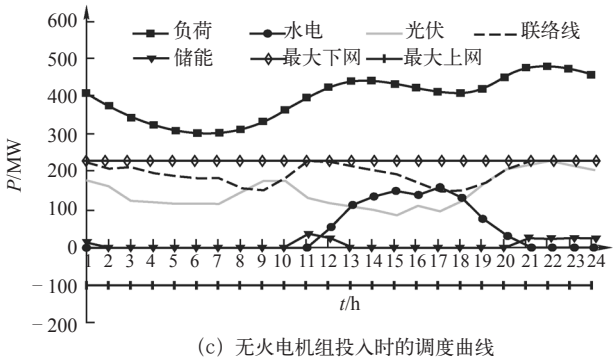
在限负荷校验时，负荷选择峰谷差最大的类，水电选择类别Ⅱ，光伏选择类别Ⅲ,然后选择峰谷差最小的数据，双机的调度曲线如图7a所示，单机刚好满足条件的调度曲线如图7b所示,无火电机组投入时刚好满足条件的调度曲线如图7c所示。



（a）双机的调度曲线



（b）单机的调度曲线



（c）无火电机组投入时的调度曲线

图7　改进的限负荷校验

Fig. 7 Improved load limit check

火电双机运行时可以满足所有的条件；单机投入运行时，不满足条件的天数占比例

为1.2%；在无火电机组投入时，不满足条件的水电约90.2%。

4 结束语

1）基于K均值聚类的算法在地区电网分析时，只能判断是否有风险，不能判断出风险的概率。

2）基于K均值算法在分析时，由于只选取数量最多的类，因此有的风险不能计算出来。

3）本文提出的算法和实际数据具有较小的偏差。

参考文献

[1] 韩兆洲，姚曈彤．扩展的“克强指数”构造：基于广东省的实证[J]．统计与决策，2018，34(11)：90-94．

[2] 刘振亚．全球能源互联网[M]．北京：中国电力出版社，2015．

[3] 舒印彪，张智刚，郭剑波，等．新能源消纳关键因素分析及解决措施研究[J]．中国电机工程学报，2017，37(1)：1-9．

[4] 范高锋，张楠，梁志锋，等．我国“三北”地区弃风弃光原因分析[J]．华北电力技术，2016(12)：55-59．

[5] 樊国伟，樊国旗，蔺红，等．含高载能负荷的光-火地区调度研究[J]．青海电力，2019，38(2)：1-5．

[6] 王锡凡，王秀丽．随机生产模拟及其应用[J]．电力系统自动化，2003(8)：10-15，31．

[7] 马彦宏，姜继恒，鲁宗相，等．基于随机生产模拟的火电机组深度调峰提升新能源消纳能力评估方法[J]．全球能源互联网，2019，2(1)：35-43．

[8] 林俐，费宏运，刘汝琛，等．基于分层聚类算法的地区风电出力典型场景选取方法[J]．电力系统保护与控制，2018，46(7)：1-6．

[9] 樊国旗，蔺红，程林，等．基于K均值模式划分改进模糊聚类与BP神经网络的风力发电预测研究[J]．智慧电力，2019，47(5)：38-42，83．

作者简介

任 冲

任冲（1984－），男，硕士，高级工程师，主要研究方向为电力系统稳定及控制。

卫琳（1980－），女，硕士，高级工程师，主要研究方向为电力系统稳定分析。

(收稿日期：2019－01－18)EA

公示

根据国家新闻出版署《关于2019年全国统一换发新闻记者证的通知》(国新出发[2019]39号)要求，我单位《电气应用》杂志已对换发记者证人员的资格进行严格审核，现将我单位拟换发新闻记者证人员名单进行公示。

换发新闻记者证人员：史海疆。举报电话：010-88379714。中国记者网新闻记者证核发办公室举报电话：010-83138953。

公示期：2019年12月21日至2019年12月31日。