

基于改进集成聚类 and BP 神经网络的电压偏差预测

王知芳¹, 杨 秀¹, 潘爱强², 陈甜甜², 谢真桢¹

(1. 上海电力学院电气工程学院, 上海 200090;
2. 上海市电力公司电力科学研究院, 上海 200437)

摘要: 电能质量问题日益严重, 其中电压偏差的危害最为明显。本文提出一种基于改进集成聚类和 BP 神经网络的电压偏差预测方法。针对单一聚类算法的不足, 将 AP 聚类算法与经典 K-means 聚类算法结合, 形成改进集成聚类算法, 实现两类算法的优势互补, 该算法包括 PCA 降维、AP 聚类、K-means 聚类三步。选择与待预测点相似相近的样本数据集, 采用改进集成聚类算法对数据集的气象数据进行聚类, 提取训练样本, 最后采用 BP 神经网络算法建立预测模型。结果表明, 该方法预测结果平均相对误差为 2.987%, 优于传统 BP 神经网络预测模型以及结合 PCA 降维的 BP 神经网络预测模型。

关键词: 电压偏差; 改进集成聚类; BP 神经网络; 气象因素

DOI: 10.12067/ATEEE1708083 **文章编号:** 1003-3076(2018)05-0073-08 **中图分类号:** TM711

1 引言

随着新型电力电子化装置的大量接入, 电能质量问题日益严重。越来越多变电站升级电能质量监测系统, 得到海量的实时电能质量监测数据。通过挖掘海量监测数据, 可预测电能质量的变化趋势, 及时做出预警及应对措施, 以保证电网安全、稳定、经济运行^[1]。

电能质量问题包括稳态电能质量问题和暂态电能质量问题。稳态电能质量问题包括电压偏差、频率偏差、电压波动与闪变、谐波及三相不平衡度五项指标^[2]。其中, 电压偏差的危害最为明显。电压过高造成设备过电压, 威胁绝缘, 减少使用寿命; 电压过低导致用户设备不能正常使用。因此, 预测电压偏差变化趋势并及时采取应对措施对于电力系统的正常运行具有很高的应用价值。由于不同的电能质量指标数据具有不同的特征, 不能把五项稳态指标用一个预测模型来预测, 否则会造成精度差的问题。本文将对电压偏差建立预测模型。

常用的预测方法主要有时间序列法、灰色预测法、支持向量机预测法等。时间序列法注重数据的

拟合, 对影响电能质量的不确定性因素考虑不足; 灰色预测法对数据的离散程度有要求, 当数据离散程度较大时, 预测精度较差; 支持向量机预测法在处理随机波动性较强的数据时, 精度较差, 且数据集规模过大时, 训练时间长, 速度慢^[3]。

目前, 对电能质量预测的研究文献较少, 而负荷预测方法已比较成熟, 可参考负荷预测的方法来进行电能质量预测。文献[4-6]采用基于小波分析的预测方法, 将负荷数据进行分解, 能够改善预测精度, 但是预测精度过多依赖于历史相似日的选取, 预测结果不稳定。文献[7, 8]采用基于决策树的负荷预测模型, 模式简单, 易于理解, 但是在选取属性进行扩展时, 对属性的内部相关性考虑欠缺, 且易受噪音影响。文献[9-11]采用综合预测方法, 结合多种预测模型的有效信息, 优势互补, 改善了预测精度, 但是模型的权重确定困难, 建模难度大。文献[12-14]采用偏最小二乘回归分析法进行负荷预测, 通过提取成分, 最大限度地利用了数据信息, 但是如何提高成分解释能力仍然难度较大。

电能质量预测的方法尚不成熟, 研究文献极少。文献[15]提出一种基于 Monte Carlo 抽样的电铁电

收稿日期: 2017-08-31
基金项目: 上海市科委地方能力建设计划项目(16020500900)
作者简介: 王知芳(1993-), 女, 山东籍, 硕士研究生, 研究方向为大数据在智能配用电中的应用;
杨 秀(1972-), 男, 江苏籍, 教授, 研究方向为分布式发电与微电网的运行与仿真。

能质量预测方法,即先用 Monte Carlo 抽样方法对牵引负荷进行随机波动处理,然后再对负序、谐波等指标进行预测,但是此方法预测对象有限制,仅适用于具有单相独立性、随机波动性等电铁一类的电能质量预测。文献[16]提出了一种间接预测方法,即先建立时间序列预测模型对有功功率进行预测,再通过有功功率与电能质量稳态指标的相关性建立神经网络预测模型,但是有功功率的预测结果较大程度影响电能质量的预测结果。文献[17]采用神经网络的预测模型对变电站和风电、光伏等用户的电能质量进行预测,但是此预测方法过于单一,且对预测精度没有做具体分析和说明。

本文针对电压偏差非线性、周期性的特点并结合实时气象数据提出一种基于改进集成聚类和 BP 神经网络的电压偏差预测方法。首先对气象数据采用主成分分析法(Principal Component Analysis, PCA)降维;然后采用改进集成聚类对气象主成分聚类,根据聚类结果提取与待预测点同类的历史数据;最后建立 BP 神经网络预测模型。将本文预测方法应用到上海市某电站电压偏差数据,结果证明本文提出的预测方法精度更高。

2 改进集成聚类算法描述

改进集成聚类算法的基本思想可表述如下:对气象数据进行 PCA 降维得到其主成分;对主成分进行 AP(Affinity Propagation)聚类,得到聚类数 K 和聚类中心,并将其作为 K-means 聚类的聚类数和初始聚类中心,进而得到最终聚类结果。

2.1 PCA 降维

由于气象因素较多,且包含无关变量,在应用 BP 神经网络进行电压偏差预测时,若将全部气象因素作为输入,会导致 BP 神经网络学习效率低甚至不能完成学习,因此有必要对气象数据进行降维处理。

PCA 降维算法将多个具有相关性的变量,重新组合成一组相互独立的主成分变量,其表达式为^[18]:

$$\mathbf{V} = \text{eigvec}[\text{cov}(\mathbf{X})] \tag{1}$$

式中, \mathbf{X} 为归一化的数据矩阵; $\text{cov}(\mathbf{X})$ 为协方差矩阵; \mathbf{V} 的列向量为 $\text{cov}(\mathbf{X})$ 的正交单位特征向量。

将 \mathbf{X} 降至 L 维后的数据矩阵为:

$$\mathbf{P} = \mathbf{X}\mathbf{V}_L \tag{2}$$

式中, \mathbf{V}_L 为矩阵 \mathbf{V} 的前 L 列。

2.2 AP 聚类分析

AP 聚类是 Frey 和 Dueck 于 2007 年提出的一种新的聚类算法^[19,20]。AP 聚类过程在信息传递的过程,吸引信息矩阵 \mathbf{R} 和归属信息矩阵 \mathbf{A} 在数据点之间交换信息,不断迭代更新,直到结束。 $r(i, k)$ 、 $a(i, k)$ 分别为 i 点与 k 点之间的吸引信息矩阵元素和归属信息矩阵元素, $r(i, k)$ 、 $a(i, k)$ 的值越大,则 k 点成为聚类中心的可能性就越大,并且 i 点隶属于以 k 点为聚类中心的类簇的可能性也越大。其计算公式为:

$$r(i, k) = s(i, k) - \max_{k' \neq k} \{a(i, k') + s(i, k')\} \tag{3}$$

式中, $s(i, k)$ 为 i 点与 k 点之间的相似度。

$$a(i, k) = \begin{cases} \min\{0, r(k, k) + \sum_{i' \neq \{i, k\}} \max\{0, r(i', k)\}\} & i \neq k \\ \sum_{i' \neq k} \max\{0, r(i', k)\} & i = k \end{cases} \tag{4}$$

由于 AP 聚类算法迭代过程容易产生振荡,所以每次迭代都加一个阻尼系数 λ , $\lambda \in (0, 1)$, 即

$$\begin{cases} r_{i+1}(i, k) = \lambda r_i(i, k) + (1 - \lambda) r_{i+1}^{\text{old}}(i, k) \\ a_{i+1}(i, k) = \lambda a_i(i, k) + (1 - \lambda) a_{i+1}^{\text{old}}(i, k) \end{cases} \tag{5}$$

相对于传统聚类, AP 聚类算法有如下优点:①无需事先指定聚类个数 K ;②无需随机选取初值,多次执行算法得到的结果相同;③误差平方低于其他单一聚类方法;④通过输入相似度矩阵来启动算法,因此对距离矩阵的对称性没要求,数据适用范围大。AP 算法的流程图如图 1 所示。

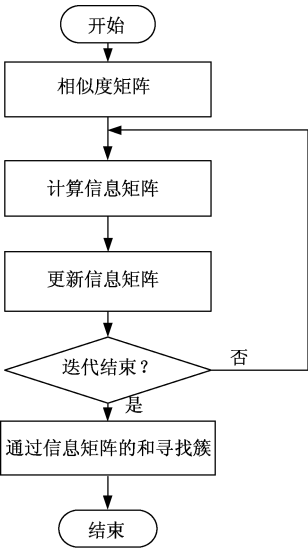


图 1 AP 算法流程图

Fig. 1 Flowchart of AP algorithm

2.3 K-means 聚类分析

K-means 聚类算法原理简单、操作快速。其核心思想为:以数据对象到该类聚类中心的距离平方和最小为目标,把 n 个数据对象按照特定标准划分为 K 个聚类^[21]。具体步骤如下:

(1)从 n 个数据对象 $X = \{x_1, x_2, \dots, x_n\}$ 中,随机选择 K 个对象作为初始聚类中心 $\mu_k (k = 1, 2, \dots, K)$ 。

(2)对于未选中的其他对象,则根据某种评价相似的标准,分别将他们划分给与其最相似的类别,形成 K 个簇 $C = \{c_k, k = 1, 2, \dots, K\}$ 。

(3)计算各类数据对象到该类别聚类中心 μ_k 的距离平方和 $J(C)$,直至最小。计算类内所有数据的均值作为新的聚类中心,其表达式为:

$$J(C) = \sum_{k=1}^K J(c_k) = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$$
$$= \sum_{k=1}^K \sum_{i=1}^n d_{ki} \|x_i - \mu_k\|^2 \tag{6}$$

式中,若 $x_i \in c_i, d_{ki} = 1$;若 $x_i \notin c_i, d_{ki} = 0$ 。

(4)判断聚类中心和 $J(C)$ 值是否发生变化,若变化转步骤(2),若不变,则聚类结束。

衡量相似性的标准有欧氏距离、余弦相似度等。本文选择欧氏距离衡量气象数据间的相似性。

3 BP 神经网络

针对电压偏差非线性、周期性的特点,本文选取神经网络建立预测模型。神经网络具备很强的学习能力,能更好地适应数据空间的变化和各种复杂的数据关系。

BP 神经网络的特点是信号前向传递,误差反向传递。BP 神经网络凭借其良好的性能广泛应用于分类识别、逼近、回归、压缩等领域^[22,23]。其结构如图 2 所示。

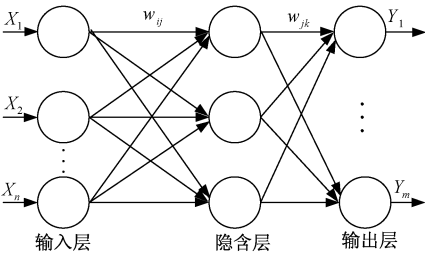


图2 BP神经网络结构模型

Fig.2 BP neural network structure model

4 案例分析

4.1 电压偏差预测流程

本文选取 2012 年 4 月 10 日~5 月 15 日、2013 年 4 月 10 日~5 月 15 日、2014 年 4 月 1 日~4 月 26 日共 2352 个整时点气象数据和电压偏差数据作为历史数据,对接下来的工作日 2014 年 4 月 27 日、4 月 28 日以及节假日 5 月 1 日整时点的电压偏差进行预测。本文气象数据来自于上海浦东国际机场使用的气象数据,电压偏差数据来自上海市某变电站的真实数据,组成如下的数据集记录:[整点时刻 气温 露点 湿度 气压 风向 风速 状况 电压偏差]。

基于改进集成聚类和 BP 神经网络的电压偏差预测方法流程如图 3 所示。其步骤如下。

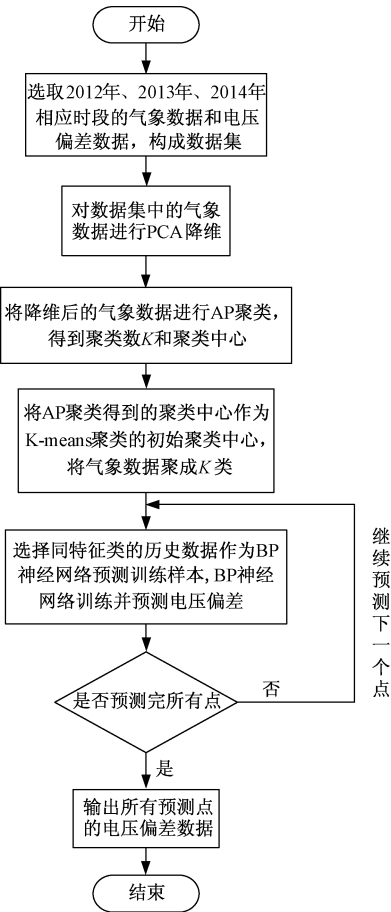


图3 本文改进方法预测流程图

Fig.3 Flowchart of voltage deviation forecasting

(1)为减小数据集的规模,提炼有效的气象数据信息,首先对预处理后的气象数据进行 PCA 降维处理。

(2)将降维后的气象数据进行 AP 聚类分析,得到聚类数 K 和 K 个聚类中心 $\mu_k(k=1,2,\cdots,K)$ 。

(3)将步骤(2)得到的 K 个聚类中心 $\mu_k(k=1,2,\cdots,K)$ 作为 K-means 聚类的初始聚类中心,对步骤(1)得到的气象主成分进行 K-means 聚类,得到 K 个类簇。

(4)根据步骤(3)的结果选择与待测点的气象因素同类历史数据集作为训练样本数据集,运用 BP 神经网络进行训练并预测电压偏差。

4.2 改进集成聚类算法

(1)PCA 降维处理

由于影响电压偏差的因素很多,为避免信息冗余现象,需对影响因素进行降维。分析数据集中数据的相关性,相关系数如表 1 所示。绘制相关性强度图,如图 4 所示。由表 1 和图 4 可知,气象因素中气温、湿度、风速、露点与电压偏差的相关性较高,但同时气象因素内部间的相关性也较高。所以,为避免信息重复现象,将气象数据进行 PCA 降维处理,得到一组相互独立又信息量大的主成分。根据贡献率,本文将原始气象数据 PCA 降维为 4 维。

表 1 气象因素与电压偏差的相关系数

Tab.1 Correlation coefficient of meteorological factor and voltage deviation

气象因素	气温	露点	湿度	气压	风向	风速	状况
相关系数	-0.47	-0.17	0.21	0.11	-0.02	-0.2	-0.06

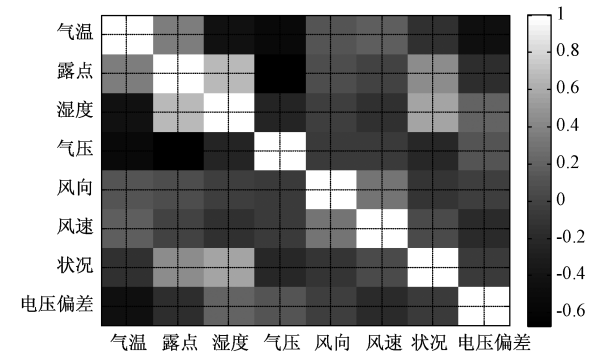


图 4 相关性强度图

Fig.4 Correlation strength

(2)AP 聚类

AP 聚类的输入为数据点之间的相似度矩阵 S , 本文数据点 i 和 j 之间的相似度为 $s(i,j) = -(x_i - x_j)^2$ 。图 5 为 AP 聚类迭代过程。

将降维后的气象数据进行 AP 聚类,由图 5 可知,迭代约 10 次便趋于稳定,得到聚类数 K 为 21,并且得到 21 个高质量聚类中心。

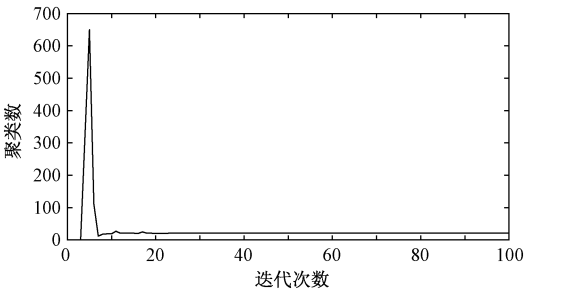


图 5 AP 聚类迭代过程

Fig.5 Iteration of AP clustering

(3)K-means 聚类

将 AP 聚类得到的 K 和聚类中心作为 K-means 聚类的聚类数和初始聚类中心,反复迭代,直到聚类中心不再发生变化,得到最终聚类结果。

(4)聚类评价指标 SSE

误差平方和 (SSE) 指标 I_{SSE} 的表征量为各样本到其所在类簇聚类中心的欧氏距离和,其表达式为:

$$I_{SSE} = \sum_{i=1}^K \sum_{x \in X_i} d^2(c_i, x)$$

(7)

式中, c_i 为类簇 X_i 的聚类中心; $d(c_i, x)$ 为向量间的欧氏距离。

各种聚类方法 SSE 值对比如表 2 所示。传统 K-means 聚类简单且效率高,以类簇内对象均值为聚类中心,目标为得到紧凑独立的簇,但是需事先给定聚类数 K 值,且受初始聚类中心的影响大,导致 K-means 聚类算法不稳定。AP 聚类无需随机选取初值,重复执行 AP 算法得到的结果相同,稳定性好,但其聚类中心为原始数据集中的点,此聚类中心不一定为最优聚类中心。本文提出的改进集成聚类算法将 AP 聚类和 K-means 聚类相结合,实现两类算法的优势互补,由表 2 可知,改进集成聚类算法的误差平方和最小,类簇间的分散性高,类簇内的紧凑性好。

表 2 各种聚类方法 SSE 值对比

Tab.2 SSEs of various clustering algorithms

聚类方法	AP 聚类算法	K-means 聚类算法	改进集成聚类算法
SSE 指标	3105.2168	3212.1584	2856.2801

4.3 BP 神经网络预测

(1)BP 神经网络设计

输入向量为与待测点的气象因素同类历史数据集,为 4 维向量;输出向量为整点时刻的电压偏差,为 1 维向量。本文取 10 次预测结果的均值作为

最终预测结果。

(2) 预测结果

本文提出的改进预测方法与传统 BP 神经网络预测方法、结合 PCA 降维的 BP 神经网络预测方法程序均在 MATLAB 平台上成功运行,将其结果进行对比。分别预测工作日 2014 年 4 月 27 日、4 月 28 日以及节假日 5 月 1 日的电压偏差,以测试工作日与节假日期间由于人们社会活动的不同,在电能质量发生较大波动的情况下,算法的稳定性是否良好。工作日 2014 年 4 月 27 日和 4 月 28 日预测结果如图 6、图 7 所示,节假日 2014 年 5 月 1 日预测结果如图 8、图 9 所示。表 3 为预测结果对比,并计算其平均相对误差:

$$e_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{L_i - F_i}{L_i} \right| \times 100\% \quad (8)$$

式中, n 为样本量; L_i 为实际值; F_i 为预测值。

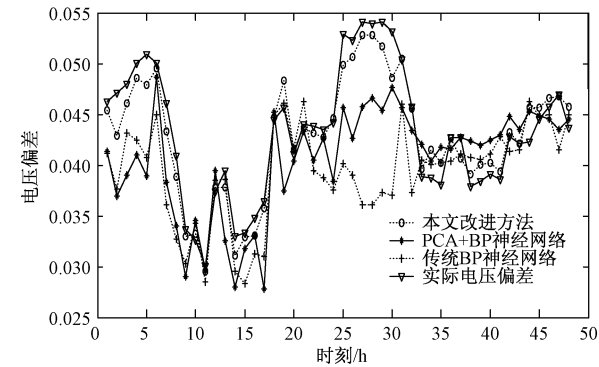


图 6 工作日预测结果对比
Fig. 6 Forecasting results in Apr. 27, 2014 and Apr. 28, 2014

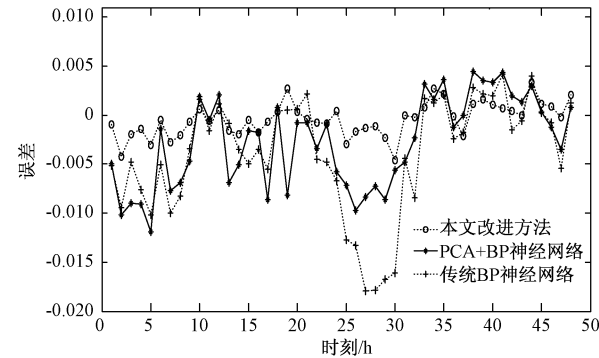


图 7 工作日预测误差对比
Fig. 7 Forecasting errors in Apr. 27, 2014 and Apr. 28, 2014

由图 6 ~ 图 9 和表 3 可知,本文改进方法比传统 BP 神经网络、PCA + BP 神经网络预测结果的平

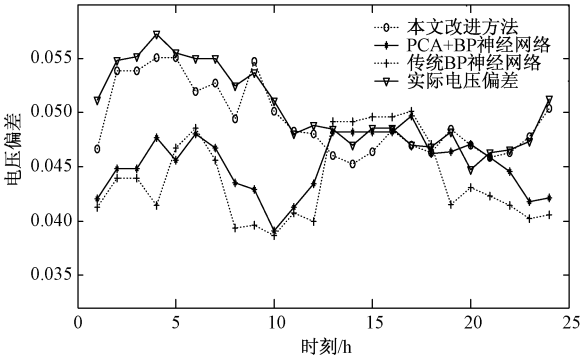


图 8 节假日预测结果对比
Fig. 8 Forecasting results in May 1, 2014

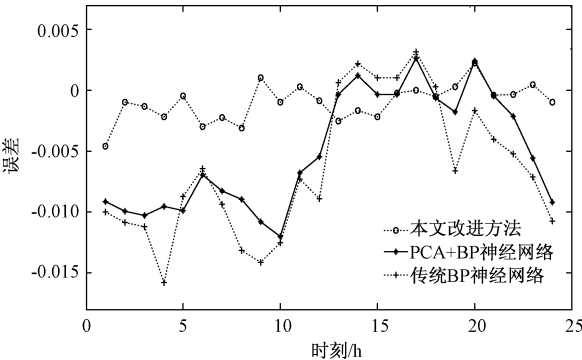


图 9 节假日预测误差对比
Fig. 9 Forecasting errors in May 1, 2014

表 3 预测方法误差对比

预测方法		传统 BP 神经网络	PCA + BP 神经网络	本文改进方法
平均 相对 误差 (%)	4 月 27 日	10.379	11.160	3.077
	4 月 28 日	11.969	8.158	3.224
	5 月 1 日	13.818	10.753	2.657

均相对误差都低,说明本文的改进方法中 PCA 降维和 AP + K-means 聚类两个步骤均能够提高预测精度;节假日较工作日的电压偏差整体有增大,但是本文方法仍有较高的预测精度,说明本文改进方法在外界环境有一定变化时,仍有较高的稳定性。

将 4 月 27 日、4 月 28 日、5 月 1 日三天三种方法预测结果的相对误差落在某一范围内的频数用直方图表示,得到误差分布直方图,如图 10 ~ 图 12 所示。从图 10 ~ 图 12 的直方图中可以看出,本文方法相对误差在 0 附近更为集中。最终预测结果如表 4 所示。本文改进方法平均相对误差为 2.987%,比传统神经网络方法降低了 9.069%,而且传统神经网络方法和 PCA + BP 神经网络方法的预测结果相

对误差控制在 3% 以内的概率分别为 16.67%、23.61%,而本文改进方法相对误差控制在 3% 以内的概率达到 59.72%,极端相对误差只是小概率事件。综上所述,本文的改进方法在电压偏差预测中效果更好,可为下一步电压偏差预警及电压偏差问题策略制定提供可靠依据。

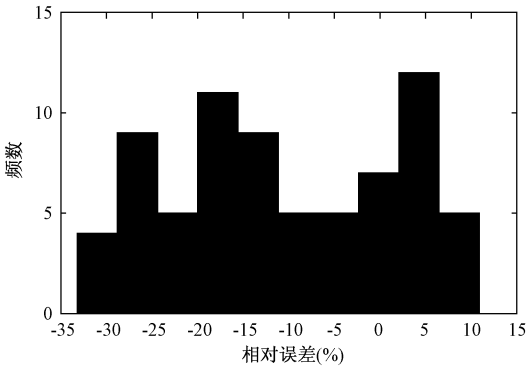


图 10 传统 BP 神经网络法误差分布图
Fig. 10 Distribution histogram of traditional BP model

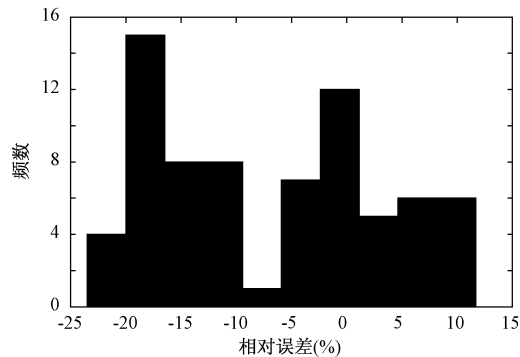


图 11 PCA + BP 神经网络法误差分布图
Fig. 11 Distribution histogram of BP model with PCA

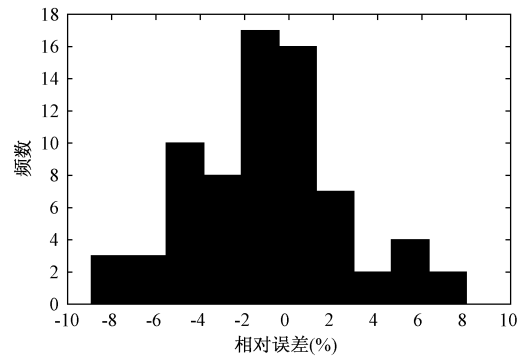


图 12 本文改进方法误差分布图
Fig. 12 Distribution histogram of proposed method

表 4 预测方法误差分析

预测方法	传统 BP 神经网络	PCA + BP 神经网络	本文改进方法
平均相对误差 (%)	12.056	10.024	2.987
误差控制在 3% 以内概率 (%)	16.67	23.61	59.72

5 结论

本文针对电压偏差非线性、周期性的特点,同时考虑气象因素,将聚类算法与 BP 神经网络结合,建立了基于改进集成聚类与 BP 神经网络的电压偏差预测模型。

(1)由于影响电压偏差预测精度的因素过多,为避免信息干扰和信息重复,将气象数据进行 PCA 降维,得到一组相互独立又信息量大的主成分,对气象主成分做聚类分析。

(2)将 K-means 聚类与 AP 聚类算法相结合,形成改进集成聚类算法,实现两类算法的优势互补,结果证明改进集成聚类算法效果更好。

(3)选择与待预测整时点同类的历史数据集作为训练样本,运用 BP 神经网络进行训练并预测,将其预测结果与传统 BP 神经网络、PCA + BP 神经网络预测结果相比较,结果证明本文提出的方法整体预测指标更好。

虽然本文提出的预测模型能够较好地预测短期电压偏差,但仍有提高的空间。例如,AP 聚类算法的输入为 $N \times N$ 矩阵,内存开销大,且算法复杂度高;BP 神经网络算法样本依赖性强且初始权重为随机给定,导致其预测结果稳定性欠佳。因此可基于以上问题,对本文预测模型做进一步改进。

参考文献 (References):

[1] 丁泽俊, 刘平, 欧阳森, 等 (Ding Zejun, Liu Ping, Ouyang Sen, et al.). 电能质量预测与预警机制及其应用 (Mechanism of power quality forecast and early warning and their application) [J]. 电力系统及其自动化学报 (Proceedings of the CSU-EPSCA), 2015, 27 (10): 87-92.

[2] 胡铭, 陈珩 (Hu Ming, Chen Heng). 电能质量及其分析方法综述 (Survey of power quality and its analysis methods) [J]. 电网技术 (Power System Technology), 2000, 24 (2): 36-38.

- [3] Jigoria-Opera D, Lustrea B, Kilyeni St, et al. Daily load forecasting using recursive artificial neural network vs. classic forecasting approaches [A]. 5th International Symposium on Applied Computational Intelligence and informatics [C]. Timisoara, Romania, 2009. 487-490.
- [4] 邵能灵, 侯志俭, 李涛, 等 (Tai Nengling, Hou Zhi-jian, Li Tao, et al.). 基于小波分析的电力系统短期负荷预测方法 (New principle based on wavelet transform for power system short-term load forecasting) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2003, 23 (1): 46-51.
- [5] 曲正伟, 张坤, 王云静, 等 (Qu Zhengwei, Zhang Kun, Wang Yunjing, et al.). 基于 PSO-OMP 优化的 WD-ASD 超短期负荷预测 (Short-term load forecasting based on WD-ASD optimized by PSO-OMP) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2017, 36 (12): 39-45.
- [6] 李如琦, 褚金胜, 王宗耀, 等 (Li Ruqi, Chu Jinsheng, Wang Zongyao, et al.). 小波分析及其在短期负荷预测中的应用 (Wavelet analysis and its use in short-term load forecasting) [J]. 现代电力 (Modern Electric Power), 2009, 26 (3): 63-67.
- [7] 孙英云, 何光宇, 翟海青, 等 (Sun Yingyun, He Guangyu, Zhai Haiqing, et al.). 一种基于决策树技术的短期负荷预测算法 (A short-term load forecasting method based on decision-tree approaches) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2004, 23 (3): 55-58, 75.
- [8] 高亦凌, 胡翼, 辛洁晴, 等 (Gao Yiling, Hu Yi, Xin Jieqing, et al.). 基于决策树主变最佳负荷预测研究 (Study of transformer optimal load forecasting method based on decision tree) [J]. 电网与清洁能源 (Power System and Clean Energy), 2014, 30 (3): 93-97.
- [9] 康重庆, 夏清, 沈瑜, 等 (Kang Chongqing, Xia Qing, Shen Yu, et al.). 电力系统负荷预测的综合模型 (Integrated model of power system load forecasting) [J]. 清华大学学报 (自然科学版) (Journal of Tsinghua University (Science and Technology)), 1999, 39 (1): 8-11.
- [10] 程旭, 康重庆, 夏清, 等 (Cheng Xu, Kang Chongqing, Xia Qing, et al.). 短期负荷预测的综合模型 (Integrated model of short-term load forecasting) [J]. 电力系统自动化 (Automation of Electric Power Systems), 2000, 24 (9): 42-44.
- [11] 李春祥, 牛东晓, 孟丽敏, 等 (Li Chunxiang, Niu Dongxiao, Meng Limin, et al.). 基于三指标量的中长期负荷预测综合模型 (Comprehensive models for mid-long term load forecast based on three target quantities) [J]. 华东电力 (East China Electric Power), 2008, 36 (6): 10-14.
- [12] 毛李帆, 江岳春, 龙瑞华, 等 (Mao Lifan, Jiang Yuechun, Long Ruihua, et al.). 基于偏最小二乘回归分析的中长期电力负荷预测 (Medium- and long-term load forecasting based on partial least squares regression analysis) [J]. 电网技术 (Power System Technology), 2008, 30 (19): 71-77.
- [13] 蔡金錠, 王慧 (Cai Jinding, Wang Hui). 非线性偏最小二乘回归在电力负荷预测中的应用 (Application of non-linear partial least square regression in electricity load prediction) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2006, 25 (2): 15-17, 58.
- [14] 季泽宇, 袁越, 邹文仲 (Ji Zeyu, Yuan Yue, Zou Wenzhong). 改进偏最小二乘回归在电力负荷预测中的应用 (Application of improved partial least square regressive model in power load forecasting) [J]. 电力需求侧管理 (Power Demand Side Management), 2011, 13 (1): 10-14.
- [15] 秦浩庭, 李群湛, 刘燕, 等 (Qin Haoting, Li Qunzhan, Liu Yan, et al.). 基于 Monte Carlo 的电铁电能质量预测方法 (Prediction of the power quality caused by electrified railways based on Monte Carlo) [J]. 电力系统保护与控制 (Power System Protection and Control), 2011, 39 (13): 64-70, 77.
- [16] 苏卫卫, 马素霞, 齐林海 (Su Weiwei, Ma Suxia, Qi Linhai). 基于 ARIMA 和神经网络的电能质量稳态指标预测 (Predicting of power quality steady indicators based on ARIMA and neural network) [J]. 计算机技术与发展 (Computer Technology and Development), 2014, 24 (3): 163-167.
- [17] 刘可, 薛俊茹, 宋锐, 等 (Liu Ke, Xue Junru, Song Rui, et al.). 青海电网数字化在线监测平台短期电能质量预测功能应用与分析 (Application and analysis of short-term power quality prediction of digital on-line monitoring platform for Qinghai power grid) [J]. 科技信息 (Science and Technology Information), 2012, 34 (15): 690-691.
- [18] 张斌, 庄池杰, 胡军, 等 (Zhang Bin, Zhuang Chijie, Hu Jun, et al.). 结合降维技术的电力负荷曲线集成聚类算法 (Ensemble clustering algorithm combined with dimension reduction techniques for power load profiles) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2015, 35 (15): 3741-3749.
- [19] Fery B J, Dueck D. Clustering by passing messages between data points [J]. Science, 2007, 305 (5814): 972-976.

- [20] Dueck D, Frey B J. Non-metric affinity propagation for unsupervised image categorization [A]. IEEE International Conference on Computer Vision [C]. Rio de Janeiro, Brazil, 2007. 1-8.
- [21] Parmeet K. Outlier detection using k means and fuzzy min max neural network in network data [A]. 2016 8th International Conference on Computational Intelligence and Communication Networks [C]. Jabalpur and Dehradun, India, 2016. 693-696.
- [22] 刘波, 张焰, 陈煜, 等 (Liu Bo, Zhang Yan, Chen Yu, et al.). 基于GA-改进BP神经网络算法在大电网短路电流预测中的应用 (Short-circuit current forecast application of big electrical network based on improved BP artificial neural network combined with genetic algorithm) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2006, 25 (4): 43-46.
- [23] 安晨帆, 杜志叶, 李慧慧, 等 (An Chenfan, Du Zhiye, Li Huihui, et al.). 基于组合赋权和BP神经网络的500kV交流输电线路电磁环境评估方法研究 (Study of 500kV of AC transmission line electromagnetic environment evaluation method based on combination empowerment and BP neural network) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2016, 35 (3): 62-68.

Voltage deviation forecasting based on improved ensemble clustering and BP neural network

WANG Zhi-fang¹, YANG Xiu¹, PAN Ai-qiang², CHEN Tian-tian², XIE Zhen-zhen¹

(1. College of Electric Engineering, Shanghai University of Electric Power, Shanghai 200090, China;

2. State Grid Shanghai Electric Power Research Institute, Shanghai 200437, China)

Abstract: This paper proposes a voltage deviation forecasting method based on improved ensemble clustering and BP neural network. The ensemble algorithm was a combination of principal component analysis (PCA), affinity propagation clustering and K-means clustering. Firstly the PCA algorithm was used for lowering the dimensions of the meteorological data, and then the improved ensemble cluster analysis is performed on the principal components, and finally the BP neural network for voltage deviation forecasting is adopted. The practical application results prove that the e_{MAPE} of the proposed method is 2.987%, which is obviously better than that of the traditional BP model and BP model with PCA.

Key words: voltage deviation; improved ensemble clustering; BP neural network; meteorological factor