

基于 GSA 的电力系统不良数据辨识算法

杨伟, 胡军, 吴军基

(南京理工大学动力学院, 江苏 南京 210094)

摘要: 随着数字化技术在电力系统中的广泛应用及对电力系统运行可靠性要求的不断提高, 不良数据的辨识显得越来越重要。目前广泛应用的状态估计法, 存在残差污染等缺点。论文研究了建立在神经网络和聚类分析基础上的 GSA 不良数据辨识算法, 该算法运用神经网络完成对测量数据的预处理, 然后由 GSA 算法对聚类分析后的结果进行判断, 完成不良数据的辨识。论文借助 Matlab 及 C 语言对 GSA 算法进行了仿真, 通过一个具体的网络不良数据辨识将此算法与状态估计算法进行了比较, 验证了该算法的有效性及其实用性, 有效地避免了不良数据的漏检、误检。

关键词: GSA; 数据挖掘; 不良数据辨识

中图分类号: TM73

文献标识码: A

文章编号: 1003-4897(2005)22-0041-04

0 引言

现代电力系统中, 数据的质量变得越来越重要, 系统操作中的不良数据可能会对系统的安全运行造成重大的影响。

目前国内外常用基于状态估计的方法检测不良数据, 此方法的缺点是很可能出现残差污染和残差淹没现象, 从而造成漏检或误检, 影响了辨识的效果。

论文研究了基于 GSA 的不良数据辨识算法, 此算法以数理统计为基础, 融合神经网络技术和聚类算法的优点, 有效地避免了不良数据的漏检、误检。

1 基于 GSA 的不良数据算法

GSA 算法主要由神经网络和聚类分析两个模块构成, 如图 1 所示^[1]。

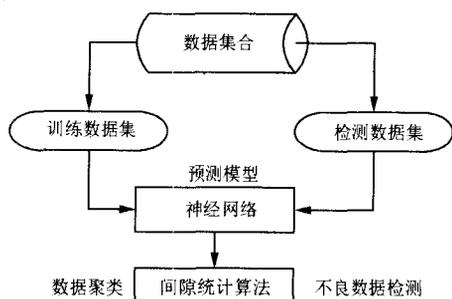


图 1 GSA 算法模块组成图

Fig 1 GSA algorithm module composition

论文选用 BP 神经网络, 神经网络将对 200 行训练数据进行网络训练 (原始数据表共有 220 行, 前 200 行用来训练, 后 20 行用作测试), 考虑到网络

训练的速度问题以及计算中的实际需求, 神经网络的输入选为 $u_1, u_2, u_3, p_1, p_2, p_3, q_1, q_2, q_3$, 输出选为 $u_1, u_2, p_1, p_2, q_1, q_2$ 。由此设置系统精度为 0.000 001, 最大训练次数设为 10 000, 步长设为 0.1, 输入层数目设为 9, 隐含层数目设为 18, 输出层数目设为 6, 激活函数选择 $Tansig()$ 。取期望输出和实际输出之差的平方和为误差函数, 则有: $e = \frac{1}{2} \sum_i (X_i^m - Y_i)^2$, 如果实际输出不能满足要求, 则把误差信号反向传回, 对每个隐层的各个神经元的权系数进行修改, 直至误差函数值最小, 即输出满足系统的精度要求^[2]。

神经网络的输出数据作为聚类分析模块的待测数据输入。聚类模块采用 k-means 算法^[3], 对于每一次输入的聚类数 k , 都需要计算一次, 得到以均方差为标准的误差函数 $W_k = \sum_{j=1}^k \sum_{i \in C_j} |i_i - w_j|^2$, 其中, w_j 取聚类 C_j 中各空间点的平均值, 代表聚类 C_j 的聚类中心。

GSA 算法的基本思想是将测量数据的 $\ln(W_s(k))$ 与它相应的参考数据的 $\ln(W_{ref}(k))$ 的数学期望值作比较。如果在某个 k 值点, $\ln(W_s(k))$ 的曲线与参考曲线相比下降的幅度最大, 即两条曲线之间的间隙 (gap) 最大, 说明 k 值就是最佳的聚类个数的值^[1, 4, 5]。定义

$$Gap(k) = E\{\ln(W_{ref_j}(k))\} - \ln(W_s(k))$$

式中: $E\{\ln(W_{ref_j}(k))\} = \frac{1}{F} \sum_{j=1}^F \ln(W_{ref_j}(k))$ 表示 F 组分布样本大小为 n 的 p 维参考数据的数学期望

值 (这个期望值的大小与 n 有关,因为在计算 w_k 的过程中隐含了与 n 有关的信息)。

作为比较基准,参考数据集通过在待检测数据范围内产生均匀分布的随机值的方法得到。对于每组待测数据,产生 n 组参考数据集。具体来说,参考数据采用以下两种方法形成:

- a 在观察值的范围内以均匀分布方式产生参考数据。
- b 在观察值的主分量组成的集合中,用均匀分布方式产生参考数据。

具体来说,假设 X 是一个 $n \times p$ 的矩阵,可以对其进行奇异值分解 $X = UDV^T$,首先进行矩阵变换 $X = XV$,接着在 X 列值范围内产生均匀分布的 Z (使用 a 中的方法),最后通过矩阵变化 $Z = ZV^T$,得到参考数据集 Z 。

最佳的聚类个数的值 k 是满足下式的最小值:

$$Gap(k) = \frac{Gap(k+1) - (k+1)}{\sqrt{(1 + \frac{1}{F}) [I_{ref}(k)]^2}}$$

$$[I_{ref}(k)]^2 = E\{[\ln(W_{ref,j}(k))]^2\} - E^2\{\ln(W_{ref,j}(k))\}$$

式中:引入 (k) 主要是考虑到每一组参考数据中的内部偏移量对聚类的影响。

2 GSA 算法流程图

图 2 给出了基于 GSA 算法的程序流程图^[1],论文利用 C 语言根据此流程图编制了 GSA 算法程序,以下有关 GSA 算法的数据均由此程序计算得出。

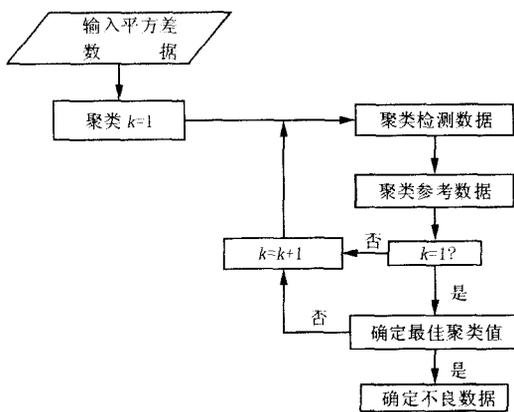


图 2 GSA 算法程序流程图

Fig 2 GSA algorithm program flow

GSA 算法中神经网络模块的输入量为系统中各节点的测量数据与安全可靠运行数据之间的差值 e_i , 经过神经网络后得到相应的输出量 O_i 。同时,计

算得到各节点的平方值 $(e_i - O_i)^2$, 这组数据值作为聚类分析模块的待检测数据输入。在聚类模块中,通过对方差数据集 $(e_i - O_i)^2$ 进行 GSA 判定,确定出最佳聚类 k 。当最佳的聚类值是 1 时,意味着所有待检测数据都可以被认为是正常数据。然而,一旦最佳聚类值不为 1 时,意味着所观察的待检测值中有不良数据。接着,计算每个聚类的 $(e_i - O_i)^2$ 的平均值。某个具有最小 $(e_i - O_i)^2$ 平均值的聚类被鉴定为正常数据所在的类,而其它的都被认为是不良数据类。通过上述步骤,正常数据被归并到一个类中而不良数据被分离开来。

3 算例仿真与结果分析

3.1 算例仿真

目前,在电力系统中广泛采用状态估计法检测不良数据。状态估计是利用实时测量系统的冗余度来提交数据精度的,自动排除随机干扰所引起的错误信息,估计或预测系统的运行状态 (或轨迹)^[6]。

假设在图 3 所示的电力系统中,节点 34 的 a 相有功功率出现不良数据 0.2, 分别用状态估计模糊 ISODATA 算法与 GSA 算法进行检测,检测结果如表 1 所示。其中模糊 ISODATA 算法选取 0.5 作为其检测的门槛值^[7,8]。

表 1 模糊 ISODATA 算法检测结果

Tab 1 Vague ISODATA algorithm checking result

测点	P34a	P34b	P34c	P45a	P45b	P45c	P46a
隶属度	0.659 1	0.448 1	0.267 1	0.279 6	0.251 2	0.260 3	0.009 2
测点	P46b	P46c	P52a	P52b	P52c	P71a	P71b
隶属度	0.119 3	0.163 3	0.227 7	0.201 7	0.111 8	0.403 3	0.315 1
测点	P71c	P75a	P75b	P75c	P92a	P92b	P92c
隶属度	0.121 6	0.257 3	0.182 8	0.113 8	0.409 3	0.181 7	0.116 1
测点	P911a	P911b	P911c	P32-71a	P32-71b	P32-71c	
隶属度	0.325 0	0.121 6	0.090 5	0.430 0	0.426 8	0.296 8	

表 2 GSA 算法中的 $Gap(k)$ 和 (k) 值

Tab 2 $Gap(k)$ and (k) value in GSA algorithm

聚类数	$k=1$	$k=2$	$k=3$
$Gap(k)$	-2.487 0	7.419 2	6.794 5
(k)	0.004 9	0.021 0	0.114 8

表 3 GSA 算法聚类结果

Tab 3 Clustering result of GSA algorithm

聚类数	$k=1$	$k=2$	$k=3$
聚类点	所有点	34 其余点	34 52, 71, 92, 911 其余点
$(e_i - O_i)^2$ 平均值	0.228 7	0.572 6	0.185 7 0.572 6 0.185 95 0.185 54

根据表 2 中的数据绘制了如图 4 所示的曲线图。

从图 4 中可以看出在聚类数 $k=2$ 时,参考数据曲线与测量数据曲线之间的间隙值达到最大,因此

$k=2$ 时就是最佳聚类。表 3 通过对各聚类 $(e_i - O_i)^2$ 平均值的计算显示了在聚类数 $k=2$ 时,不良数据节点 34 被检测出来。

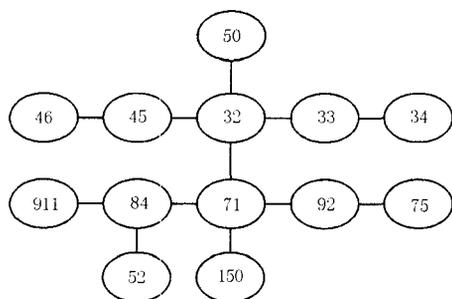


图 3 电力系统拓扑图

Fig 3 Power system topology

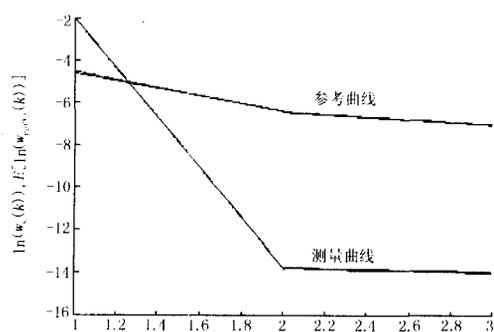


图 4 GSA 算法检测效果图

Fig 4 GSA algorithm test effect

3.2 结果分析

从仿真结果中可以看出,模糊 ISODATA 算法能够准确地检测出图 3 系统中出现的不良数据,但是当系统规模比较大以及不良数据的相关性较大时,模糊 ISODATA 算法的检测标准——隶属度将会接近所选取的阈值 0.5,如果系统规模进一步扩大或者当阈值选择不恰当时,将会出现漏检、错检的情况。然而 GSA 算法的检测效果取决于神经网络以及聚类算法,只要神经网络构造得好,训练精度高并且聚类分析算法选择恰当,对于各种规模的电力系统,不论不良数据的形式如何,GSA 算法都能够将其辨识出来。

可以看出,GSA 算法与状态估计法相比,具有以下几个特点:

1)检测精度高。GSA 算法不受系统规模以及数据性质的限制,摆脱了阈值选取主观性和经验性的困扰,极大地提高了检测的准确性。

2)过程简单。由于状态估计法是在假设加权残差或标准残差满足一定概率分布的前提下,对其

进行假设检验,进而寻找不良数据,所以相比之下,GSA 算法牵涉到的计算简单,可以有效地提高不良数据的检测效率。

4 结论

GSA 算法是一种新的不良数据检测算法。它结合了神经网络技术与聚类算法在不良数据检测方面的优点,为检测不良数据提供了一条新的途径。本文论述了 GSA 算法的基本理论及其在电力系统中的实用性,并在仿真结果上与状态估计法进行了比较,验证了 GSA 算法在检测不良数据的优越性。

参考文献:

- [1] HUANG Shy-jier, L N Jeur-min Enhancement of Power System Data Debugging Using GSA-based Data-mining Technique[J]. IEEE Trans on Power Systems, 2002, 17 (4): 1022-1026
- [2] 闻新,周露. Matlab神经网络仿真与应用[M]. 北京:科学出版社,2003.
WEN Xin, ZHOU Lu The Imitation and Application of the Neural Network by Matlab [M]. Beijing: Science Press, 2003.
- [3] 邵峰晶,于忠清. 数据挖掘原理与算法[M]. 北京:中国水利水电出版社,2003.
SHAO Feng-jing, YU Zhong-qing The Principle and Algorithm of Data Mining[M]. Beijing: China Water Power Press, 2003.
- [4] Tibshirini R, Walther G, Hastie T. Estimating the Number of Clusters in a Dataset via the Gap Statistic[R]. New York: Stanford University, 2000.
- [5] LUO Yu-hui, Jonathon C. Active Source Selection Using Gap Statistics for Underdetermined Blind Source Separation[A]. Signal Processing and Its Applications Proceedings, Seventh International Symposium. 2003. 137-140.
- [6] 李钊年. 电力系统状态估计中的不良数据辨识[J]. 青海大学学报(自然科学版), 2001, (1): 49-51.
LI Zhao-nian The Identification of the Bad Data in the State Estimate of the Power System [J]. Journal of Qinghai University, Sci & Tech, 2001, (1): 49-51.
- [7] 张云岗. 三相状态估计中不良数据检测与辨识的研究(硕士学位论文)[D]. 南京:河海大学,2001.
ZHANG Yun-gang The Research of the Examination and Identification of the Bad Data in the Three-phase State Estimate, Thesis[D]. Nanjing: Hohai University, 2001.
- [8] 卫志农,张云岗,郑玉平. ISODATA 方法在配电网状态估计不良数据辨识中的应用[J]. 河海大学学报,

2003, (2): 97-100.

WEI Zhi-nong, ZHANG Yun-gang, ZHENG Yu-ping
The Application of ISODATA in the Bad Data Identification of the State Estimate of the Distribution Power System [J]. Journal of Hohai University, 2003, (2): 97-100.

作者简介:

杨伟 (1965 -),男,副教授,从事电力系统及其自动化方面的教学与研究工作; E-mail: yangwei807@vip.sina.com
胡军 (1982 -),男,硕士,研究方向为配电网自动化;
吴军基 (1955 -),男,教授,博士生导师,从事电力系统及其自动化、电力市场等方面的教学与研究工作。

收稿日期: 2005-03-03; 修回日期: 2005-04-06

The identification algorithm of bad data in power system based on GSA

YANG Wei, HU Jun, WU Jun-ji

(School of Power Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: With the wide application of numeric technology in power system and increasingly higher requirements of the power system operation reliability, the recognition of the bad data seems more and more important. The state estimating algorithm (be widely used at present) has disadvantage of residual pollution. This paper researches the GSA bad data algorithm based on neural network and clustering analysis. The algorithm accomplishes the pretreatment of surveying data via the neural network, and then judges the result of clustering analysis so as to finish the recognition of the bad data. The paper employs Matlab and C Language to simulate the GSA algorithm and compares the GSA algorithm with the state estimating algorithm via a concrete network, which shows that GSA algorithm can effectively remove residual pollution so as to accomplish the recognition of the bad data.

Key words: GSA; data mining; bad data recognition

(上接第 36 页 continued from page 36)

- JU Ping. Power System Non-linear Identification [M]. Nanjing: Hohai University Press, 1999.
- [10] 李刚, 张仙平, 李志伟, 等. 小型制冷设备低压启动探析 [J]. 郑州纺织工学院学报, 2000, 11 (2): 12-16.
- LI Gang, ZHANG Xian-ping, LI Zhi-wei, et al. Investigation and Analysis on Low-voltage Starting of Small Refrigeration Equipment [J]. Journal of Zhengzhou Textile Institute, 2000, 11 (2): 12-16.
- [11] 李刚. 小型制冷设备启动装置研究 [J]. 制冷, 2000, 19 (1): 60-63.
- LI Gang. Study on Starting Device of Small Refrigeration

- Equipments [J]. Refrigeration, 2000, 19 (1): 60-63.
- [12] 黄立培. 电动机控制 [M]. 北京: 清华大学出版社, 2003.
- HUANG Li-pei. Motor Control [M]. Beijing: Tsinghua University Press, 2003.

收稿日期: 2005-03-22; 修回日期: 2005-06-16

作者简介:

张利民 (1958 -),男,高级工程师,从事电网规划和电气设备方面的科研和教学工作;
鞠平 (1962 -),男,博士,教授,博士生导师,主要研究方向为电力系统建模与控制。E-mail: pju@hhu.edu.cn

Relationship between a air conditioner load and temperature and its influences on voltage stability

ZHANG Limin¹, FU Hong-jun², LI Yu-yan³, DA IQI¹, HAN Jing-dong¹, JU Ping¹

(1. Hohai University, Nanjing 210098, China; 2. Henan Power Company, Zhengzhou 450052, China;
3. Nanjing Automation Factory, Nanjing 210003, China)

Abstract: Air conditioner load is increasing and should be paid more attention. A model is set up for predicting the air conditioner load varying with the temperature in Henan Power Grid. The influence of air conditioner load, especially the re-starting process, on power system voltage stability is studied. The results show that air conditioner load is bad to power system voltage stability.

This paper is supported by Natinal Natural Science Foundation of China (No. 5059412).

Key words: air conditioner; bad modeling; load characteristic; stall