

利用自组织映射 SOM 实现电力系统暂态稳定 评估结果可视化

李亦宁¹, 王晓茹¹, 向丽萍¹, 王建², 谢桦³, 肖丹雄¹

(1. 西南交通大学电气工程学院, 四川 成都 610031; 2. 清华大学电机工程与应用电子技术系, 北京 100084;
3. 北京交通大学电气工程学院, 北京 100044)

摘要: 自组织映射 SOM (self-organizing map) 具有拓扑关系保持的优良性能, 可应用于降维可视化。该文基于 SOM 实现暂态稳定评估结果的可视化。采用经典的 SOM 算法分别对原始特征和特征选取后的 IEEE16 机暂态稳定分类数据进行可视化, 表明了可视化的可行性以及特征选取的重要性。采用一种改进的 SOM 算法—DPSOM 算法, 提高了可视化的性能, 得到了满意的暂态稳定评估分类可视化结果。

关键词: 暂态稳定评估; 可视化; SOM; DPSOM

SOM based visualization of power system transient stability assessment results

LI Yi-ning¹, WANG Xiao-ru¹, XIANG Li-ping¹, WANG Jian², XIE Hua³, XIAO Dan-xiong¹

(1. Southwest Jiaotong University, Chengdu 610031, China; 2. Tsinghua University, Beijing 100084, China;
3. Beijing Jiaotong University, Beijing 100044, China)

Abstract: Due to the topology-preserving nature, the SOM (self-organizing map) algorithm can be used to visualize the high-dimensional data. In this paper, the visualization of power system transient stability assessment results which based on SOM is realized. Firstly, for the IEEE 16 generators transient stability classified data, the original feature data and the feature selected data are visualized respectively by using the classic SOM algorithm, then the practicability of visualization and the importance of feature selection are explained. Finally, an improved SOM algorithm—DPSOM algorithm is adopted, which improves the performance of visualization and gets the satisfied results of visualization.

This project is supported by Special Fund of the National Basic Research Program of China (No. G1998010301).

Key words: transient stability assessment; data visualization; SOM; DPSOM

中图分类号: TM712 文献标识码: A 文章编号: 1674-3415(2009)05-0041-06

0 引言

数据挖掘为电力系统暂态稳定评估提供了新的方法^[1]。特征集数据通过训练好的分类器可获得暂态稳定或不稳定的分类结果。如果该分类结果能够在有稳定域边界的二维平面上表示出来, 有助于调度员直观了解系统的安全稳定状态, 实施更有效的调度和控制。

文献[2]提出一种割集功率空间上电压稳定域边界局部近似算法, 并在此基础上实现了静态电压稳定域的可视化, 为电力系统电压稳定性分析和监

视控制提供了支持。在暂态稳定评估问题上, 需要对暂态稳定数据进行可视化, 得到明确的分类可视化结果。

本文根据 IEEE16 机系统暂态稳定特征和分类数据, 采用经典的降维可视化方法—SOM 算法, 分别对原始数据与特征选取后的数据进行可视化, 阐明了可视化的可行性以及特征选取的重要性; 然后采用改进的 SOM 算法—DPSOM 算法进行可视化, 解决了原始 SOM 算法中存在的问题, 得到了满意的分类可视化结果。

1 暂态稳定评估特征和分类数据

数据来自于 IEEE16 机仿真系统: 采用仿真软件 PST2.0, 对 IEEE16 机 86 节点系统进行时域仿真计算, 建立原始特征数据库^[3]。暂态稳定评估的初始

基金项目: 国家 973 重点基础研究发展规划资助项目 (G1998010301)

特征集, 如表 1 所示。

表 1 初始输入特征
Tab.1 Primary input feature

1: 系统中发电机机械输入功率的平均值
2: 故障切除时刻所有发电机转子动能的平均值
3: 系统总的能量调整
4: 故障切除时对系统的冲击
5: 故障切除时 coi 坐标下转速的和
6: 故障切除时与 coi 相差最大的转速
7: 故障切除时与 coi 相差最大的功角
8: 故障切除时领前机与殿后机的功角之差
9: 故障切除时最大的加速度之差
10: 故障切除时最大的加速度的变化率之差
11: 故障切除时最大的转子动能之差
12: 故障切除时最大的转子动能变化率之差
13: $TZ13= TZ12/TZ10$
14: $TZ14= TZ12/TZ11$
15: $TZ15= TZ11/TZ8$
16: $TZ16= TZ12/TZ8$
17: $TZ17= TZ11/TZ10$
18: 故障瞬间所有发电机转子初始加速度的最大值
19: 具有最大初始加速度发电机的初始角度
20: 所有转子初始加速度的最小值
21: 故障切除时刻所有转子动能的最大值
22: 故障切除时具有最大动能发电机的转子角度
23: 所有发电机初始加速度的方差
24: 故障切除时具有最大转子角度发电机的转子动能
25: 所有发电机初始加速功率的均值
26: 所有发电机初始加速功率的方差
27: 所有发电机相对初始加速功率的均值
28: 所有发电机相对初始加速功率的方差
29: 所有发电机初始加速度的均值
30: 故障瞬间发电机所受的最大有功冲击
31: 故障瞬间发电机所受的最小有功冲击
32: 所有发电机相对初始加速度的平均值
33: 所有发电机相对初始加速度的方差
34~38: 取 t_0 、 $0.5 t_0$ 和 t_c (t_0 为故障发生时刻, t_c 为故障切除时刻) 3 个时刻系统中各发电机大轴间的最大相对摇摆角 δ_i ($i=0, 1, 2$) 及它们的变化率 V_i ($i=1, 2$), 共 5 个变量
39~70: 故障切除时各发电机有功功率和无功功率
71: 故障瞬间发电机发出的有功功率之和
72: 故障切除时发电机发出的有功功率之和
73: 故障切除时所有发电机转子动能的平均值
74: 所有发电机转子初始加速度的均方根误差

仿真实验中, 所有故障设为三相短路, 近端切除故障, 然后远端切除故障, 形成永久故障, 仿真时间为 3 s。设置了 80%, 100%, 110%, 120% 四个负荷水平, 近端-远端切除故障时间分别设定为 0.15~0.2 s, 0.25~0.3 s, 0.3~0.35 s。假设发电机机械功率和故障发生时刻系统中各发电机轴间的最大相对摇摆角保持不变, 消去上述 74 个原始特征中的特征 1 和特征 34, 对剩下的 72 个特征进行分析^[3]。共获得 305 个样本, 将其作为初始特征集数据用于可视化试验。

文献[4]通过特征选取从上述初始特征集中选取特征 4, 12, 14, 16, 21, 31 作为特征选取后特征集。用支持向量机对原始特征和特征选取后的 IEEE16 机暂态稳定数据分别进行分类, 正确率分别为 100% 与 99%^[4]。组成特征选取后的 IEEE16 机暂态稳定数据样本 305 个, 用支持向量机训练样本进行分类, 共有 7 个样本分类错误。剔除错误的分类结果, 剩余 298 个包含 6 特征的数据作为特征选取后的特征集数据用于可视化试验。

以上数据均为分类完毕的二分类数据。每个数据都有一个分类特征: 暂态稳定的数据用 0 表示, 暂态失稳的数据用 1 表示。本文通过可视化技术, 在二维平面上将该分类信息表现出来。

2 可视化算法

数据可视化方法大致分为以下几类^[5]:

1) 对原有的数据维, 在低维空间对其重新排列, 如星型坐标系 (star coordinates) 和平行坐标系 (parallel coordinates)。

2) 采用具有多个可视特征的图标, 用每个可视特征代表一个数据维。

3) 按照所有的数据维对低维空间进行层次划分, 如 dimensional stacking 和 Treemap。

4) 针对数据维的不同组合, 采用多个子窗口分别表示: 如散列图矩阵 (scatterplot matrices)。

5) 高维数据映射到低维空间中, 尽量保持数据间的某种关系不变, 如 PCA (principal component analysis), MDS (multidimensional scaling), SOM (self-organizing map) 等方法。

比较而言, 在对高维复杂数据进行可视化的情况下, 降维映射方法能更好地展现数据结构, 本文采用的 SOM 可视化方法就属于这一类。

2.1 SOM 算法

2.1.1 SOM 概述

自组织映射 SOM (self-organizing map) 是荷兰赫尔辛基大学神经网络专家 Kohonen 教授于 1982

年提出的竞争式神经网络, 因此也称为Kohonen自组织特征映射。它模拟大脑神经元细胞的学习过程, 是一种无监督的学习方法。SOM算法在对数据矢量量化的同时还能实现数据的降维映射, 该映射具有拓扑关系保持的优良性能, 从而广泛应用于降维可视化领域。

SOM网络结构如图1所示, 网络只有输入层和输出层两层。设输入向量 X^r 的维数为 n 维, 则输入层有 n 个节点, 分别为输入向量 X^r 的各个分量; 输出层由 m 个神经元组成二维平面阵列, 其排列方式可以是矩形排列、六边形排列或者随机排列, 如图2所示; 输入层各节点与输出层各神经元之间完全连接^[6]。

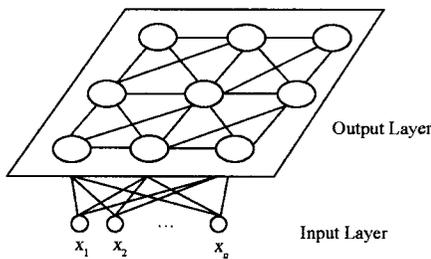


图1 SOM网络基本结构

Fig.1 Basic structure of SOM network

简单地说, SOM就是针对每一个输入向量, 都在低维输出平面上(一般为二维)找出一个点与之对应, 这种对应关系建立以后, 就称该输入向量激活了一个输出神经元。

SOM映射的两个主要目的是^[7]:

- 1) 将任意维的输入向量映射到由有限个输出神经元构成的一维或二维空间。
- 2) 尽可能使相邻的输入向量激活位置相邻的神经元。

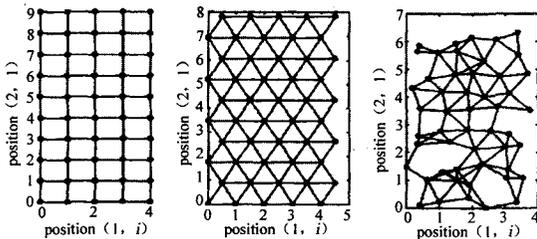


图2 输出神经元排列方式

Fig.2 Arrangement method of output neurons

2.1.2 SOM算法流程

若输入向量为 n 维, 则任一输出神经元 j ($1 \leq j \leq m$) 对应 n 维权向量 $w^j = \{w_{j1}, \dots, w_{jn}\}$ 。具体算法如下:

- 1) 设置邻域半径 $ND(0)$ 的初始值为一较大的

正数, 一般取输出神经元在输出平面上的最大距离; 设定初始学习率 $\alpha(0)$ 的值, 一般为接近1的小数, 例如0.9, 这里0表示第0次训练时的值。 ND 与 α 都是随着训练次数的增加而不断减小的正数。

- 2) 将所有输出神经元的权值 w^j 初始化为随机的小数, 计迭代次数 $t=0$ 。

- 3) 计输入向量的个数 $num=0$ 。

- 4) 随机给定一个输入 X^r , 按照以下公式计算 X^r 与所有输出神经元之间的距离

$$d_{rj} = \sqrt{\sum_{i=1}^n [x_{ri} - w_{ji}]^2}, \quad j \in (1, m)$$

选取获胜神经元 v , 使得 $d_{rv} = \min d_{rj}$ 成立。

- 5) 采用邻域学习方式, 对于获胜神经元邻域以内的所有神经元, 均按下式更新其权重

$$w^j(t+1) = w^j(t) + \alpha(t) \Lambda_{vj}(t) (X^r - w^j(t))$$

其中 $\alpha(t)$ 为第 t 步迭代时的学习率, $\Lambda_{vj}(t)$ 表示第 t 步迭代时的邻域函数, 根据Kohonen的理论^[8], $\Lambda_{vj}(t)$ 的定义有多种, 例如可取 $\Lambda_{vj}(t) = 1$ 。本文

采用如下的形式: $\Lambda_{vj}(t) = e^{-\frac{d_{vj}^2}{2ND(t)^2}}$, 其中 $ND(t)$ 为第 t 步迭代时的邻域半径。

- 6) 令 $num = num + 1$, 如果 $num < MAXnum - 1$, 则跳到第4步, 否则跳到第7步, 其中 $MAXnum$ 为输入向量总的个数。

- 7) 令 $t = t + 1$, ND 与 α 递减。

- 8) 如果 $t < MAXt - 1$, 则跳到第3步, 否则结束, 其中 $MAXt$ 为约定的最大迭代次数。

SOM采用邻域学习的方法, 从而使得输出平面上位置相邻的神经元, 其权值越来越相似; 同时数据相似的输入, 其激活的神经元也是相近的, 这就是SOM映射结果可以保持数据在高维空间的拓扑结构的原理。

3 基于原始SOM的可视化实验

3.1 仿真条件

用Matlab7.1构建SOM网络分别对初始特征集数据(305个, 每个包含72特征)与经过特征选取的数据(298个, 每个包含6特征)进行可视化实验。迭代次数均取1000次, 学习速率从0.9降至0.02。

3.2 初始特征集数据实验

构建SOM网络的结构为 20×20 , 25×25 两种, 以观察算法的参数不同对可视化结果产生的影响。实验结果分别如图3, 图4所示, 图中“o”表示故障发生以后暂态稳定的数据, “x”表示故障发生以后暂

态失稳的数据。

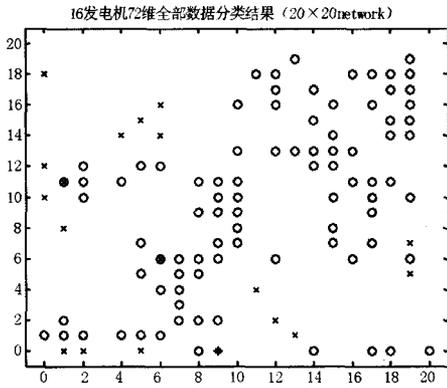


图3 16机系统72维数据分类图 (20×20)

Fig.3 Classified map of 72-dimensions data in 16-generators system (20×20)

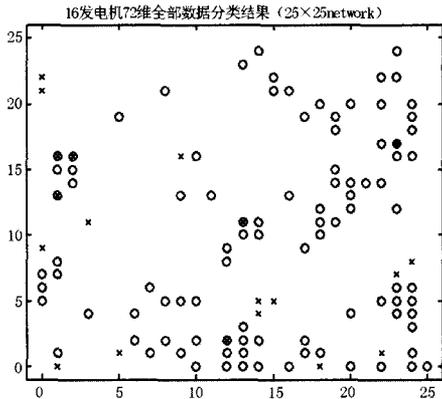


图4 16机系统72维数据分类图 (25×25)

Fig.4 Classified map of 72-dimensions data in 16-generators system (25×25)

从图中可以看出，无论是20×20的网络结构，还是25×25的网络结构，可视化结果均不太理想，某些稳定的点和失稳的点混杂在一起，无法展现出数据的分类信息。

3.3 特征选取后的数据实验

同样构建20×20，25×25两种网络结构。实验结果分别如图5、图6所示，“o”表示故障发生以后暂态稳定的数据，“x”表示故障发生以后暂态失稳的数据。

从图5、图6可以看出，稳定与失稳的点明显地分成两类，边界较清晰。结果基本展示出了分类信息。

3.4 分析与思考

3.4.1 特征选取对于可视化的意义

SOM是一种无监督的算法，可以在低维空间上较好地反映出数据在高维空间的拓扑结构；但是，在降

维映射的过程中，数据结构要产生一定的扭曲。数据维数越高，数据越复杂，降维过程中扭曲的程度也就越严重。72维数据的可视化结果不尽如人意，而经特征提取的6维数据，其可视化结果比较理想。因此，在可视化以前对数据进行特征选取是很有必要的。

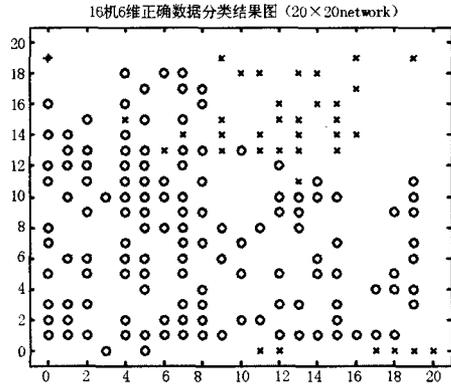


图5 16机系统6维数据分类图 (20×20)

Fig.5 Classified map of 6-dimensions data in 16-generators system (20×20)

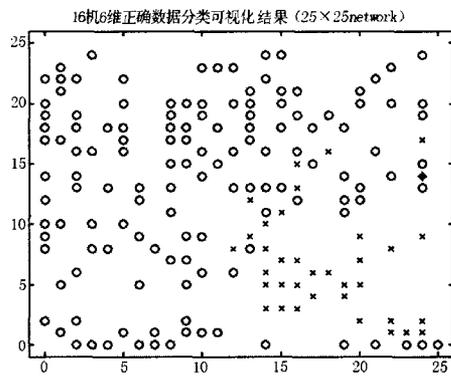


图6 16机系统6维数据分类图 (25×25)

Fig.6 Classified map of 6-dimensions data in 16-generators system (25×25)

3.4.2 算法参数对可视化结果的影响

分析图5、图6发现：图5的分类结果不如图6理想，图5的失稳数据大多分布在左上一块，但有一小部分失稳数据分布在右下，与稳定数据混淆。究其原因，是神经元数目太少所致，因此，算法参数的选择是会直接影响可视化结果的。

3.4.3 原始SOM算法的可行性与局限性

图6的结果相对比较理想，但仍然有多个输入数据激活了同一个神经元的情况。经比较计算后发现，出现这种情况的原因是输入数据相同或相似所致。

1) 例如，第289与第290个输入均激活了神经元(1,8)，而第289与第290个输入数据，其值均为(0.3532, 0.19862, 0.19954, 0.25867, 0.24553,

0.44462), 即这2个数据在高维空间就是重合的。这正好反映了SOM算法分类的正确性, 因此, 用SOM进行可视化实验是可行的。

2) 但是, 有些数据相似的点也激活了同一个点, 如第106个输入 (1.6154, 1.3784, 1.3845, 1.125, -1.032, 0.15183) 与第108个输入 (2.6531, 1.4365, 1.4423, 1.2551, -1.0032, 0.21238) 均激活了神经元 (0, 10)。显然, 这两个输入在原始六维空间中是有一定距离的, 但是间距并不大, 在映射到二维空间以后, 他们本该分布在神经元 (0, 10) 的附近, 但可供选择的神经元只有 (0, 10) 一个, 导致他们最终均激活了这一个点。这个问题的根源是神经元被固定在网格上, 无法移动, 这也是原始SOM算法很大的一个不足。

4 DPSOM算法的引入

4.1 DPSOM算法简介

神经元的位置可调^[9]是目前对原始 SOM 算法改进的热点。相关的改进算法比较多, 如 GNSOM^[10], ACSOM^[11], PNSOM^[9], DPSOM^[5]等, 本文采用的是 DPSOM 算法。它在原始 SOM 算法的基础上, 附加了一步位置调节操作, 而无需添加额外的参数, 且位置调节因子是伸缩的, 从而避免了神经元的过度收缩问题。

算法流程总体与原始 SOM 算法类似, 不过在第5步与第6步之间添加了一步位置调节:

$$p_j(t+1) = p_j(t) + \alpha(t) \left(1 - \frac{d_{vj}}{\gamma_{vj}} \right) \cdot (p_v(t) - p_j(t))$$

其中: $p_j(t)$ 表示神经元 j 在 t 时刻的坐标, $\alpha(t)$ 为 t 时刻的学习率, d_{vj} , γ_{vj} 分别表示神经元 v 与 j 在高维空间的距离 (即代表两个神经元的权值之间的欧式距离) 与在输出平面上的距离。所有参数都可以在原始 SOM 算法中得到。

4.2 基于 DPSOM 算法的可视化实验

对经过特征选取的数据, 构建SOM网络为 25×25 , 迭代次数取1000次, 学习速率从0.9降至0.02。实验结果如图7所示。

和图6比较, 可以发现结果有明显改善。两类数据的分类信息更加清晰, 而且那些数据相似的点, 现在终于不再激活同一个点了。观察第106和第108个输入数据发现: 106激活的是 (1.03, 22.61), 而108激活的是 (0.68, 20.91)。

另外, 用DPSOM算法对初始特征集数据 (72特征) 进行可视化, 发现结果也得到了很大的改善, 如图8所示。

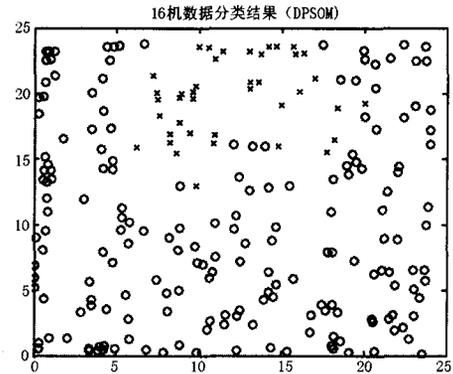


图7 16机6维数据分类图byDPSOM (25×25)

Fig.7 Classified map of 6-dimensions data in 16-generators system by DPSOM (25×25)

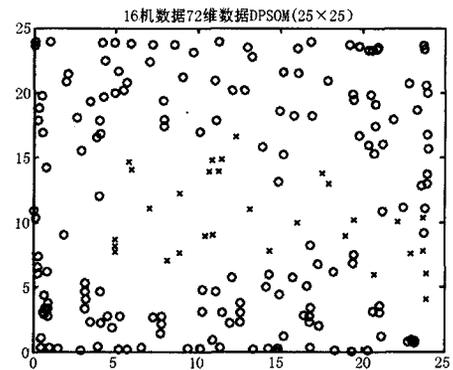


图8 16机72维数据分类图byDPSOM (25×25)

Fig.8 Classified map of 72-dimensions data in 16-generators system by DPSOM (25×25)

可见, DPSOM算法对于可视化结果的改进, 其帮助是相当大的。可以说, 现在的结果更加真实地反映了数据在高维空间的距离信息和拓扑结构。

4.3 基于DPSOM的可视化步骤

综合上面的工作, 可以提出针对电力系统暂态稳定评估结果的可视化步骤:

1) 对原始暂态稳定数据, 采用特征提取技术提取特征子集, 该特征子集必须满足两个条件: 一是可以尽量表征原数据的所有状态; 二是特征的个数尽可能地少, 以减少降维映射中的畸变。

2) 以特征提取后的数据作为输入, 采用DPSOM算法, 设定合适的参数进行可视化。

5 结论

本文采用经典的SOM算法对IEEE16机系统暂态稳定数据进行降维可视化, 对原始数据与特征提取后的数据分别实验, 验证了SOM算法用于研究的可行性, 以及特征提取对于可视化的重要性, 同时分析了原始SOM算法的局限性。在此基础上, 采用一

种新的基于 SOM 的算法——DPSOM 算法进行实验, 并成功地得到了比较满意的可视化结果: 不但分类结果更加完善, 而且将原始 SOM 算法不能区分的点分离开来, 从而达到了可视化的目的, 为电力系统暂态稳定评估提供了支持。

参考文献

[1] 李军, 刘艳, 顾雪平. 基于信息熵的属性离散化算法在暂态稳定评估中的应用[J]. 电力系统自动化, 2005,29(8): 26-31.
LI Jun, LIU Yan, GU Xue-ping. An Entropy-based Discretization Method Applied in Transient Stability Assessment[J]. Automation of Electric Power Systems, 2005, 29 (8): 26-31.

[2] 王成山, 许晓菲, 余贻鑫. 基于割基功率空间上的静态电压稳定域局部可视化方法[J]. 中国电机工程学报, 2004, 24 (9): 13-18.
WANG Cheng-shan, XU Xiao-fei, YU Yi-xin. Visualization of Power System Static Voltage Stability Region in Cut-set Space[J]. Proceedings of the CSEE, 2004, 24 (9): 13-18.

[3] 方怡, 王君, 王晓茹. 暂态稳定评估的数据挖掘预处理[J]. 机电工程, 2007, 24 (10): 46-50.
FANG Yi, WANG Jun, WANG Xiao-ru. Data Preprocess of Association Rules Based on Transient Stability Assessment[J]. Mechanical & Electrical Engineering Magazine, 2007, 24(10): 46-50.

[4] 向丽萍, 王晓红, 王建. 基于支持向量机的暂态稳定分类中的特征选择[J]. 继电器, 2007, 35 (9): 17-21.
XIANG Li-ping, WANG Xiao-hong, WANG Jian. Feature Selection for SVM Based Transient Stability Classification[J]. Relay, 2007, 35 (9): 17-21.

[5] 邵超, 黄厚宽. 一种新的基于SOM的数据可视化算法[J]. 计算机研究与发展, 2006, 43 (3): 429-435.
SHAO Chao, HUANG Hou-kuan. A New Data Visualization Algorithm Based on SOM[J]. Journal of Computer Research and Development, 2006, 43 (3): 429-435.

[6] 涂晓芝, 颜学峰, 钱锋. 基于SOM网络的基因表达数据聚类分析[J]. 华东理工大学学报(自然科学版), 2006, 32 (8): 992-996.
TU Xiao-zhi, YAN Xue-feng, QIAN feng. Cluster Analysis of Gene Expression Data Based on SOM Network[J]. Journal of East China University of Science and Technology(Natural Science Edition), 2006, 32 (8): 992-996.

[7] 白耀辉, 陈明. 利用自组织特征映射神经网络进行可视化聚类[J]. 计算机仿真, 2006, 23 (1): 180-183.
BAI Yao-hui, CHEN ming. Visualizing Clusters Based on the Self-Organizing Map[J]. Computer Simulation, 2006, 23 (1): 180-183.

[8] Kohonen T. The Self-organizing map[J]. Proceedings of the IEEE, 1990, 78(9): 1464-1480.

[9] Shao C, Huang H K. Improvement of Data Visualization Based on SOM[A]. In: Proc. Int'l Symposium on Neural Networks[C]. Berlin: Springer-Verlag, 2004. 707-712.

[10] Merk D, Rauber A. Alternative Ways for Cluster Visualization in Self-organizing Maps[A]. In: The Workshop on Self-Organizing Maps(Wsom'97)[C]. Helsinki University of Technology, 1997.

[11] Wang D, Ressom H, Musavi H, et al. Double Self-organizing Maps to Cluster gene Expression Data[A]. In: The 10th European Symposium on Artificial Neural Networks (ESANN'02) [C]. Bruges: 2002.

收稿日期: 2008-05-10; 修回日期: 2008-06-08

作者简介:

李亦宁(1982-), 男, 硕士研究生, 研究方向为数据挖掘与电力系统紧急控制; E-mail: liyning84@163.com
王晓茹(1962-), 女, 教授, 博士生导师, 主要研究方向为电力系统保护与安全稳定控制、变电站自动化;
向丽萍(1981-), 女, 硕士研究生, 研究方向为数据挖掘与电力系统紧急控制。

(上接第 13 页 continued from page 13)

XU Zheng. Steady State Characteristics Analysis of Thyristor Controlled Series Compensation[J]. Power Electronics, 1998,(2):32-35.

[3] 徐政. 可控串补的可控范围及串容与电抗关系的研究[J]. 电工技术学报, 1998, 13(3):13-17.
XU Zheng. The Controllable Impedance Range of TCSC and the Parameter Selection Method of Its TCR[J]. Transactions of China Electrotechnical Society, 1998, 13(3):13-17.

[4] 胡国文, 王仲鸿, 韩英铎. 可控串补基频等效阻抗与TCR基频电抗关系的仿真和动模实验研究[J]. 电工技术学报, 2002, 17(4):93-99.
HU Guo-wen, WANG Zhong-hong, HAN Ying-duo. Research of the Numerical Emulation and Dynamic Simulation of the Relation Between the Basic Frequency

Equivalent Impedance of the TCSC and Basic Frequency Reactance of the TCR[J]. Transactions of China Electrotechnical Society, 2002, 17(4):93-99.

[5] 韩彦华, 李琥, 施围. 可控串补等值阻抗的计算方法研究[J]. 电工技术学报, 2003, 18(11):96-99.
HAN Yan-hua, LI Hu, SHI Wei. An Approach to Calculating the Equivalent Impedance of TCSC[J]. Transactions of China Electrotechnical Society, 2003, 18(11):96-99.

收稿日期: 2008-04-22; 修回日期: 2008-06-22

作者简介:

王小立(1980-), 男, 硕士, 主要从事电网调度运行工作; E-mail: wxl_972116@163.com
张一工(1956-), 男, 教授, 主要研究方向为电力电子技术应用。