

基于快速密度聚类的电力通信网节点重要性评估

狄立¹, 郑征¹, 夏旻², 胡凯²

(1. 国网河南省电力公司经济技术研究院, 河南 郑州 450052; 2. 江苏省大数据分析技术重点实验室, 南京信息工程大学, 江苏 南京 210044)

摘要: 电力通信网的节点重要性评估是电力通信研究的一个重要议题。针对目前电力通信网节点重要性评估存在的连接权值单一以及评价指标单一等问题, 利用电力通信网的带宽和距离作为权值, 计算电力通信网节点的多种评价指标: 节点强度、节点紧密度以及节点的介数。基于电力通信网节点的多种评价指标, 利用快速密度聚类方法建立电力通信网的节点重要性评估模型, 为电网通信的规划做支撑。通过快速密度聚类方法进行无监督的分类, 将节点分为若干个重要性等级。该方法可以有效地改善基于距离的无监督分类方法的不足。利用某省的实际电网通信数据进行检验, 验证了该方法在电力通信网中的实用性。

关键词: 电力通信网; 节点重要性; 快速密度聚类; 无监督分类; 节点特性

Node importance evaluation of electric power communication network based on fast density clustering

DI Li¹, ZHENG Zheng¹, XIA Min², HU Kai²

(1. Economics and Technology Research Institute of State Grid Henan Electric Power Company, Zhengzhou 450052, China; 2. Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: Node importance evaluation of electric power communication network is an important topic of power communication studies. For the problem of single connection weights and single evaluation index, the bandwidth and the distance of the electric power communication network are used for the network's weight, computing the evaluating indicators of power communication network node, including joint strength, precision and node betweenness. Evaluation based on a variety of power communication network node and the fast density clustering method are used to establish the importance of the assessment model for power node communication network, supporting for the planning of grid communications. Using unsupervised classification by fast density clustering method, the node is divided into several fragile levels. This method can effectively improve the unsupervised classification method. The actual data grid communications are tested to verify the usefulness of this method in electric power communication network.

This work is supported by National Natural Science Foundation of China (No. 61105115).

Key words: electric power communication networks; node importance; fast density clustering; unsupervised classification; node characteristic

0 引言

电力通信网是专门为电力行业服务的专用网络, 由电厂、变电站以及各级电力系统相互通信传输系统组成, 目前已经成为电网的重要组成部分。由于电力系统的不断发展壮大, 电力通信网承载的业务量也变得越来越, 增长速度惊人。随着电力

通信网络规模的不断扩大, 电力通信网的安全性、可靠性研究也变得越来越重要^[1]。而电力通信网中的重要节点在很大程度上决定了电力通信网的稳定性和安全性^[2-4], 对重要节点的重点维护和管理成为电力通信网的一个重要任务^[5]。因此, 如何在复杂的电力通信网中找到关键节点成为电力通信网研究的一个重要议题。

电力通信网中的重要节点决定了网络的稳定性和安全性, 电力通信网中节点的重要性表明了该节

点在可能的故障情况下对网络性能的影响,使用定量的方法来对网络节点进行评估已经成为电力通信网研究的热点。目前国内外对电力通信网节点的重要性评估主要使用以下几种方法^[6-10]:1)利用节点的度作为节点重要性的指标。该方法中某节点连接的边越多,那么该节点的重要性就越大。2)基于节点删除的评价。该方法通过删除某个节点,看网络连通性的变化来分析节点的重要性。3)基于节点的介数的评价,如果该节点的介数越大那么该节点越重要。4)通过节点紧密度来对节点进行评价,紧密度越大则节点的重要性越强。目前的电力通信网节点重要性评估主要采用以上几种方法,但是以上的方法都具有有一些缺陷^[11-12]:以节点的度作为评价指标存在一定的片面性,例如桥节点的度虽然不大,但重要性却很强;基于节点删除的方式中如果删除的是末梢节点,那么就不能实现节点重要性的客观评估;介数和紧密度体现的是节点对于全局的影响。节点介数定义为网络中所有最短路径中经过该节点的路径的数目占最短路径总数的比例。紧密度指标 C 用于刻画网络中的节点通过网络到达其他节点的难易程度。介数和紧密度没有体现出该节点对局域网络的影响,因此基于节点的介数和紧密度的方法的评估不能全面反映节点的重要性。而节点度的指标能够刻画节点对局部网络的影响。因此,本文中综合利用节点的度、介数和紧密度作为节点重要性的评价指标。

目前有关电力通信网节点重要性评估的工作基本上都是以单一连接权值为主,将距离和带宽都考虑进评估模型的还没有^[13-14]。而且以往的电力通信网的节点重要性评估往往只利用单个的评价指标,而没有有效地综合多种评价指标,并且很多方法只针对网络结构本身,而忽略了网络的连接权重^[15-16]。针对目前电力通信网评估存在的各种问题,本文利用网络带宽和距离作为加权,通过多种评价指标的综合来对节点的重要性进行评估。采用多指标的方式进行节点重要性的评估分类,并且这种分类是一种无监督的分类^[17]。目前针对电力通信网节点重要性评估的无监督分类主要为K-means方法。K-means方法为无监督分类中的主要传统方法,但是因为该方法基于距离,故结果是圆形的聚类形状,并不能很好地对复杂特征的融合进行分类。针对该问题,本文采用了一种快速密度聚类的方法^[18],该方法发表在2014年的《science》期刊上,该方法不仅利用距离作为分类的度量,还利用密度作为分类的指标,有效解决了K-means方法的不足。在加权网络的基础上,本文基于快速密度聚类的方法,采用节点度、

节点紧密度以及节点的介数作为评价指标,对电力通信网的节点重要性进行评估。利用某省的实际电网通信数据进行检验,验证了本文方法在电力通信网中的实用性。

1 节点重要性相关指标分析

节点数为 n ,边数为 m 的通信网有权网络数学模型可以用一个加权矩阵来描述。

$$W_G = \begin{pmatrix} w_{G11} & \cdots & w_{G1n} \\ \vdots & & \vdots \\ w_{Gm1} & \cdots & w_{Gmn} \end{pmatrix} \quad (1)$$

边权邻接矩阵的矩阵元为:如果节点 i 和节点 j 之间有连接,那么 $w_{Gij} = B_{ij}$;如果节点 i 和节点 j 之间没有连接,就有 $w_{Gij} = \infty$ 。式中, B_{ij} 为节点 i 与节点 j 之间线路的权值。边权代表节点间相互作用的强度或难易程度。如果节点之间的距离作为权值,权值越大表示两点间的距离越大,作用越弱。如果节点之间的带宽作为权值,则权值越大作用越大。在本研究中将距离和带宽的权值进行归一化整合,距离和带宽的权值各为50%。

网络中节点的重要性不但与节点本身的度有一定关系,而且与节点的邻居节点的度大小、节点的紧密度、节点的介数也存在一定的关联,这些指标体现了节点的影响力。节点的影响力包括直接影响力和间接影响力。直接影响力反映节点本身影响其他节点的能力,如节点的度。间接影响力反映节点通过网络对其他节点施加影响的能力,如紧密度和介数。本文利用节点度、节点紧密度、节点的介数作为节点重要性的评价指标。

节点的度:节点 i 的度数 k_i 是指与该节点直接连接的边数,反映的是一个节点对于网络中其他节点的直接影响力。具有 n 个节点的网络中,节点的度不会超过 $n-1$ 。通常情况下,加权后节点 i 的强度 S_i 定义为

$$S_i = k_i / \max(k_j), j=1, 2, \dots, n \quad (2)$$

节点的紧密度:紧密度指标 C 用于刻画网络中的节点通过网络到达其他节点的难易程度,反映节点的间接影响力。其值定义为该节点到达所有其他节点的距离之和的倒数。具有 n 个节点的网络中,节点到达所有其他节点的距离之和不会小于 $n-1$,则归一化的紧密度指标 C_i 为

$$C_i = \min(d) / \sum_{j=1}^N d_{ij} \quad (3)$$

对于加权网络,两节点间的距离 d_{ij} 一般定义为连接两节点间的最短路径上边的权值之和。

介数：选取复杂网络理论中合适的静态几何参数进行网络拓扑脆弱性的计算。本文采用的是介数指标。节点介数定义为网络中所有最短路径中经过该节点的路径的数目占最短路径总数的比例。每个节点的介数用参数 b_i 表示，反映节点在网络连通中的重要程度。 b_i 也反映了节点拓扑脆弱性， V_i 表示节点在整个网络中的脆弱程度 $V_i = \frac{b_i}{b_{\max}} \times 100\%$ 。

$\frac{b_i}{b_{\max}}$ 代表的是节点的相对拓扑重要度，测度了故障造成的潜在影响程度，能够直接找到最脆弱点。

本文中以中国某省的电网通信实际数据作为分析来源，实现节点重要性的评估。

2 快速密度聚类

针对目前非监督聚类存在的一些问题，本文采用快速密度聚类的方法来进行电力通信网的节点重要性评估分类。快速密度聚类算法的核心思想是：类簇中心被具有较低局部密度的邻居点包围，且与具有更高密度的任何点有相对较大的距离。基于这种思想，对于任意数据点 i ，需要计算两个量：局部密度值 ρ_i 以及点 i 到具有更高局部密度点的距离 δ_i 。 ρ_i 和 δ_i 的值都与数据点之间的距离 d_{ij} 有关(d_{ij} 代表点 i 和点 j 之间的距离，本文中用欧式距离表示)。任意数据点 i 的局部密度 ρ_i 定义如式(4)。

$$\rho_i = \sum_j \chi(d_{ij} - d_c) \quad (4)$$

其中， d_c 是截断距离，如果 $d_{ij} - d_c < 0$ ，那么 $\chi(d_{ij} - d_c) = 1$ ，如果 $d_{ij} - d_c \geq 0$ ，那么就有 $\chi(d_{ij} - d_c) = 0$ 。数据点 i 的局部密度 ρ_i 本质上为数据点 i 距离为 d_c 内的点的个数。实验表明分类结果对 d_c 的大小选择具有很高的鲁棒性， d_c 的大小对分类结果影响不大。在本文算法中对于 d_c 的选择规则为：使得点的平均邻居数大概是数据集中点的总数的 1%~2%。点 i 到具有更高局部密度点的距离 δ_i 定义为

$$\delta_i = \min_{j: \rho_j > \rho_i} (d_{ij}) \quad (5)$$

在本算法中具有高 ρ_i 和相对较高的 δ_i 的点可被认为是类簇中心，类簇中心找到后，剩余的每个点被归属到它的有更高密度的最近邻所属类簇。类簇分配只需一步即可完成，不像其他算法要对目标函数进行迭代优化。图 1 中的简单示例展示了算法的核心思想，下图为样本点的聚类，上图为聚类中心的寻找。

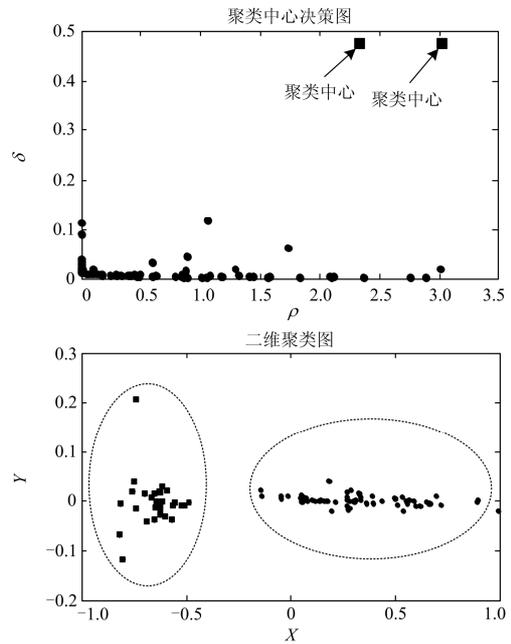


图 1 算法实例展示

Fig. 1 Algorithm examples figure

3 节点重要性分类结果

本文利用节点度、节点紧密度、节点的介数作为节点重要性的评价指标。节点度的大小非常直观，反映了节点本身影响其他节点的能力。一般情况，度越大的节点重要性越大，但是对于某些桥节点来说不正确。间接影响力反映节点通过网络对其他节点施加影响的能力，如紧密度和介数。本文利用节点度、节点紧密度、节点的介数这三个要素，基于快速密度聚类方法对通信网节点的重要性进行评估。本文中的要素为3个，那么分类空间为三维。在分类过程中，点 i 和点 j 之间的距离 d_{ij} 利用欧式距离表示。然后按照式(4)和式(5)分别计算节点 $i(i=1, 2, 3, \dots, n)$ 的局部密度值 ρ_i 以及点 i 到具有更高局部密度点的距离 δ_i 。然后寻找 ρ 和 δ 都大的节点作为聚类中心。本文中利用 $\rho \cdot \delta$ 来作为寻找聚类中心的依据。由于本文将节点的重要性分为4类，因此本文寻找 $\rho \cdot \delta$ 最大的4个点作为聚类中心，然后进行聚类。

图2是网络节点的紧密度分析，从图2可以看出，紧密度大的点它的度未必很大，但是该点在数据传输中的作用却很大。图3为网络节点的脆弱性分析，节点的脆弱性代表了该节点在网络连通方面的能力，如果连通能力强，那么该节点遭到损毁后对网络的影响就越大，即认为脆弱性强。从图3可以看出，节点的脆弱性并不和节点的度分布和紧密度成完全的正相关关系。

图 7 给出了电力通信网骨干网的节点重要性分类结果，从图中可以很直观地看出我们采用的方法可以比较好地对重要节点进行分类。在本研究中将距离和带宽的权值进行归一化整合，距离和带宽的权值各为 50%。因此节点重要性的评估能够反映出距离和带宽的影响。对于重要性一般的节点和重要性比较弱的节点来说，由于本文中采用了距离和带宽作为权值的一部分，因此一般重要性节点在距离和带宽的加权值比重要性比较弱的节点大。这也就是图 7 中一般重要性节点和重要性比较弱的节点的区别。

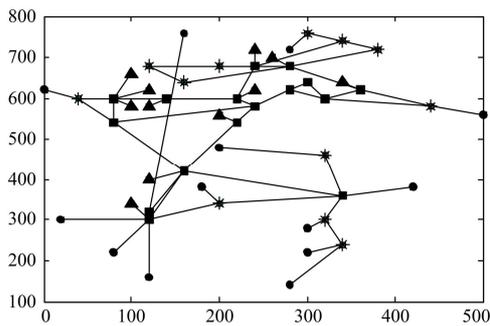


图 7 电力通信网节点的重要性分类图

Fig. 7 Importance classification diagram of power communication network

4 结论

电力通信网的节点重要性评估是电力通信研究的一个重要议题^[19]。本文针对以往电力通信网节点重要性评估中存在使用单一因素的问题，采用节点强度、节点精密度以及节点的介数作为节点重要性的评价指标，并以电力通信网的带宽和距离作为权值，基于快速密度聚类方法建立电力通信网的节点重要性评估模型，为电网通信的规划做支撑。本文通过快速密度聚类方法进行无监督的分类，将节点分为 4 个重要性等级。本文利用某省的实际电网通信数据进行检验，验证了本文方法在电力通信网中的实用性，证明本文方法可以有效地应用于实际电力通信网评估当中。

参考文献

[1] 李瑞生. 一种基于虚拟节点网络拓扑结构适用于架空线路主动配电网的纵联保护方案[J]. 电力系统保护与控制, 2015, 43(2): 70-75.
LI Ruisheng. An aerial line active distribution network pilot protection scheme based on topology of virtual node[J]. Power System Protection and Control, 2015, 43(2): 70-75.

[2] 薛浩然, 张珂珩, 李斌, 等. 基于布谷鸟算法和支持向量机的变压器故障诊断[J]. 电力系统保护与控制, 2015, 43(8): 8-13.
XUE Haoran, ZHANG Keheng, LI Bin, et al. Fault diagnosis of transformer based on the cuckoo search and support vector machine[J]. Power System Protection and Control, 2015, 43(8): 8-13.

[3] 李蓉蓉, 陈曦, 吴延琳, 等. 基于状态转移的高压直流输电系统可靠性分析[J]. 高压电器, 2015(12): 66-71.
LI Rongrong, CHEN Xi, WU Yanlin, et al. Reliability analysis of HVDC transmission system based on state transition[J]. High Voltage Apparatus, 2015(12): 66-71.

[4] 班淑珍, 韩俊峰. 改进关联规则方法在电力设备故障预测中的应用[J]. 电网与清洁能源, 2015, 31(10): 83-88.
BAN Shuzhen, HAN Junfeng. Application of improved association rules method in prediction of electric power equipment faults[J]. Power System and Clean Energy, 2015, 31(10): 83-88.

[5] 国家能源局. DL/T 光纤通道传输保护信息通用技术条件[S]. 2010.

[6] GUNGOR V C, LAMBERT F C. A survey on communication networks for electric system automation[J]. Computer Networks, 2006, 50(7): 877-897.

[7] GAO Huisheng, RAN Jingxue, SUN Yiqun. Risk evaluation of communication network of electric power based on restriction coefficient[C] // International Conference on Wireless Communications, Networking and Mobile Computing, 2007: 3127-3134.

[8] 谢琼瑶, 邓长虹, 赵红生, 等. 基于有权网络模型的电力网节点重要度评估[J]. 电力系统自动化, 2009, 33(4): 21-24.
XIE Qiongyao, DENG Changhong, ZHAO Hongsheng, et al. Evaluation method for node importance of power grid based on the weighted network model[J]. Automation of Electric Power Systems, 2009, 33(4): 21-24.

[9] 张恺伦, 江全元. 基于攻击树模型的 WAMS 通信系统脆弱性评估[J]. 电力系统保护与控制, 2013, 41(7): 116-122.
ZHANG Kailun, JIANG Quanyuan. Vulnerability assessment on WAMS communication system based on attack tree model[J]. Power System Protection and Control, 2013, 41(7): 116-122.

[10] 蒋康明, 曾瑛, 邓博仁, 等. 基于业务的电力通信网风险评价方法[J]. 电力系统保护与控制, 2013, 41(24): 101-106.
JIANG Kangming, ZENG Ying, DENG Boren, et al. Risk evaluation method of electric power communication

- network based on services[J]. Power System Protection and Control, 2013, 41(24): 101-106.
- [11] 郭静, 王东蕊. 基于复杂网络理论的电力通信网脆弱性分析[J]. 电力系统通信, 2009, 30(9): 6-10.
GUO Jing, WANG Dongrui. Vulnerability analysis on power communication network based on complex network theory[J]. Telecommunications for Electric Power System, 2009, 30(9): 6-10.
- [12] 程文清, 马庆峰, 赵建之. 电力光纤传输网可靠性评估方法的研究[J]. 电力系统通信, 2011, 32(9): 11-15.
CHENG Wenqing, MA Qingfeng, ZHAO Jianzhi. Research on the reliability evaluation of power optical fiber transmission network[J]. Telecommunications for Electric Power System, 2011, 32(9): 11-15.
- [13] 高会生, 王旭蕊, 王慧芳. 一种基于贝叶斯网络的电力通信网 SDH 设备综合安全评估模型[J]. 现代电力, 2014, 31(3): 68-73.
GAO Huisheng, WANG Xurui, WANG Huifang. A comprehensive security evaluation model of SDH equipments in power communication network based on Bayesian network[J]. Modern Electric Power, 2014, 31(3): 68-73.
- [14] 梅鲁海. 基于 SDH 光网络的分层区域式保护通信系统的可靠性研究[J]. 电力系统保护与控制, 2014, 42(21): 81-85.
MEI Luhai. Reliability research of layered regional protective communication system based on SDH optical network[J]. Power System Protection and Control, 2014, 42(21): 81-85.
- [15] 蔡泽祥, 王星华, 任晓娜. 复杂网络理论及其在电力系统中的应用研究综述[J]. 电网技术, 2012, 32(11): 114-119.
CAI Zexiang, WANG Xinghua, REN Xiaona. A review of complex network theory and its application in power systems[J]. Power System Technology, 2012, 32(11): 114-119.
- [16] 杨蕾, 黄小庆, 曹丽华, 等. 考虑区域性的复杂电力网络演化模型[J]. 电力系统及其自动化学报, 2012, 24(2): 5-10.
YANG Lei, HUANG Xiaoqing, CAO Lihua. Novel evolving network model for complex power grid considering rationality[J]. Proceedings of the CSU-EPSA, 2012, 24(2): 5-10.
- [17] HE Xi, WANG Wei, LIU Xinyu, et al. Risk assessment of communication network of power company based on rough set theory and multiclass SVM[J]. Physics Procedia, 2012, 24: 1226-1231.
- [18] ALEX R, LAIO A. Clustering by fast search and find of density peaks[J]. Science, 2014: 1492-1496.
- [19] 姚致清, 于飞, 赵倩, 等. 基于模块化多电平换流器的大型光伏并网系统仿真研究[J]. 中国电机工程学报, 2013, 33(36): 27-33.
YAO Zhiqing, YU Fei, ZHAO Qian, et al. Simulation research on large-scale PV grid-connected systems based on MMC[J]. Proceedings of the CSEE, 2013, 33(36): 27-33.

收稿日期: 2015-08-06; 修回日期: 2015-10-19

作者简介:

狄立(1987-), 男, 硕士, 工程师, 研究方向为电力通信网络评估; E-mail: 137886399@qq.com

郑征(1979-), 男, 硕士, 工程师, 研究方向为电力通信系统规划等; E-mail: 342100300@qq.com

夏旻(1983-), 男, 通信作者, 博士, 副教授, 研究方向为大数据分析及机器学习理论, 电力通信网评估等. E-mail: xiamin@nuist.edu.cn

(编辑 葛艳娜)