



基于自动聚类模型的输电线路外力破坏预警预测

马大燕

(国网电子商务有限公司, 北京 100053)

摘要: 外力破坏事件已成为严重威胁架空输电线路安全稳定运行的主要因素, 给防御、预警工作带来一定的困难。针对传统的聚类方法聚类中心难以准确确定、易受异常点影响的问题, 提出了一种基于自动聚类模型的输电线路外力破坏数据分析方法, 对外力破坏数据从时间和空间纬度进行分析。该算法首先通过 Canopy 算法初始聚类中心, 采用削弱不符合正态分布的异常数据权值的思想, 利用优化的 K -means 算法进行聚类处理, 最终通过实验分析证明了该算法的有效性、高效性。本文算法能够应用于电力信息系统的 GIS 模块, 实现分析结果的时空可视化, 为找到输电线路外力破坏发生原因、进行预警预测提供有力的决策支持。

关键词: 外力破坏; 自动聚类; Canopy; K -means; 数据分析

中图分类号: TM715

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019053

Early warning prediction of external force destruction in transmission lines based on automatic clustering model

MA Dayan

State Grid Electronic Commerce Co., Ltd., Beijing 100053, China

Abstract: The external force destruction has become a major threat to the safe and stable operation of overhead transmission lines, bringing difficulties to the defense and early warning work. In order to solve the problem that the traditional clustering center is difficult to accurately determined and susceptible to abnormal points, an automatic clustering method for data analysis work of transmission lines was presented, and external damage data was analyzed from time and space latitude. Firstly, the cluster center was initialized in this method by using Canopy algorithm. Then, the optimized K -means algorithm was used to perform clustering. Finally, the effectiveness of this method was proved by experimental analysis. This method will be applied to the GIS module in the power information system, which can realize the spatio-temporal visualization of the analysis results and provide powerful decision support for finding cause of the external force damage of the transmission line.

Key words: external force destruction, automatic clustering, Canopy, K -means, data analysis

收稿日期: 2018-12-12; 修回日期: 2019-02-28

基金项目: 国家电网公司科技项目 (国网分布式光伏云网深化研究与应用)

Foundation Item: Science and Technology Project of State Grid Corporation of China (Deepening Research and Application of PVCloud)



1 引言

输电线路是电力系统的重要组成部分,具有线路距离长、覆盖范围大、安全可靠要求高等特点^[1-3]。随着国家生态环境保护力度的增大及城镇化建设进程的加快,国土植被覆盖范围不断扩大、施工作业数量增多且分布面积增加,架空输电线路通道内树竹、山火、施工机械“三大”外部隐患显著上升,由此导致的跳闸、被迫停电等事故发生的概率呈上升趋势。据统计,全国近 1/3 的停电事故都是由外力破坏引起的,因此可见,外破事件已成为严重威胁架空输电线路安全稳定运行的主要因素^[4-5]。

针对架空输电线路因外力破坏导致的跳闸以及由此造成故障停运率居高不下的问题,国内线路防外力破坏工作分别从技术和管理两方面着手。管理上输电单位采取了“政企联合”“企企联合”,建立群众护线网络等一系列的组织措施^[6],建立以设备管理人员为主体,由乡镇负责人、属地供电所人员、林业管理人员等共同参与的群众护线网络,同时加大电力犯罪打击力度,推动电力执法建设。技术上主要通过视频实时监控、防范输电通道内各种潜在的外力破坏风险隐患^[7-8],但是,由于输电线路分布广阔,全程进行监控难度大,同时线路中缺乏有效的预警预测功能,实际应用效果一般。而在具体的防输电线路受外力破坏工作中,投入大量的人力巡视、值守等仍是主要的护电手段,由此不仅带来人员需求量大、工作效率低等一系列问题,也为护电工作的有效开展造成了被动局面。因此开展输电线路外破故障的预警预测,可显著提高供电可靠性,并将提升电网运行智能化水平^[9-11]。

本文实现了一种自动聚类模型,并将其应用于外力破坏事件分析工作中,解决了 *K-means* 算法中初始聚类个数难以确定、仅适用于凸形状的簇即对于异常点(离群值)较为敏感等问题^[12-13],

并以北京地区历史输电线路外力破坏数据为例,从时间和空间纬度进行分析,最终通过实验证明了本文算法的有效性。本文算法实现了对输电线路、走廊进行各类外力破坏数据的有效分析,为预测预警工作提供了决策支持。

2 自动聚类的模型介绍

本文将输电线路的外力破坏故障主要分为施工撞线、异物短路、异物缺陷 3 方面,即本文算法主要针对该 3 类历史外破数据进行分析。首先,对输电线路外破历史数据进行抽取和清理,删除缺失或不规范的数据,根据线路及杆塔的 GIS 信息定位设备坐标并作为算法的样本输入;然后,采用 *Canopy* 算法计算初始聚类中心,作为 *K-means* 算法的初始聚类中心;同时,由于 *K-means* 算法对于孤立点较为敏感,因此采用优化的 *K-means* 算法对数据进行聚类,获得更加准确的输电线路外破数据分析结果。本文算法流程如图 1 所示。

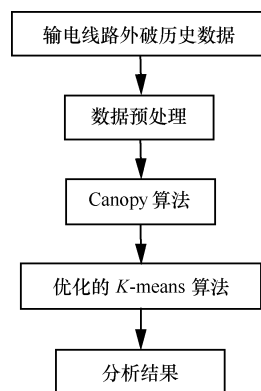


图 1 自动聚类算法流程

2.1 预处理

在进行数据分析之前,需要对数据进行预处理操作。由于外力破坏数据多是工作人员手动记录的数据,因此主要需要解决数据缺失和数据内容不规范问题。前者的处理方式是直接删除缺失信息的数据;后者的问题主要是由于业务人员对于输电线路与杆塔信息输入不规范,从而无法从

国家电网的统一 GIS 平台定位到具体的杆塔位置,即不能够定位到外力破坏事件发生的具体位置,因此处理的方式主要是人工校正或直接删除。

2.2 Canopy 算法

与传统的聚类算法不同, K -means 算法、Canopy 算法最大的特点是不需要事先指定 K 值,即聚类中心的个数,仅将距离阈值作为初始参数,更符合实际应用需要,因此应用场景较多。Canopy 算法的主要思想是把聚类分为两个阶段:阶段一,通过使用一个简单、快捷的距离计算方法把数据(样本)分为可重叠的子集,称为“Canopy”;阶段二,通过使用一个精准、严密的距离计算方法来计算出现在阶段一中同一个 Canopy 的所有数据向量的距离。Canopy 算法的特别之处在于使用了两种距离计算方式,同时由于只计算了重叠部分的数据向量,所以达到了减少计算量的目的。

具体来说,Canopy 算法的步骤如下。

步骤 1 将数据集向量化得到一个 List (列表)后放入内存,选择两个距离阈值: T_1 和 T_2 , 其中 $T_1 > T_2$ (T_1 和 T_2 的值可以用交叉校验来确定)。

步骤 2 从 List 中任取一点 P , 用低计算成本方法快速计算点 P 与所有 Canopy 之间的距离。如果不存在 Canopy, 将点 P 作为一个 Canopy; 如果点 P 与某个 Canopy 距离在 T_1 以内, 则将点 P 加入这个 Canopy。

步骤 3 如果点 P 与某个 Canopy 的距离在 T_2 以外, 则需要把点 P 从 List 中删除, 此时认为点 P 与这个 Canopy 较近, 所以不会加入其他 Canopy。

步骤 4 重复步骤 2、步骤 3, 直到 List 为空时结束。

Canopy 算法是一种“粗”聚类, 执行速度快, 但精度较低, 本文用于 K -means 算法进行“细”聚类操作之前, 缓解了 K -means 算法对于初始聚类中心点敏感的问题。

2.3 优化的 K -means 算法

K -means 算法是一种常用的基于划分的聚类算法, 其扩展性较好, 面对大规模数据集, 具有较高的效率与准确度。 K -means 算法以 K 为参数, 把数据样本分成 K 个簇, 使簇内具有较高的相似度, 而簇间的相似度较低。 K -means 的处理过程如下。

步骤 1 首先, 随机选择 K 个样本作为初始 K 个簇的中心(聚类中心)。

步骤 2 将其余样本根据其与其各个簇的中心的距离分配到最近的簇。

步骤 3 计算各个簇的中心。

步骤 4 重复步骤 2、步骤 3, 直到达到某个终止条件。

聚类中心由式(1)求得:

$$z_j = \frac{1}{M_j} \sum_{i=1}^{M_j} x_i \quad (1)$$

其中, j 表示类别, z_j 表示聚类中心, M_j 表示属于 j 类的数据点的个数, i 表示 j 类中的一个样本索引。式(1)表示属于某个簇的所有点的算术平均值, 即该簇的中心, 样本到中心的距离一般采用欧氏距离。在具体实现时, 步骤 4 通常需要定义一个最大迭代次数作为终止条件。

K -means 算法力求找出使平方误差函数值最小的 K 个划分, 当数据分布满足正态分布时, 即簇与簇之间区别明显时, 其效果较好。因此, K -means 算法不适用于处理非凸面形状的簇, 或者大小差别很大的簇, 而且其对于噪音和孤立点等异常数据是敏感的。因此, 本文采用给样本加权的方式, 降低异常样本的影响, 权值计算如下:

$$z_j = \frac{1}{M} \sum_{i=1}^{M_j} w_{i,j} x_j; w_{i,j} = \frac{N(x_i | \mu_j, \varepsilon_j)}{\sum_{i=1}^K N(x_i | \mu_j, \varepsilon_j)} \quad (2)$$

其中, $w_{i,j}$ 表示第 i 个样本 x_i 属于第 j 个类别高斯分布的概率, N 为高斯概率密度求解函数, μ_j 表示均值, ε_j 表示方差。这个方法本质上来讲是在



计算聚类中心的时候,提高符合正太分布样本的权值,降低不符合正太分布样本的权值。

此外,由于 K -means 算法的 K 值需要提前给出,其初始聚类中心是随机选择的,而不同的初始中心对聚类结果有很大的影响,导致算法通常会终止于局部最优解。因此,本文采用第 2.2 节的 Canopy 算法求解 K 值,并计算出初始簇中心的具体位置,在一定程度上加快了 K -means 算法的收敛速度,提高了聚类分析的准确性和计算效率。

3 实验仿真与分析

在北京市区范围内,对来自于 500 kV、220 kV、110 kV 的线路设备的近 5 年的外力破坏数据采用本文算法进行处理,具体实验包括以下步骤。

(1) 对外力破坏数据进行筛选,包括时间(年月)、外力破坏类型(异物短路、施工碰线、异物缺陷)、电压等级(500 kV、220 kV、110 kV)等。

(2) 执行 Canopy 算法,该算法将一堆混乱的数据分类成有一定规则的数据簇,由于 Canopy 算法本身的目的只是将混乱的数据划分成大概几个类别,所以它是不太准确的。执行该算法后,将聚类结果作为 K -means 算法的初始聚类中心点。本步骤需要设定参数 T_1 、 T_2 ($T_1 > T_2$),即 $T_1 = 5\ 000\text{ m}$, $T_2 = 2\ 000\text{ m}$ 。

(3) 执行优化的 K -means 算法,该算法采用空间距离作为相似性的评价指标,距离簇中心越近,其相似度就越大,而异常数据对于簇中心的计算影响几乎消失。在输入 k 个初始聚类中心点的具体位置(来自步骤 2 的结果)后, K -means 算法快速收敛,从而获得最优的聚类中心以及属于该簇的样本,即容易受到外破影响的设备位置。

下面以相同方式对 3 类外破数据进行分析。

对施工碰线类型外破数据进行分析,根据受到外破影响的设备数量对簇进行降序排序。以第一个簇为例,包含样本最多的簇的聚类中心在王君一线路上 21 号杆塔附近,受到影响的线路包

括:通*,安*(500 kV);西*二,华*一,华*一等(220 kV);王*二,王*,君*一等(110 kV)。实验总共分析 278 条数据,16 个聚类中心,聚类覆盖数据 144 条。

异物短路类型数据的聚类结果中,包含样本最多的簇的聚类中心在阳*二线路上 14 号杆塔附近,共 272 条数据、18 个聚类中心,聚类覆盖数据 147 条。

异物缺陷类型数据的聚类结果中,包含样本最多的簇的聚类中心在安*一线路上 76 号杆塔附近,共 267 条数据、18 个聚类中心,聚类覆盖数据 195 条。

此外,在本文的实验中(CPU Intel 双核主频 2.66 GHz,4 GB 内存),本文算法在系统中前端 GIS 页面显示外部破坏数据和聚类分析结果响应时间整体不到 2 s,证明了本文算法的高效性。这是由于 Canopy 算法虽然计算聚类中心精度较低,但其在速度上有较大优势,当其得到 K 个“粗”聚类中心后,再通过 K -means 算法进行“细”聚类操作,同时也加快了 K -means 算法的收敛速度,在本文对于 3 类外部破坏数据的处理实验中,在迭代 3 次左右后聚类中心即稳定下来。

4 结束语

外力破坏指的是人们无意或者有意导致的电力线路部件非正常状态、破坏和损毁的行为,会导致线路出现安全问题。本文提出了一种基于自动聚类模型的输电线路外破数据分析方法,解决了 K -means 算法无法直接确定最优 K 值的问题,以及消除了异常数据对于计算结果的影响。通过该方法分析研究外力破坏情况下的时空分布特征,进而制定有效的防治措施,确保线路安全运行,为线路防外力破坏的预测预警工作提供了有力辅助。

参考文献:

[1] 林飞飞.特高压输电线路继电保护算法的研究[D].秦皇岛:

- 燕山大学, 2017.
- LIN F F. Research on relay protection algorithm for UHV transmission lines [D]. Qinhuaangdao: Yanshan University, 2017.
- [2] 李天, 石鑫, 李永倩. 基于输电线路监测的无线传感器网络路由优化技术研究[J]. 电测与仪表, 2015, 21(10): 6-10.
- LI T, SHI X, LI Y Q. Research on routing optimization technique of wireless sensor networks based on power lines monitoring[J]. Electrical Measurement and Instrumentation, 2015, 21(10): 6-10.
- [3] 胡毅, 刘凯, 吴田, 等. 输电线路运行安全影响因素分析及防治措施[J]. 高电压技术, 2014, 40(11): 3491-3499.
- HU Y, LIU K, WU T, et al. Analysis of influential factors on operation safety of transmission line and countermeasures[J]. High Voltage Engineering, 2014, 40(11): 3491-3499.
- [4] 巢亚锋, 徐志强, 段建家, 等. 架空输电线路防外力破坏智能管控系统设计[J]. 湖南电力, 2017, 37(5): 18-28.
- CHAO Y F, XU Z Q, DUAN J J, et al. Design of anti-external damage intelligent control system for overhead transmission lines [J]. Hunan Electric Power, 2017, 37(5): 18-28.
- [5] 冷雪敏, 高阳, 许傲然. 架空输电线路防外力破坏系统研究[J]. 电子世界, 2017(13): 45-46.
- LENG X M, GAO Y, XU A R. Study on anti external force damage system of overhead transmission lines[J]. Electronics World, 2017(13): 45-46.
- [6] 杨振伟, 易伟, 陈键. 输电线路外力破坏故障的分析与防治[J]. 吉林电力, 2015, 43(5): 48-50.
- YANG Z W, YI W, CHEN J. Analysis and countermeasures of damage from outside destroy in transmission line[J]. Jilin Electric Power, 2015, 43(5): 48-50.
- [7] 谢兵, 许金彤. 基于 3G 网络的输电线路防外力破坏智能视频监控研究[J]. 华东电力, 2010, 38(11): 1729-1732.
- XIE B, XU J T. Study of destruction of intel-ligent video surveillance system based on transmission lines against external 3G network[J]. East China Power, 2010, 38(11): 1729-1732.
- [8] 许金明, 江志蓉. 嘉兴电网外力破坏事故和预防措施[J]. 电力建设, 2009, 30(7): 106-107.
- XU J M, JIANG Z R. Jiaxing Power Grid's external damage accidents and preventive measures[J]. Electric Power Construction, 2009, 30(7): 106-107.
- [9] 刘振亚. 中国特高压交流输电技术创新[J]. 电网技术, 2013, 37(3): 566-574.
- LIU Z Y. Innovation of UHVAC transmission technology in China[J]. Power System Technology, 2013, 37(3): 566-574.
- [10] 王继业. 智能电网大数据[M]. 北京: 中国电力出版社, 2016.
- WANG J Y. Big data of smart grid[M]. Beijing: China Electric Power Press, 2016.
- [11] CHEESEMAN P, STUTZ J. Bayesian classification (AutoClass): theory and results[M]. Cambridge: AAAI/MIT Press, 1996: 153-180.
- [12] LI F H, ZHAO Z F, WANG Z Y. Hierarchical clustering based on single-pass for breaking topic detection and tracking[J]. High Technology Letters, 2018(4).
- [13] SUN J Z, LING B L. Density PSO-based software module clustering algorithm[J]. The Journal of China Universities of Posts and Telecommunications, 2018(4).

[作者简介]



马大燕(1984-), 女, 博士, 国网电子商务有限公司工程师, 主要从事分布式电源与微电网、能源互联网技术方面的研究工作。