

DOI:10.3969/j.issn.1001-4551.2015.10.022

考虑历史缺陷文本信息的断路器状态评价研究^{*}

马润泽,王龙响,余佳文,王慧芳^{*},邱 剑
(浙江大学 电气工程学院,浙江 杭州 310027)

摘要:针对当前断路器进行状态评价基于其运行信息,但未涉及能反映断路器部件可靠性的断路器历史缺陷情况的特点,以及断路器中文缺陷文本的处理分析问题,提出了融合运行信息与历史缺陷信息的断路器状态综合评价模型,即在现有评价导则基础上添加了专门针对历史缺陷文本的评价规则,并提出了基于 KNN 算法的缺陷文本分类方法,方法先通过分词、词频统计、去停用词和文本向量化等适合短文本的处理步骤进行中文文本预处理,然后采用 KNN 分类算法对缺陷文本进行归类计算。通过算例,对缺陷文本分类方法以及考虑历史缺陷信息后的断路器状态综合评价模型进行了验证。研究结果表明,缺陷文本分类方法是可行的,且加入历史缺陷信息后,断路器的状态评价结果得到合理性的提升。

关键词:断路器;状态评价;缺陷文本;KNN 算法

中图分类号:TM56

文献标志码:A

文章编号:1001-4551(2015)10-1375-05

Circuit breakers' condition evaluation considering the information in historical defect texts

MA Run-ze, WANG Long-xiang, YU Jia-wen, WANG Hui-fang, QIU Jian
(College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China)

Abstract: Aiming at the feature that the condition evaluation of the circuit breaker (CB) is based on the operation information currently, while it does not involve CB's historical defects which reflect the reliability of CB's components, and the problem of the treatment and analysis of CBs' Chinese defect texts, a comprehensive evaluation model of the CBs' condition integrating the CBs' operation information and historical defect information was proposed, which adds the evaluation rules for historical defect texts to the existing evaluation guidelines, and also a classification method of the defect texts was proposed based on KNN algorithm. The steps of the method is firstly the Chinese text pre-processing was done by doing word segmentation, word frequency statistics, deleting stop word and text vectorization, which were suitable for short texts, then the KNN algorithm was used to classify the defect texts. An example was used to verify the defect text classification method and the comprehensive evaluation model considering the information in historical defect texts. The results indicate that the classification method is feasible and the rationality of the CB's evaluation results is improved.

Key words: circuit breaker; condition evaluation; defect texts; KNN algorithm

0 引 言

断路器是一种容易出现缺陷的电力设备^[1]。相较于传统的事后检修或定期检修,状态检修是基于运行

状态制定检修策略,具有检修效率高、时效性好、成本低等优点,是目前检修的发展方向^[2]。健康状态准确评价是实施状态检修的前提,因此,对断路器的状态评价进行研究具有十分重要的意义。

收稿日期:2015-05-04

基金项目:国家电网公司科技资助项目(GY71-14-075)

作者简介:马润泽(1993-),男,浙江建德人,主要从事电力设备状态评估研究方面的研究. E-mail:631787404@qq.com

通信联系人:王慧芳,女,副教授. E-mail:huifangwang@zju.edu.cn

目前针对断路器的状态评价,国网公司于 2008 年发布了 Q/GDW 171—2008《SF₆ 高压断路器状态评价导则》^[3],作为制定状态检修计划的依据。此外一些专家学者采用断路器状态量数据,将导则或专家意见作为断路器状态评价的规则,针对断路器状态评价模型的构建进行了探究,采用的算法主要有层次分析法^[4]、人工神经网络法^[5]、模糊综合评判法^[6]、聚类分析法^[7]、物元法^[8]、雷达图法^[9]等,一些文献将这些算法与证据理论^[10-11]、可拓论^[12]、突变理论^[13]相结合,以提升算法性能。

以上研究在进行断路器状态评价时所采用的依据均为当前测得的状态量运行数据,未涉及断路器的历史缺陷情况。断路器的历史缺陷情况能够反映断路器部件的可靠性。如果一台断路器在一定时间内出现多种缺陷或多次出现同一种缺陷,说明该断路器在运行中更易出现缺陷,在进行断路器的状态评价时需考虑上述情况。因此,可将断路器的历史缺陷情况作为断路器状态评价的一部分。断路器的历史缺陷情况一般由断路器的缺陷文本记录。缺陷文本通常采用自然语言表述,采用机器分析具有难度。目前鲜见电网文本处理的相关研究。缺陷文本评判的核心是确定缺陷的严重程度,即缺陷等级。以往缺陷等级的分类一般通过运维人员经验判断,虽可以在大多数情况下保证分类正确性,但费时费力,且受人的主观性影响。因此需对缺陷文本库进行信息挖掘研究,提出缺陷等级评判方法。

本研究将断路器的历史缺陷情况加入到断路器状态评价模型之中,提出了基于断路器运行状态量与断路器历史缺陷的断路器状态综合评价模型,并针对断路器缺陷文本进行研究,提出了基于 KNN 算法的缺陷文本分类方法。最后对分类方法与相应的评价模型进行了算例验证。

1 问题的提出与模型的建立

1.1 状态量的量化评分标准与历史消缺信息作为状态量的依据

Q/GDW 171—2008《SF₆ 高压断路器状态评价导则》是国家电网公司为规范与有效开展断路器状态检修而制定的。导则将断路器分为本体、操动机构、并联电容器、合闸电阻 4 个部件,每一部件的具体状态量视对于断路器安全运行的影响程度,从低到高分为 1~4 共 4 个权重,对于状态量的劣化程度,从低到高分为 I、II、III 和 IV 级 4 个等级,分别对应基本扣分值 2、4、8、10,断路器各部件扣分值等于该部件各状态量劣化

程度扣分与权重乘积之和。根据断路器各个部件的扣分和评价标准,可得到部件的状态评价结果。断路器整体评价状态为各部件评价状态中最严重的状态。

导则是在大范围参考其他标准并结合大量 SF₆ 高压断路器缺陷分析及运行检修管理经验的基础上制定的,具有清晰、明确、合理的优点,可作为评价的整体框架。因此,本研究在建立断路器状态评价模型时,针对断路器运行信息,采用导则中的规则进行评价,而针对历史缺陷信息,则在导则中添加专门针对历史缺陷文本的规则进行评价。

断路器的缺陷文本一般包含设备型式、电压等级、名称、所在变电所、厂家、出厂时间、缺陷部位、类型、具体内容、缺陷等级、消缺状态、技术原因、处理结果,发现过程及消缺流程中的相关时间与相关责任人等信息。其中缺陷等级一般被归为“紧急”、“严重”、“一般”3 类。断路器的缺陷文本能反映断路器曾经出现过的缺陷情况并且能提供相应的消缺记录,是对断路器可靠性进行判断的有力依据,本研究将其作为断路器状态评价的依据之一。

1.2 历史消缺信息状态量的扣分模型

待建立的状态评价模型从数据来源看为一个综合运行数据与历史缺陷文本的状态评分模型。对于断路器运行数据,根据导则打分;对于缺陷文本,则通过中文文本处理技术将其向量化后,采用分类算法与之前建立好的缺陷文本库对文本进行缺陷程度的分类,并根据在导则中添加的评价规则进行相应打分;最后两部分打分合计。之后根据导则完成状态评价的余下步骤,得到最终结果。对导则的添加内容具体为:在导则的每一个部件分类下添加状态量描述“在评价时间范围内含有历史缺陷”,权重为 1,根据缺陷等级不同,分别为:

(1) 不扣分,判断依据为“[0,0.4) 次/年的‘一般’等级缺陷,[0,0.2) 次/年的‘严重’与‘紧急’等级缺陷”;

(2) 劣化程度为 II 级,基本扣分为 4,判断依据为“[0.2,0.4) 次/年‘严重’等级缺陷;[0.4,0.6) 次/年‘一般’等级缺陷”;

(3) 劣化程度为 III 级,基本扣分为 8,判断依据为“[0.2,0.4) 次/年‘紧急’等级缺陷;[0.4,0.6) 次/年的缺陷,其中至少有 0.2 次/年‘严重’等级缺陷;0.6 次/年及以上(含 0.6 次/年,下同)的‘一般’等级缺陷”;

(4) 劣化程度为 IV 级,基本扣分为 10,判断依据为“0.4 次/年及以上的缺陷,其中至少有 0.2 次/年

‘紧急’等级缺陷;0.6 次/年及以上的缺陷,其中至少有 0.2 次/年‘严重’等级缺陷”。

上述评价规则可实现:在设备正常时,设备的状态评价结果仍然为正常状态;在设备出现缺陷时,尤其该缺陷历史上已出现过,设备的状态评价结果将适当严重化,从而在检修中提升对该部件乃至该设备的注意程度。

2 基于 KNN 算法的中文缺陷文本分类

2.1 缺陷文本预处理过程

对文本信息进行分类等操作无法直接进行,在进行该类操作之前,必须去除文本中的无用信息,并将文本中的有用信息转化为向量表示。这一转化过程被称作文本预处理。

中文文本的预处理与英文文本有一定区别。英文文本词与词之间通过空格隔开,具有自然分界,中文文本则不具这一特点。一般而言,中文文本的预处理需要经过文档分段、分句、分词、去除停用词、特征选择、词频统计、文本向量化^[14-17]几个步骤,这样的预处理流程适合长文本。而断路器的缺陷文本一般为短文本,根据这一特点,其预处理仅保留分词、词频统计、去除停用词和文本向量化步骤,文本预处理流程如图 1 所示。

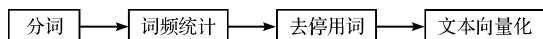


图 1 文本预处理流程

步骤 1:分词。分词即将文本中的字符串合理地切分成词语序列,是中文文本信息处理必备且最为基础的环节。

步骤 2:词频统计。即是对文本中所有词进行出现频次的统计并据此排序。完整的词频统计结果即包括了文本中出现的所有词,可用这些词构成文本向量的向量空间。每个词对应向量空间中一维。

步骤 3:去除停用词。即去除无法表征文本的词,执行中考虑先建立停用词表,再将词频统计结果和停用词表中的词进行对照,删去其中的停用词。

步骤 4:文本向量化。是把经过上述处理的文本转化为数字向量的表示结果。每一条文本对应一个特征向量,特征向量的每一维对应词频统计结果中的每一个词。具体方法为根据文本向量的向量空间,再根据分词结果,将每条文本中的词与向量空间每一维对应的词作对照,如果相同,则在该文本对应的特征向量中,将该词对应的维记为 1,否则为 0。

2.2 缺陷文本的 KNN 分类模型

对于已经向量化后的缺陷文本,需要通过文本分

类的方式来对其进行归类。通过将待归类的单条缺陷文本与缺陷文本库中的已归类的缺陷文本进行对比,寻找与缺陷文本库中相近的文本,然后将该缺陷文本归入最接近的一类当中。

文本库中的缺陷文本已经过文本预处理,构建好了向量空间,缺陷等级也已按“紧急”、“严重”、“一般”三类分好,可认为其中大部分分类结果是正确的。为便于计算,本研究将上述缺陷等级分别量化为 3、2、1。待归类的缺陷文本则经过文本预处理后,将采用 KNN 算法进行归类。

KNN 算法又名 k-最近邻算法,其核心思想为如果一个样本在特征空间内的若干个最相邻的样本中的大多数属于某一个类别,则该样本也属于这个类别,并具有这个类别上样本的特性^[18-19]。相比 Kmeans、Ada-boost 等其他分类算法,KNN 算法既不是通过在大范围内寻找区域中心,然后判断样本与各区域中心的相近程度,也不是通过对所有样本分布规律进行总结的划分规则来进行样本分类,它不仅适合缺陷文本在大范围内混杂,在小范围内聚集的特点,而且不需建立复杂的划分规则。

采用 KNN 算法进行缺陷文本归类的操作步骤为:

(1)计算待归类文本与文本库中每个文本的文本相近程度。

关于两向量化后的文本相近程度的定义,可采用以下表示的相似度公式:

$$S_{ij} = \frac{\sum_{l=1}^M W_{il} \times W_{jl}}{\sqrt{\sum_{l=1}^M W_{il}^2} \sqrt{\sum_{l=1}^M W_{jl}^2}} \quad (1)$$

式中: S_{ij} —待求的待测文本 i 与文本库中文本 j 之间的相似度; d_i —文本 i 的特征向量; d_j —文本 j 的特征向量; M —向量的维数; W_{il} —向量 d_i 的第 l 维的值; W_{jl} —向量 d_j 的第 l 维的值。

(2)按照文本相似度,在文本库中选出与待归类文本最相似的 k 条文本,其中 k 可设置。

(3)利用式(2)计算待归类缺陷文本 i 描述的缺陷严重程度 y_i :

$$y_i = \frac{\sum_{l=1}^k S_{il} y_l}{\sum_{l=1}^k S_{il}} \quad (2)$$

式中: y_l —第 l 条缺陷文本描述的缺陷严重程度。

(4)根据计算结果将该缺陷的严重程度归类,如计算结果为 $y_i \in (1, 1.5]$,则归为一般类,计算结果为

$y_i \in (1.5, 2.5)$, 则归为严重类, 结果为 $y_i \in (2.5, 3]$, 则归为紧急类。

缺陷文本归类后, 根据 1.2 节中介绍的不同缺陷等级的扣分模型进行后续状态评价。

3 算例分析

本研究以华东地区某电力公司的断路器缺陷文本为例, 对断路器各部件缺陷文本的缺陷严重程度进行了重新分类, 并与初始结果进行了比较; 笔者通过对一台断路器进行状态评价, 验证了本研究所提出的断路器状态评价方法的可行性。算例过程中的分词步骤由 ICTCLAS 完成, 并复制结果于文本文档中保存, 词频统计由 Microsoft Excel 中编写宏程序完成, 结果复制后保存于文本文档中, 其余步骤通过 Matlab 完成。

ICTCLAS 是由中科院计算所开发, 基于层次隐马尔科夫模型的中文词法分析系统, 被广泛应用于中文信息处理的各个领域。但 ICTCLAS 本身无法识别电气行业的专业词汇, 需要进行新词添加, 为此本研究首先建立了“断路器缺陷文本词库”。

3.1 中文缺陷文本 KNN 算法的测试

本研究对断路器各部件缺陷文本进行分类计算, KNN 算法中取 $k=3$, 分类结果如图 2 所示。

由图 2 所示, 上方图像为初始分类情况, 即为不同的责任人分类的, 下方图像为 KNN 分类计算情况。图像中每一个点代表一条缺陷文本, 极轴表示缺陷严重程度 y_i , 3 个圆环由内到外分别表示缺陷等级为一般、严重、紧急。由此可见, 本体、操动机构和并联电容的缺陷文本分类计算结果与初始分类有一定数量的差别, 具体统计结果如表 1 所示。

表 1 各部件缺陷文本的分类统计结果

部件名称	所有缺陷文本数量	分类正确数量	分类正确率
本体	235	212	0.902 1
操动机构	318	287	0.902 5
并联电容	184	174	0.945 7
合闸电阻	7	7	1.000 0

表 1 得到的分类正确率是合理满足要求的, 可以认为通过 2 中介绍的方法对缺陷文本进行分类是可行的。

3.2 融合运行信息与历史缺陷信息的断路器状态综合评价模型测试

某断路器 5 年内的缺陷情况如表 2 所示, 其缺陷文本分类结果也列于表 2 内。

表 2 某断路器 5 年内的历史缺陷情况

缺陷所在部件	缺陷内容	缺陷等级
操动机构	打压超时信号动作, 无法复归信号。	紧急
操动机构	油泵无法建压, 打压超时信号动作	紧急
本体	接线绝缘	紧急

根据 1.2 节的历史消缺信息扣分模型, 则其本体部分有 0.2 次/年紧急缺陷, 追加扣分值 8; 操动机构部分有 0.4 次/年紧急缺陷, 追加扣分为 10。此时:

(1) 若根据运行信息的状态评价结果为所有部件正常, 即本体、操动机构、并联电容、合闸电阻四部件原扣分均为 0, 按原评价方法判定为所有部件均为正常状态, 总评为正常状态; 融入历史缺陷信息后, 四部件扣分分别为 8、10、0、0, 评价结果为所有部件均为正常状态, 总评为正常状态, 评价等级未发生变化;

(2) 若根据运行信息的状态评价结果, 操动机构由于打压超时信号动作, 根据当前导则扣 8 分, 其他部件没有扣分, 按原评价方法判定为所有部件均为正常状态, 总评为正常状态; 融入历史缺陷信息后, 四部件的扣分分别为 8、18、0、0, 评价结果为本体、并联电容器、合闸电阻为正常状态, 操动机构为注意状态, 总评为注意状态, 评价等级严重程度提升。

从以上算例可以得出, 当断路器当前状态完全正常时, 本研究所提出的评价方式不会使得评价等级恶化; 当断路器出现缺陷, 且缺陷出现在历史已出现过多次缺陷的部件中时, 该部件的状态评价结果会比原来

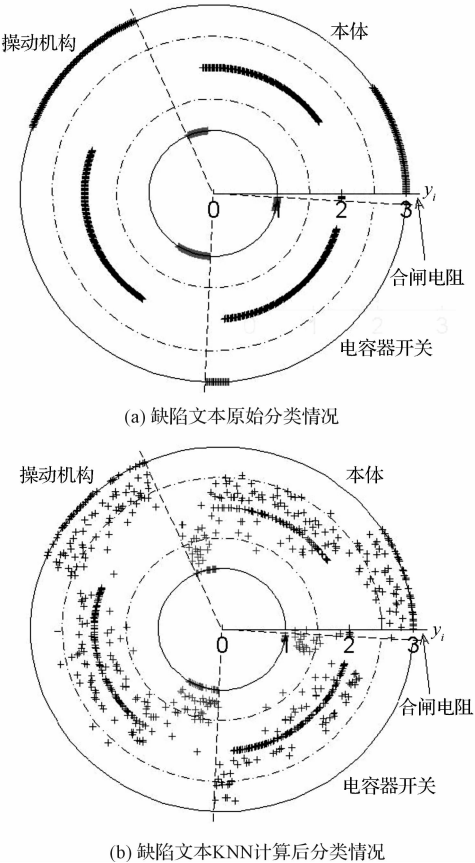


图 2 断路器缺陷文本分类情况

严重,从而使断路器整体评价下降,引起检修人员注意。因此状态评价中引入历史缺陷信息,是既合理又可行的。

4 结束语

本研究在断路器状态评价中加入历史缺陷信息进行综合评判,并提出了基于 KNN 算法的缺陷文本分类方法以对断路器的缺陷文本进行分析处理,通过算例验证后得到以下结论:

(1)引入历史缺陷文本信息用于断路器状态评价,是对当前的状态评价方法的有益补充,能促进断路器状态评价结果更为完善与合理。

(2)采用中文文本处理方法对缺陷文本进行处理,是一种有益的尝试,算例结果表明,基于 KNN 算法对文本进行分类以确定缺陷等级是可行的。考虑到电网中设备运行缺陷都有相应的文本记录,在进一步研究中,可将此方法用于电网其他设备,尤其是不适合装设状态监测装置的设备的状态评价。

参考文献 (References):

- [1] JANSSEN A, MAKAREINIS D, SÖLVER C. International surveys on circuit-breaker reliability data for substation and system studies[J]. *IEEE Transactions on Power Delivery*, 2014, 29(2): 808-814.
- [2] ENDRENYI J, ABORESHEID S, ALLAN R N, et al. The present status of maintenance strategies and the impact of maintenance on reliability[J]. *IEEE Transactions on Power Systems*, 2001, 16(4): 638-646.
- [3] 国网电网公司. Q/GDW 171—2008, SF6 高压断路器状态评价导则[S]. 国网电网公司, 2008.
- [4] LIN Xin, QI Hong-wei, TENG Yun, et al. Evaluation Method of High-Voltage Circuit Breaker Running State Based on Matter-Element Theory and AHP[C]//*E-Product E-Service and E-Entertainment (ICEEE)*, 2010 International Conference on ;IEEE, 2010: 1-4.
- [5] 徐建源, 张 彬, 林 莘, 等. 能谱熵向量法及粒子群优化的

的 RBF 神经网络在高压断路器机械故障诊断中的应用[J]. *高电压技术*, 2012, 38(6): 1299-1306.

- [6] 陈伟根, 魏延芹, 廖瑞金. 高压断路器运行状态的变权模糊综合评判方法[J]. *高压电器*, 2009, 45(3): 73-77.
- [7] 张国钢, 李 宇, 汤 翔, 等. 模糊聚类分析用于断路器状态评估因素分类[J]. *高电压技术*, 2008, 34(2): 350-354.
- [8] 陈伟根, 魏延芹, 王有元. 带变权的高压断路器状态物元评估方法[J]. *重庆大学学报*, 2009, 32(2): 151-157.
- [9] 李海英, 冯 冬, 宋建成. 中压真空断路器状态的雷达图法评估模型[J]. *电网技术*, 2013, 37(7): 2053-2059.
- [10] LIN Po-chun, GU Jyh-chen, YANG Ming-ta. Intelligent maintenance model for condition assessment of circuit breakers using fuzzy set theory and evidential reasoning[J]. *IET Generation, Transmission & Distribution*, 2014, 8(7): 1244-1253.
- [11] 魏延芹. 基于物元与证据理论相结合的高压断路器状态评估方法研究[D]. 重庆: 重庆大学电气工程学院, 2008.
- [12] 马国立. 基于可拓理论的高压断路器状态评估[D]. 北京: 华北电力大学电气工程学院, 2009.
- [13] 陈伟根, 吴 娅, 刘 强. 基于突变理论的断路器运行状态模糊综合评判方法[J]. *高压电器*, 2007, 43(2): 127-130.
- [14] 何金凤. 基于中文信息检索的文本预处理研究[D]. 成都: 电子科技大学计算机科学与工程学院, 2008.
- [15] 钟茂生. 基于内容相关度计算的文本结构分析方法研究[D]. 上海: 上海交通大学电子信息与电气工程学院, 2010.
- [16] 张 宁. 基于语义的中文文本预处理研究[D]. 西安: 西安电子科技大学管理学院, 2011.
- [17] 廖一星. 文本分类及其特征降维研究[D]. 杭州: 浙江大学计算机科学与技术学院, 2012.
- [18] 胡 元. 基于区域划分的 KNN 文本快速分类算法研究[D]. 济南: 山东大学计算机科学与技术学院, 2012.
- [19] 张孝飞, 黄海燕. 一种采用聚类技术改进的 KNN 文本分类方法[J]. *模式识别与人工智能*, 2009, 22(6): 936-940.

[编辑: 张 豪]

本文引用格式:

马润泽, 王龙响, 余佳文, 等. 考虑历史缺陷文本信息的断路器状态评价研究[J]. *机电工程*, 2015, 32(10): 1375-1379.

MA Run-ze, WANG Long-xiang, YU Jia-wen, et al. Circuit breakers' condition evaluation considering the information in historical defect texts[J]. *Journal of Mechanical & Electrical Engineering*, 2015, 32(10): 1375-1379.

《机电工程》杂志: <http://www.meem.com.cn>