DOI: 10.3969/j.issn.1006-4729.2013.06.003

变压器缺陷管理系统中不确定性数据的处理

袁仲雄

(上海电力学院 计算机科学与技术学院, 200090)

摘要: 针对变压器缺陷管理中缺陷定级时专家定级数据的不确定性, 利用不确定数据 top-k 查询的相关理论, 建立了变压器缺陷定级不确定性数据模型——可能世界模型, 并确定专家定级数据排序, 排在首位的定级结果作为该缺陷的定级结果. 然后根据定级结果安排消缺任务.

关键词: 不确定 top-k 查询; 变压器缺陷; 不确定性数据管理; 缺陷定级

中图分类号: TP274; TM41

文献标志码: A

文章编号: 1006-4729(2013)06-0523-04

Top-k Query Processing on Uncertain Data for Transformer Fault Management System

YUAN Zhongxiong

(School of Computer Science and Technology, Shanghai University of Electric Power, Shanghai 200090, China)

Abstract: Because of the expert fault diagnosis data uncertainty in transformer fault management system, the transformer fault diagnosis uncertain data model and possible world model are established by the use of top-k query theory on uncertain data. The experts fault diagnosis data are sorted and the first in the queue considered as the final fault diagnosis result. The eliminating fault task is arranged according to the result of the fault diagnosis result.

Key words: uncertain top-k query; transformer fault; uncertain data management; fault diagnosis

电力变压器是电力系统中的重要设备之一, 其安全可靠运行对电力系统的安全稳定运行具有关键的作用. 在长时间的运行过程中, 变压器出现设备故障是不可避免的, 正确判断变压器故障并在必要时对故障进行维修, 对保证变压器安全和经济运行具有重要意义. 变压器故障诊断方法较多, 常用的有油中溶解气体分析技术(DGA)、概率推理、贝叶斯分类等,^[1-4] 这些方法通常是根据故障征兆来确定故障类型. 在变压器缺陷管理中, 当变压器有故障征兆时, 管理人员通过设备缺陷

管理系统, 录入故障, 然后确定缺陷的等级, 再确定检修的时间和方法. 其中缺陷等级对确定变压器检修的时间很重要, 紧急缺陷需要马上安排抢修, 重要缺陷也需要尽快安排维修, 对于一般缺陷, 可以暂时不维修, 等到整个系统维修(A修B修)时再进行处理.

对变压器管理系统来说, 如何根据故障征兆来确定缺陷等级是一个难题. 在变压器的运行监测、检修和预试中, 监测到的缺陷征兆区分往往不是很清晰, 很难做出结论, 而且针对同一故障征

收稿日期: 2013-09-24

通讯作者简介: 袁仲雄(1956-), 男, 副教授, 上海人. 主要研究方向为电力信息化. E-mail: yzx19700701@163.com.

兆,不同的专家会给出不同的定级.因此,根据故障特征兆确定缺陷等级通常是一个不确定性问题.

1 基于不确定数据模型 top-k 技术

1.1 不确定性数据模型

描述不确定数据的模型主要有关系型数据的数据模型、半结构化数据模型、数据流模型、多维数据模型.^[5]在变压器缺陷管理系统中主要采用关系型数据的数据模型,该模型将不确定性分为存在级不确定性和属性级不确定性.在属性级不确定性中,元组是确定存在的,元组中的某个属性具有不确定性,用概率来表示该属性取某值的概率;在存在级不确定性中,元组的属性取值是确定的而整个元组具有不确定性,元组整体用一个概率值表示该元组的存在概率.

表1为不确定数据.表1中,元组 t_1 为存在级不确定性, $A=2$ 的概率是0.8;元组 t_2 为属性级不确定性, $B=3$ 的概率为0.2, $B=5$ 的概率为0.8;元组 t_3 既有存在性不确定性,又有属性级不确定性.

表1 不确定性数据

元组	A	B	概率
t_1	2		0.8
t_2	3	<3,0.2>, <5,0.8>	
t_3	4	<4,0.3>, <9,0.7>	0.5

对不确定数据建模,一般采用可能世界模型,利用可能世界模型将不确定数据元组存在的概率转化为可能世界实例出现的概率.不确定数据模式 D 由一系列的关系 $R_1, R_2, R_3, \dots, R_m$ 组成,对于每个关系 $R_i(i=1, 2, 3, \dots, m)$ 属性中的一个属性 s ,表示元组的出现概率, s 为 $[0, 1]$.在模式 D 上的可能世界 $W = \{W_1, W_2, W_3, \dots, W_n\}$,并且 $P: W \rightarrow [0, 1]$ 是一个概率分布,且对于任意一个 $P(W_k > 0)$, $\sum_{k=1, n} P(w_k) = 1$, W_k 为一个可能世界实例.一个可能世界实例是所有那些不确定性属性满足约束条件的元组的任意一种组合.^[6]

1.2 top-k 查询

top-k 查询通常对表中多个属性建立起一个评分模型,在数据空间中返回满足查询条件的前 k 个对象.在包含 m 个属性的数据空间 D 中,需

要查询的问题有 g 个特征 $d_j(1 \leq j \leq g)$.根据特征的重要程度,为每个特征 d_j 设定一个权值 $w_j(w_j \geq 0)$,某些权重可以设为零.评分函数为:

$$F = \sum_{i=1}^g w_i d_i$$

对于 $F(t_i) \geq F(t_j)$,top-k 查询返回 D 的 k 个子集 $Q, t_i \in Q, t_j \in (D - Q)$.

1.3 不确定数据模型 top-k 查询

目前,针对不确定数据的 top-k 查询的技术主要有 U-topK, U-kRanks, PT-k^[7]等.

1.3.1 U-topK

对于不确定性数据模式 D 和可能世界模型 $W = \{W_1, W_2, W_3, \dots, W_n\}$, $T_k(W_i)$ 为可能世界模型实例 W_i 中的前 k 个元组,且 $T_k(W_i)$ 是 k 个元组按照评分函数 F 排序后的元组向量.^[7]则有 $T_k(W_i) \in T_k$ 满足:

$$T_k(W_i) = \operatorname{argmax}_{T_k(W_i) \in T_k} \left(\sum_{w \in T_k(W_i)} P(w) \right)$$

U-topK 查询实际上是要查找一个长度为 k 的元组向量 $T_k(W_i)$, $T_k(W_i)$ 中的可能世界实例的概率值之和最大.

1.3.2 U-kRanks

对于不确定性数据模式 D 和可能世界模型 $W = \{W_1, W_2, W_3, \dots, W_n\}$, $T_k(W_i)$ 为可能世界模型实例 W_i 中的前 k 个元组, T_k 中的所有元组按照排序函数 F 对其进行从大到小排序.基于 F 的 U-kRanks,返回 k 个记录.^[7]即 $\{t_i^* | i=1, 2, 3, \dots, k\}$,满足:

$$t_i^* = \operatorname{arg max}_i \left(\sum_{w \in w(t_i)} P(w) \right)$$

U-kRanks 查询就是对出现在位置 i 的可能世界概率求和,取概率最大的作为返回结果的第 i 个记录.

1.3.3 PT-k

对于不确定数据模式 D 和可能世界模型 $W = \{W_1, W_2, W_3, \dots, W_n\}$, $T_k(W_i)$ 为可能世界模型实例 W_i 中的前 k 个元组,且 $T_k(W_i)$ 是 k 个元组按照评分函数 F 排序后的元组向量.如果有些可能世界实例所包含的元组的个数小于 k , $T_k(W_i)$ 就为这些实例中所有元组的集合.^[6]在 T_k 的基础上计算每个记录位于 top-k 的概率:

$$p(t) = \sum_{t \in T_k(w_i)} P(w_i) \geq p$$

设定概率阈值 $p(0 < p \leq 1)$,PT-k 查询对每

个元组出现在 top-k 位置的可能世界概率求和, 返回的元组排在前 k 位的所有可能世界实例概率值之和都大于等于阈值 p .

2 不确定性数据 top-k 查询在变压器缺陷管理系统中的应用

2.1 变压器缺陷征兆和定级准确率

在变压器缺陷实际管理中, 一般根据变压器故障征兆初步判定缺陷, 经过消缺后, 在消缺验收环节根据变压器打开后实际检查和维修情况对变压器缺陷等级进行评级.

依据变压器缺陷实际管理的情况和不确定性 top-k 原理, 可将变压器缺陷管理系统中的缺陷分为紧急、严重、一般 3 个等级, 用 1, 2, 3 表示; 系统中缺陷定级可分为初步定级和验收定级, 初步定级根据缺陷征兆确定, 验收定级用于确定初步定级是否准确. 系统根据初步定级安排消缺计划.

由于初步定级是专家根据缺陷征兆来确定的, 存在一定的不确定性, 而验收定级是变压器打开检查维修后的定级, 比较准确. 用前后两次定级的准确度的差和总的定级次数来表示初步定级是否准确的概率. 例如: 在系统数据库中某专家对变压器油位超标缺陷定级记录为 22 次, 22 次的初始定级都为 3 (一般缺陷), 但验收定级中 16 次为 3 (一般缺陷), 6 次为 2 (严重缺陷), 则油位超标的初始定级为 $\langle 3, 0.73 \rangle$ 和 $\langle 2, 0.27 \rangle$, 即该专家油位超标初始定级为一般缺陷的概率是 0.73, 定级为严重缺陷的概率为 0.27.

2.2 top-k 查询的变压器缺陷等级确定

由于一种故障的变压器缺陷往往有几个专家参与定级, 不同的专家会给出不同的初始定级, 并且初始定级是个不确定数, 而是否安排维修又由初始定级确定, 所以需要通过对不确定性数 top-k 查询来确定初始定级.

2.2.1 不确定数据模型

根据缺陷的种类、缺陷等级和专家参与同类缺陷评级次数 (≥ 3 次), 从数据库中选择任意 5 个专家, 由他们对同一个缺陷变压器进行评级, 得到某变压器 E_i 缺陷的定级模型, 见表 2.

由于专家的评级存在一定的偶然性, 因此该模型是一个不确定性模型.

表 2 某变压器 E_i 缺陷定级数据模型

序号	专家	缺陷初始定级/级	同类缺陷定级次数/次	定级历史准确率/%
s_1	R_1	3	20	73
s_2	R_1	2	20	27
s_3	R_2	3	3	100
s_4	R_3	3	16	40
s_5	R_4	2	30	60

2.2.2 可能世界模型

组成可能世界的限制条件: 同一专家不可能同时把一个缺陷确定为不同等级, 因此可以将同一专家纪录成互斥关系.

由此得到上述数据模型中的元组依赖规则 $s_3, s_1 \oplus s_2, s_4, s_5$, 根据元组依赖规则可以建立如下可能世界模型实例.

每个可能世界模型实例的准确率是其所包含的元组准确率之积, 所有的可能世界模型准确率之和为 1. 所有的可能世界模型如表 3 所示.

表 3 可能世界模型

可能世界	准确率/%	可能世界	准确率/%
$w_1 = \{s_3, s_1, s_4\}$	29.2	$w_5 = \{s_3, s_4\}$	40
$w_2 = \{s_3, s_1, s_5\}$	43.8	$w_6 = \{s_3, s_5\}$	60
$w_3 = \{s_3, s_2, s_4\}$	10.8	$w_7 = \{s_4, s_5\}$	24
$w_4 = \{s_3, s_2, s_5\}$	16.2		

2.2.3 不确定性 U-topK 查询

由于专家的定级存在一定的偶然性, 根据常识, 专家参与定级次数越多, 经验越丰富, 越有可能准确定级, 所以按照专家同类缺陷定级次数进行排序. 在上述可能世界模型中, 对专家缺陷定级进行排序, 选出最有可能的缺陷定级, 见表 4.

表 4 U-topK 查询结果

可能世界	U-top2 查询	准确率/%
$w_1 = \{s_3, s_1, s_4\}$	$\langle s_1, s_4 \rangle$	29.2
$w_2 = \{s_3, s_1, s_5\}$	$\langle s_1, s_5 \rangle$	43.8
$w_3 = \{s_3, s_2, s_4\}$	$\langle s_2, s_4 \rangle$	10.8
$w_4 = \{s_3, s_2, s_5\}$	$\langle s_2, s_5 \rangle$	16.2
$w_5 = \{s_3, s_4\}$	$\langle s_3, s_4 \rangle$	40.0
$w_6 = \{s_3, s_5\}$	$\langle s_3, s_5 \rangle$	60.0
$w_7 = \{s_4, s_5\}$	$\langle s_4, s_5 \rangle$	24.0

即: $P(1,4) = P(w_1) = 0.292$

$$P(1,5) = P(w_2) = 0.438$$

$$P(2,4) = P(w_3) = 0.108$$

$$P(2,5) = P(w_4) = 0.162$$

$$P(3,4) = P(w_5) = 0.400$$

$$P(3,5) = P(w_6) = 0.600$$

$$P(4,5) = P(w_7) = 0.240$$

$P(3,5)$ 和 $P(1,5)$ 的准确率最高, 因此 U-top2 返回的结果为 $\langle 3,5 \rangle$ 和 $\langle 1,5 \rangle$. 结合专家的评级次数, 最终确定 s_5 对应的专家 R_4 的评级结果(等级 2, 严重缺陷)为本次缺陷定级的结果.

3 结 语

在变压器缺陷管理中, 由于专家定级存在不确定性, 而实际生产中又是根据专家的初始定级来安排消除缺陷的维修任务, 因此利用不确定数据 top-k 查询的相关理论, 建立了变压器缺陷专家定级不确定性数据模型——可能世界模型, 利用 top-k 查询方法来确定哪个专家的定级最准确, 并以此作为该缺陷的初始定级, 再根据初始定级结果安排消缺任务. 该模型的建立可为变压器缺陷管理系统中的定级处理提供了一种新方法.

参考文献:

- [1] 朱永利, 吴立增, 李雪玉. 贝叶斯分类器与粗糙集相结合的变压器综合故障诊断[J]. 中国电机工程学报, 2005, 25(10): 159-165.
- [2] 廖瑞金, 廖玉祥, 杨丽君, 等. 多神经网络与证据理论融合的变压器故障综合诊断方法研究[J]. 中国电机工程学报, 2006, 26(3): 119-124.
- [3] 张彼德, 周爱华, 方春恩, 等. 免疫遗传法结合概率因果模型诊断变压器故障[J]. 高电压技术, 2008, 34(11): 2 344-2 348.
- [4] 陈伟根, 钱国超, 黎明. 使用多参量的变压器故障综合诊断技术[J]. 高电压技术, 2007, 33(8): 92-97.
- [5] 周傲英, 金澈清, 王国仁, 等. 不确定性数据管理技术研究综述[J]. 计算机学报, 2009, 32(1): 1-16.
- [6] 李文凤, 彭智勇, 李德毅. 不确定性 Top-k 查询处理[J]. 软件学报, 2012, 23(6): 1 542-1 560.
- [7] SOLIMAN M A, ILYAS IF, CHANG K C C. Top-k query processing in uncertain databases [C] // Proceedings of the 23rd IEEE International Conference on Data Engineering (ICDE'07), 2007: 896-905.

(编辑 胡小萍)

(上接第 517 页)

- [16] 张蓓, 王晓明. 带关键词检索的公钥加密方案[J]. 计算机工程, 2012, 36(6): 155-157.
- [17] LIU Q, WANG G, WU J. An efficient privacy preserving keyword search scheme in cloud computing [C] // Proceeding of 2009 International Conference on Computational Science and Engineering, 2009: 715-720.
- [18] BOLDYREVA A, CHENETTE N, ONEILL A. Order-preserving encryption revisited; improved security analysis and alternative solutions [C] // Proceedings of Advances in Cryptology (CRYPTO'11), Springer, 2011: 578-595.
- [19] HOREB, MEHROTRA S, CANIM M, *et al.* Secure multi-dimensional range queries over outsourced data [J]. The VLDB Journal, 2012, 21(3): 333-358.
- [20] SHI E, VETHENCOURT J, CHAN T, *et al.* Multidimensional range query over encrypted data [C] // Proceedings of the IEEE Symposium on Security and Privacy (SP'07), 2007: 350-364.
- [21] 蔡克, 张敏, 冯登国. 基于单断言的安全的密文区间检索[J]. 计算机学报, 2011, 34(11): 2 093-2 103.

(编辑 胡小萍)