

基于知识库的主动式学习电网运行经验库的研究

刘 晶¹, 周强强¹, 邹庆年¹, 徐 岩², 黄国林²

(1. 广州供电局有限公司 变电管理一所, 广东 广州 510245;

2. 华北电力大学 电气与电子工程学院, 河北 保定 071003)

摘要: 运用面向对象技术与产生式规则知识表达方式, 将电网中的运行经验按工作性质分类, 建立一个拥有主动学习能力的电网运行经验知识库。运用基于权重的贝叶斯分类模型来对统计数据进行分类, 从而识别语义, 以此达到电网运行经验知识与用户的反馈交流的目的。利用知识库强大的学习能力, 使电网运行经验知识库在与用户反馈交流中优化自身知识体系结构, 有地区针对性地进一步扩充与完善运行经验库。该方法可以有效地为工作人员提供技术支持, 提高工作人员的业务素质, 加强电网的安全稳定运行。

关键词: 知识库; 主动式学习; 电网运行经验库

中图分类号: TM732 文献标识码: A DOI: 10.3969/j.issn.1672-0792.2014.04.009

0 引言

电力安全始终是电力运行的重中之重, 为实现电力系统的经济安全运行, 在目前电力系统的元件和自动装置可靠性一定的条件下, 只能努力提高工作人员的业务水平。因此, 建立电网运行经验知识库, 将电网现有的运行经验与电力员工在工作中积累的设备运行维护、故障判断及处理经验等知识编码化、有序化, 加快知识流动, 促进知识共享, 有利于知识管理, 提高企业的竞争力。

文献[1~2]介绍的广西电网电气设备知识库管理系统, 实现了电网企业隐性知识到显性知识的转换, 构建了一个知识共享学习平台, 但该系统提供的学习是单方面的, 即只有人向现有的已存在数据库中的知识学习, 而无系统主动向人学习功能, 无法实现人与系统的反馈交流; 另一方面, 目前广泛应用于电力系统的面向对象技术仅局限于电气元件本身固有属性的继承化与模块化^[3~8], 并没有考虑到元件在电网中运行经验这一外延属性。文献[9]仅将知识库技术应用在变电站监测系统中, 应用范围小。文献[10~

11] 仅是对贝叶斯推理功能的应用, 没有与知识库技术相结合并充分发挥知识库的自主学习能力。本文介绍一种以电网运行经验为电气元件属性, 并能进行主动学习的电网运行经验知识库, 实现电网运行经验与工作人员的反馈交流, 使知识库在运行使用过程中不断自我更新完善。

1 电网运行经验库体系结构设计

知识库系统是构建学习主动式电网运行经验库的重要支撑, 任何领域的专家系统、智能化服务都需要可靠的领域知识, 而主动式电网运行经验库模型的建立也是智能化专题应用的一个重要分支。

1.1 基于知识库的主动式学习电网运行经验库系统

本文从知识库设计的原理入手, 重点研究电网运行经验库的知识表示方法和学习方法, 并且基于主动式服务的要求, 提出了面向主动式电网运行经验库模型的知识库系统应用方法, 最终建立了该模型, 如图1所示。

从图1可以看出, 电网运行经验知识库系统的主要作用是在用户需求到电网运行经验之间进行辅助选择, 通过各种条件下的智能化判断, 给

收稿日期: 2013-09-22。

作者简介: 刘晶(1982-), 女, 工程师, 从事继电保护工作, E-mail: emliu_jing_1130@163.com。

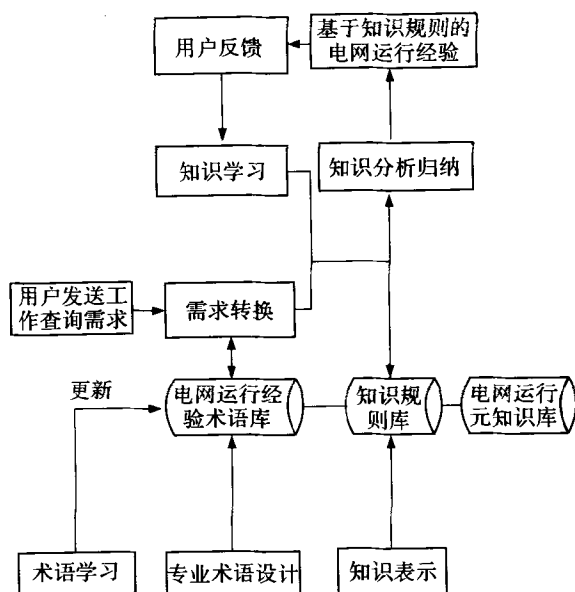


图1 电网运行经验知识库系统结构示意图

出当前知识所认为的最适用的电网运行经验，即基于知识规则的运行经验。主要过程包括以下几个模块：用户发送工作查询需求，利用电网运行经验术语库进行需求转换得到标准的需求术语，知识规则库对需求的处理，知识分析归纳以及方案搜索，最终得到基于知识规则的电网运行经验。最后，由用户对该方案进行评价与信息反馈，知识库系统通过用户反馈信息的学习对知识规则库的规则做出更新。

1.2 知识的表示方法

知识表示可以看作是数据存储结构及其逻辑处理机制的综合，即知识表示 = 数据存储结构 + 逻辑处理机制。因此，知识表示研究的主要问题是设计各种适当的数据结构，研究以知识的形式进行表示的方法、知识表示与逻辑控制的关系、知识表示与知识推理的关系以及知识表示与其他领域的关系。

目前，面向对象知识表达方式以其良好的继承性与鼓励类的重用性，使其在电力系统上有广泛的应用，但仍存在以下局限性：面向对象技术表达的仅仅是电气元件的物理属性，如：电压等级、节点类型等类型的知识。而在实际工作中，工作人员需要与待操作电气设备工作经验之类的相关知识。本文对电网中大量工作经验进行分类，运用面向对象技术将其表达。

(1) 电网运行元知识的表示方式

本文将电网的运行经验按工作性质分为7大类：作业规范、缺陷管理、反事故措施、装置说明书、事故分析与处理、仪器仪表的操作方法、保护配置二次回路典型设计、技术规范。由上述分类可以看出，每类经验都对应一类设备，同一类设备都有相同的运行经验。如图2所示，以变压器为例，在变压器的运行经验属性中必然有这7项，变压器类是双绕组变压器类和三绕组变压器类的父类，可将变压器的通用属性映射为父类表的属性，双绕组和三绕组变压器的特殊属性映射为子类表的属性，设备ID号相同即可保持对象的继承关系。因此，对于电网运行经验本文采用面向对象的知识表达方式^[12]，继承性允许和鼓励类的重用，使系统的程序设计简化，提高创建数据库的速度。

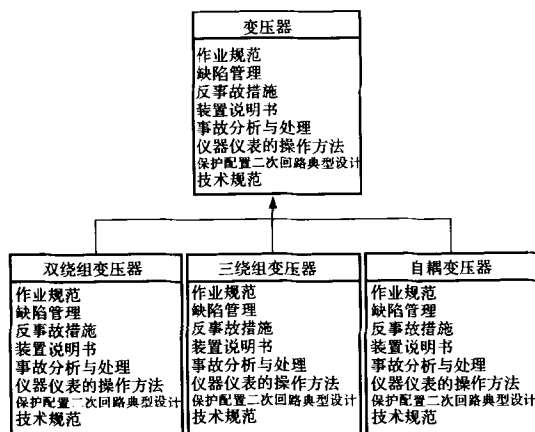


图2 电网运行经验元知识表达方式示意图

(2) 知识规则库的知识表示方法

知识规则库中每一条规则都是该库中的最小粒度知识，每一个知识点都可以作为电网运行经验查询的规则来存储。由于知识规则库中涉及大量的电气元件的运行维护等处理方法，再根据电网运行经验查询的特点，结合知识表示的具体方法，使用产生式规则来表示知识规则库中的规则，即 if 查询术语 then 电网运行经验的方式。如 if 变压器套管头部过热 then 故障分析与处理：采用变铜铝过渡为银铜接触。

2 知识学习及更新策略

随着技术的发展，电网中的设备也会有相应

的变更,新的设备会投入运行,而新设备在投入后的一段时间内,是该设备运行经验的空白期。专家不可能把所有的运行经验都事先录入知识库中,因此,知识库在使用中需要靠与用户的交流来实现自我的更新。

在本知识库系统中,采用的是自主式学习,将经典的学习方法与人工智能相结合。在知识表示的基础上,研究电网运行经验库重点知识的学习方法以及重点知识的更新策略。主要从电网运行经验库知识学习方法、知识系统更新策略两方面进行分析,最后,得到电网运行经验库的学习系统。

2.1 电网运行经验知识库的学习方法

常见的知识学习方法,包括:机械式学习、指导式学习、类比学习、解释学习和示例学习等,这些学习方法在不同的领域中具有各自的优势。电网运行经验库中的知识主要分为两大类:专家知识与用户反馈的知识。专家知识中属于规则一类的可直接存储到知识库中。本文的知识学习主要是针对用户反馈的知识。对这类知识的学习过程,要求系统能主动学习用户的偏好(即用户单位常需要查询使用的电网运行经验),并要能根据用户的反馈对知识进行校正与更新。用户选择某一信息特征频数越多,该信息的权重也就越大,系统推荐给用户的信息会根据权重从大到小给用户做出优先的方案。如果某一信息在一段时间内都没有被用户使用过,则这条信息的权重为零,将该信息放在可删除的一类信息中。这种学习是可重复的,而且对于每一种特征都是相互独立的,即不同的属性不会相互影响,属于离散的单一学习。利用基于参数权重的朴素贝叶斯模型来统计用户反馈的某一知识的频率,以此提高系统的可靠性与客观性。学习方法之间的相互关联关系及其和朴素贝叶斯模型之间的关系如图3所示。

2.2 基于权重的朴素贝叶斯分类模型

通过图1中所示的需求转换环节将用户输入的查询语言,转换为电网运行经验库能识别的专业术语,将每一条术语作为一个词条,这个词条的计算属性为词频与反文档频率;将每一条电网运行经验作为一个小文档,处理电网运行经验实质上是处理一个一个小文本,这样以文本分类的形式来建立贝叶斯模型。

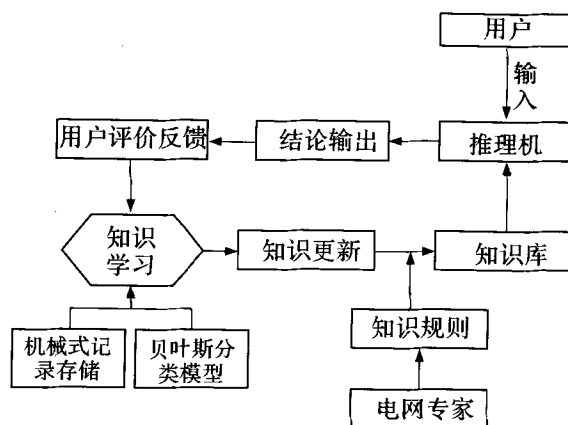


图3 电网运行经验知识库的知识学习方法

2.2.1 朴素贝叶斯模型

对文本分类,假设各个词条 W_i 和 W_j 之间两两独立。设训练样本集分为 k 类,记为 $C = \{C_1, C_2, \dots, C_k\}$, 则每个类 C_i 的先验概率为 $P(C_i)$, $i = 1, 2, \dots, k$, 其值为 C_i 类的样本数除以训练集总样本数 n 。对于新样本 d , 其属于 C_i 类的条件概率是 $P(d/C_i)$ 。

从贝叶斯公式直观看来,文本的类别与特征词的词频、文档频率有很大关系,词频、文档频率越大,说明该词越具有代表性。但实际情况并非如此,词频及文档频率并不足以完全代表某个词在分类中的重要程度。这就导致分类系统不能有效地找到那些能够标识一个类别的特征,从而造成应用朴素贝叶斯的分类效率的低下。

为了有效地表示特征词(电网运行经验库术语)与每个类别的关系,使用了权重来代替词频或文档频率计算特征词属于某类的先验概率,并认为权重较大的特征词对分类的影响较大。

某个词条或某几个词条的组合在一条电网运行经验中出现频率很高,而在其他电网运行经验中出现频率比较低,则认为此词语或此词语的组合有很好的类区别能力,其相应的权重值应该较高。

2.2.2 权重计算方法

目前,常见而又高效率的权重计算方式为TFIDF模型,计算公式为: $TF \times IDF$ 。TF为词频,表示电网运行经验专业术语(词条) i_k 在一条电网运行经验(文档) d 中出现的频率。IDF为反文档频率,计算公式为 $IDF = \log(N/n)$ 。其中, N 为全部文档数, n 表示包含词条 t_k 的文档频率。

由于训练集的偏差, 如, 每类训练集数目不同、训练集文本内容大小不相等, 通过 TFIDF 计算词条对训练集的权重时, 可能会出现某些词条对几类训练集具有相似的权重, 这就不能体现出该词条对其他类别的权重状态。为了解决同一词条对多个类别有类似的贡献度的问题, 需要对 TFIDF 计算所得的权重进行改进。在这里, 采用信息增益 IG 来改进权重, IG 可以用下式表示:

$$IG(t) = - \sum_{i=1}^m P(C_i) \log P(C_i) + P(t) \sum_{i=1}^m P(C_i/t) \log P(C_i/t) + P(\bar{t}) \sum_{i=1}^m P(C_i/\bar{t}) \log P(C_i/\bar{t}) \quad (1)$$

式中: $P(C_i/t)$ 表示文本中出现特征 t 时, 文本属于 C_i 的概率; $P(C_i/\bar{t})$ 表示文本中不出现单词 t 时, 文本属于 C_i 的概率; $P(C_i)$ 表示类别出现的概率; $P(t)$ 表示 t 在整个文本训练集中出现的概率。将权重归一化如下:

$$W(t_i, d_j) = \frac{tf_{ik} \times \log(N/n + 0.01) \times IG_k}{\sqrt{\sum_{k=1}^i (f_{ik} \times \log(N/n + 0.01) \times IG_k)^2}} \quad (2)$$

$W(t_i, d_j)$ 为计算词条 t_i 在类别 d_j 所得的权重, 这样有效地改善了在某类中权重较大而在其他类别中权重较小的词对分类结果的影响。

对于在电气设备下的 7 大属性中的某个属性的电网运行经验文档中, 总会出现的高频词条, 这个词条对类内的文档分类没有任何贡献, 只是能区别不同的设备属性, 采用公式 (3) 来计算先验概率:

$$P(w_j/C_i) = \frac{1 + W(w_j/C_i)}{|V| + \sum_{k=0}^v (w_j/C_i)} \quad (3)$$

式中: $|V|$ 表示特征词表中总词条数; (w_j/C_i) 表示该词条的权重。

2.2.3 基于权重的朴素贝叶斯模型

将权重的概念引入到朴素贝叶斯分类方法中, 则基于权重的朴素贝叶斯分类模型的公式为:

$$P(C_i/d) = P(C_i) \prod_{j=1}^m P(w_j/C_i) \times W(w_j/C_i) = P(C_i) \prod_{j=1}^m \frac{1 + W(w_j/C_i)}{|V| + \sum_{k=0}^v (w_j/C_i)} \times W(w_j/C_i) \quad (4)$$

公式 (4) 基于如下思想: 该词条在某类中先验概率越大, 权重越大, 它对区分文档类别的贡献也就越大^[13]。

2.3 电网运行知识库的更新策略

知识库系统是整个智能系统的重要组成部分, 一般情况下, 知识库的建设过程不是一次就建设完成的, 而是需要多次反复知识修正设置才能完成。建设知识库的核心问题就是如何保证知识库的协调性和正确性, 既不能从知识库中推导出矛盾的结论, 也不能让规则本身出现问题。当出现矛盾时, 必须尽早消除矛盾, 这就是知识库的知识更新。

知识库的更新分为两部分: 专家指导下的知识更新与用户评价反馈后系统的知识更新。专家指导下的知识更新可由知识库的维护人员定期进行。对于用户评价反馈知识的更新采用贝叶斯算法的概率原理, 计算用户评价反馈结果的可靠性, 若可靠性低于设计的最小概率则放弃对此项内容的更新, 反之, 则更新知识。这样既可以保证知识库内的知识与时俱进, 又能为知识库节省出更多的空间。

3 知识库的维护

知识库的维护包括对知识的删除、添加和修改。设定不同的管理权限, 修改权限可对所有用户开放, 而添加与删除权限只为固定的管理员开放。由于本系统具有自主学习能力, 所以知识库的维护相对来说比较简单。电力系统引入新设备后, 可由电网专家向知识库中添加相应的新知识, 并将那些不适用于电网的旧知识删除。而在知识库的日常运行中, 知识库可根据用户的反馈, 运用基于权重的朴素贝叶斯模型, 运用自身的学习能力, 对知识库内的知识进行分类、排序与优化。

4 结论

(1) 利用知识库拥有的学习能力, 实现知识的自主学习。某一条电网运行经验的频率与用户对这条运行经验的评分都可以作为权重, 基于权重的贝叶斯分类模型会重新计算某一运行经验与用户要查询内容的关联度, 并对这些运行经验按

关联度的大小进行排序, 优先向用户提供关联度最大的运行经验。

(2) 实现人与机器的反馈交流, 机器主动学习人的偏好。

(3) 电网运行经验库的运行与完善同步进行。电网运行经验库的使用, 即是运行经验库的优化与维护过程。

(4) 运用面向对象的知识表达方法, 把电网运行经验作为设备的属性, 突破以前只能将电气设备的物理特点作为设备属性的现状。可快速为工作人员提供电气设备相关工作经验。

(5) 电网运行经验知识库在长期的使用过程中, 能形成具有对地区电网的运行维护、故障处理、缺陷处理等知识有针对性的知识库。

参考文献:

- [1] 韦秋丽, 张蕾, 甘一君, 等. 广西电网电气设备知识库管理系统推介 [J]. 广西电力, 2010, 33 (6): 30-32.
- [2] 韦秋丽, 张蕾, 甘一君, 等. 电网电气设备知识库管理系统的设计与实现 [J]. 广西电力, 2010, 33 (5): 3-5.
- [3] 王萍, 罗颖昕, 杨培龙, 等. 基于面向对象的知识库的电网图智能 CAD 系统 [J]. 电力系统自动化学报, 2004, 16 (5): 9-13.
- [4] 范文涛, 薛禹胜, 慕志恒. 面向对象技术及其在电力系统中的应用 [J]. 电力系统自动化, 1998, 12 (22): 72-76.
- [5] 邹燕, 刘金官, 莫来恩, 等. 面向对象技术在调度员培训系统中的应用 [J]. 电网技术, 1998, 22 (1): 22-24.
- [6] 杨正东. 配电网知识库的建立 [D]. 保定: 华北电力大学, 2011.
- [7] 刘世欣. 基于故障树的变电设备故障诊断专家系统 [D]. 保定: 华北电力大学, 2006.
- [8] 刘玮, 曹丽娟, 余南华, 等. 基于知识库的电力培训网络考评系统设计与实现 [J]. 中国电力教育, 2013, 2.
- [9] 杨皓然, 刘琦, 郑连清. 变电站智能告警专家系统设计 [J]. 电力科学与工程, 2010, 27 (5): 7-10, 21.
- [10] 赵建立, 高会生, 赵生岗. 贝叶斯网络在可靠性评估中的应用 [J]. 电力科学与工程, 2008, (2): 51-53.
- [11] 李鹏鹏, 彭显刚, 孟安波, 等. 小波贝叶斯神经网络在冲击负荷地区短期负荷预测中的应用 [J]. 电力科学与工程, 2012, 28 (11): 7-12.
- [12] 万小云. 面向对象知识库管理系统中知识对象模型的设计 [J]. 上海海运学院学报, 2000, 21 (1): 53-59.
- [13] 代磊, 马卫东, 王凌楠, 等. 基于权重的朴素贝叶斯分类器设计与实现 [J]. 情报理论与实践, 2008, 31 (3): 440-442.

Study of Knowledge-based Active Learning of Operation Experience Base of Grid

Liu Jing¹, Zhou Qiangqiang¹, Zou Qingnian¹, Xu Yan², Huang Guolin²

(1. Guangzhou Power Supply Bureau Co., Ltd., Guangzhou 510245, China; 2. School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China)

Abstract: By using of object-oriented technology and the knowledge representation of the production rule, this paper classifies the operation experience of grid according to the nature and builds an operation experience of grid with active learning capability. Through application of Bayesian classifier model that based on weight, it classifies the statistical data and identifies the semantic, thus it realizes the exchange of the operation experience of grid and the feedback of the user. Using the powerful learning ability of knowledge base, it can make the operation experience of the grid optimize the knowledge system structure of itself in the process of exchanging the operation experience and the feedback of the user, and improve and expand the operation experience of grid further which is region targeted. This method can provide technical support and improve the quality of the stuff, so it can strengthen the safety stability operation of the grid.

Keywords: knowledge base; active learning; operation experience of grid