文章编号:1671-4598(2017)05-0035-03 DOI:10.16526/j.cnki.11—4762/tp.2017.05.011 中图分类号:TP3 文献标识码:A

融合语义网的故障检索系统设计与构建

任 勇
(南京电子技术研究所,南京 210039)

摘要:针对雷达故障数据和经验性技术支持文档的信息提取,提出了一种融合语义网技术的故障检索系统的设计与实现方法;该系统采用故障树组织结构化的历史故障数据,将故障树转化为二元决策图从而获取顶层故障事件的最小割集,而后通过语义匹配获取精确的故障原因信息;同时,借助开源搜索引擎 Lucene 检索技术文档库,获得与用户检索条目相关的支持文档;其中,为了提升中文分词和文本匹配的精确性,该系统还利用本体语言构建了专业词汇语义网,考量了词语之间的语义相关度;该系统已应用于某型机载有源相控阵雷达电源单元的故障检索,取得了很好的技术辅助效果。

关键词:故障检索;语义网;故障树;二元决策图;开源搜索引擎

Design and Implementation of a Fault Searching System Combined with Semantic Web

Ren Yong
(Nanjing Research Institute of Electronics Technology, Nanjing 210039, China)

Abstract: As for the information retrieval of radar fault data and empirically technical support documents, we propose a design and implementation method of a fault searching system combined with semantic web technology. The system uses fault trees to organize structurized historical fault data, and converts them into binary decision diagrams to obtain minimal cut sets of top fault events, and then acquires accurate fault reasons via semantic matching. Meanwhile, it obtains support documents related to users' query items by using an open source search engine Lucene to search from technical document database. In the system, it uses ontology language to build a semantic web of terminologies, for improving the accuracy of Chinese words segmentation and text matching, so as to involve semantic correlation between different words. The system has been used in the fault searching of the power unit of an AESA radar and has obtained good effect for technological assistance.

Keywords: fault searching; semantic web; fault tree; binary decision diagram; Lucene

0 引言

现代雷达系统日趋复杂,对故障诊断和相关技术支持信息的检索提出了更高的要求[1-2]。自然语言处理技术,特别是语义网技术,为传统经验性故障诊断的自动检索提供了新的思路,可提升故障定位和排除的效率。

目前,交互式电子手册和专家系统在雷达的故障诊断中已有广泛的应用。交互式电子手册(interactive electronic technical manual, IETM)是具备交互功能的信息支持软件,突破了纸质技术文档的限制,能够为装备保障人员提供高效的技术信息支持[3]。IETM 通常包括数据模块、公共源数据库等组件,以交互式软件的形式将多媒体信息精确地呈现给维修或操作人员。然而,IETM 更多的是对产品技术文档进行标准化组织和可视化呈现,对全周期产品维修保障中的故障信息积累与查询并未涉及。专家系统是人工智能的一个分支领域,它将某一领域的专家知识和经验构建成知识库,通过推理机模拟人类专家的问题求解和决策过程[4]。在雷达专家系统的设计中,通常使用故障树来组织历史故障数据,使用产生式表示法表示知识。然而,专家系统的知识库构建一般需要大规模的人工编制,不能充分地利用现有的经验性数据和技术文档。

此外,自然语言处理技术在故障诊断系统中的应用也日益广泛,不过现有的方法并没有过多地考量专业词汇语义在信息检索中的重要作用[5]。

本文提出了一种雷达故障信息检索系统,利用故障树组织历史故障数据,为用户提供最直接的数据支持。根据用户的检索请求,该系统能够对经验性技术文档按照相关度进行排序,供用户参考。在文本语义匹配方面,该系统利用本体语言将专业词汇组织成语义网,从而提升检索的准确性。

1 雷达故障检索系统结构

雷达故障检索系统对用户输入的故障问题检索条目进行解析,首先对其进行中文分词,然后按照设定的规则解析为结构化的检索请求。系统包含故障树数据库和技术文档数据库,作为排故信息检索的数据来源。其中,语义网为中文分词和文本匹配做支撑。检索系统的架构如图 1 所示。

2 故障问题检索条目解析

故障问题条目的解析是检索的第一步,对搜索返回的结果集有很大影响。用户在检索框输入的通常为自然语言形式的条目,例如“数字子阵电源绝缘不好”,因而需要对其进行结构化解析。

2.1 中文分词

对故障检索条目进行分词是进一步分析的基础。该系统采

收稿日期:2016-11-18; 修回日期:2017-01-05。

作者简介:任 勇(1989-),男,河北邯郸人,助理工程师,工学硕士,主要从事雷达故障检索技术方向的研究。

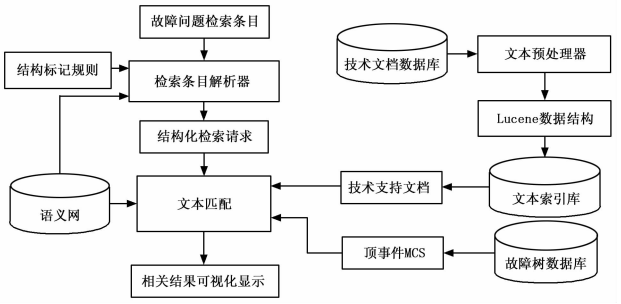

图 1 故障检索系统框架

用中国科学院计算技术研究所开发的 ICTCLAS 中文分词系统，该程序基于层叠隐马尔科夫模型进行汉语词法分析，在中文分词、词性标注等方面具有卓越的性能[6]。然而，雷达领域含有大量的分词系统无法有效切分的未登录专业词汇或缩略词，此外，企业内部往往也会有产品代号或约定俗成的术语，这给识别主体词造成了困难。这里采用本体语言语义网组织雷达专业词汇。

2.2 专业术语的本体表示

用本体语言表示专业术语的优势在于不仅可以表示事物或事物群，而且可以表示事物之间的关系。这样可以清晰地表示雷达产品中层级式的系统组成关系，而且还可以将专业术语与其同义的缩略词等相关联。雷达专业词汇本体建模方式包括以下几种：

（1）类和实例，例如 ClassAssertion（：雷达 ：某型机载有源相控阵雷达），表示归属关系；

（2）类层次关系，例如 SubClassOf（：相控阵雷达 ：雷达），表示类和子类的关系；

（3）对象属性，例如 ObjectPropertyAssertion（：hasUnit ：某型机载有源相控阵雷达 ：电源单元），表示个体与个体之间如何关联；

（4）等价和非等价个体，例如 SameIndivi －dual（：固态放大链 ：固放），表明两个名称指向同一个个体，或是排除两个名称是相同个体。

例如，与“数字子阵电源”相关的本体建模实例如下：

ClassAssertion（：组件 ：数字子阵电源）

ObjectPropertyAssertion（：hasUnit ：电源单元 ：数字子阵电源）

SameIndividual（：数字子阵电源 ：数字子阵）

由此可构建用于雷达故障检索系统的语义网，从而构建一个专业性的知识体系，该系统语义网的部分本体及关系如图 2 所示。

将通过本体语言组织的雷达专业词汇加入分词系统的用户词典，可以提高分词的准确性。例如，“数字子阵电源绝缘不好”经过分词，可得到“数字子阵电源/绝缘/不好”。此外，本系统还包含了一个小型常用汉语词汇语义网，用于补充词义解析。鉴于雷达产品的层级式结构，需要将检索条目进一步按照产品、单元（分系统）、组件、故障问题来解析。因此，需要进行文本相似度计算。

2.3 文本相似度计算

计算文本相似度的常用方法有夹角余弦、最长公共子串等

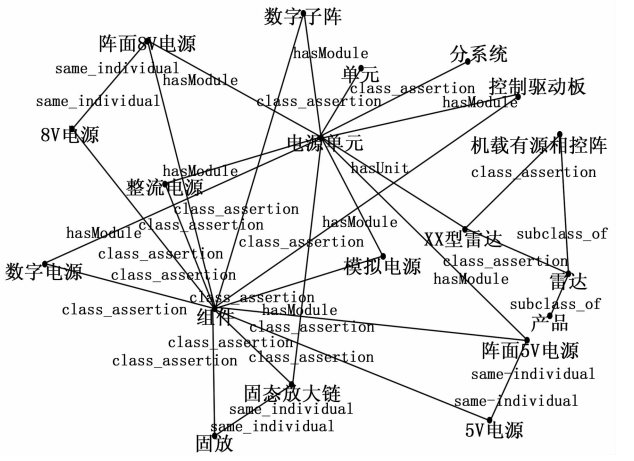

图 2 专业词汇语义网实例

算法。由于夹角余弦没有体现词序信息，故采用经过改进的最长公共子串来衡量文本的相似度。最长公共子串利用动态规划进行计算，同时根据本体建模加入词语之间的语义关系，用于考量词汇之间的语义相似度。

算法 1：加入语义匹配的最长公共子串算法

```
输入:经过分词的词序列 s1 和 s2
输出:s1 和 s2 的最长公共子串
m ← length(s1)
n ← length(s2)
for i ← 1 to m
for j ← 1 to n
    if IsSimilar(s1[i−1], s2[j−1]) then
      num[i][j] ← 1 + num[i−1][j−1]
    else
     num[i][j] ← Max(num[i−1][j],
num[i][j−1])
while m ! = 0 and n ! = 0
   if IsSimilar(s1[m−1], s2[n−1]) then
    result.append(s1[m−1])
    m−−
    n−−
   else if num[m][n−1] >= num[m−1][n] then
    n−−
else
   m−−
Reverse(result)
return result
```

其中判定两个词汇相似的函数 IsSimilar 还根据词汇语义网进行考量，具有 SameIndividual 特性的两个词汇也会判定两个词汇相似。

获得两个字符串的最长公共子串后可用下式计算文本相似度：

$$Sim(s_1,s_2)=\frac{l_{LCS}(W_1s_1,W_2s_2)}{\min(l_{W_1s_1},l_{W_2s_2})}.$$

式中，W_1、W_2 为两个字符串的加权系数矩阵，因为在字符串中数字往往承载着重要信息，因而将数字词的权重设定为普通词汇的两倍；l_{LCS} 为最长公共子串的长度。由此可计算得到归一化的两个字符串文本相似度。表 1 列出了检索字符串的相似

度值，可见上述方法对于计算文本的相似度是有效的。

2.4　检索条目结构化解析

用户键入的检索语句通常会包含有关雷达产品的信息，一般按照“产品－单元（分系统）－组件－故障问题”的顺序，可能在某些层级会有缺失。检索条目结构化解析首先需要对语句进行中文分词，并滤去部分不重要的助词等，然后按顺序将分词所得词语与专业术语库中的本体实例进行文本相似度计算，将相似度超过阈值（例如大于 0.75）的词语按照本体类别进行判定。经过以上处理后，用户检索语句会得到结构化解析，例如，“数字子阵电源绝缘不好”被解析为＜product＝"" ＞＜unit＝"" ＞＜module＝" 数字子阵电源" ＞＜fault＝" 绝缘不好" ＞，可用于后续进一步分析。

表 1　文本相似度计算实例

字符串 1	字符串 2	相似度
阵面 8V 电源无法正常开机	8V 不能开机	1.00
电源单元固态放大链功率不达标	固放频谱指标不达标	0.60
模拟电源无法下载程序	数字电源通信软件无法烧写	0.25
电源单元的 18 端口接触不良	电源单元的 15 端口安装问题	0.43

3　故障树信息检索

雷达历史故障数据可通过故障树形式组织。故障树分析（Fault Tree Analysis，FTA）通过树状逻辑关系对系统的失效原因进行诊断，而且可以计算系统发生故障的概率[7]。

3.1　故障树建模

以某型机载有源相控阵雷达电源单元为例，FTA 可将历史积累的故障数据按照系统结构的逻辑关系进行组织，以获知系统运行的薄弱环节，进而指导维修和检测。电源单元的一个故障树实例如图 3 所示。故障树底事件失效概率由发生频次计算得出，发生次数越多的事件失效概率越高。

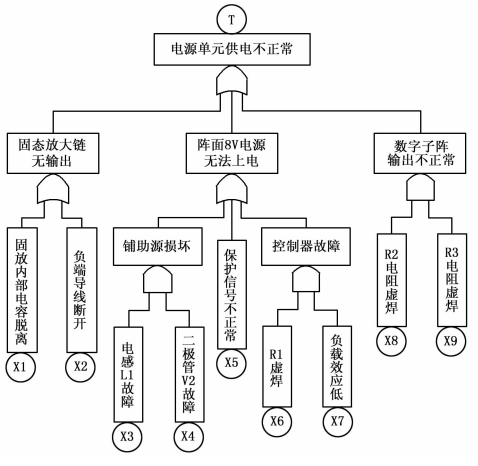

图 3　某型雷达电源单元故障树实例

3.2　故障树分析

故障树构建完成后，需要对其进行分析以便于检索，从而为用户提供有效的排故信息。故障树分析一种常用的方法是使用二元决策图（Binary Decision Diagram，BDD），使用 ite（if － then－else）将故障树转化为 BDD 结构，再通过遍历 BDD 结构直接获取割集[7-8]。故障树转化为 BDD 结构的规模与底事件的排序有关，而寻找最优排序属于 NP 完全问题[8]。

为了获取较小规模的树结果，在转化之前对底事件的排序依次做如下处理[7]：（1）计算故障树结构中底事件到顶事件的距离，距离较短者排在靠前的位置；（2）重复次数较多的底事件排在靠前的位置；（3）故障树结构中影响同一个中间事件的底事件在转化中应排在靠近的位置。根据以上的排序准则，图 3 所示的故障树实例底事件的排序为 X1＜X2＜X5＜X8＜X9＜X3＜X4＜X6＜X7。

利用 ite 将图 3 所示的故障树转化为 BDD 结构如图 4 所示。

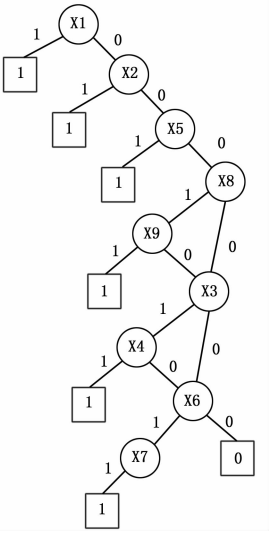

图 4　BDD 结构

根据故障树的 BDD 结构可求解最小割集（MCS）[9]。根据 MCS 定义，首先对 BDD 进行遍历可获得从根节点到终节点的解，然后找出所有子集均不是解的 MCS，对于图 3 而言即为 {X1}、{X2}、{X5}、{X3，X4}、{X6，X7}、{X8，X9}。

3.3　故障树检索

构建故障树的最终目的在于排故指引，能通过故障条目检索到问题发生的原因，即某个顶事件的最小割集。检索过程的实现包括以下步骤：（1）对检索条目进行结构化解析，明确故障归属的产品、单元、组件等信息；（2）顺序搜索故障树顶事件，通过文本相似度计算进行语义匹配；（3）显示与检索目标匹配的故障树顶事件最小割集及相应发生概率。

4　技术支持文档检索模块

故障树归属于结构化数据，而产品在使用和维修过程中还会产生大量的经验性专业技术文档，例如维修手册、产品说明书等，这些文档对于故障排除都有参考价值，因而系统应包含技术文档数据库的检索。

本系统采用开源搜索引擎 Lucene 进行全文检索。Lucene 运行主要包括两部分，即建立索引库和检索索引库。在创建索引库过程中，Lucene 首先对原始文本进行分词，形成 Token 流（Token 是 Lucene 中定义的词语的抽象概念），本系统采用 ICTCLAS 进行中文分词，然后，原始文档就可映射到 Lucene 的存储接口中。索引库构建完成后，就可对其进行检索，对于

（下转第 41 页）

连接模块，登陆模块，数据异常点计模块，告警模块组成。对编号为 0001 的光伏组件进行遮挡试验，过 2 分钟就能在软件界面显示异常信息，显示数据准确，编号相符，符合设计要求。

图 8　PC 终端软件监控界面

6　结论

该设计通过检测旁路二极管温度的方法，间接的监控了 PV 组件中单体电池失效的情况，方法简单独特，并通过 ZigBee 自组网的方式，实现数据实时上传到服务器，解决了当前光伏电站安装偏远区域监控不便的问题，并能及时发现、定位失效组件故障，在一定程度上减轻了维护压力，减少了电站维护费用，并在设计中充分注意了数据异常的算法，最后验证了系统监测稳定、可靠。

参考文献：

[1] 熊志金．基于无线传感器网络的列车货物安全监测系统 [J]．计算机测量与控制，2012 (8)：2102－2104.

[2] 陈建新．用于固体激光器泵浦的大功率脉冲 LD 驱动电路的设计与实现 [D]．泉州：华侨大学，2013.

[3] 翟载腾，程晓舫，丁金磊，等．被部分遮挡的串联光伏组件输出特性 [J]．中国科学技术大学学报，2009 (4)：398－402.

[4] 海　涛，梁挺兴，黄曲达，等．一种高效的光伏监控方案及发电量预测 [J]．计算机测量与控制，2015 (8)：2637－2639.

[5] 张经炜，丁　坤，卞新高，等．一种户外光伏组件测试平台研制 [J]．电子测量技术，2013 (7)：93－96.

(上接第 37 页)

检索字段的文本分析必须采用相同的分词系统。Lucene 系统会根据检索字段对存储结构中的文档进行排序，从而返回最优的匹配结果。

Lucene 的文档评分算法如下[10]：

$$Score = \sum_{t\_in\_q} [tf(t\_in\_d) \cdot idf(t)/norm(d,t) + tf(t\_in\_q) \cdot idf(t)/norm(q)]$$

其中：$tf(t\_in\_d)$ 表示词 t 在文档 d 中的词频；$idf(t)$ 表示 t 在整个文档库中的倒排频率；$norm(d, t)$ 表示查询文档中 t 的权值；$tf(t\_in\_q)$ 表示词 t 在查询字段中的词频；$norm(q)$ 表示查询字段中查询词的权值。

经过评分和排序，Lucene 模块便会返回匹配检索要求的文档。

5　试验结果与分析

该故障检索系统按照上述的构建方式从故障树数据库和技术支持文档数据库获取相关信息，列出与用户输入的搜索条目匹配的结果。例如，搜索条目为“阵面 8V 互锁”，则检索系统会显示以下结果：

故障出现的可能原因：

监控模块故障　　出现概率：37.5%

输出保护模块故障　　出现概率：37.5%

辅助源上管理电源开机的相应电路　　出现概率：25%

请参考以下文档：

《某型雷达电源阵面 8V 互锁故障分析及处理方法》；

《某型雷达电源单元互锁功能失效故障》。

经过测试，检索系统可以获得相关度较高的故障信息，从而可用来辅助排故。故障树分析模块根据失效概率对检索到顶事件的 MCS 进行排序，Lucene 文本检索模块根据算法评分对相关技术支持文档进行排序。从检索结果的相关度可见，语义网的加入能够提高搜索精度。

6　结论

本文在融合语义网的基础上提出了一种故障检索系统的构建方法。该检索系统将产品研制、使用及维修过程中产生的历史故障数据以故障树形式组织，并将故障树转化为 BDD 结构以快速获得顶事件的最小割集，从而指导产品维修和排故。在文本语义的处理方面，该系统针对某型雷达建立了专业词汇语义网，对于提高文本语义匹配的准确性有重要意义。同时融合开源文本搜索引擎，用于检索技术支持文档，从而为故障排除提供综合参考信息。

在这种故障检索系统构建方法的基础上，随着大数据的积累，将来可增加知识推理模块，提高诊断和排故指引的智能化程度。

参考文献：

[1] 彭　为．装备一体化精益保障架构及发展趋势 [J]．现代雷达，2013，35 (8)：2－4.

[2] 王　众，曾　静．复杂电子装备军地一体化保障平台建设 [J]．国防科技，2016，37 (1)：82－85.

[3] 吴永明，叶海生．基于 IETM 的装备故障诊断系统技术研究 [J]．计算机测量与控制，2011，19 (10)：2377－2379.

[4] 孙福安，刘辉峰，段方振．一种雷达故障诊断专家系统设计 [J]．现代雷达，2014，36 (9)：74－48.

[5] 陈　勇，王昌明．基于自然语言理解的故障诊断方法研究 [J]．计算机测量与控制，2012，20 (3)：610－613.

[6] 刘　群，张华平，俞鸿魁，等．基于层叠隐马模型的汉语词法分析 [J]．计算机研究与发展，2004，41 (8)：1421－1429.

[7] 闵　苹，童节娟，奚树人．利用二元决策图求解故障树的基本事件排序 [J]．清华大学学报 (自然科学版)，2005，45 (12)：1646－1649.

[8] 高　巍，张琴芳．基于二叉决策图的故障树求解法 [J]．核技术，2011，34 (10)：791－795.

[9] Kohda T. A Simple Method to Derive Minimal Cut Sets for a Non－coherent Fault Tree [J]. International Journal of Automation and Computing, 2006, 2, 151－156.

[10] 周登朋，谢康林．Lucene 搜索引擎 [J]．计算机工程，2007，33 (18)：95－118.