

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2016 年 12 月 1 日 (01.12.2016)



(10) 国际公布号
WO 2016/188279 A1

- (51) 国际专利分类号:
G06F 11/22 (2006.01)
- (21) 国际申请号: PCT/CN2016/080015
- (22) 国际申请日: 2016 年 4 月 22 日 (22.04.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510272657.8 2015 年 5 月 25 日 (25.05.2015) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 英属开曼群岛大开曼乔治城资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。
- (72) 发明人: 及
- (71) 申请人 (仅对美国): 刘迅 (LIU, Xun) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.); 中国北京市朝阳区小关北里甲 2 号渔阳置业大厦 B 座 605, Beijing 100029 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE,

[见续页]

(54) Title: GENERATING METHOD AND DEVICE FOR FAULT SPECTRA, AND DETECTING METHOD AND DEVICE BASED ON FAULT SPECTRA

(54) 发明名称: 一种故障谱的生成、基于故障谱的检测方法和装置

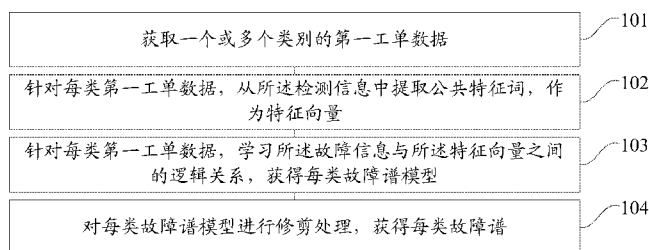


图 1

- 101 Acquire one or more categories of first work order data
102 Extract a public feature word from detection information as a feature vector aiming at each category of first work order data
103 Learn a logic relationship between fault information and the feature vector aiming at each category of first work order data, and acquire a model about each category of fault spectra
104 Perform trimming treatment on the model about each category of fault spectra, and acquire each category of fault spectra

(57) Abstract: A generating method and device for fault spectra, and a detecting method and device based on fault spectra. The generating method comprises: acquiring one or more categories of first work order data (101), wherein each piece of first work order data comprises fault information and detection information; extracting a public feature word from the detection information as a feature vector aiming at each category of first work order data (102); learning a logic relationship between the fault information and the feature vector aiming at each category of first work order data, and acquiring a model about each category of fault spectra (103); and performing trimming treatment on the model about each category of fault spectra, and acquiring each category of fault spectra (104). By establishing fault spectra, concurrency detection is supported to be performed according to one or more detecting paths during follow-up detection, thereby reducing the time consumed for detection, and increasing detection efficiency; and meanwhile, the detection operation for applying the fault spectra is simple, thereby greatly reducing the participation frequency of manpower, and reducing the consumption of vigor of a user.

(57) 摘要:

[见续页]



WO 2016/188279 A1



IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO,
RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD,
TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种故障谱的生成、基于故障谱的检测方法和装置，该生成方法包括：获取一个或多个类别的第一工单数据(101)；每个第一工单数据中包括故障信息与检测信息；针对每类第一工单数据，从所述检测信息中提取公共特征词，作为特征向量(102)；针对每类第一工单数据，学习所述故障信息与所述特征向量之间的逻辑关系，获得每类故障谱模型(103)；对每类故障谱模型进行修剪处理，获得每类故障谱(104)。通过建立故障谱，使得后续检测时支持并发依据该一个或多个检测路径进行检测，减少了检测耗时，提高检测的效率，同时，应用故障谱的检测操作简单，大大减少了人工参与的频次，减少用户精力的耗费。

一种故障谱的生成、基于故障谱的检测方法和装置

技术领域

本申请涉及计算机的技术领域，特别是涉及一种故障谱的生成方法、一种基于故障谱的检测方法、一种故障谱的生成装置和一种基于故障谱的检测装置。

背景技术

随着科技的快速发展，各种产品，如虚拟主机、云平台等等，广泛进入人们的生活、学习、工作等领域。

通常，在产品出现故障时，用户会向工单系统提交工单，进行检测、维护，进而解决故障。

现有的工单系统主要由两个子系统组成：自主解答系统和客服解答系统。

在工单系统中，用户需要自己查阅帮助中心文档或根据向导提示来解决故障。

由于用户需要根据文档或提示一步一步操作排查，即串行排查，耗费较多的时间，故障检测的速度慢；并且，工单系统所累积的技术文档一般数量很多，操作复杂，需要耗费用户大量的精力；此外，阅读技术文档需要多需对领域内的知识有积累，技术门槛较高，对于技术功底弱的用户或客服很难独自解决问题。

发明内容

鉴于上述问题，提出了本申请实施例以便提供一种克服上述问题或者至少部分地解决上述问题的一种故障谱的生成方法、一种基于故障谱的检测方法和相应的一种故障谱的生成装置、一种基于故障谱的检测装置。

为了解决上述问题，本申请实施例公开了一种故障谱的生成方法，包括：获取一个或多个类别的第一工单数据；每个第一工单数据中包括故障信

息与检测信息;

针对每类第一工单数据,从所述检测信息中提取公共特征词,作为特征向量;

5 针对每类第一工单数据,学习所述故障信息与所述特征向量之间的逻辑关系,获得每类故障谱模型;

对每类故障谱模型进行修剪处理,获得每类故障谱。

优选地,所述故障谱中包括相连的根节点与叶子节点,所述根节点表征故障信息,所述叶子节点表征检测信息,至少部分叶子节点之间具有逻辑关系,所述叶子节点具有一个或多个父节点。

10 优选地,所述从所述检测信息中提取公共特征词的步骤包括:

对所述检测信息进行分词处理,获得一个或多个第一分词;

统计所述第一分词的词频;

通过所述第一分词的词频计算所述第一分词的权重;

按照所述权重提取至少部分第一分词作为公共特征词。

15 优选地,所述从所述检测信息中提取公共特征词的步骤还包括:

采用所述一个或多个第一分词在预置的停用词库中进行匹配;

移除匹配成功的第一分词。

优选地,所述对每类故障谱模型进行修剪处理的步骤包括:

20 在所述故障谱模型中查找相同的子树;所述子树为一个或多个叶子节点的集合;

当查找到时,将相同的子树的父节点连接至其中一个子树;

在相同的子树中,剪去已连接的子树之外的其他的子树。

优选地,所述对每类故障谱模型进行修剪处理的步骤还包括:

按照预设的剪枝方式对所述故障谱模型进行修剪处理。

25 优选地,所述对每类故障谱模型进行修剪处理的步骤还包括:

从所述故障谱模型剪去逻辑关系不合法的叶子节点。

本申请实施例还公开了一种基于故障谱的检测方法,包括:

当接收到第二工单数据时，从所述第二工单数据中提取关键词；
查找所述第二工单数据所属类别对应的故障谱；
在所述故障谱中，根据所述关键词查找一个或多个检测路径；
依据所述一个或多个检测路径进行检测，获得检测结果。

- 5 优选地，所述故障谱中包括相连的根节点与叶子节点，所述根节点表征故障信息，所述叶子节点表征检测信息，至少部分叶子节点之间具有逻辑关系，所述叶子节点具有一个或多个父节点。

优选地，所述从所述第二工单数据中提取关键词的步骤包括：
对所述第二工单数据进行分词处理，获得一个或多个第二分词；
10 识别所述一个或多个第二分词的词性；
按照所述词性从所述一个或多个第二分词中提取关键词。
优选地，所述从所述第二工单数据中提取关键词的步骤还包括：
采用所述一个或多个第一分词在预置的停用词库中进行匹配；
移除匹配成功的第二分词。

- 15 优选地，所述在所述故障谱中，根据所述特征词查找一个或多个检测路径的步骤包括：

在所述故障谱中，查找与所述关键词匹配的根节点；
遍历与所述根节点相连的一个或多个叶子节点，获得一个或多个检测路径。

- 20 优选地，所述依据所述一个或多个检测路径进行检测，获得检测结果的步骤包括：

针对每个检测路径，获取所述检测路径中的一个或多个叶子节点表征的检测信息；

按照当前叶子节点表征的检测信息进行检测，获得候选检测结果；

- 25 查找逻辑关系与所述候选检测结果匹配的下一叶子节点，返回执行按照当前叶子节点表征的检测信息进行检测的步骤，直至执行至最终的叶子节点；

将最终的叶子节点的候选检测结果设置为检测结果。

优选地，所述故障谱通过以下方式生成：

获取一个或多个类别的第一工单数据；每个第一工单数据中包括故障信息与检测信息；

5 针对每类第一工单数据，从所述检测信息中提取公共特征词，作为特征向量；

针对每类第一工单数据，学习所述故障信息与所述特征向量之间的逻辑关系，获得每类故障谱模型；

对每类故障谱模型进行修剪处理，获得每类故障谱。

10 本申请实施例还公开了一种故障谱的生成装置，包括：

工单数据获取模块，用于获取一个或多个类别的第一工单数据；每个第一工单数据中包括故障信息与检测信息；

公共特征词提取模块，用于针对每类第一工单数据，从所述检测信息中提取公共特征词，作为特征向量；

15 故障谱模型学习模块，用于针对每类第一工单数据，学习所述故障信息与所述特征向量之间的逻辑关系，获得每类故障谱模型；

故障谱模型修剪模块，用于对每类故障谱模型进行修剪处理，获得每类故障谱。

20 优选地，所述故障谱中包括相连的根节点与叶子节点，所述根节点表征故障信息，所述叶子节点表征检测信息，至少部分叶子节点之间具有逻辑关系，所述叶子节点具有一个或多个父节点。

优选地，所述公共特征词提取模块包括：

第一分词处理模块，用于对所述检测信息进行分词处理，获得一个或多个第一分词；

25 词频统计模块，用于统计所述第一分词的词频；

权重计算模块，用于通过所述第一分词的词频计算所述第一分词的权重；

第一分词提取子模块，用于按照所述权重提取至少部分第一分词作为公

共特征词。

优选地，所述公共特征词提取模块还包括：

第一匹配子模块，用于采用所述一个或多个第一分词在预置的停用词库中进行匹配；

5 第一移除子模块，用于移除匹配成功的第一分词。

优选地，所述故障谱模型修剪模块包括：

子树查找子模块，用于在所述故障谱模型中查找相同的子树；所述子树为一个或多个叶子节点的集合；

10 连接子模块，用于在查找到时，将相同的子树的父节点连接至其中一个子树；

第一修剪子模块，用于在相同的子树中，剪去已连接的子树之外的其他的子树。

优选地，所述故障谱模型修剪模块还包括：

15 第二修剪子模块，用于按照预设的剪枝方式对所述故障谱模型进行修剪处理。

优选地，所述故障谱模型修剪模块还包括：

第三修剪子模块，用于从所述故障谱模型剪去逻辑关系不合法的叶子节点。

20 本申请实施例还公开了一种基于故障谱的检测装置，包括：

关键词提取模块，用于在接收到第二工单数据时，从所述第二工单数据中提取关键词；

故障谱查找模块，用于查找所述第二工单数据所属类别对应的故障谱；

25 检测路径查找模块，用于在所述故障谱中，根据所述关键词查找一个或多个检测路径；

检测模块，用于依据所述一个或多个检测路径进行检测，获得检测结果。

优选地，所述故障谱中包括相连的根节点与叶子节点，所述根节点表征故障信息，所述叶子节点表征检测信息，至少部分叶子节点之间具有逻辑关

系, 所述叶子节点具有一个或多个父节点。

优选地, 所述关键词提取模块包括:

第二分词处理子模块, 用于对所述第二工单数据进行分词处理, 获得一个或多个第二分词;

5 词性识别子模块, 用于识别所述一个或多个第二分词的词性;

第二分词提取子模块, 用于按照所述词性从所述一个或多个第二分词中提取关键词。

优选地, 所述关键词提取模块还包括:

10 第二匹配子模块, 用于采用所述一个或多个第一分词在预置的停用词库中进行匹配;

第二移除子模块, 用于移除匹配成功的第二分词。

优选地, 所述检测路径查找模块包括:

根节点匹配子模块, 用于在所述故障谱中, 查找与所述关键词匹配的根节点;

15 叶子节点遍历子模块, 用于遍历与所述根节点相连的一个或多个叶子节点, 获得一个或多个检测路径。

优选地, 所述检测模块包括:

检测信息获取子模块, 用于针对每个检测路径, 获取所述检测路径中的一个或多个叶子节点表征的检测信息;

20 候选检测结果获取子模块, 用于按照当前叶子节点表征的检测信息进行检测, 获得候选检测结果;

叶子节点查找子模块, 用于查找逻辑关系与所述候选检测结果匹配的下一叶子节点, 返回调用候选检测结果获取子模块, 直至执行至最终的叶子节点;

25 检测结果设置子模块, 用于将最终的叶子节点的候选检测结果设置为检测结果。

优选地, 所述故障谱通过调用以下模块生成:

工单数据获取模块, 用于获取一个或多个类别的第一工单数据; 每个第

一工单数据中包括故障信息与检测信息;

公共特征词提取模块,用于针对每类第一工单数据,从所述检测信息中提取公共特征词,作为特征向量;

故障谱模型学习模块,用于针对每类第一工单数据,学习所述故障信息与所述特征向量之间的逻辑关系,获得每类故障谱模型;

故障谱模型修剪模块,用于对每类故障谱模型进行修剪处理,获得每类故障谱。

本申请实施例包括以下优点:

本申请实施例通过建立故障谱,使得后续检测时支持并发依据该一个或多个检测路径进行检测,减少了检测耗时,提高检测的效率,同时,应用故障谱的检测操作简单,大大减少了人工参与的频次,减少用户精力的耗费,同时,利用海量的工单数据所形成的知识库中的知识点处理故障,大大降低了技术门槛,方便技术功底弱的用户或客服独自解决问题。

附图说明

图 1 是本申请的一种故障谱的生成方法实施例的步骤流程图;

图 2A 和图 2B 是本申请的一种故障谱模型的修剪示例图;

图 3A 和图 3B 是本申请的一种故障谱模型的修剪示例图;

图 4 是本申请的一种基于故障谱的检测方法实施例的步骤流程图;

图 5 是本申请的一种检测路径的示例图;

图 6A 是现有的一种检测示例图;

图 6B 是本申请的一种检测示例图;

图 7 是本申请的一种故障谱的生成装置实施例的结构框图;

图 8 是本申请的一种基于故障谱的检测装置实施例的结构框图。

具体实施方式

为使本申请的上述目的、特征和优点能够更加明显易懂,下面结合附图和具体实施方式对本申请作进一步详细的说明。

参照图 1, 示出了本申请的一种故障谱的生成方法实施例的步骤流程图, 具体可以包括如下步骤:

步骤 101, 获取一个或多个类别的第一工单数据;

5 在实际应用中, 可以对历史中海量的第一工单数据进行存储, 对该海量的第一工单数据分析总结后, 把典型的第一工单数据写成知识点, 保存在知识库中。

一般而言, 每个第一工单数据可以包括: 日期、用户 ID、产品、问题分类、问题 (故障信息)、解决办法 (检测信息)、沟通记录等要素。

10 其中, 故障信息可以为记载所发生的故障的信息, 检测信息可以为记载如何进行检测解决该故障的信息, 两者是相对应的。

例如, 在某个工单数据中, 故障信息为 “DB (Database, 数据库) 访问慢”, 检测信息为 “请您首先对网络拥塞检测吧”。

15 通过问题分类可以提取足够数量的、同属一个类别的第一工单数据, 作为训练样本。

步骤 102, 针对每类第一工单数据, 从所述检测信息中提取公共特征词, 作为特征向量;

公共特征词, 为该类中部分第一工单数据中所共有的词, 可以用于表征检测信息的特征, 作为训练样本的参数。

20 在本申请的一种优选实施例中, 步骤 102 可以包括如下子步骤:

子步骤 S11, 对所述检测信息进行分词处理, 获得一个或多个第一分词; 在具体实现中, 可以通过以下方式进行分词处理:

1、基于字符串匹配的分词方法: 是指按照一定的策略将待分析的汉字串与一个预置的机器词典中的词条进行匹配, 若在词典中找到某个字符串, 则匹配成功 (识别出一个词)。

25 2、基于特征扫描或标志切分的分词方法: 是指优先在待分析字符串中识别和切分出一些带有明显特征的词, 以这些词作为断点, 可将原字符串分为较小的串再来进机械分词; 或者将分词和词类标注结合起来,

利用丰富的词类信息对分词决策提供帮助，并且在标注过程中又反过来对分词结果进行检验、调整。

3、基于理解的分词方法：是指通过让计算机模拟人对句子的理解，达到识别词的效果。其基本思想就是在分词的同时进行句法、语义分析，
5 利用句法信息和语义信息来处理歧义现象。

4、基于统计的分词方法：对语料中相邻共现的各个字的组合的频度进行统计，计算它们的互现信息，以及计算两个汉字 X、Y 的相邻共现概率。互现信息可以体现汉字之间结合关系的紧密程度。当紧密程度高于某一个阈值时，便可认为此字组可能构成了一个词。

10 当然，上述分词处理方式只是作为示例，在实施本申请实施例时，可以根据实际情况设置其他分词处理方式，本申请实施例对此不加以限制。另外，除了上述分词处理方式外，本领域技术人员还可以根据实际需要采用其它分词处理方式，本申请实施例对此也不加以限制。

子步骤 S12，统计所述第一分词的词频；

15 子步骤 S13，通过所述第一分词的词频计算所述第一分词的权重；

在实际应用中，可以通过 TF-IDF (term frequency-inverse document frequency, 一种用于资讯检索与资讯探勘的常用加权技术) 计算第一分词的权重。

20 具体而言，TF-IDF 可以用于评估一字词对于一个文件集或一个语料库中的其中一份文件的重要程度，字词的重要性随着它在文件中出现的次数成正比增加，但同时会随着它在语料库中出现的频率成反比下降。

子步骤 S14，按照所述权重提取至少部分第一分词作为公共特征词。

若通过 TF-IDF 计算第一分词的权重，则可以提取权重最高的前 N(N 为正整数，如 10) 个第一分词作为公共特征词。

25 对于整体而言，会获得每个分类的故障信息及其公共特征词。

例如：对于连接失败的分类，提取的公共特征词如下：

拦截，错误日志，部分失败，...，白名单；

报错，验证失败，...，密码；

连接失败，访问拒绝，...，端口。

在本申请的另一种优选实施例中，步骤 102 还可以包括如下子步骤：

子步骤 S15，采用所述一个或多个第一分词在预置的停用词库中进行匹配；

5 子步骤 S16，移除匹配成功的第一分词。

停用词库中可以存储出现频率很高，但实际意义又不大的词，主要指副词、虚词、语气词等，如“是”、“而是”等。

在本申请实施例中，在子步骤 S12 之前，可以通过停用词滤去第一分词中无意义的词。

10 例如，检测信息“请您首先对网络拥塞检测吧”可以划分为“请”、“您”、“首先”、“对”、“网络拥塞”、“检测”、“吧”等第一分词，通过停用词库，可以去除“请”、“您”、“首先”、“对”、“吧”等无意义的词。

步骤 103，针对每类第一工单数据，学习所述故障信息与所述特征向量之间的逻辑关系，获得每类故障谱模型；

15 应用本申请实施例，可以预先设置训练器，可以用于学习各个维度的数据（即故障信息、特征向量）的逻辑关系，如支持向量机（Support Vector Machine, SVM）、决策树（Decision Tree）、随机森林（Random Forest）等等，本申请实施例对此不加以限制。

20 其中，支持向量机是通过一个非线性映射 p ，把样本空间映射到一个高维乃至无穷维的特征空间中（Hilbert 空间），使得在原来的样本空间中非线性可分的问题转化为在特征空间中的线性可分的问题。

25 随机森林，是用随机的方式建立一个森林，森林里面有很多的决策树组成，随机森林的每一棵决策树之间是没有关联的。在得到森林之后，当有一个新的输入样本进入的时候，就让森林中的每一棵决策树分别进行一下判断，看看这个样本应该属于哪一类（对于分类算法），然后看看哪一类被选择最多，就预测这个样本为那一类。

决策树是在已知各种情况发生概率的基础上，通过构成决策树来求取净现值的期望值大于等于零的概率，评价项目风险，判断其可行性的决策分析

方法，是直观运用概率分析的一种图解法。

在训练器训练时，可以对故障谱模型进行拟合，若误差（CP）小于一个预先设定的误差阈值（如 0.001），停止拟合，所训练处的故障谱模型如图 2A 所示，是树形结构，包括根节点与叶子节点，至少部分叶子节点之间具有逻辑关系。

如图 2A 所示，如“2.5”、“3.1”为节点（包括根节点、叶子节点），表征特征向量，“ $mmax < 6100$ ”、“ $syct > = 360$ ”等表征逻辑关系。

此外，根节点表征故障信息，叶子节点表征检测信息。

具体而言，如图 3A 所示，“A：DB 连接失败”为根节点，表征故障信息，“E：本地检测”、“B：网络断”为叶子节点，表征检测信息，与根节点“A：DB 连接失败”不存在逻辑关系。

而“H：配置排查”、“C：修复网络”为“B：网络断”的子节点，即“B：网络断”为“H：配置排查”、“C：修复网络”的父节点，“H：配置排查”、“C：修复网络”与“B：网络断”存在逻辑关系（图上未示出）。

步骤 104，对每类故障谱模型进行修剪处理，获得每类故障谱。

在实际应用中，可以按照实际需求对故障谱模型进行修剪处理，获得故障谱。

其中，故障谱中可以包括相连的根节点与叶子节点，根节点可以表征故障信息，叶子节点可以表征检测信息，至少部分叶子节点之间可以具有逻辑关系，叶子节点可以具有一个或多个父节点。

修剪好的故障谱，可以存储在故障谱仓库（数据库）中。

在本申请的一种优选实施中，步骤 104 可以包括如下子步骤：

子步骤 S21，在所述故障谱模型中查找相同的子树；

其中，所述子树可以为一个或多个叶子节点的集合；

子步骤 S22，当查找到时，将相同的子树的父节点连接至其中一个子树；

子步骤 S23，在相同的子树中，剪去已连接的子树之外的其他的子树。

在本申请实施例，由于某些子树可能有重复，因此，可以递归检查有重复的子树，发现后，在一个节点指向另一个子树，同时删除本身的子树，

使得某些叶子节点具有多个父节点（表示一个现象可能由多种原因造成），形成类树形结构，即有向无环图（DAG，指一个有向图无法从某个顶点出发经过若干条边回到该点）。

5 类树形结构的故障谱并非树形结构，如二叉树，二叉树中某些子树可能会其他子树重复，造成二叉树的结构冗长，分支过多，逻辑关系不清晰；但在 DAG 中不存在重复的子树，因为如果重复了，父节点可以删除本子树，并指向其他重复的子树，因此，相对于二叉树等树形结构层次较少，逻辑较清晰。

10 需要说明的是，相同（即重复）是指叶子节点相同、叶子节点之间的逻辑关系相同。

如图 3A 所示，“H：配置排查”、“F：白名单”、“D：检测密码”和“J：检测端口”这四个叶子节点组成的子树重复，如图 3B 所示，可以删除其中一个子树，使得该子树的父节点“B：网络断”和“E：本地检测”指向同一个子树。

15 在本申请的另一种优选实施中，步骤 104 还可以包括如下子步骤：

子步骤 S24，按照预设的剪枝方式对所述故障谱模型进行修剪处理。

一般情况下，训练出来的故障谱模型可能含有较深的层次，可能造成检测的步骤繁多。

20 在本申请实施例中，可以通过预设的剪枝方式，如 `prune()` 函数，对故障谱模型进行修剪处理，在可接受检测误差的范围内，把故障谱模型的层次减低，减少故障谱模型的复杂度，减少检测的步骤。

例如，如图 2A 所示的故障谱模型的层次为 6 层，通过剪枝方式修剪之后，获得如图 2B 所示的 4 层故障谱模型。

在本申请的另一种优选实施中，步骤 104 还可以包括如下子步骤：

25 子步骤 S25，从所述故障谱模型剪去逻辑关系不合法的叶子节点。

在本申请实施例中，可以通过人工确认或者通过合法逻辑关系排查，剪去对一些逻辑关系不合法的叶子节点，提高准确率。

参照图 4，示出了本申请的一种基于故障谱的检测方法实施例的步骤流程图，具体可以包括如下步骤：

步骤 401，当接收到第二工单数据时，从所述第二工单数据中提取关键词；

5 本申请实施例可以应用在虚拟客服系统中，该虚拟客服系统可以使用流式实时处理框架 strom，保证在极小的延迟完成检测。

当然，除了 strom，还可以应用 S4（Simple Scalable Streaming System）、MillWheel、Kinesis 等流式实时处理框架中，本申请实施例对此不加以限制。

在具体实现中，用户可以通过浏览器、独立的应用等方式提交第二工单数据给虚拟客服系统，虚拟客服系统可以会去除无关干扰信息、净化第二工单数据。

通常，该第二工单数据中允许包含一个故障信息，解决一个问题，若用户提出了一个问题，与此问题无关的信息可以认为是干扰信息。

15 进一步而言，在某种情况下，第二工单数据可以根据产品分类，与当前产品无关的信息可以第二工单数据。

例如，在某个第二工单数据中，用户询问：“我的 SQL 中的 UDF 函数为什么不能执行？”虚拟客服系统回答：“因为现在对外没有开放 UDF 的权限，因此你的 UDF 不能执行。”此外，用户再次询问：“明白。另一个问题：我的日志怎么下载呀？”

20 此示例中，用户的第 2 个问题与第一个问题无关，属于干扰信息。

对于过滤干扰信息后的第二工单数据，可以提取关键词，该关键词可以为体现第二工单数据（即故障信息）特征的信息。

在本申请的一种优选实施例中，步骤 401 可以包括如下子步骤：

25 子步骤 S31，对所述第二工单数据进行分词处理，获得一个或多个第二分词；

在具体实现中，可以通过以下方式进行分词处理：

- 1、基于字符串匹配的分词方法。
- 2、基于特征扫描或标志切分的分词方法。

3、基于理解的分词方法。

4、基于统计的分词方法。

当然，上述分词处理方式只是作为示例，在实施本申请实施例时，可以根据实际情况设置其他分词处理方式，本申请实施例对此不加以限制。另外，除了上述分词处理方式外，本领域技术人员还可以根据实际需要采用其它分词处理方式，本申请实施例对此也不加以限制。

子步骤 S32，识别所述一个或多个第二分词的词性；

子步骤 S33，按照所述词性从所述一个或多个第二分词中提取关键词。

在本申请实施例中，可以对第二分词进行词性分析，获得各第二分词的词性，如名词、动词、形容词、副词、介词、连词、助词等等。

其中，可以通过名词与动词组成关键词，名词可以用于确定目标对象，动词可以推测出主要语义。

例如，在某个第二工单数据中，用户 A 询问“有没有最大值的 UDF 函数”，此示例中，动词为“有没有”，名词为“UDF 函数”，即用户 A 的问题（即关键词）是问，“有没有 UDF 函数”。

又例如，在某个第二工单数据中，用户 B 询问“如何调试 UDF 函数”，此示例中，动词为“调试”，名词为“UDF 函数”，即用户 B 的问题（即关键词）是问，“调试 UDF 函数”的方法。

在本申请的一种优选实施例中，步骤 401 还可以包括如下子步骤：

子步骤 S34，采用所述一个或多个第一分词在预置的停用词库中进行匹配；

子步骤 S35，移除匹配成功的第二分词。

在本申请实施例中，在子步骤 S32 之前，可以通过停用词滤去第二分词中无意义的词。

步骤 402，查找所述第二工单数据所属类别对应的故障谱；

应用本申请实施例，可以预先训练故障谱，存储在故障谱仓库（数据库）中。

在实际应用中，可以通过文本相似度查找第二工单数据所属类别对应的

故障谱，即在故障谱仓库中查找与第二工单数据中的关键词相似的故障谱。

例如，如果关键词是：“DB”、“查询”、“等待”、“慢”等，则可以根据文本相似度匹配上“DB 访问慢”这个类目的故障谱。

在本申请实施例中，故障谱中可以包括相连的根节点与叶子节点，根节点可以表征故障信息，叶子节点可以表征检测信息，至少部分叶子节点之间可以具有逻辑关系，叶子节点可以具有一个或多个父节点。

步骤 403，在所述故障谱中，根据所述关键词查找一个或多个检测路径；在具体实现中，该检测路径可以记载检测方式及检测顺序的信息。

在本申请的一种优选实施例中，步骤 403 可以包括如下子步骤：

10 子步骤 S41，在所述故障谱中，查找与所述关键词匹配的根节点；

子步骤 S42，遍历与所述根节点相连的一个或多个叶子节点，获得一个或多个检测路径。

在本申请实施例中，故障谱是一个类树形结构（有向无环图 DAG），因此，检索检测路径时可以从上致下地进行使用“节点”匹配。

15 根据关键词可以在故障谱中定位根节点，是从该根节点向下遍历所经过的所有叶子节点，组成检测路径。

例如，如图 3B 所示的故障谱中存在两个检测路径，分布为“A→E→H→F/D/J”和“A→B→C/H→F/D/J”。

20 由于根节点与下一层的叶子节点一般不存在逻辑关系，因此，该下一层叶子节点的数量与检测路径的数量一般相同。

需要说明的是，相连可以指根节点与子节点直接相连，也可以指根节点与子节点间接相连。

25 例如，如图 3B 所示，叶子节点“E：本地检测”与根节点“A：DB 连接失败”直接相连，叶子节点“H：配置排查”与根节点“A：DB 连接失败”间接相连。

步骤 404，依据所述一个或多个检测路径进行检测，获得检测结果。

在申请实施例中，支持并发依据该一个或多个检测路径进行检测，减少了检测耗时，提高检测的效率。

在本申请的一种优选实施例中，步骤 404 可以包括如下子步骤：

子步骤 S51，针对每个检测路径，获取所述检测路径中的一个或多个叶子节点表征的检测信息；

子步骤 S52，按照当前叶子节点表征的检测信息进行检测，获得候选检测结果；

子步骤 S53，查找逻辑关系与所述候选检测结果匹配的下一叶子节点，返回执行子步骤 S51，直至执行至最终的叶子节点；

子步骤 S54，将最终的叶子节点的候选检测结果设置为检测结果。

在本申请实施例中，逻辑关系和叶子节点表征的检测信息可以称之为规则，即当符合某个条件（逻辑关系）就做执行某个操作（检测信息）。

例如，当 DB 访问慢时，就检查网络的带宽及流量，这就是一条逻辑关系。

规则是预先定义在规则引擎（rule engine）中的，如 JBoss Rules（业务规则引擎），一旦条件触发，规则引擎会执行这条规则，如：执行网络状态检测命令 ifstat。

在具体实现中，在检测路径按照逻辑关系逐层检测，直至最终的叶子节点，可以避免执行不符合逻辑关系的部分子节点。其中，最终的叶子节点可以指没有下一层叶子节点的叶子节点，并不一定是检测路径中最底层的叶子节点。

例如，在如图 5 所示的检测路径中，根节点“A：DB 访问慢”表征故障信息，要素是网络，DB 负载，SQL（Structured Query Language，结构化查询语言）。

首先要按照子节点“B：网络拥塞检测”确定网络是否有问题，若网络有问题（即“Y”），其余的要素都很难起作用（即无需执行“C：DB 负载检测”），按照子节点“H：联系网工”进行人工解决；其次，如果网络没有问题（即“N”），则按照子节点“C：DB 负载检测”判断 DB 负载是否高，如果“DB 负载高”，即使 SQL 本身没有问题（即无需执行“D：慢 SQL 检测”），对外也会展现出访问慢的情况，则按照子节点“J：SQL 线程检测”确认 SQL

线程是否运行正常；如果“DB 负载低”，则按照子节点“D：慢 SQL 检测”进行检测，分布为叶子节点“K：索引”的检测、“M：执行计划”的检测和“N：锁”的检测。

若执行到子节点“H：联系网工”、“J：SQL 线程检测”、“K：索引”、
5 “M：执行计划”和“N：锁”，则可以终止检测，获得检测结果。

据某项数据统计表明，70%的工单数据是很基本的常见问题，有较大的重复性，以现有的工单系统处理这些问题需占用 50%以上的客服资源，花费大量重复劳动，效率很低，本申请实施例的目的之一是自动化解决这 70%的
10 常见问题。

现有的工单系统一般是用户根据自己的经验来引导问题排查，即导向方向为“人→问题”，操作复杂，需要耗费用户大量的精力，并且，技术门槛高。

本申请实施例是让系统引导排查过程，是由海量的工单数据生成的故障
15 谱来指导用户排查问题，是逆向引导，即导向方向为“系统→问题→人”，操作简单，大大减少了人工参与的频次，减少用户精力的耗费，同时，利用海量的工单数据锁形成的知识库中的知识点处理问题，大大降低了技术门槛，方便技术功底弱的用户或客服独自解决问题。

如图 6A 所示，现有的工单系统在串行排查时，即没有任何内在逻辑关
20 系地进行排查，需要对图 6A 中所有的节点一一进行乱序排查，如：A→B→C→D→E，直到发现问题。因此，时间复杂度为 $O(N)$ ，N 为节点的数量，即所有的节点都要进行排查。

在本申请实施例中，如图 6B 所示，通过故障谱进行并发排查（如并发
25 执行 B、C）时，因为节点间是有逻辑关系的，一般不需要对所有的节点一一排查，节省了很多排查步骤，如，若在 C 检测导向 D，则不需要执行 E，反之，若在 C 检测导向 E，则不需要执行 D。对于故障谱的结构，利用分支结构（每次除以 2），可以省去很多排查步骤，排查复杂度最低可达 $O(\log_2^N)$ ，N 为节点的数量。

在本申请的一种优选实施例中，所述故障谱可以通过以下方式生成：

子步骤 S61，获取一个或多个类别的第一工单数据；每个第一工单数据中包括故障信息与检测信息；

5 子步骤 S62，针对每类第一工单数据，从所述检测信息中提取公共特征词，作为特征向量；

子步骤 S63，针对每类第一工单数据，学习所述故障信息与所述特征向量之间的逻辑关系，获得每类故障谱模型；

子步骤 S64，对每类故障谱模型进行修剪处理，获得每类故障谱。

10 在实际应用中，所述故障谱中可以包括相连的根节点与叶子节点，所述根节点可以表征故障信息，所述叶子节点可以表征检测信息，至少部分叶子节点之间可以具有逻辑关系，所述叶子节点可以具有一个或多个父节点。

在本申请的一种优选实施例中，子步骤 S62 可以包括如下子步骤：

15 子步骤 S621，对所述检测信息进行分词处理，获得一个或多个第一分词；

子步骤 S622，统计所述第一分词的词频；

子步骤 S623，通过所述第一分词的词频计算所述第一分词的权重；

子步骤 S624，按照所述权重提取至少部分第一分词作为公共特征词。

在本申请的另一种优选实施例中，子步骤 S62 还可以包括如下子步骤：

20 子步骤 S625，采用所述一个或多个第一分词在预置的停用词库中进行匹配；

子步骤 S626，移除匹配成功的第一分词。

在本申请的一种优选实施例中，子步骤 S64 可以包括如下子步骤：

25 子步骤 S641，在所述故障谱模型中查找相同的子树；所述子树为一个或多个叶子节点的集合；

子步骤 S642，当查找到时，将相同的子树的父节点连接至其中一个子树；

子步骤 S643，在相同的子树中，剪去已连接的子树之外的其他的子树。

在本申请的另一种优选实施例中，子步骤 S64 还可以包括如下子步骤：
子步骤 S644，按照预设的剪枝方式对所述故障谱模型进行修剪处理。

在本申请的另一种优选实施例中，子步骤 S64 还可以包括如下子步骤：
子步骤 S645，从所述故障谱模型剪去逻辑关系不合法的叶子节点。

5

需要说明的是，对于方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请实施例并不受所描述的动作顺序的限制，因为依据本申请实施例，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作并不一定是本申请实施例所必须的。

10

参照图 7，示出了本申请的一种故障谱的生成装置实施例的结构框图，具体可以包括如下模块：

工单数据获取模块 701，用于获取一个或多个类别的第一工单数据；每个第一工单数据中包括故障信息与检测信息；

15

公共特征词提取模块 702，用于针对每类第一工单数据，从所述检测信息中提取公共特征词，作为特征向量；

故障谱模型学习模块 703，用于针对每类第一工单数据，学习所述故障信息与所述特征向量之间的逻辑关系，获得每类故障谱模型；

20

故障谱模型修剪模块 704，用于对每类故障谱模型进行修剪处理，获得每类故障谱。

在具体实现中，所述故障谱中可以包括相连的根节点与叶子节点，所述根节点可以表征故障信息，所述叶子节点可以表征检测信息，至少部分叶子节点之间可以具有逻辑关系，所述叶子节点可以具有一个或多个父节点。

25

在本申请的一种优选实施例中，所述公共特征词提取模块 702 可以包括如下子模块：

第一分词处理模块，用于对所述检测信息进行分词处理，获得一个或多个第一分词；

词频统计模块，用于统计所述第一分词的词频；

权重计算模块，用于通过所述第一分词的词频计算所述第一分词的权重；

5 第一分词提取子模块，用于按照所述权重提取至少部分第一分词作为公共特征词。

在本申请的一种优选实施例中，所述公共特征词提取模块 702 还可以包括如下子模块：

匹配子模块，用于采用所述一个或多个第一分词在预置的停用词库中进行匹配；

10 移除子模块，用于移除匹配成功的第一分词。

在本申请的一种优选实施例中，所述故障谱模型修剪模块 704 可以包括如下子模块：

子树查找子模块，用于在所述故障谱模型中查找相同的子树；所述子树为一个或多个叶子节点的集合；

15 连接子模块，用于在查找到时，将相同的子树的父节点连接至其中一个子树；

第一修剪子模块，用于在相同的子树中，剪去已连接的子树之外的其他的子树。

20 在本申请的一种优选实施例中，所述故障谱模型修剪模块 704 还可以包括如下子模块：

第二修剪子模块，用于按照预设的剪枝方式对所述故障谱模型进行修剪处理。

在本申请的一种优选实施例中，所述故障谱模型修剪模块 704 还可以包括如下子模块：

25 第三修剪子模块，用于从所述故障谱模型剪去逻辑关系不合法的叶子节点。

参照图 8，示出了本申请的一种基于故障谱的检测装置实施例的结构框

图，具体可以包括如下模块：

关键词提取模块 801，用于在接收到第二工单数据时，从所述第二工单数据中提取关键词；

5 故障谱查找模块 802，用于查找所述第二工单数据所属类别对应的故障谱；

检测路径查找模块 803，用于在所述故障谱中，根据所述关键词查找一个或多个检测路径；

检测模块 804，用于依据所述一个或多个检测路径进行检测，获得检测结果。

10 在具体实现中，所述故障谱中可以包括相连的根节点与叶子节点，所述根节点可以表征故障信息，所述叶子节点可以表征检测信息，至少部分叶子节点之间可以具有逻辑关系，所述叶子节点可以具有一个或多个父节点。

在本申请的一种优选实施例中，所述关键词提取模块 801 可以包括如下子模块：

15 第二分词处理子模块，用于对所述第二工单数据进行分词处理，获得一个或多个第二分词；

词性识别子模块，用于识别所述一个或多个第二分词的词性；

第二分词提取子模块，用于按照所述词性从所述一个或多个第二分词中提取关键词。

20 在本申请的一种优选实施例中，所述关键词提取模块 801 还可以包括如下子模块：

第二匹配子模块，用于采用所述一个或多个第一分词在预置的停用词库中进行匹配；

第二移除子模块，用于移除匹配成功的第二分词。

25 在本申请的一种优选实施例中，所述检测路径查找模块 803 可以包括如下子模块：

根节点匹配子模块，用于在所述故障谱中，查找与所述关键词匹配的根节点；

叶子节点遍历子模块，用于遍历与所述根节点相连的一个或多个叶子节点，获得一个或多个检测路径。

在本申请的一种优选实施例中，所述检测模块 804 可以包括如下子模块：

检测信息获取子模块，用于针对每个检测路径，获取所述检测路径中的

5 一个或多个叶子节点表征的检测信息；

候选检测结果获取子模块，用于按照当前叶子节点表征的检测信息进行检测，获得候选检测结果；

叶子节点查找子模块，用于查找逻辑关系与所述候选检测结果匹配的下一叶子节点，返回调用候选检测结果获取子模块，直至执行至最终的叶子节点；

检测结果设置子模块，用于将最终的叶子节点的候选检测结果设置为检测结果。

在本申请的一种优选实施例中，所述故障谱可以通过调用以下模块生成：

15 工单数据获取模块，用于获取一个或多个类别的第一工单数据；每个第一工单数据中包括故障信息与检测信息；

公共特征词提取模块，用于针对每类第一工单数据，从所述检测信息中提取公共特征词，作为特征向量；

故障谱模型学习模块，用于针对每类第一工单数据，学习所述故障信息与所述特征向量之间的逻辑关系，获得每类故障谱模型；

故障谱模型修剪模块，用于对每类故障谱模型进行修剪处理，获得每类故障谱。

在本申请的一种优选实施例中，所述公共特征词提取模块 702 可以包括如下子模块：

25 第一分词处理模块，用于对所述检测信息进行分词处理，获得一个或多个第一分词；

词频统计模块，用于统计所述第一分词的词频；

权重计算模块，用于通过所述第一分词的词频计算所述第一分词的权

重；

第一分词提取子模块，用于按照所述权重提取至少部分第一分词作为公共特征词。

在本申请的一种优选实施例中，所述公共特征词提取模块还可以包括如

5 下子模块：

匹配子模块，用于采用所述一个或多个第一分词在预置的停用词库中进行匹配；

移除子模块，用于移除匹配成功的第一分词。

在本申请的一种优选实施例中，所述故障谱模型修剪模块可以包括如下

10 子模块：

子树查找子模块，用于在所述故障谱模型中查找相同的子树；所述子树为一个或多个叶子节点的集合；

连接子模块，用于在查找到时，将相同的子树的父节点连接至其中一个子树；

15 第一修剪子模块，用于在相同的子树中，剪去已连接的子树之外的其他的子树。

在本申请的一种优选实施例中，所述故障谱模型修剪模块还可以包括如下子模块：

20 第二修剪子模块，用于按照预设的剪枝方式对所述故障谱模型进行修剪处理。

在本申请的一种优选实施例中，所述故障谱模型修剪模块还可以包括如下子模块：

第三修剪子模块，用于从所述故障谱模型剪去逻辑关系不合法的叶子节点。

25

对于装置实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。

本领域内的技术人员应明白，本申请实施例的实施例可提供为方法、装置、或计算机程序产品。因此，本申请实施例可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

在一个典型的配置中，所述计算机设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM) 和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括非持续性的电脑可读媒体(transitory media)，如调制的数据信号和载波。

本申请实施例是参照根据本申请实施例的方法、终端设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、

专用计算机、嵌入式处理机或其他可编程数据处理终端设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理终端设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

- 5 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理终端设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

10 这些计算机程序指令也可装载到计算机或其他可编程数据处理终端设备上，使得在计算机或其他可编程终端设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程终端设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

15 尽管已描述了本申请实施例的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例做出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本申请实施例范围的所有变更和修改。

20 最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者终端设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下，由语句“包括一个.....”限定的要素，并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。

25

以上对本申请所提供的一种故障谱的生成方法、一种基于故障谱的检测

方法、一种故障谱的生成装置和一种基于故障谱的检测装置，进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

5

权利要求书

1、一种故障谱的生成方法，其特征在于，包括：

获取一个或多个类别的第一工单数据；每个第一工单数据中包括故障信息与检测信息；

5 针对每类第一工单数据，从所述检测信息中提取公共特征词，作为特征向量；

针对每类第一工单数据，学习所述故障信息与所述特征向量之间的逻辑关系，获得每类故障谱模型；

对每类故障谱模型进行修剪处理，获得每类故障谱。

10 2、根据权利要求 1 所述的方法，其特征在于，所述故障谱中包括相连的根节点与叶子节点，所述根节点表征故障信息，所述叶子节点表征检测信息，至少部分叶子节点之间具有逻辑关系，所述叶子节点具有一个或多个父节点。

15 3、根据权利要求 1 或 2 所述的方法，其特征在于，所述从所述检测信息中提取公共特征词的步骤包括：

对所述检测信息进行分词处理，获得一个或多个第一分词；

统计所述第一分词的词频；

通过所述第一分词的词频计算所述第一分词的权重；

按照所述权重提取至少部分第一分词作为公共特征词。

20 4、根据权利要求 3 所述的方法，其特征在于，所述从所述检测信息中提取公共特征词的步骤还包括：

采用所述一个或多个第一分词在预置的停用词库中进行匹配；

移除匹配成功的第一分词。

25 5、根据权利要求 1 或 2 或 4 所述的方法，其特征在于，所述对每类故障谱模型进行修剪处理的步骤包括：

在所述故障谱模型中查找相同的子树；所述子树为一个或多个叶子节点的集合；

当查找到时，将相同的子树的父节点连接至其中一个子树；

在相同的子树中，剪去已连接的子树之外的其他的子树。

6、根据权利要求 5 所述的方法，其特征在于，所述对每类故障谱模型进行修剪处理的步骤还包括：

5 按照预设的剪枝方式对所述故障谱模型进行修剪处理。

7、根据权利要求 5 所述的方法，其特征在于，所述对每类故障谱模型进行修剪处理的步骤还包括：

从所述故障谱模型剪去逻辑关系不合法的叶子节点。

10 8、一种基于故障谱的检测方法，其特征在于，包括：

当接收到第二工单数据时，从所述第二工单数据中提取关键词；

查找所述第二工单数据所属类别对应的故障谱；

在所述故障谱中，根据所述关键词查找一个或多个检测路径；

依据所述一个或多个检测路径进行检测，获得检测结果。

15 9、根据权利要求 8 所述的方法，其特征在于，所述故障谱中包括相连的根节点与叶子节点，所述根节点表征故障信息，所述叶子节点表征检测信息，至少部分叶子节点之间具有逻辑关系，所述叶子节点具有一个或多个父节点。

20 10、根据权利要求 8 所述的方法，其特征在于，所述从所述第二工单数据中提取关键词的步骤包括：

对所述第二工单数据进行分词处理，获得一个或多个第二分词；

识别所述一个或多个第二分词的词性；

按照所述词性从所述一个或多个第二分词中提取关键词。

25 11、根据权利要求 10 所述的方法，其特征在于，所述从所述第二工单数据中提取关键词的步骤还包括：

采用所述一个或多个第一分词在预置的停用词库中进行匹配；

移除匹配成功的第二分词。

12、根据权利要求 9 所述的方法，其特征在于，所述在所述故障谱中，

根据所述特征词查找一个或多个检测路径的步骤包括:

在所述故障谱中, 查找与所述关键词匹配的根节点;

遍历与所述根节点相连的一个或多个叶子节点, 获得一个或多个检测路径。

- 5 13、根据权利要求 12 所述的方法, 其特征在于, 所述依据所述一个或多个检测路径进行检测, 获得检测结果的步骤包括:

针对每个检测路径, 获取所述检测路径中的一个或多个叶子节点表征的检测信息;

按照当前叶子节点表征的检测信息进行检测, 获得候选检测结果;

- 10 查找逻辑关系与所述候选检测结果匹配的下一叶子节点, 返回执行按照当前叶子节点表征的检测信息进行检测的步骤, 直至执行至最终的叶子节点;

将最终的叶子节点的候选检测结果设置为检测结果。

- 15 14、根据权利要求 8 或 9 或 10 或 11 或 12 或 13 所述的方法, 其特征在于, 所述故障谱通过以下方式生成:

获取一个或多个类别的第一工单数据; 每个第一工单数据中包括故障信息与检测信息;

针对每类第一工单数据, 从所述检测信息中提取公共特征词, 作为特征向量;

- 20 针对每类第一工单数据, 学习所述故障信息与所述特征向量之间的逻辑关系, 获得每类故障谱模型;

对每类故障谱模型进行修剪处理, 获得每类故障谱。

15、一种故障谱的生成装置, 其特征在于, 包括:

- 25 工单数据获取模块, 用于获取一个或多个类别的第一工单数据; 每个第一工单数据中包括故障信息与检测信息;

公共特征词提取模块, 用于针对每类第一工单数据, 从所述检测信息中提取公共特征词, 作为特征向量;

故障谱模型学习模块，用于针对每类第一工单数据，学习所述故障信息与所述特征向量之间的逻辑关系，获得每类故障谱模型；

故障谱模型修剪模块，用于对每类故障谱模型进行修剪处理，获得每类故障谱。

- 5 16、根据权利要求 15 所述的装置，其特征在于，所述故障谱中包括相连的根节点与叶子节点，所述根节点表征故障信息，所述叶子节点表征检测信息，至少部分叶子节点之间具有逻辑关系，所述叶子节点具有一个或多个父节点。

- 10 17、根据权利要求 15 或 16 所述的装置，其特征在于，所述公共特征词提取模块包括：

第一分词处理模块，用于对所述检测信息进行分词处理，获得一个或多个第一分词；

词频统计模块，用于统计所述第一分词的词频；

- 15 权重计算模块，用于通过所述第一分词的词频计算所述第一分词的权重；

第一分词提取子模块，用于按照所述权重提取至少部分第一分词作为公共特征词。

18、根据权利要求 17 所述的装置，其特征在于，所述公共特征词提取模块还包括：

- 20 第一匹配子模块，用于采用所述一个或多个第一分词在预置的停用词库中进行匹配；

第一移除子模块，用于移除匹配成功的第一分词。

19、根据权利要求 15 或 16 或 18 所述的装置，其特征在于，所述故障谱模型修剪模块包括：

- 25 子树查找子模块，用于在所述故障谱模型中查找相同的子树；所述子树为一个或多个叶子节点的集合；

连接子模块，用于在查找到时，将相同的子树的父节点连接至其中一个子树；

第一修剪子模块，用于在相同的子树中，剪去已连接的子树之外的其他的子树。

20、根据权利要求 19 所述的装置，其特征在于，所述故障谱模型修剪模块还包括：

5 第二修剪子模块，用于按照预设的剪枝方式对所述故障谱模型进行修剪处理。

21、根据权利要求 19 所述的装置，其特征在于，所述故障谱模型修剪模块还包括：

10 第三修剪子模块，用于从所述故障谱模型剪去逻辑关系不合法的叶子节点。

22、一种基于故障谱的检测装置，其特征在于，包括：

关键词提取模块，用于在接收到第二工单数据时，从所述第二工单数据中提取关键词；

15 故障谱查找模块，用于查找所述第二工单数据所属类别对应的故障谱；

检测路径查找模块，用于在所述故障谱中，根据所述关键词查找一个或多个检测路径；

检测模块，用于依据所述一个或多个检测路径进行检测，获得检测结果。

20 23、根据权利要求 22 所述的装置，其特征在于，所述故障谱中包括相连的根节点与叶子节点，所述根节点表征故障信息，所述叶子节点表征检测信息，至少部分叶子节点之间具有逻辑关系，所述叶子节点具有一个或多个父节点。

24、根据权利要求 22 所述的装置，其特征在于，所述关键词提取模块包括：

25 第二分词处理子模块，用于对所述第二工单数据进行分词处理，获得一个或多个第二分词；

词性识别子模块，用于识别所述一个或多个第二分词的词性；

第二分词提取子模块，用于按照所述词性从所述一个或多个第二分词中

提取关键词。

25、根据权利要求 24 所述的装置，其特征在于，所述关键词提取模块还包括：

5 第二匹配子模块，用于采用所述一个或多个第一分词在预置的停用词库中进行匹配；

第二移除子模块，用于移除匹配成功的第二分词。

26、根据权利要求 23 所述的装置，其特征在于，所述检测路径查找模块包括：

10 根节点匹配子模块，用于在所述故障谱中，查找与所述关键词匹配的根节点；

叶子节点遍历子模块，用于遍历与所述根节点相连的一个或多个叶子节点，获得一个或多个检测路径。

27、根据权利要求 25 所述的装置，其特征在于，所述检测模块包括：

15 检测信息获取子模块，用于针对每个检测路径，获取所述检测路径中的一个或多个叶子节点表征的检测信息；

候选检测结果获取子模块，用于按照当前叶子节点表征的检测信息进行检测，获得候选检测结果；

20 叶子节点查找子模块，用于查找逻辑关系与所述候选检测结果匹配的下一叶子节点，返回调用候选检测结果获取子模块，直至执行至最终的叶子节点；

检测结果设置子模块，用于将最终的叶子节点的候选检测结果设置为检测结果。

28、根据权利要求 22 或 23 或 24 或 25 或 26 或 27 所述的装置，其特征在于，所述故障谱通过调用以下模块生成：

25 工单数据获取模块，用于获取一个或多个类别的第一工单数据；每个第一工单数据中包括故障信息与检测信息；

公共特征词提取模块，用于针对每类第一工单数据，从所述检测信息中提取公共特征词，作为特征向量；

故障谱模型学习模块，用于针对每类第一工单数据，学习所述故障信息与所述特征向量之间的逻辑关系，获得每类故障谱模型；

故障谱模型修剪模块，用于对每类故障谱模型进行修剪处理，获得每类故障谱。

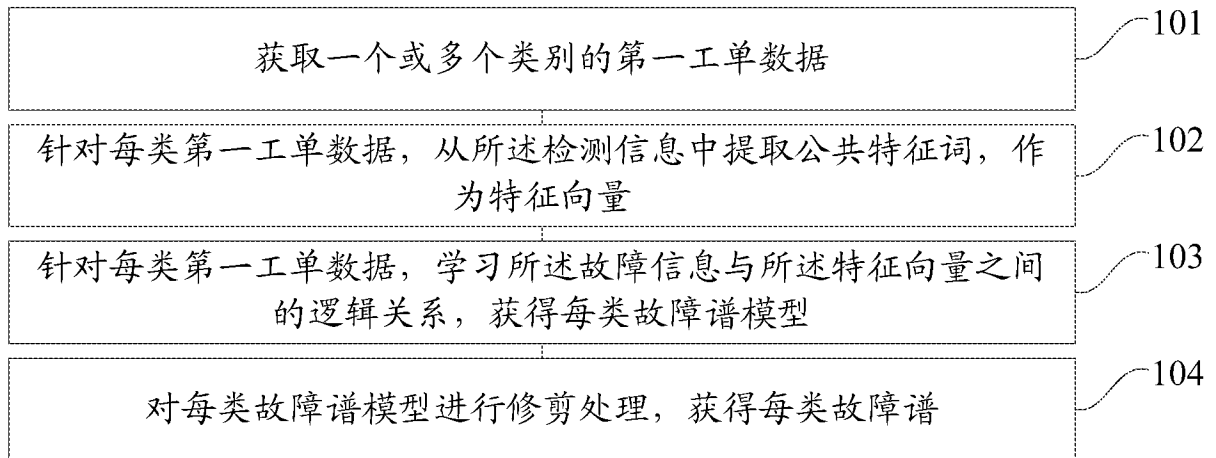


图 1

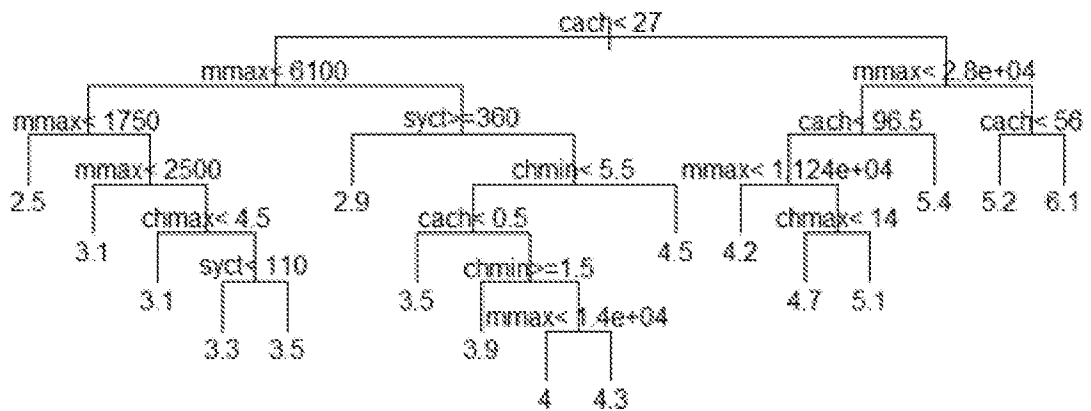


图 2A

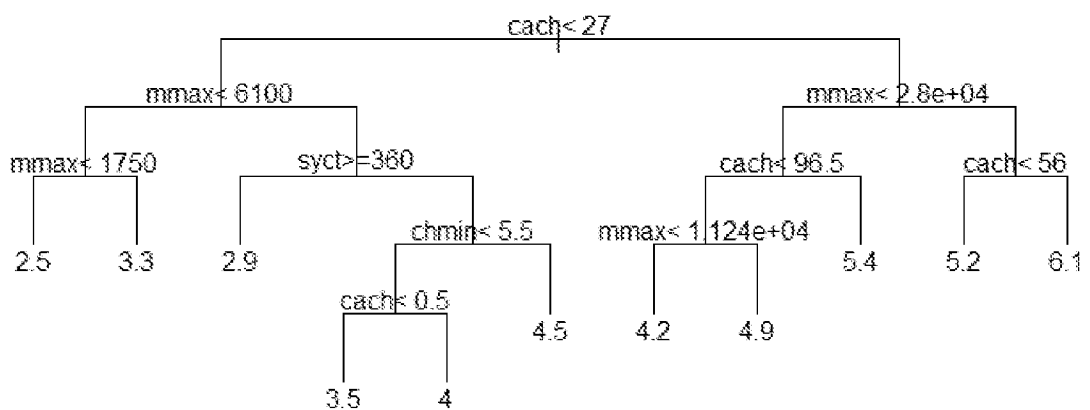


图 2B

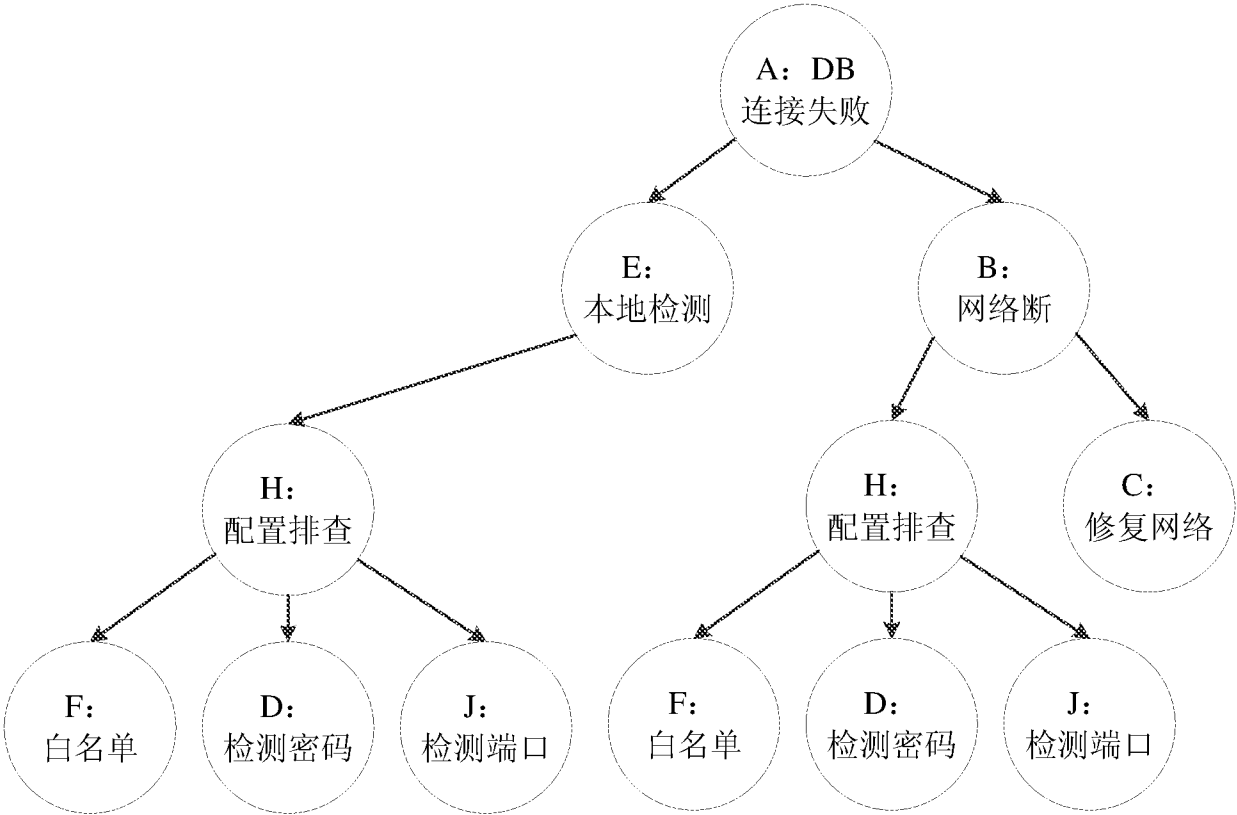


图 3A

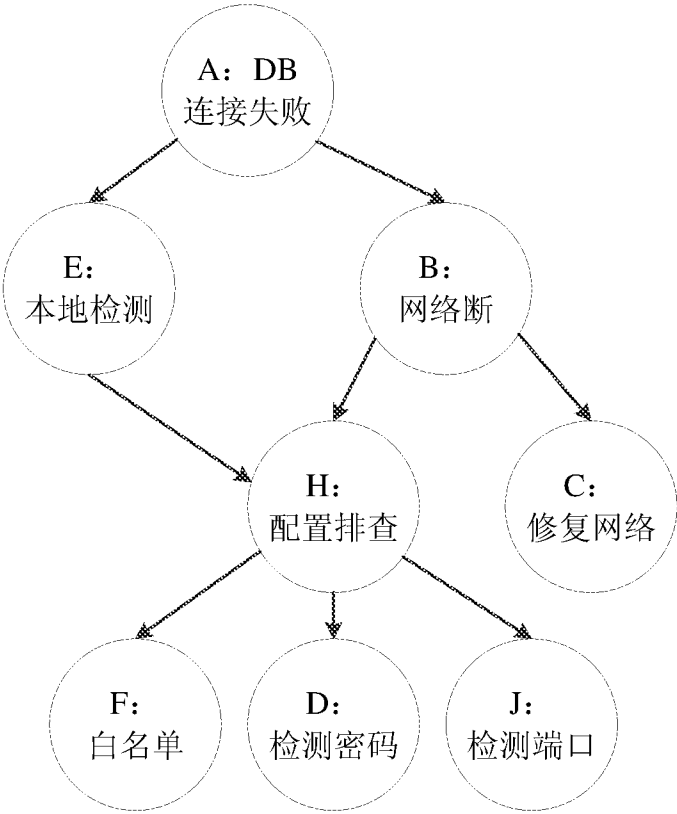


图 3B

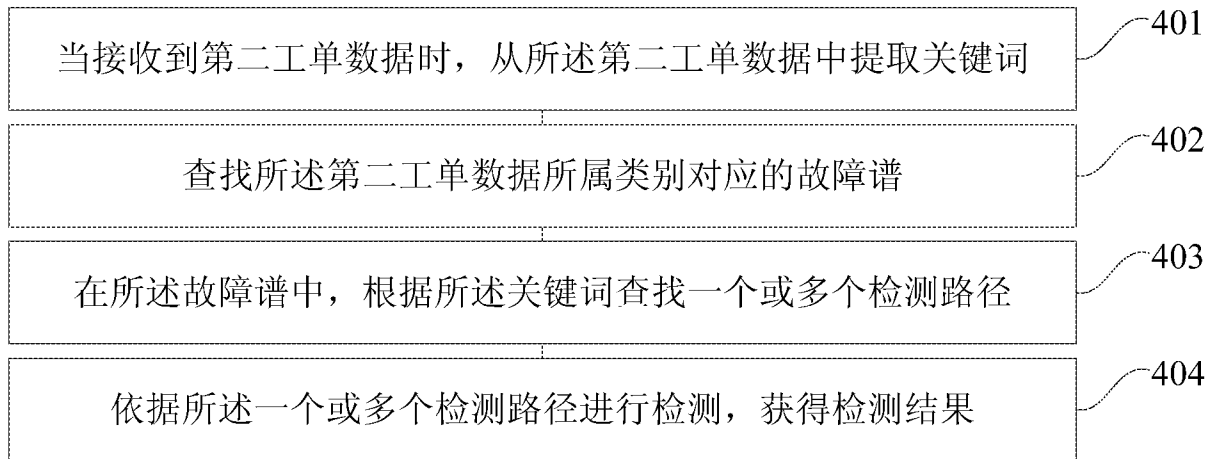


图 4

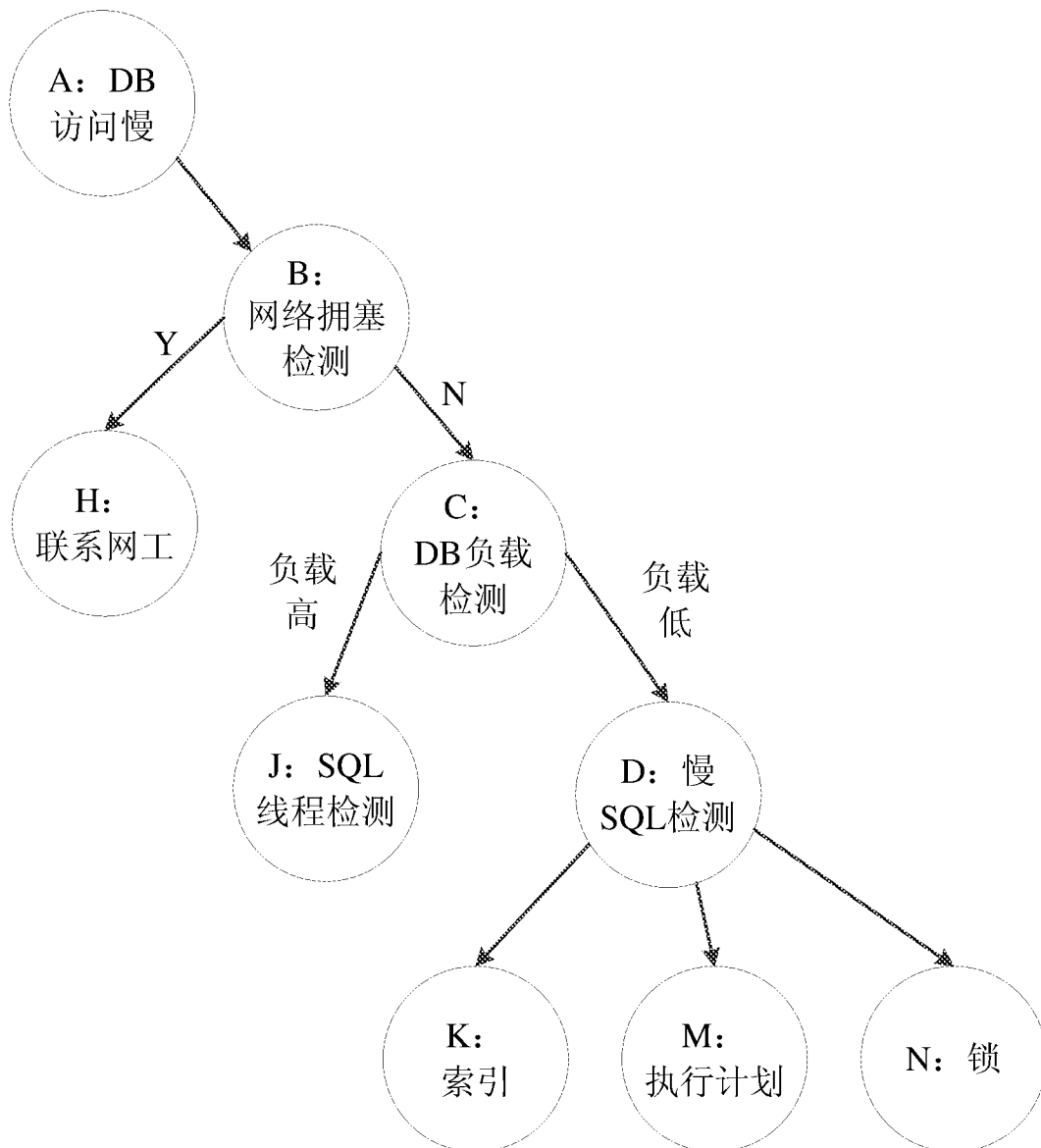


图 5

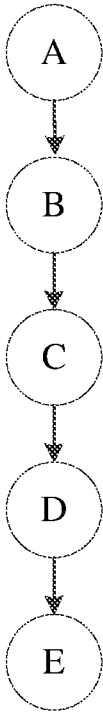


图 6A

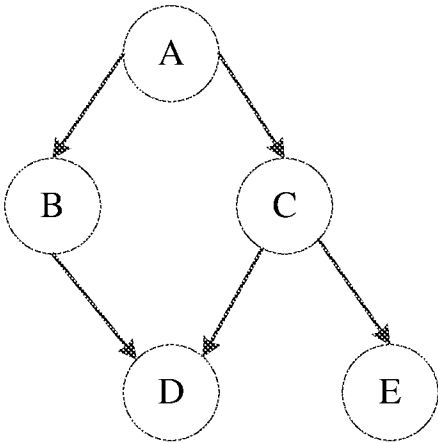


图 6B

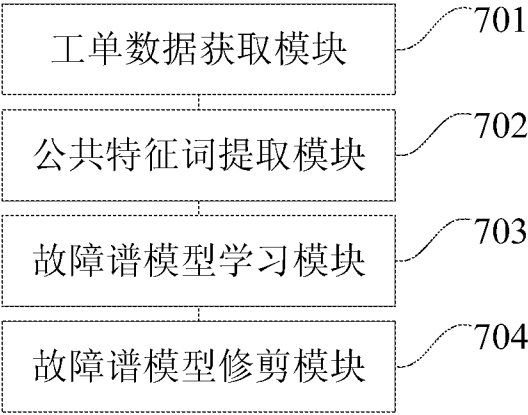


图 7

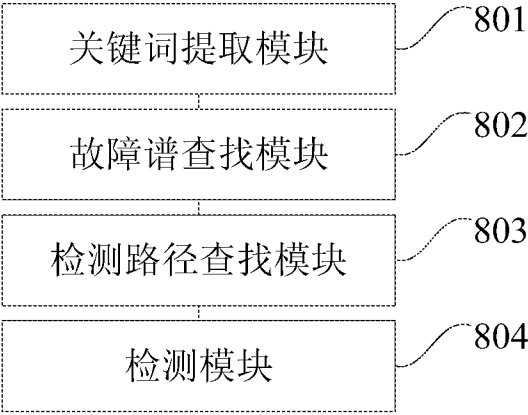


图 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/080015

A. CLASSIFICATION OF SUBJECT MATTER

G06F 11/22 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; H04W; G06Q; G05B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNPAT, CNKI, IEEE: fault, tree, generate, group, detect, word, repair, extract, character, keyword, train, learn, vector, trim, failure, worksheet, work order, class, type, lop

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 104376033 A (ENG ACADEMY ARMORED FORCES PLA) 25 February 2015 (25.02.2015) description, paragraphs [0020] to [0033], and figures 1 and 3	8-13, 22-27
X	CN 101742540 A (HUAWEI TECHNOLOGIES CO., LTD.) 16 June 2010 (16.06.2010) description, pages 2 to 6	8-13, 22-27
A	CN 103310389 A (CHINA SOUTH POWER GRID ELECTRIC POWER RES INST CO., LTD. et al.) 18 September 2013 (18.09.2013) description, page 2	1-28
A	CN 101846992 A (UNIVERSITY SHANGHAI SCIENCE & TECH) 29 September 2010 (29.09.2010) the whole document	1-28

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 01 July 2016	Date of mailing of the international search report 26 July 2016
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer JU, Bo Telephone No. (86-10) 62413661

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/080015

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104376033 A	25 February 2015	None	
CN 101742540 A	16 June 2010	None	
CN 103310389 A	18 September 2013	None	
CN 101846992 A	29 September 2010	None	

A. 主题的分类 G06F 11/22 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																	
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06F H04W G06Q G05B 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) WPI, EPDOC, CNPAT, CNKI, IEEE: 故障, 谱, 树, 生成, 产生, 类, 工单, 检测, 维修, 解决, 提取, 特征, 关键, 词, 训练, 学习, 向量, 修剪, 剪枝, fault, tree, generate, group, detect, word, repair, extract, character, keyword, train, learn, vector, trim, failure																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 104376033 A (中国人民解放军装甲兵工程学院) 2015年 2月 25日 (2015 - 02 - 25) 说明书第0020-0033段、附图1、3</td> <td>8-13、22-27</td> </tr> <tr> <td>X</td> <td>CN 101742540 A (华为技术有限公司) 2010年 6月 16日 (2010 - 06 - 16) 说明书第2-6页</td> <td>8-13、22-27</td> </tr> <tr> <td>A</td> <td>CN 103310389 A (南方电网科学研究院有限责任公司 等) 2013年 9月 18日 (2013 - 09 - 18) 说明书第2页</td> <td>1-28</td> </tr> <tr> <td>A</td> <td>CN 101846992 A (上海理工大学) 2010年 9月 29日 (2010 - 09 - 29) 全文</td> <td>1-28</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 104376033 A (中国人民解放军装甲兵工程学院) 2015年 2月 25日 (2015 - 02 - 25) 说明书第0020-0033段、附图1、3	8-13、22-27	X	CN 101742540 A (华为技术有限公司) 2010年 6月 16日 (2010 - 06 - 16) 说明书第2-6页	8-13、22-27	A	CN 103310389 A (南方电网科学研究院有限责任公司 等) 2013年 9月 18日 (2013 - 09 - 18) 说明书第2页	1-28	A	CN 101846992 A (上海理工大学) 2010年 9月 29日 (2010 - 09 - 29) 全文	1-28
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	CN 104376033 A (中国人民解放军装甲兵工程学院) 2015年 2月 25日 (2015 - 02 - 25) 说明书第0020-0033段、附图1、3	8-13、22-27															
X	CN 101742540 A (华为技术有限公司) 2010年 6月 16日 (2010 - 06 - 16) 说明书第2-6页	8-13、22-27															
A	CN 103310389 A (南方电网科学研究院有限责任公司 等) 2013年 9月 18日 (2013 - 09 - 18) 说明书第2页	1-28															
A	CN 101846992 A (上海理工大学) 2010年 9月 29日 (2010 - 09 - 29) 全文	1-28															
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																	
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																	
国际检索实际完成的日期 2016年 7月 1日		国际检索报告邮寄日期 2016年 7月 26日															
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10) 62019451		授权官员 鞠博 电话号码 (86-10) 62413661															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/080015

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	104376033	A	2015年 2月 25日	无	
CN	101742540	A	2010年 6月 16日	无	
CN	103310389	A	2013年 9月 18日	无	
CN	101846992	A	2010年 9月 29日	无	