



(12) 发明专利

(10) 授权公告号 CN 113961708 B

(45) 授权公告日 2024. 04. 23

(21) 申请号 202111324168.4

(22) 申请日 2021.11.10

(65) 同一申请的已公布的文献号

申请公布号 CN 113961708 A

(43) 申请公布日 2022.01.21

(73) 专利权人 北京邮电大学

地址 100876 北京市海淀区西土城路10号

专利权人 中国电力科学研究院有限公司

国网天津市电力公司电力科学研
究院

国网天津市电力公司

国家电网有限公司

(72) 发明人 李睿凡 张煜松 焦飞 陆明聪

刘力卿 谈元鹏 李志坚 唐庆华

张弛 魏菊芳 冯军基 马昊

(74) 专利代理机构 天津盛理知识产权代理有限

公司 12209

专利代理师 王利文

(51) Int. Cl.

G06F 16/35 (2019.01)

G06F 40/216 (2020.01)

G06F 40/30 (2020.01)

G06F 40/289 (2020.01)

G06N 3/042 (2023.01)

G06N 3/0464 (2023.01)

G06N 3/08 (2023.01)

G06Q 10/20 (2023.01)

G06Q 50/06 (2024.01)

(56) 对比文件

CN 110717047 A, 2020.01.21

CN 111767397 A, 2020.10.13

CN 112529071 A, 2021.03.19

CN 113205039 A, 2021.08.03

CN 113343701 A, 2021.09.03

WO 2019210820 A1, 2019.11.07

王伟;唐庆华;刘力卿;李敏;谢军.基于加权
综合损失优化深度学习和DGA的变压器故障诊断
方法.南方电网技术.2020, (03), 全文. (续)

审查员 邵金

权利要求书2页 说明书6页 附图2页

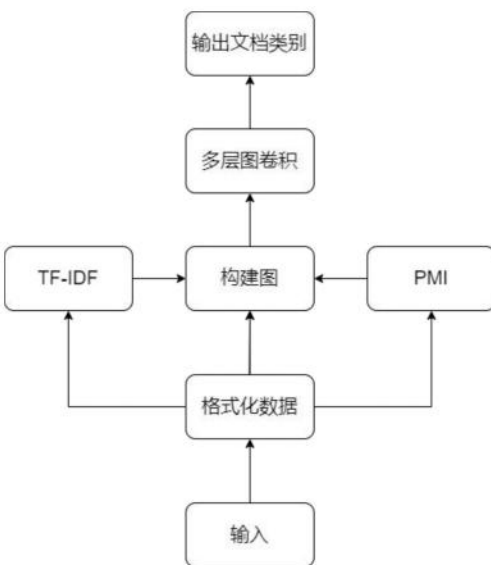
(54) 发明名称

一种基于多层次图卷积网络的电力设备故障溯源方法

(57) 摘要

本发明涉及一种基于多层次图卷积网络的电力设备故障溯源方法,其技术特点是:从电力工单系统中收集电力设备的故障文本信息,并计算故障文本信息中的TF-IDF指标和PMI指标;根据TF-IDF指标和PMI指标构建电力工单图G(V,E),V和E分别为节点集和边集;根据电力工单图G(V,E)构建多层图卷积网络,并训练多层图卷积网络;使用训练好的多层图卷积网络对电力工单系统中电力设备的故障文本信息进行识别,确定电力设备故障位置。本发明可以从电力工单系统的故障描述中准确预测故障位置,有效解决了基于中文文本的电力设备故障溯源问题,具有较高

的鲁棒性和容错率。



CN 113961708 B

[转续页]

[接上页]

(56) 对比文件

张弛;谢伟;王卫斌;杨波.电力故障抢修管

理系统(TCM)的建设与试点研究.华东电力
.2010, (004), 全文.

1. 一种基于多层次图卷积网络的电力设备故障溯源方法,其特征在于:包括以下步骤:
 步骤1、从电力工单系统中收集电力设备的故障文本信息,并计算故障文本信息中的TF-IDF指标和PMI指标;
 步骤2、根据TF-IDF指标和PMI指标构建电力工单图 $G(V, E)$, V 和 E 分别为节点集和边集;
 步骤3、根据电力工单图 $G(V, E)$ 构建多层次图卷积网络,并训练多层次图卷积网络;
 步骤4、使用训练好的多层次图卷积网络对电力工单系统中电力设备的故障文本信息进行识别,确定电力设备故障位置;

所述步骤2构建的电力工单图 $G(V, E)$ 中的邻接矩阵 A 的构造规则为:

$$A_{ij} = \begin{cases} PMI(i, j) & i, j \text{ are words, } PMI(i, j) > 0 \\ TF-IDF(i, j) & i \text{ is document, } j \text{ is word} \\ 1 & i = j \\ 0 & otherwise \end{cases}$$

其中,

$$PMI(i, j) = \log \frac{p(i, j)}{p(i)p(j)}$$

$$p(i, j) = \frac{\#W(i, j)}{\#W}$$

$$p(i) = \frac{\#W(i)}{\#W}$$

式中, $\#W(i)$ 是包含单词 i 的滑动窗口的数量, $\#W(i, j)$ 是包含单词 i 和 j 的滑动窗口的数量, $\#W$ 是滑动窗口的总数。

2. 根据权利要求1所述的一种基于多层次图卷积网络的电力设备故障溯源方法,其特征在于:所述步骤1中的TF-IDF指标用于提取故障文本信息中的关键词,该TF-IDF指标的计算方法为:首先分别计算词频、计算逆文档频率,然后通过下式计算TF-IDF指标:

$$TF-IDF = TF \times IDF.$$

3. 根据权利要求1所述的一种基于多层次图卷积网络的电力设备故障溯源方法,其特征在于:所述步骤1中的PMI指标用于计算词语件的语义相似度,该PMI指标通过下式计算:

$$PMI(a, b) = \log\left(\frac{P(a, b)}{P(a)P(b)}\right)$$

其中, $P(a, b)$ 表示两个词语 a 与 b 共同出现的概率, $P(a)$ 与 $P(b)$ 分别表示两个词语单独出现的概率。

4. 根据权利要求1所述的一种基于多层次图卷积网络的电力设备故障溯源方法,其特征在于:所述步骤3的具体实现方法为:

首先,构建一个包括词节点和文档节点的词-文档图;

然后,初始化特征矩阵 $X \in \mathbb{R}^{n \times m}$,该特征矩阵包含所有节点及其特征,特征向量的维数

为 m , 设置 $X=I$, 每个单词和文档都表示为一个独热向量作为输入;

$$L^{(j+1)} = \rho(\tilde{A}L^{(j)}W_j)$$

式中, ρ 是激活函数, $\tilde{A}=D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 是拉普拉斯矩阵, 初始化 $L^{(0)}=X$;

设置交叉熵函数作为损失函数, 该损失函数表示如下:

$$L = -\sum_i Y_i \ln Z_i$$

式中, Z 为模型输出, Y 为真实类别标签, Z 和 Y 都是 m 维的向量, 且 m 等于类别数目, i 为训练样本的下标;

最后, 使用训练样本进行多层次图卷积网络训练。

5. 根据权利要求4所述的一种基于多层次图卷积网络的电力设备故障溯源方法, 其特征在于: 所述多层次图卷积网络为四层卷积神经网络。

6. 根据权利要求4所述的一种基于多层次图卷积网络的电力设备故障溯源方法, 其特征在于: 激活函数为ReLU激活函数。

一种基于多层次图卷积网络的电力设备故障溯源方法

技术领域

[0001] 本发明属于电力设备技术领域,涉及电力故障溯源方法,尤其是一种基于多层次图卷积网络的电力设备故障溯源方法。

背景技术

[0002] 电力故障溯源指的是根据故障的表现和各种实际情况确定故障位置的过程。在实际电力系统运行过程中,故障位置一般由现场的维修工人根据设备的各种情况通过分析得出。由于上述故障分析过程对维修工人的经验和知识储备都有要求,有经验的维修工人可以在短时间内分析出电力设备的故障位置,因此人工的电力设备故障溯源依赖于维修工人的经验和知识,具有不稳定性,而且需要工人实地或者远程沟通分析,比较占用人力。

[0003] 传统的自动电力故障溯源工作大多都是从电力专业角度出发,根据电力设备的各种状态信息(温度,电压,电流等)进行实时的电力设备故障监测,虽然通过这种方法监测故障往往比较准确,还可能提前预测故障,但是这种方法往往需要电力设备是智能化的,或者在变电站安装独立的传感器进行信息采集,难以大范围推广,并且对所有的设备都进行实时监测的各种硬件成本也很昂贵,因此短时间内推广的可能性不大。

[0004] 在电力系统的运行过程中,有一类由报修行为产生的文本形式的缺陷描述信息,这种文本信息是已经存在于电力工单系统内的,而且这些文本中蕴含了一定量的故障信息,通过对这些文本信息的智能化分析,可以通过文本对故障部件有初步的判断。因为可能故障的部件是离散且有限的,所以可将其作为一个文本分类问题来研究,可以通过如图1所示的模型来进行分析,该模型输入是缺陷描述,输出是缺陷部件。

[0005] 文本分类问题是自然语言处理领域的主要问题之一,传统的文本分类算法主要是基于统计的,通过提取文本中的共现概率,词频等统计学特征,设计算法来对文本进行分类。近年来,随着单词嵌入技术的发展和深度神经网络的广泛应用,两种典型的文本分类神经网络是递归神经网络(RNN)和卷积神经网络(CNN)。RNN包括长短时记忆(LSTM)、双向LSTM和其他变形方法,CNN衍生出深金字塔卷积神经网络(DPCNN)。此外,为了提高分类性能,注意力机制已被引入该领域,作为文本分类方法的一个组成部分。此外,基于图形神经网络(GCN)的方法将文本和单词提取到图节点中,结合特定关系建立单词-文本图,并构建用于文本分类的网络模型,这种基于GCN的方法取得了良好的效果。

[0006] 但是,在电力文本分析领域,传统的自然语言处理算法面临很多挑战,难点在于文本语料是电力相关的,具有高度的领域特征,即中文电力语料和通用的文本分类算法研究的中文语料的词典,用词习惯等有着很大的不同,例如,一个缺陷描述文本如:“中新#1#2线773、776组合电器空气压力低闭锁重合闸”。从中我们可以看出,在电力语料中,省略了大部分语法结构以提升文字的信息量,从而提升沟通的效率,这为电力相关文本处理任务带来了困难,并且近年来非常流行的预训练模型也无法很好地解决这个问题。

发明内容

[0007] 本发明的目的在于克服现有技术的不足,提供一种设计合理、快速准确的基于多层次图卷积网络的电力设备故障溯源方法。

[0008] 本发明解决现有的技术问题是采取以下技术方案实现的:

[0009] 一种基于多层次图卷积网络的电力设备故障溯源方法,包括以下步骤:

[0010] 步骤1、从电力工单系统中收集电力设备的故障文本信息,并计算故障文本信息中的TF-IDF指标和PMI指标;

[0011] 步骤2、根据TF-IDF指标和PMI指标构建电力工单图G(V,E),V和E分别为节点集和边集;

[0012] 步骤3、根据电力工单图G(V,E)构建多层次图卷积网络,并训练多层次图卷积网络;

[0013] 步骤4、使用训练好的多层次图卷积网络对电力工单系统中电力设备的故障文本信息进行识别,确定电力设备故障位置;

[0014] 所述步骤2构建的电力工单图G(V,E)中的邻接矩阵A的构造规则为:

$$[0015] \quad A_{ij} = \begin{cases} PMI(i, j) & i, j \text{ are words, } PMI(i, j) > 0 \\ TF-IDF(i, j) & i \text{ is document, } j \text{ is word} \\ 1 & i = j \\ 0 & otherwise \end{cases}$$

[0016] 其中,

$$[0017] \quad PMI(i, j) = \log \frac{p(i, j)}{p(i)p(j)}$$

$$[0018] \quad p(i, j) = \frac{\#W(i, j)}{\#W}$$

$$[0019] \quad p(i) = \frac{\#W(i)}{\#W}$$

[0020] 式中,#W(i)是包含单词i的滑动窗口的数量,#W(i, j)是包含单词i和j的滑动窗口的数量,#W是滑动窗口的总数。

[0021] 进一步,所述步骤1中的TF-IDF指标用于提取故障文本信息中的关键词,该TF-IDF指标的计算方法为:首先分别计算词频、计算逆文档频率,然后通过下式计算TF-IDF指标:

[0022] $TF-IDF = TF \times IDF$ 。

[0023] 进一步,所述步骤1中的PMI指标用于计算词语件的语义相似度,该PMI指标通过下式计算:

$$[0024] \quad PMI(a, b) = \log \left(\frac{P(a, b)}{P(a)P(b)} \right)$$

[0025] 其中,P(a, b)表示两个词语a与b共同出现的概率,P(a)与P(b)分别表示两个词语

单独出现的概率。

[0026] 进一步,所述步骤3的具体实现方法为:

[0027] 首先,构建一个包括词节点和文档节点的词-文档图;

[0028] 然后,初始化特征矩阵 $X \in \mathbb{R}^{n \times m}$,该特征矩阵包含所有节点及其特征,特征向量的维数为m,设置 $X=I$,每个单词和文档都表示为一个独热向量作为输入;

$$[0029] \quad L^{(j+1)} = \rho(\tilde{A}L^{(j)}W_j)$$

[0030] 式中, ρ 是激活函数, $\tilde{A}=D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 是拉普拉斯矩阵,初始化 $L^{(0)}=X$;

[0031] 设置交叉熵函数作为损失函数,该损失函数表示如下:

$$[0032] \quad L = -\sum_i Y_i \ln Z_i$$

[0033] 式中,Z为模型输出,Y为真实类别标签,Z和Y都是m维的向量,且m等于类别数目,i为训练样本的下标;

[0034] 最后,使用训练样本进行多层次图卷积网络训练。

[0035] 进一步,所述多层次图卷积网络为四层卷积神经网络。

[0036] 进一步,激活函数为ReLU激活函数。

[0037] 本发明的优点和积极效果是:

[0038] 1、本发明设计合理,其从工单系统中收集了电力设备的故障文本信息,利用词频-逆文档频率(Term Frequency-Inverse Document Frequency,TF-IDF)和点互信息(Pointwise Mutual Information,PMI)指标构建电力工单图,根据电力工单图建立了多层次图卷积网络并应用于电力文本分类任务,可以从电力工单系统的故障描述中准确预测故障位置,实现溯源任务,具有较高的鲁棒性和容错率,将其集成在电力工单系统中,具有很强的应用价值。

[0039] 2、本发明采用多层次图卷积电力设备故障溯源模型(Multi-layer Fault Recognition Graph Convolutional Network,MFRGCN),可有效解决基于中文文本的电力设备故障溯源问题,可以用于电力故障文本分类任务的评价和训练。

[0040] 3、本发明在收集到的数据集上进行了实验,定量地比较了本发明和国内外先进文本分类模型的效果,实验结果表明本发明在故障溯源上表现好于国内外先进的文本分类模型。

附图说明

[0041] 图1为文本分类问题分析模型;

[0042] 图2为本发明构建的词-文档结构图;

[0043] 图3为本发明的模型工作流程图。

具体实施方式

[0044] 以下结合附图对本发明实施例做进一步详述。

[0045] 本发明的设计思想是:

[0046] 电力设备故障溯源任务的目标是通过对各种信息的分析自动化地输出电力设备的故障部位等信息。我们通过对电力工单系统的研究认为：由于图卷积模型使用的参数较少，不容易过拟合，是一个具有很强的鲁棒性的算法，对于电力文本处理是一个很适合的模型，因此，本发明提出一种基于电力工单文本的多层次图卷积电力设备故障溯源模型 (Multi-layer Fault Recognition Graph Convolutional Network, MFRGCN)，该MFRGCN通过对工单中的文本信息进行分析，利用工单中的文本信息定位电力设备的故障。具体来讲，MFRGCN通过将电力工单中的缺陷描述文本 (词) 和文档 (工单) 作为节点构建图，定义图的两端词-词边和词-文档边，从而构建图，然后在这个图上通过分类任务训练。我们将分类任务定义为将文档分类到故障部件中，从而将电力设备故障溯源任务转化为文本分类任务。

[0047] 基于上述设计思想，本发明提出一种基于多层次图卷积网络的电力设备故障溯源方法，如图3所示，包括以下步骤：

[0048] 步骤1、从电力工单系统中收集电力设备的故障文本信息，并计算故障文本信息中的TF-IDF指标和PMI指标。

[0049] 在本步骤中，通过TF-IDF (词频-逆文档频率) 提取故障文本信息中的关键词。

[0050] TF-IDF是一种用于信息检索与数据挖掘的常用加权技术，常用于挖掘文章中的关键词。TF-IDF有两层意思，一层是“词频” (Term Frequency, 缩写为TF)，另一层是“逆文档频率” (Inverse Document Frequency, 缩写为IDF)。

[0051] 通过TF可以将文档中最常见但对结果毫无帮助、需要的词过滤掉。当高频词过滤后就只需考虑剩下的有实际意义的词。

[0052] 通过IDF可以给常见的词较小的权重，其大小与一个词的常见程度成反比。

[0053] 当有TF和IDF后，将这两个词相乘，就能得到一个词的TF-IDF的值。某个词在文档中的TF-IDF越大，那么一般而言这个词在文档中的重要性会越高，所以通过计算文档中各个词的TF-IDF，由大到小排序，排在最前面的几个词，提取文档中的关键词。

[0054] 在本步骤中，TF-IDF指标的计算方法包括以下步骤：

[0055] (1) 计算词频：

[0056] 词频 (TF) = 某个词在文档中出现的次数，考虑到文档有长短之分，为了便于不同文档的比较，进行“词频”标准化。

$$[0057] \quad TF_{\omega} = \frac{\omega_{freq}}{sum} \quad (1)$$

[0058] 其中 ω_{freq} 为文档中词 ω 的出现次数，sum为该文档中词语总数。

[0059] (2) 计算逆文档频率：

[0060] 这时，需要一个语料库 (corpus)，用来模拟语言的使用环境。

$$[0061] \quad IDF = \log\left(\frac{filenum}{file_{\omega} + 1}\right) \quad (2)$$

[0062] 其中filenum为文档总数，分母为包含该词的文档数量+1。如果一个词越常见，那么分母就越大，逆文档频率就越小越接近0。分母之所以要加1，是为了避免分母为0 (即所有文档都不包含该词)。

[0063] (3) 计算TF-IDF：

$$[0064] \quad TF-IDF = TF \times IDF \quad (3)$$

[0065] 可以看到,TF-IDF与一个词在文档中的出现次数成正比,与该词在整个语言中的出现次数成反比。所以,自动提取关键词的算法就很清楚了,就是计算出文档的每个词的TF-IDF值,然后按降序排列,取排在最前面的几个词。

[0066] TF-IDF具有简单快速、容易理解的特点。

[0067] 在本步骤中,采用PMI(点间互信息)计算词语间的语义相似度。其基本思想是统计两个词语在文本信息中同时出现的概率,如果概率越大,其相关性就越紧密,关联度越高。两个词语的PMI值计算公式如下式所示:

$$[0068] \quad PMI(a,b) = \log\left(\frac{P(a,b)}{P(a)P(b)}\right) \quad (4)$$

[0069] 其中, $P(a,b)$ 表示两个词语a与b共同出现的概率,即a与b共同出现的文档数, $P(a)$ 与 $P(b)$ 分别表示两个词语单独出现的概率,即词语出现的文档数。若两个词语在数据集的某个小范围内共现概率越大,表明其关联度越大;反之,关联度越小。 $P(a,b)$ 与 $P(a)P(b)$ 的比值是a与b两个词语的统计独立性度量。其值可以转化为3种状态:

[0070] $PMI(a,b) > 0$:两个词语是相关的;值越大,相关性越强。

[0071] $PMI(a,b) = 0$:两个词语是统计独立的,不相关也不互斥。

[0072] $PMI(a,b) < 0$:两个词语是不相关的,互斥的。

[0073] 步骤2、根据TF-IDF指标和PMI指标构建电力工单图 $G(V,E)$ 。

[0074] 在本步骤中,使用数据集构建 $G(V,E)$,其中V和E分别为图中的节点集和边集,该图是根据TF-IDF指标和PMI指标进行构建的。该图的邻接矩阵A的构造规则如下:

$$[0075] \quad A_{ij} = \begin{cases} PMI(i,j) & i, j \text{ are words, } PMI(i,j) > 0 \\ TF-IDF(i,j) & i \text{ is document, } j \text{ is word} \\ 1 & i = j \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

[0076] 此处PMI值计算规则定义如下:

$$[0077] \quad \begin{aligned} PMI(i,j) &= \log \frac{p(i,j)}{p(i)p(j)} \\ p(i,j) &= \frac{\#W(i,j)}{\#W} \\ p(i) &= \frac{\#W(i)}{\#W} \end{aligned} \quad (6)$$

[0078] 其中 $\#W(i)$ 是包含单词i的滑动窗口的数量, $\#W(i,j)$ 是包含单词i和j的滑动窗口的数量, $\#W$ 是滑动窗口的总数。我们使用固定尺寸为15的滑动窗口。如果PMI值为正值,则表示对应的单词高度相关。然后我们在它们之间添加一条边到图中。否则,相应的单词之间没有边相连。

[0079] 步骤3、根据电力工单图 $G(V,E)$ 构建多层次图卷积网络,并训练多层次图卷积网络

(GCN)。

[0080] 在本步骤中,首先构建一个词-文档图,其中包含词节点和文档节点,如图2所示,图中圆圈节点是文档节点,其他节点是词节点,圆圈节点和方框节点之间的实线边是词-文档边,方框节点之间的虚线边是词之间的边,最右边的四个圆圈是文档类别(仅显示四个示例类)。

[0081] 初始化特征矩阵 $X \in \mathbb{R}^{n \times m}$,该矩阵包含所有节点及其特征,特征向量的维数为m。我们设置 $X=I$,即单位矩阵。这意味着每个单词和文档都表示为一个独热向量作为输入。

$$[0082] \quad L^{(j+1)} = \rho(\tilde{A}L^{(j)}W_j) \quad (7)$$

[0083] ρ 是激活函数,通常被设置为ReLU激活函数。 $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 是拉普拉斯矩阵,并且我们初始化 $L^{(0)} = X$ 。我们在卷积神经网络中堆叠了四层。我们设置交叉熵函数作为损失函数,即:

$$[0084] \quad L = -\sum_i Y_i \ln Z_i \quad (8)$$

[0085] 其中Z为模型输出,Y为真实类别标签。Z和Y都是m维的向量,且m等于类别数目,此处的i为训练样本的下标。

[0086] 步骤4、使用训练好的多层次图卷积网络对电力工单系统中电力设备的故障文本信息进行识别,确定电力设备故障位置。

[0087] 需要强调的是,本发明所述的实施例是说明性的,而不是限定性的,因此本发明包括并不限于具体实施方式中所述的实施例,凡是由本领域技术人员根据本发明的技术方案得出的其他实施方式,同样属于本发明保护的范围。

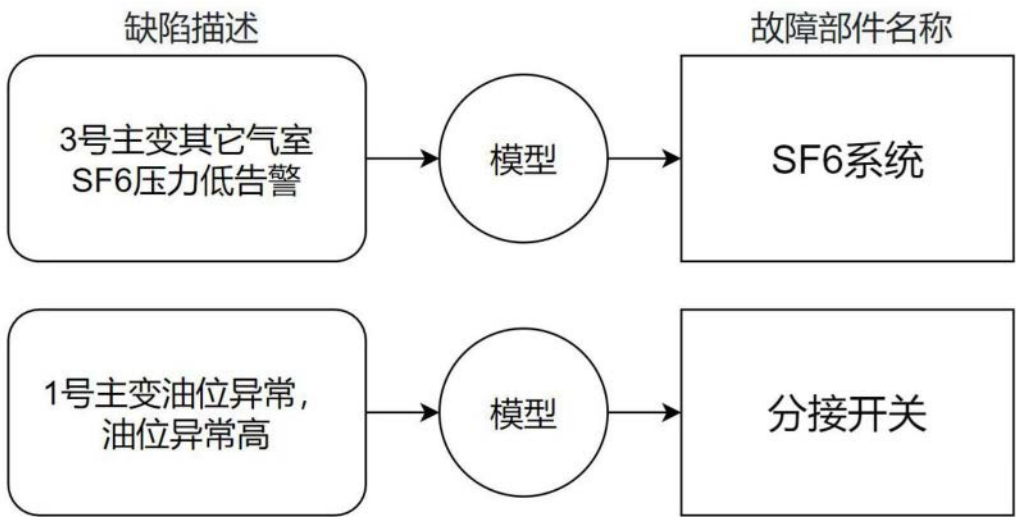


图1

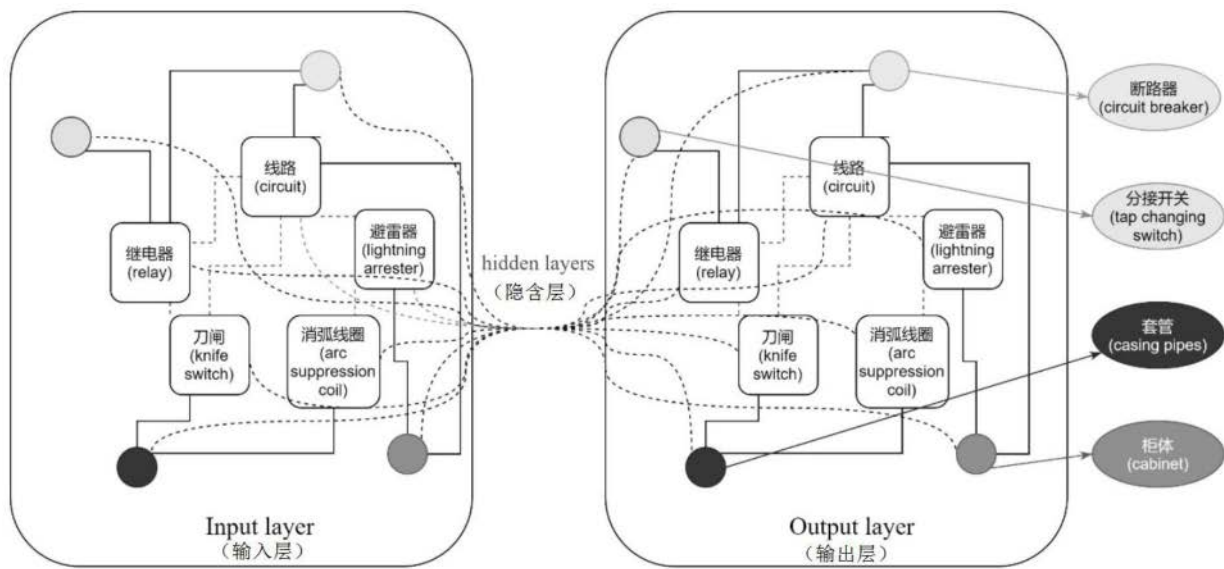


图2

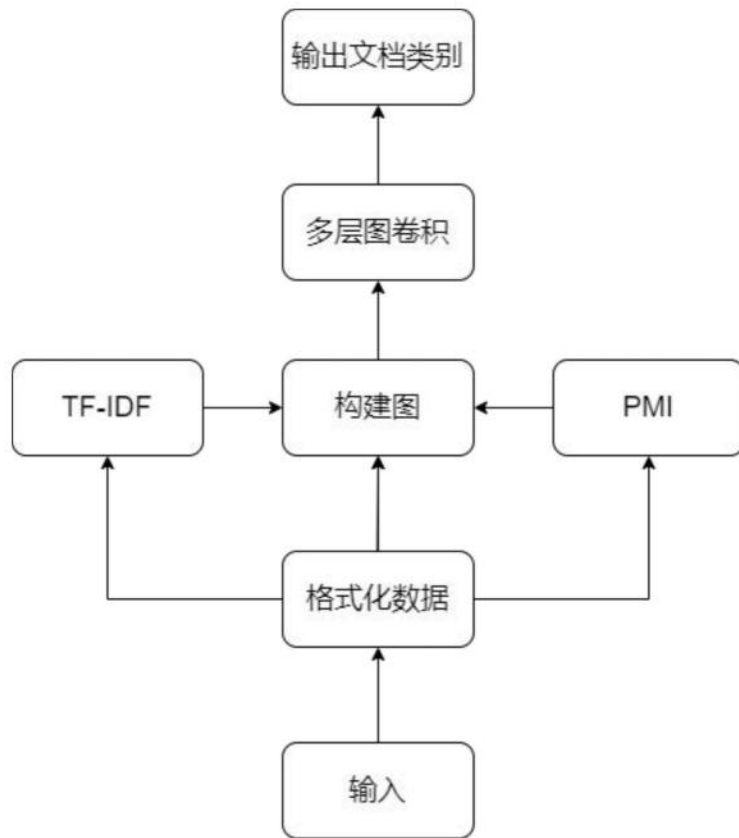


图3