(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 117375845 B
(45)授权公告日 2024.08.20

(21)申请号 202311345134.2

(22)申请日 2023.10.17

(65)同一申请的已公布的文献号
申请公布号 CN 117375845 A

(43)申请公布日 2024.01.09

(73)专利权人 中国电子科技集团公司第十五研究所
地址 100083 北京市海淀区北四环中路211号

(72)发明人 任传伦 张先国 杨天长 刘策越 李宝静 尹誉衡 唐然 郭强

(74)专利代理机构 北京丰浩知识产权代理事务所(普通合伙) 11781
专利代理师 李强

(51)Int.Cl.
H04L 9/32(2006.01)

(56)对比文件
CN 114579832 A,2022.06.03
邱云飞.基于网络结构和文本内容的群体画像构建方法研究.图书情报工作.2019,第63卷(第22期),第3.5节.

审查员 黄俊云

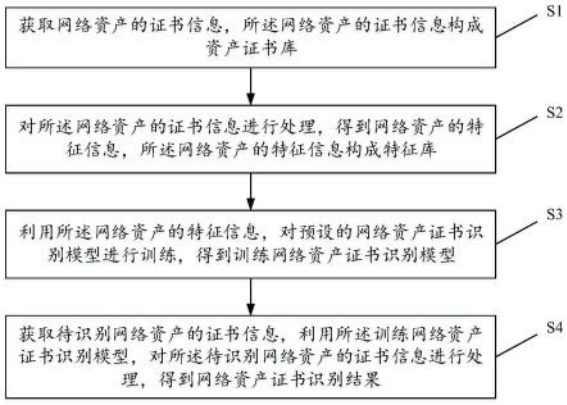
权利要求书4页 说明书14页 附图3页

(54)发明名称

一种网络资产证书识别方法及装置

(57)摘要

本发明公开了一种网络资产证书识别方法及装置,该方法包括:获取网络资产的证书信息,所述网络资产的证书信息构成资产证书库;对所述网络资产的证书信息进行处理得到网络资产的特征信息,所述网络资产的特征信息构成特征库;利用所述网络资产的特征信息,对预设的网络资产证书识别模型进行训练得到训练网络资产证书识别模型;获取待识别网络资产的证书信息,利用所述训练网络资产证书识别模型,对所述待识别网络资产的证书信息进行处理,得到网络资产证书识别结果。本发明将两种相似性度量方法相结合,从证书的内容和结构两方面进行相似性比对;同时,结合自主学习反馈机制,探索自动化程度高、准确率更高的证书相似性度量和识别查询方法。



CN 117375845 B

1.一种网络资产证书识别方法,其特征在于,所述方法包括:

S1,获取网络资产的证书信息,所述网络资产的证书信息构成资产证书库;

S2,对所述资产证书库中的网络资产的证书信息进行处理,得到网络资产的特征信息,所述网络资产的特征信息构成特征库,包括:

S21,对所述网络资产的证书信息进行结构相似度计算,得到结构相似度信息,包括:

S211,获取网络资产A的证书信息,所述网络资产A的信息项数量为m;

S212,获取网络资产B的证书信息,所述网络资产B的信息项数量为n;

S213,对所述网络资产A的证书信息和所述网络资产B的证书信息进行处理,得到所述网络资产A和所述网络资产B都包含的信息项数量x;

S214,利用结构相似度计算模型,对所述网络资产A的信息项数量为m、所述网络资产B的信息项数量为n和所述信息项数量x进行处理,得到结构相似度信息;

所述结构相似度计算模型为:

$$str(A,B)=\frac{x}{m+n-x}$$

其中str(A,B)为结构相似度信息;

S22,对所述网络资产的证书信息进行内容相似度计算,得到内容相似度信息,包括:

S221,对网络资产A的证书信息进行处理,得到网络资产A的证书信息项形式化表示$content_A=(a_1,a_2,\cdots,a_x)$;

S222,对网络资产B的证书信息进行处理,得到网络资产B的证书信息项形式化表示$content_B=(b_1,b_2,\cdots,b_x)$;

S223,利用内容相似度信息计算模型,对所述网络资产A的证书信息项形式化表示$content_A=(a_1,a_2,\cdots,a_x)$和所述网络资产B的证书信息项形式化表示进行处理,得到内容相似度信息;

所述内容相似度计算模型为:

$$con(A,B)=\frac{1}{1+d(A,B)}$$

其中,con(A,B)为内容相似度信息,$d(A,B)=\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}$;

S23,对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息,包括:

利用相似度信息融合模型,对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息;

所述相似度信息融合模型为:

$$total(A,B)=str(A,B)*con(A,B)$$

$$=\frac{x}{m+n-x}*\frac{1}{1+\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}}$$

其中,total(A,B)为网络资产的特征信息,str(A,B)为结构相似度信息,con(A,B)为内

容相似度信息；

S3,利用所述特征库中的网络资产的特征信息,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型；

所述预设的网络资产证书识别模型的获取方法为：

采用离散小波分解对网络资产的特征信息进行处理；

网络资产的特征信息被分解到不同频带后计算其能量占比,归一化后利用主成分分析对特征降维,构成的特征向量集作为最小二乘支持向量机输入；

利用混合粒子群算法对最小二乘支持向量机初始参数寻优,从而搭建预设的网络资产证书识别模型；

混合粒子群算法将快速模拟退火法和粒子群算法结合,粒子群算法用于全局搜索区域的探索,当粒子群算法搜索到当前迭代次数的全体最优解时使用快速模拟退火法对粒子群算法找到的最佳位置进行调整得到新解,将新解与粒子群算法得出的最优解进行对比,如果新解优于最优解,则将新解作为当前最优解,否则以一定概率接受新解；

S4,获取待识别网络资产的证书信息,利用所述训练网络资产证书识别模型,对所述待识别网络资产的证书信息进行处理,得到网络资产证书识别结果,包括：

S41,获取待识别网络资产的证书信息；

S42,对所述待识别网络资产的证书信息进行处理,得到待识别网络资产的特征信息；

S43,根据所述待识别网络资产的特征信息,得到首次识别结果；

根据首次识别结果,对与所述首次识别结果不同的网络资产证书进行标注,得到正例证书和反例证书；

将所述反例证书加入未标记证书样本中,构成最优训练样本集；

利用所述最优训练样本集,对预设的网络资产证书识别模型进行训练,得到优化网络资产证书识别模型；

S44,根据所述首次识别结果,利用所述优化网络资产证书识别模型,对所述待识别网络资产的特征信息进行处理,得到网络资产证书识别结果。

2.根据权利要求1所述的网络资产证书识别方法,其特征在于,所述利用所述网络资产的特征信息,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型,包括：

S31,对所述网络资产的特征信息进行划分,得到标注证书样本和未标记证书样本；

S32,将所述未标记证书样本作为训练样本,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型。

3.一种网络资产证书识别装置,其特征在于,所述装置包括：

数据获取模块,用于获取网络资产的证书信息,所述网络资产的证书信息构成资产证书库；

特征提取模块,用于对所述资产证书库中的网络资产的证书信息进行处理,得到网络资产的特征信息,所述网络资产的特征信息构成特征库,包括：

S21,对所述网络资产的证书信息进行结构相似度计算,得到结构相似度信息,包括：

S211,获取网络资产A的证书信息,所述网络资产A的信息项数量为m；

S212,获取网络资产B的证书信息,所述网络资产B的信息项数量为n；

S213,对所述网络资产A的证书信息和所述网络资产B的证书信息进行处理,得到所述网络资产A和所述网络资产B都包含的信息项数量x;

S214,利用结构相似度计算模型,对所述网络资产A的信息项数量为m、所述网络资产B的信息项数量为n和所述信息项数量x进行处理,得到结构相似度信息;

所述结构相似度计算模型为:

$$str(A,B)=\frac{x}{m+n-x}$$

其中str(A,B)为结构相似度信息;

S22,对所述网络资产的证书信息进行内容相似度计算,得到内容相似度信息,包括:

S221,对网络资产A的证书信息进行处理,得到网络资产A的证书信息项形式化表示$content_A=(a_1,a_2,\cdots,a_x)$;

S222,对网络资产B的证书信息进行处理,得到网络资产B的证书信息项形式化表示$content_B=(b_1,b_2,\cdots,b_x)$;

S223,利用内容相似度信息计算模型,对所述网络资产A的证书信息项形式化表示$content_A=(a_1,a_2,\cdots,a_x)$和所述网络资产B的证书信息项形式化表示进行处理,得到内容相似度信息;

所述内容相似度计算模型为:

$$con(A,B)=\frac{1}{1+d(A,B)}$$

其中,$con(A,B)$为内容相似度信息,$d(A,B)=\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}$;

S23,对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息,包括:

利用相似度信息融合模型,对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息;

所述相似度信息融合模型为:

$$total(A,B)=str(A,B)*con(A,B)$$

$$=\frac{x}{m+n-x}*\frac{1}{1+\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}}$$

其中,total(A,B)为网络资产的特征信息,str(A,B)为结构相似度信息,con(A,B)为内容相似度信息;

模型训练模块,用于利用所述特征库中的网络资产的特征信息,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型;

所述预设的网络资产证书识别模型的获取方法为:

采用离散小波分解对网络资产的特征信息进行处理;

网络资产的特征信息被分解到不同频带后计算其能量占比,归一化后利用主成分分析对特征降维,构成的特征向量集作为最小二乘支持向量机输入;

利用混合粒子群算法对最小二乘支持向量机初始参数寻优,从而搭建预设的网络资产

证书识别模型;

混合粒子群算法将快速模拟退火法和粒子群算法结合,粒子群算法用于全局搜索区域的探索,当粒子群算法搜索到当前迭代次数的全体最优解时使用快速模拟退火法对粒子群算法找到的最佳位置进行调整得到新解,将新解与粒子群算法得出的最优解进行对比,如果新解优于最优解,则将新解作为当前最优解,否则以一定概率接受新解;

证书识别模块,用于获取待识别网络资产的证书信息,利用所述训练网络资产证书识别模型,对所述待识别网络资产的证书信息进行处理,得到网络资产证书识别结果,包括:

S41,获取待识别网络资产的证书信息;

S42,对所述待识别网络资产的证书信息进行处理,得到待识别网络资产的特征信息;

S43,根据所述待识别网络资产的特征信息,得到首次识别结果;

根据首次识别结果,对与所述首次识别结果不同的网络资产证书进行标注,得到正例证书和反例证书;

将所述反例证书加入未标记证书样本中,构成最优训练样本集;

利用所述最优训练样本集,对预设的网络资产证书识别模型进行训练,得到优化网络资产证书识别模型;

S44,根据所述首次识别结果,利用所述优化网络资产证书识别模型,对所述待识别网络资产的特征信息进行处理,得到网络资产证书识别结果。

4.一种网络资产证书识别装置,其特征在于,所述装置包括:

存储有可执行程序代码的存储器;

与所述存储器耦合的处理器;

所述处理器调用所述存储器中存储的所述可执行程序代码,执行如权利要求1-2任一项所述的网络资产证书识别方法。

一种网络资产证书识别方法及装置

技术领域

[0001] 本发明网络资产身份识别技术领域,尤其涉及一种网络资产证书识别方法及装置。

背景技术

[0002] 数字证书是指在互联网通讯中标志通讯各方身份信息的一个数字认证,可以用于识别验证通讯各方的身份。证书信息通常包括以下信息项:版本号、序列号、签名算法、颁布者、有效期、主体、主体公钥、主体公钥算法、签名值等。其中版本号是证书的版本信息,每个证书都有一个唯一的证书序列号,签名算法是认证过程所使用的签名算法,颁布者是证书的发行机构名称,有效期标注证书的有效时间,主体是证书所有人的名称,主体公钥是证书所有人的公开密钥,签名值是证书发行者对证书的签名。

[0003] 数字证书是用于互联网信息活动中网络资产行为主体的身份证明,也是可以用于验证网络资产发送信息保密性和完整性的电子数据。合理准确地对网络资产的证书实施相似性度量,可以有效地对网络资产进行分类识别和查询检索。

[0004] 目前还没有直接针对证书相似性度量和识别的相关研究,考虑证书的结构包括若干个信息项(版本号、序列号、签名算法、颁布者等)和每个信息项对应的取值内容,可以参考对文本分类中文本相似性衡量的方法。文本分类中的大多数算法,比如KNN方法、支持向量机方法、K均值方法等都需要通过计算相似度来达到分类的目的。常用的传统的文本相似性度量方法有余弦相似度、Jaccard相似系数、欧式距离、曼哈顿距离、切比雪夫距离、马氏距离等。

[0005] 现有文本相似性度量方法功能单一,不可以直接应用于证书的相似性度量。鉴于证书的结构包括若干个信息项(版本号、序列号、签名算法、颁布者等)和每个信息项对应的取值内容,可以考虑将两种相似性度量方法相结合,从证书的内容和结构两方面进行相似性比对;同时,结合自主学习反馈机制,探索自动化程度高、准确率更高的证书相似性度量和识别查询方法。

发明内容

[0006] 本发明所要解决的技术问题在于,提供一种网络资产证书识别方法及装置,能够通过对不同网络资产的证书进行相似性度量来达到对网络资产进行身份识别的目的,来弥补现有技术对网络资产识别自动化程度低、准确率低的问题。为了解决上述技术问题,本发明实施例第一方面公开了一种网络资产证书识别方法,所述方法包括:

[0007] S1,获取网络资产的证书信息,所述网络资产的证书信息构成资产证书库;

[0008] S2,对所述网络资产的证书信息进行处理,得到网络资产的特征信息,所述网络资产的特征信息构成特征库;

[0009] S3,利用所述网络资产的特征信息,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型;

[0010] S4,获取待识别网络资产的证书信息,利用所述训练网络资产证书识别模型,对所述待识别网络资产的证书信息进行处理,得到网络资产证书识别结果。

[0011] 作为一种可选的实施方式,本发明实施例第一方面中,所述对所述网络资产的证书信息进行处理,得到网络资产的特征信息,包括:

[0012] S21,对所述网络资产的证书信息进行结构相似度计算,得到结构相似度信息;

[0013] S22,对所述网络资产的证书信息进行内容相似度计算,得到内容相似度信息;

[0014] S23,对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息。

[0015] 作为一种可选的实施方式,本发明实施例第一方面中,所述对所述网络资产的证书信息进行结构相似度计算,得到结构相似度信息,包括:

[0016] S211,获取网络资产A的证书信息,所述待网络资产A的信息项数量为m;

[0017] S212,获取网络资产B的证书信息,所述网络资产B的信息项数量为n;

[0018] S213,对所述网络资产A的证书信息和所述网络资产B的证书信息进行处理,得到所述网络资产A和所述网络资产B都包含的信息项数量x;

[0019] S214,利用结构相似度计算模型,对所述待网络资产A的信息项数量为m、所述网络资产B的信息项数量为n和所述信息项数量x进行处理,得到结构相似度信息;

[0020] 所述结构相似度计算模型为:

[0021] $$str(A,B)=\frac{x}{m+n-x}$$

[0022] 其中str(A,B)为结构相似度信息。

[0023] 作为一种可选的实施方式,本发明实施例第一方面中,所述对所述网络资产的证书信息进行内容相似度计算,得到内容相似度信息,包括:

[0024] S221,对网络资产A的证书信息进行处理,得到网络资产A的证书信息项形式化表示$content_A=(a_1,a_2,\cdots,a_x)$;

[0025] S222,对网络资产B的证书信息进行处理,得到网络资产B的证书信息项形式化表示$content_B=(b_1,b_2,\cdots,b_x)$;

[0026] S223,利用内容相似度信息计算模型,对所述网络资产A的证书信息项形式化表示$content_A=(a_1,a_2,\cdots,a_x)$和所述网络资产B的证书信息项形式化表示进行处理,得到内容相似度信息;

[0027] 所述内容相似度计算模型为:

[0028] $$con(A,B)=\frac{1}{1+d(A,B)}$$

[0029] 其中,con(A,B)为内容相似度信息,$d(A,B)=\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}$。

[0030] 作为一种可选的实施方式,本发明实施例第一方面中,所述对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息,包括:

[0031] 利用相似度信息融合模型,对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息;

[0032] 所述相似度信息融合模型为:

[0033] $$total(A,B)=str(A,B)*con(A,B)=\frac{x}{m+n-x}*\frac{1}{1+\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}}$$

[0034] 其中,total(A,B)为网络资产的特征信息,str(A,B)为结构相似度信息,con(A,B)为内容相似度信息。

[0035] 作为一种可选的实施方式,本发明实施例第一方面中,所述利用所述网络资产的特征信息,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型,包括:

[0036] S31,对所述网络资产的特征信息进行划分,得到标注证书样本和未标记证书样本;

[0037] S32,将所述未标记证书样本作为训练样本,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型。

[0038] 作为一种可选的实施方式,本发明实施例第一方面中,所述获取待识别网络资产的证书信息,利用所述训练网络资产证书识别模型,对所述待识别网络资产的证书信息进行处理,得到网络资产证书识别结果,包括:

[0039] S41,获取待识别网络资产的证书信息;

[0040] S42,对所述待识别网络资产的证书信息进行处理,得到待识别网络资产的特征信息;

[0041] S43,根据所述待识别网络资产的特征信息,得到首次识别结果;

[0042] S44,根据所述首次识别结果,利用所述训练网络资产证书识别模型,对所述待识别网络资产的特征信息进行处理,得到网络资产证书识别结果。

[0043] 作为一种可选的实施方式,本发明实施例第一方面中,所述方法还包括:

[0044] 获取待识别网络资产的证书信息;

[0045] 对所述待识别网络资产的证书信息进行识别,得到首次识别结果;

[0046] 根据首次识别结果,对与所述首次识别结果不同的网络资产证书进行标注,得到正例证书和反例证书;

[0047] 将所述反例证书加入所述未标记证书样本中,构成最优训练样本集;

[0048] 利用所述最优训练样本集,对预设的网络资产证书识别模型进行训练,得到优化网络资产证书识别模型。

[0049] 本发明实施例第二方面公开了一种网络资产证书识别装置,所述装置包括:

[0050] 数据获取模块,用于获取网络资产的证书信息,所述网络资产的证书信息构成资产证书库;

[0051] 特征提取模块,用于对所述网络资产的证书信息进行处理,得到网络资产的特征信息,所述网络资产的特征信息构成特征库;

[0052] 模型训练模块,用于利用所述网络资产的特征信息,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型;

[0053] 证书识别模块,用于获取待识别网络资产的证书信息,利用所述训练网络资产证书识别模型,对所述待识别网络资产的证书信息进行处理,得到网络资产证书识别结果。

[0054] 作为一种可选的实施方式,本发明实施例第二方面中,所述对所述网络资产的证书信息进行处理,得到网络资产的特征信息,包括:

[0055] S21,对所述网络资产的证书信息进行结构相似度计算,得到结构相似度信息;

[0056] S22,对所述网络资产的证书信息进行内容相似度计算,得到内容相似度信息;

[0057] S23,对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息。

[0058] 作为一种可选的实施方式,本发明实施例第二方面中,所述对所述网络资产的证书信息进行结构相似度计算,得到结构相似度信息,包括:

[0059] S211,获取网络资产A的证书信息,所述待网络资产A的信息项数量为m;

[0060] S212,获取网络资产B的证书信息,所述网络资产B的信息项数量为n;

[0061] S213,对所述网络资产A的证书信息和所述网络资产B的证书信息进行处理,得到所述网络资产A和所述网络资产B都包含的信息项数量x;

[0062] S214,利用结构相似度计算模型,对所述待网络资产A的信息项数量为m、所述网络资产B的信息项数量为n和所述信息项数量x进行处理,得到结构相似度信息;

[0063] 所述结构相似度计算模型为:

[0064] $$str(A,B)=\frac{x}{m+n-x}$$

[0065] 其中str(A,B)为结构相似度信息。

[0066] 作为一种可选的实施方式,本发明实施例第二方面中,所述对所述网络资产的证书信息进行内容相似度计算,得到内容相似度信息,包括:

[0067] S221,对网络资产A的证书信息进行处理,得到网络资产A的证书信息项形式化表示$content_A=(a_1,a_2,\cdots,a_x)$;

[0068] S222,对网络资产B的证书信息进行处理,得到网络资产B的证书信息项形式化表示$content_B=(b_1,b_2,\cdots,b_x)$;

[0069] S223,利用内容相似度信息计算模型,对所述网络资产A的证书信息项形式化表示$content_A=(a_1,a_2,\cdots,a_x)$和所述网络资产B的证书信息项形式化表示进行处理,得到内容相似度信息;

[0070] 所述内容相似度计算模型为:

[0071] $$con(A,B)=\frac{1}{1+d(A,B)}$$

[0072] 其中,con(A,B)为内容相似度信息,$d(A,B)=\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}$。

[0073] 作为一种可选的实施方式,本发明实施例第二方面中,所述对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息,包括:

[0074] 利用相似度信息融合模型,对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息;

[0075] 所述相似度信息融合模型为:

[0076] $$total(A,B)=str(A,B)*con(A,B)=\frac{x}{m+n-x}*\frac{1}{1+\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}}$$

[0077] 其中,total(A,B)为网络资产的特征信息,str(A,B)为结构相似度信息,con(A,B)为内容相似度信息。

[0078] 作为一种可选的实施方式,本发明实施例第二方面中,所述利用所述网络资产的特征信息,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型,包括:

[0079] S31,对所述网络资产的特征信息进行划分,得到标注证书样本和未标记证书样本;

[0080] S32,将所述未标记证书样本作为训练样本,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型。

[0081] 作为一种可选的实施方式,本发明实施例第二方面中,所述获取待识别网络资产的证书信息,利用所述训练网络资产证书识别模型,对所述待识别网络资产的证书信息进行处理,得到网络资产证书识别结果,包括:

[0082] S41,获取待识别网络资产的证书信息;

[0083] S42,对所述待识别网络资产的证书信息进行处理,得到待识别网络资产的特征信息;

[0084] S43,根据所述待识别网络资产的特征信息,得到首次识别结果;

[0085] S44,根据所述首次识别结果,利用所述训练网络资产证书识别模型,对所述待识别网络资产的特征信息进行处理,得到网络资产证书识别结果。

[0086] 作为一种可选的实施方式,本发明实施例第二方面中,所述方法还包括:

[0087] 获取待识别网络资产的证书信息;

[0088] 对所述待识别网络资产的证书信息进行识别,得到首次识别结果;

[0089] 根据首次识别结果,对与所述首次识别结果不同的网络资产证书进行标注,得到正例证书和反例证书;

[0090] 将所述反例证书加入所述未标记证书样本中,构成最优训练样本集;

[0091] 利用所述最优训练样本集,对预设的网络资产证书识别模型进行训练,得到优化网络资产证书识别模型。

[0092] 本发明第三方面公开了另一种网络资产证书识别装置,所述装置包括:

[0093] 存储有可执行程序代码的存储器;

[0094] 与所述存储器耦合的处理器;

[0095] 所述处理器调用所述存储器中存储的所述可执行程序代码,执行本发明实施例第一方面公开的网络资产证书识别方法中的部分或全部步骤。

[0096] 与现有技术相比,本发明实施例具有以下有益效果:

[0097] 本发明提供了一种网络资产证书识别方法及装置,通过在结构相似性和内容相似性两方面对证书样本进行相似性度量,进而获取证书的结构特征、内容特征和综合特征;采用支持向量机模型,并从如何选择最为合适的证书作为支持向量机的训练样本出发,引入

自主学习思想,构建基于主动学习反馈的支持向量机,将未标记证书样本作为训练样本,输入到支持向量机分类器进行学习;在对用户未标记的样本进行不断识别查询过程中,针对规模较大的证书库,能够避免用户标记样本数量受限影响支持向量机分类器的分类效果,或者由于进行大量标记工作花费处理时间而增大分类器计算量;同时,能够标记歧义最大的证书,结合用户标记的样本一起构成最优训练样本集,进而缩减计算时间,提高分类查询效率,提高识别查询结果质量。

附图说明

[0098] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0099] 图1是本发明实施例公开的一种网络资产证书识别方法的流程示意图;

[0100] 图2是本发明实施例公开的另一种网络资产证书识别方法的流程示意图;

[0101] 图3是本发明实施例公开的基于自主学习反馈的网络资产证书识别查询系统的结构示意图;

[0102] 图4是本发明实施例公开的样本标注示意图;

[0103] 图5是本发明实施例公开的主动学习模型示意图;

[0104] 图6是本发明实施例公开的一种网络资产证书识别装置的结构示意图;

[0105] 图7是本发明实施例公开的另一种网络资产证书识别装置的结构示意图。

具体实施方式

[0106] 为了使本技术领域的人员更好地理解本发明方案,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0107] 本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别不同对象,而不是用于描述特定顺序。此外,术语“包括”和“具有”以及它们任何变形,意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、装置、产品或设备没有限定于已列出的步骤或单元,而是可选地还包括没有列出的步骤或单元,或可选地还包括对于这些过程、方法、产品或设备固有的其他步骤或单元。

[0108] 在本文中提及“实施例”意味着,结合实施例描述的特定特征、结构或特性可以包含在本发明的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例,也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是,本文所描述的实施例可以与其它实施例相结合。

[0109] 本发明公开了一种网络资产证书识别方法及装置,能够获取网络资产的证书信息,所述网络资产的证书信息构成资产证书库;对所述网络资产的证书信息进行处理得到网络资产的特征信息,所述网络资产的特征信息构成特征库;利用所述网络资产的特征信息,对预设的网络资产证书识别模型进行训练得到训练网络资产证书识别模型;获取待识

别网络资产的证书信息,利用所述训练网络资产证书识别模型,对所述待识别网络资产的证书信息进行处理,得到网络资产证书识别结果。本发明将两种相似性度量方法相结合,从证书的内容和结构两方面进行相似性比对;同时,结合自主学习反馈机制,探索自动化程度高、准确率更高的证书相似性度量和识别查询方法。以下分别进行详细说明。

[0110] 实施例一

[0111] 请参阅图1,图1是本发明实施例公开的一种网络资产证书识别方法的流程示意图。其中,图1所描述的网络资产证书识别方法应用于网络资产证书识别系统中,本发明实施例不做限定。如图1所示,该网络资产证书识别方法可以包括以下操作:

[0112] S1,获取网络资产的证书信息,所述网络资产的证书信息构成资产证书库;

[0113] S2,对所述网络资产的证书信息进行处理,得到网络资产的特征信息,所述网络资产的特征信息构成特征库;

[0114] S3,利用所述网络资产的特征信息,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型;

[0115] S4,获取待识别网络资产的证书信息,利用所述训练网络资产证书识别模型,对所述待识别网络资产的证书信息进行处理,得到网络资产证书识别结果。

[0116] 资产证书库由网络资产的证书信息构成,包括以下信息项:版本号、序列号、签名算法、颁布者、有效期、主体、主体公钥、主体公钥算法、签名值等;

[0117] 可选的,所述对所述网络资产的证书信息进行处理,得到网络资产的特征信息,包括:

[0118] S21,对所述网络资产的证书信息进行结构相似度计算,得到结构相似度信息;

[0119] S22,对所述网络资产的证书信息进行内容相似度计算,得到内容相似度信息;

[0120] S23,对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息。

[0121] 可选的,所述对所述网络资产的证书信息进行结构相似度计算,得到结构相似度信息,包括:

[0122] S211,获取网络资产A的证书信息,所述待网络资产A的信息项数量为m;

[0123] S212,获取网络资产B的证书信息,所述网络资产B的信息项数量为n;

[0124] S213,对所述网络资产A的证书信息和所述网络资产B的证书信息进行处理,得到所述网络资产A和所述网络资产B都包含的信息项数量x;

[0125] S214,利用结构相似度计算模型,对所述待网络资产A的信息项数量为m、所述网络资产B的信息项数量为n和所述信息项数量x进行处理,得到结构相似度信息;

[0126] 所述结构相似度计算模型为:

[0127] $$str(A,B)=\frac{x}{m+n-x}$$

[0128] 其中str(A,B)为结构相似度信息。

[0129] 可选的,所述对所述网络资产的证书信息进行内容相似度计算,得到内容相似度信息,包括:

[0130] S221,对网络资产A的证书信息进行处理,得到网络资产A的证书信息项形式化表示$content_A=(a_1,a_2,\cdots,a_x)$;

[0131] S222,对网络资产B的证书信息进行处理,得到网络资产B的证书信息项形式化表示$content_B=(b_1,b_2,\cdots,b_x)$;

[0132] S223,利用内容相似度信息计算模型,对所述网络资产A的证书信息项形式化表示$content_A=(a_1,a_2,\cdots,a_x)$和所述网络资产B的证书信息项形式化表示进行处理,得到内容相似度信息;

[0133] 所述内容相似度计算模型为:

[0134] $$con(A,B)=\frac{1}{1+d(A,B)}$$

[0135] 其中,con(A,B)为内容相似度信息,$d(A,B)=\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}$。

[0136] 可选的,所述对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息,包括:

[0137] 利用相似度信息融合模型,对所述结构相似度信息和所述内容相似度信息进行融合,得到网络资产的特征信息;

[0138] 所述相似度信息融合模型为:

[0139] $$total(A,B)=str(A,B)*con(A,B)=\frac{x}{m+n-x}*\frac{1}{1+\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}}$$

[0140] 其中,total(A,B)为网络资产的特征信息,str(A,B)为结构相似度信息,con(A,B)为内容相似度信息。

[0141] 可选的,所述利用所述网络资产的特征信息,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型,包括:

[0142] S31,对所述网络资产的特征信息进行划分,得到标注证书样本和未标记证书样本;

[0143] 所述划分方法可以按照3:7的比例进行,本发明不做限制。

[0144] S32,将所述未标记证书样本作为训练样本,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型。

[0145] 可选的,所述获取待识别网络资产的证书信息,利用所述训练网络资产证书识别模型,对所述待识别网络资产的证书信息进行处理,得到网络资产证书识别结果,包括:

[0146] S41,获取待识别网络资产的证书信息;

[0147] S42,对所述待识别网络资产的证书信息进行处理,得到待识别网络资产的特征信息;

[0148] S43,根据所述待识别网络资产的特征信息,得到首次识别结果;

[0149] S44,根据所述首次识别结果,利用所述训练网络资产证书识别模型,对所述待识别网络资产的特征信息进行处理,得到网络资产证书识别结果。

[0150] 可选的,所述方法还包括:

[0151] 获取待识别网络资产的证书信息;

[0152] 对所述待识别网络资产的证书信息进行识别,得到首次识别结果;

[0153] 根据首次识别结果,对与所述首次识别结果不同的网络资产证书进行标注,得到正例证书和反例证书;

[0154] 用户根据与所述首次识别结果,设置一个阈值,标注正例和反例证书,数量自定义,可以进行二次查询识别得到结果。

[0155] 将所述反例证书加入所述未标记证书样本中,构成最优训练样本集;

[0156] 利用所述最优训练样本集,对预设的网络资产证书识别模型进行训练,得到优化网络资产证书识别模型。

[0157] 可选的,在得到结构相似度信息和内容相似度信息后,可以利用如下方法进行特征融合:

[0158] 将结构相似度信息X和内容相似度信息Y投影到一维进行线性表示,分别对应投影向量a和b,则投影后的特征矩阵变为:

[0159] $X'=a^TX, Y'=b^TY$

[0160] 使X'与Y'之间的相关系数最大化,从而得到相关系数最大时的投影向量a和b。即:

[0161] $$\underbrace{\arg\max}_{a,b}\frac{\operatorname{cov}(X',Y')}{\sqrt{D(X')}\sqrt{D(Y')}}$$

[0162] 投影前先对数据进行标准化,标准化的目的是使数据的均值为0,方差为1。这种条件下,可以得到:

[0163] $\operatorname{cov}(X',Y')=\operatorname{cov}(a^TX,b^TY)=E(<a^TX,b^TY>)=E((a^TX)(b^TY)^T)=a^TE(XY^T)b$

[0164] $D(X')=D(a^TX)=a^TE(XX^T)a$

[0165] $D(Y')=D(b^TY)=b^TE(YY^T)b$

[0166] 因为X和Y经过标准化后均值为0,所以

[0167] $D(X)=\operatorname{cov}(X,X)=E(XX^T)$

[0168] $D(Y)=\operatorname{cov}(Y,Y)=E(YY^T)$

[0169] $\operatorname{cov}(X,Y)=E(XY^T)$,$\operatorname{cov}(Y,X)=E(YX^T)$

[0170] 设$S_{XX}=\operatorname{cov}(X,X)$,则求解目标转化为:

[0171] $$\underbrace{\arg\max}_{a,b}\frac{a^TS_{XX}b}{\sqrt{a^TS_{XX}a}\sqrt{b^TS_{YY}b}}$$

[0172] 上式函数的求解方法流程为:

[0173] Step1:计算X,Y的方差S_{XX}以及S_{YY},XY和YX的协方差$S_{XY}=S_{YX}{}^T$。
Step2:计算矩阵$M=S_{XX}^{-\frac{1}{2}}S_{XY}S_{YY}^{-\frac{1}{2}}$。

[0174] Step3:求解M的奇异值,得到最大的奇异值和它的左右奇异向量u,v。

[0175] Step4:X与Y的投影向量a和b分别为:$a=S_{XX}^{-\frac{1}{2}}u, b=S_{YY}^{-\frac{1}{2}}v$。

[0176] 可选的,预设的网络资产证书识别模型采用离散小波分解对网络资产的特征信息进行处理。网络资产的特征信息被分解到不同频带后计算其能量占比,归一化后利用主成分分析对特征降维,构成的特征向量集作为最小二乘支持向量机输入。再利用混合粒子群算法对最小二乘支持向量机初始参数寻优,从而搭建预设的网络资产证书识别模型。混合

粒子群算法将FSA(快速模拟退火法)和PSO(粒子群算法)结合,PSO算法用于全局搜索区域的探索,当PSO搜索到当前迭代次数的全体最优解时使用FSA算法对PSO找到的最佳位置进行调整得到新解,将新解与PSO算法得出的最优解进行对比,如果新解优于最优解,则将新解作为当前最优解,否则以一定概率接受新解。通过这种结合得方式,避免了过早陷入局部最优解,进一步平衡了全局搜索和局部搜索之间的关系,优化了算法的效率和精度。

[0177] FSA算法流程:

[0178] 1)从初始解开始,定义初始温度T、终止温度T_{min}、降温速率r;

[0179] 2)在每个温度下,使用柯西分布对当前解进行扰动,得到新的解;

[0180] 3)对于每个新解,计算其目标函数值或成本函数值。如果新解更优,则接受该解作为当前解;否则根据Metropolis准则概率接受该解。

[0181] 4)重复步骤2)和3),直到温度下降到T_{min}。在温度下降的过程中,接受劣解的概率逐渐降低,最终只接受更优的解。

[0182] 5)当温度降到T_{min}时,算法结束,返回找到的最优解。

[0183] 实施例二

[0184] 请参阅图2,图2是本发明实施例公开的另一种网络资产证书识别方法的流程示意图。其中,图2所描述的网络资产证书识别方法应用于网络资产证书识别系统中,本发明实施例不做限定。如图2所示,该网络资产证书识别方法可以包括以下操作:

[0185] (1)获取待识别网络资产的证书信息

[0186] (2)利用所述证书相似性度量方法,通过结构和内容量化两个方面,计算证书的结构相似度和内容相似度,进而获取证书的结构特征、内容特征和综合特征;

[0187] (3)构建基于主动学习反馈的支持向量机,优化训练样本集;

[0188] (4)通过上述分类器模型进行待识别资产证书的相似度测量,对网络资产进行识别,进而建立相似网络资产库。

[0189] 基于自主学习反馈的网络资产证书识别查询系统,具体思想如下:

[0190] 1)总体结构

[0191] 检索反馈接口:包括证书查询、结果反馈。主要基于自主学习训练和认证反馈的方式,通过资产证书相似度计算,实现待识别网络资产证书和网络资产证书库中的特征对比,标识不同类别的证书;进而提供用户对选择样本证书进行再次查询的功能。

[0192] 资产证书库:已知的原始资产证书;

[0193] 特征库:结构特征、内容特征形成的综合特征库;

[0194] 2)工作原理

[0195] 用户通过证书查询模块,选择需要待查询识别的证书,通过系统显示首次识别查询结果;

[0196] 用户根据结果,标注正例和反例证书,数量自定义,可以进行二次查询识别得到结果。其中,通过主动学习反馈机制划分具有较大歧义的证书,用户可将其标注为反例证书并再次反馈证书标记结果;

[0197] 根据结果反馈模块的相关反馈结果是否符合证书识别查询预期,决策系统执行次数,直至识别结果具备一定的准确性,结束操作。

[0198] 基于自主学习反馈的网络资产证书识别查询系统原理示意图如图3。

[0199] 3)实现设计

[0200] 证书特征提取

[0201] (1)提取每个网络资产证书样本的信息项和每个信息项的取值内容;

[0202] 证书信息包括以下信息项:版本号、序列号、签名算法、颁布者、有效期、主体、主体公钥、主体公钥算法、签名值等。其中版本号是证书的版本信息,每个证书都有一个唯一的证书序列号,签名算法是认证过程所使用的签名算法,颁布者是证书的发行机构名称,有效期标注证书的有效时间,主体是证书所有人的名称,主体公钥是证书所有人的公开密钥,签名值是证书发行者对证书的签名。

[0203] 证书信息的信息项可以表示成向量的形式,主体A的证书信息可以形式化表示为

[0204] $cert_A=(A_1,A_2,\cdots,A_n)$,

[0205] 其中对于i=1到n,A_i分别表示证书的版本号、序列号、签名算法、颁布者、有效期、主体、主体公钥、主体公钥算法、签名值等。每个证书信息项都会有相应的取值,称为信息项的内容,比如,版本号可以是V3,形式化表示为A_1=V3。

[0206] (2)计算每个网络资产证书样本的与已知网络资产证书间的结构相似度,作为证书的结构特征;

[0207] 结构相似性,是指计算两个证书信息之间的结构相似性,列出两个证书的所有证书信息项,利用Jaccard相似系数计算两个证书信息之间的结构相似度。假设证书A有m个证书信息项,证书B有n个证书信息项,其中A和B均包含的信息项有x个,那么证书A和证书B的结构相似度表示为

[0208] $$str(A,B)=\frac{x}{m+n-x}$$。

[0209] 如果两个证书包含的信息项的名称和数目相同,即m=n,并且x=m=n,那么两个证书的结构相似度为1。

[0210] 计算每个网络资产证书样本的与已知网络资产证书间的内容相似度,作为证书的内容特征;

[0211] (3)内容相似性,是指计算两个证书相同信息项取值内容的相似性,列出两个证书均存在的证书信息项,对信息项的取值进行相似性度量,利用欧氏距离计算两个证书相同信息项取值内容的相似度。假设证书A和证书B均包含的证书信息项包括版本号、序列号、签名算法、颁布者、有效期、主体、主体公钥、主体公钥算法、签名值等x个信息项,证书A和证书B相同信息项的取值内容可以形式化表示为$content_A=(a_1,a_2,\cdots,a_x)$,$content_B=(b_1,b_2,\cdots,b_x)$,计算两个证书相同信息项取值内容之间的欧氏距离为

[0212] $$d(A,B)=\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}$$,

[0213] 欧氏距离越小,两个证书的相似度就越大;欧氏距离越大,两个证书相似度就越小。那么证书A和证书B的内容相似度为

[0214] $$con(A,B)=\frac{1}{1+d(A,B)}$$。

[0215] (4)基于证书的结构特征和内容特征,计算证书样本的综合特征;

[0216] 根据所述证书A和证书B的结构相似度和内容相似度,计算证书A和证书B之间的总

体相似度为

[0217] $$total(A,B)=str(A,B)*con(A,B)=\frac{x}{m+n-x}*\frac{1}{1+\sqrt{\sum_{i=1}^{x}(a_i-b_i)^2}}。$$

[0218] 主动学习反馈证书识别模型构建：

[0219] (1)证书查询

[0220] 主要基于相似度测量方法和证书综合特征进行网络资产证书识别和结果展示。考虑到证书库中与待检测资产证书类似程度高的证书情况，系统界面会基于国别地域、应用等规划展示区域。

[0221] (2)样本标注

[0222] 主要根据用户标记的正例和反例样本构成标记样本集，作为支持向量机的输入进行训练，进而确保识别查询结果的准确性。

[0223] 样本标注是实现结果反馈证书识别查询的过程操作，由用户自行考虑标记数量并按需标记相关正例证书样本和反例证书样本。同时，结合主动学习，针对识别查询结果歧义较大的证书，用户也可对这一阶段的识别结果进行正例和反例标注。如图3所示是本发明实施例公开的基于自主学习反馈的网络资产证书识别查询系统的结构示意图。

[0224] (3)基于主动学习的结果反馈

[0225] 通过采用主动学习反馈的模式，将用户大量未标记证书作为训练样本，输入到支持向量机分类器进行学习，能够得到具有较大歧义的证书结果；通过用户自行判定是否将其标记为正例证书或者反例证书，更新标记样本集，进而逐步构成最优训练样本集，提高识别查询准确性。图4是本发明实施例公开的样本标注示意图。图5是本发明实施例公开的主动学习模型示意图。

[0226] 基于主动学习的支持向量机模型具有循环执行特性，模型描述为S＝(A,H,U,M,D)，其中，A代表支持向量机分类器，H为查询函数，U为用户终端，M为标记的样本集，D为未标记的样本集。

[0227] 模型执行包括：

[0228] (1)结合相似度测量方法，获取证书结构特征、内容特征和综合特征，输出第一次证书识别查询结果；

[0229] (2)基于首次输出结果，由用户选择并标注正例样本和反例样本，获得训练样本；

[0230] (3)构造训练样本集并对训练样本进行学习，基于分类器，再次进行识别查询；

[0231] (4)在上述学习反馈结果中，自动计算输出与用户选择样本识别查询结果具有较大歧义的证书集，用户在歧义证书中自行选择、标注正例证书或者反例证书，并加入标注样本集；

[0232] (5)循环此过程，能够通过较少的样本集计算得到最优分类识别结果；

[0233] (6)计算待识别证书与证书库中每个证书的相似度距离，按照相似性排序，获取证书识别结果。

[0234] 实施例三

[0235] 请参阅图6，图6是本发明实施例公开的一种网络资产证书识别装置的结构示意

图。其中,图6所描述的网络资产证书识别装置应用于网络资产证书识别系统中,本发明实施例不做限定。如图6所示,该网络资产证书识别装置可以包括以下操作:

[0236] S301,数据获取模块,用于获取网络资产的证书信息,所述网络资产的证书信息构成资产证书库;

[0237] S302,特征提取模块,用于对所述网络资产的证书信息进行处理,得到网络资产的特征信息,所述网络资产的特征信息构成特征库;

[0238] S303,模型训练模块,用于利用所述网络资产的特征信息,对预设的网络资产证书识别模型进行训练,得到训练网络资产证书识别模型;

[0239] S304,证书识别模块,用于获取待识别网络资产的证书信息,利用所述训练网络资产证书识别模型,对所述待识别网络资产的证书信息进行处理,得到网络资产证书识别结果。

[0240] 实施例四

[0241] 请参阅图7,图7是本发明实施例公开的另一种网络资产证书识别装置的结构示意图。其中,图7所描述的网络资产证书识别装置应用于网络资产证书识别系统中,本发明实施例不做限定。如图6所示,该网络资产证书识别装置可以包括以下操作:

[0242] 存储有可执行程序代码的存储器401;

[0243] 与存储器401耦合的处理器402;

[0244] 处理器402调用存储器401中存储的可执行程序代码,用于执行实施例一或实施例二所描述的网络资产证书识别方法中的步骤。

[0245] 以上所描述的装置实施例仅是示意性的,其中作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理模块,即可以位于一个地方,或者也可以分布到多个网络模块上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0246] 通过以上的实施例的具体描述,本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件。基于这样的理解,上述技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在计算机可读存储介质中,存储介质包括只读存储器(Read-Only Memory,ROM)、随机存储器(Random Access Memory,RAM)、可编程只读存储器(Programmable Read-only Memory,PROM)、可擦除可编程只读存储器(Erasable Programmable Read Only Memory,EPROM)、一次可编程只读存储器(One-time Programmable Read-Only Memory,OTPROM)、电子抹除式可复写只读存储器(Electrically-Erasable Programmable Read-Only Memory,EEPROM)、只读光盘(Compact Disc Read-Only Memory,CD-ROM)或其他光盘存储器、磁盘存储器、磁带存储器、或者能够用于携带或存储数据的计算机可读的任何其他介质。

[0247] 最后应说明的是:本发明实施例公开的一种网络资产证书识别方法及装置所揭露的仅为本发明较佳实施例而已,仅用于说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解;其依然可以对前述各项实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这

些修改或替换,并不使相应的技术方案的本质脱离本发明各项实施例技术方案的精神和范围。

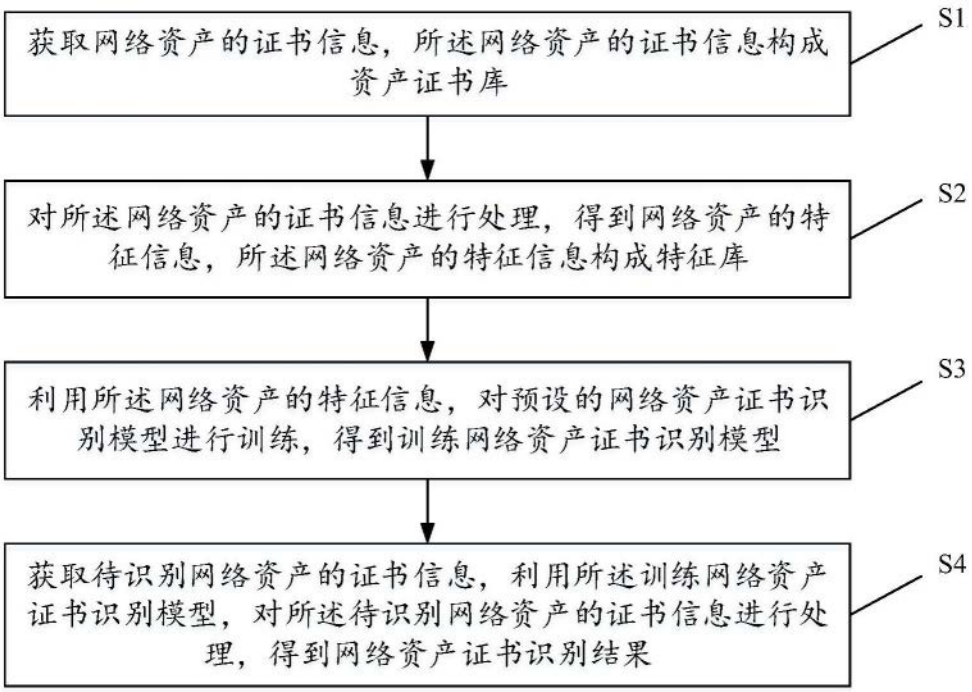

获取网络资产的证书信息，所述网络资产的证书信息构成资产证书库
S1
对所述网络资产的证书信息进行处理，得到网络资产的特征信息，所述网络资产的特征信息构成特征库
S2
利用所述网络资产的特征信息，对预设的网络资产证书识别模型进行训练，得到训练网络资产证书识别模型
S3
获取待识别网络资产的证书信息，利用所述训练网络资产证书识别模型，对所述待识别网络资产的证书信息进行处理，得到网络资产证书识别结果
S4

图1

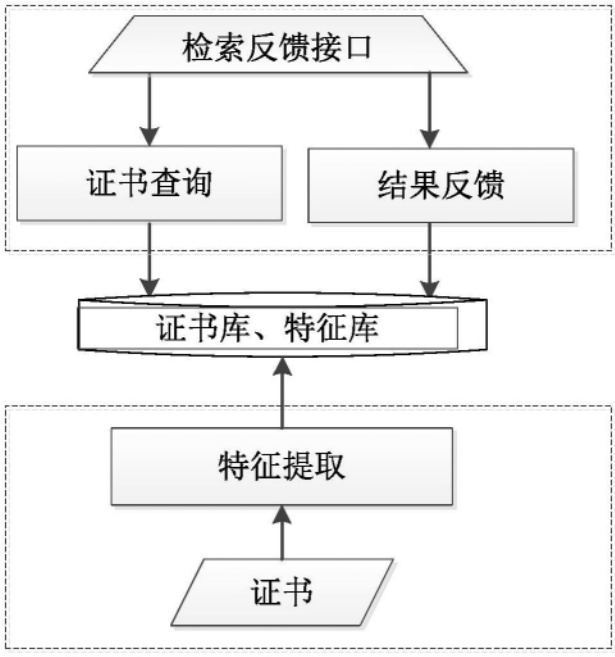

检索反馈接口
证书查询
结果反馈
证书库、特征库
特征提取
证书

图2

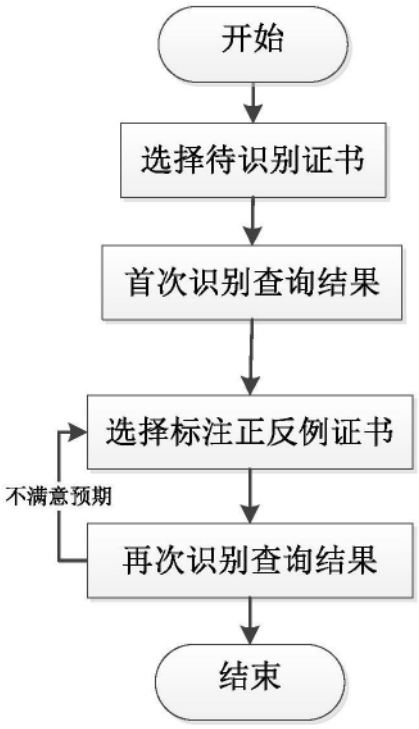

开始
选择待识别证书
首次识别查询结果
选择标注正反例证书
不满意预期
再次识别查询结果
结束

图3

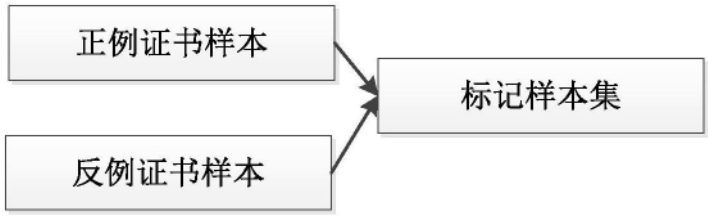

正例证书样本
反例证书样本
标记样本集

图4

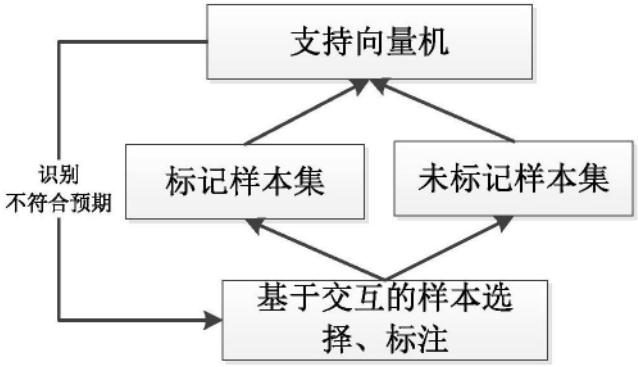

支持向量机
识别
不符合预期
标记样本集
未标记样本集
基于交互的样本选择、标注

图5

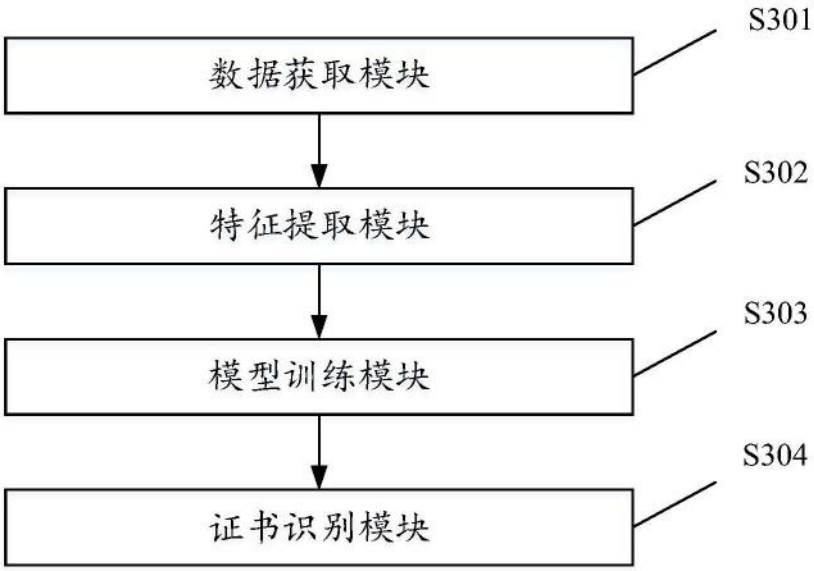

S301
数据获取模块
S302
特征提取模块
S303
模型训练模块
S304
证书识别模块

图6

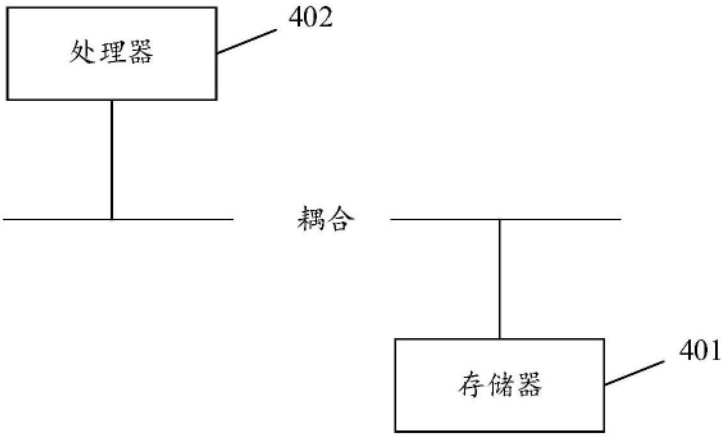

402
处理器
耦合
401
存储器

图7