



(12) 发明专利申请

(10) 申请公布号 CN 113486646 A

(43) 申请公布日 2021. 10. 08

(21) 申请号 202110722517.1

G06F 40/289 (2020.01)

(22) 申请日 2021.06.28

(71) 申请人 平安信托有限责任公司

地址 518033 广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心29层(西南、西北)、31层(3120室、3122室)、32层、33层

(72) 发明人 原琳 林达坤

(74) 专利代理机构 深圳市沃德知识产权代理事务所(普通合伙) 44347

代理人 高杰 于志光

(51) Int. Cl.

G06F 40/194 (2020.01)

G06F 40/186 (2020.01)

G06F 40/216 (2020.01)

权利要求书2页 说明书13页 附图2页

(54) 发明名称

产品报告发布方法、装置、电子设备及可读存储介质

(57) 摘要

本发明涉及智能决策领域,揭露一种产品报告发布方法,包括:从待分析产品文本信息中抽取预设字段,得到目标字段;提取报告生成请求中的产品类型,根据产品类型在预设的报告模板数据库中进行查询,得到目标报告模板;将目标字段与目标报告模板进行关联匹配,得到目标报告;对目标报告进行报告差异校验,得到目标校验差异率;获取预设反馈时间区间内的历史报告差异率,根据历史报告差异率构建校验区间;根据校验区间及目标校验差异率对目标报告进行发布评估,根据评估结果对目标报告进行发布。本发明还涉及一种区块链技术,所述目标报告模板可以存储在区块链节点中。本发明还提出一种产品报告发布装置、设备以及介质。本发明可以提高产品报告发布的效率。



1. 一种产品报告发布方法,其特征在于,所述方法包括:

根据产品报告生成请求获取待分析产品文本信息,从所述待分析产品文本信息中抽取预设字段,得到目标字段;

提取所述报告生成请求中的产品类型,根据所述产品类型在预设的报告模板数据库中进行筛选,得到目标报告模板;

将所述目标字段与所述目标报告模板进行关联匹配,得到目标报告;

对所述目标报告进行报告差异校验,得到目标校验差异率;

获取预设反馈时间区间内的历史报告差异率,根据所述历史报告差异率构建校验区间;

根据所述校验区间及所述目标校验差异率对所述目标报告进行发布评估,得到评估结果,根据所述评估结果对所述目标报告进行发布。

2. 如权利要求1所述的产品报告发布方法,其特征在于,所述根据产品报告生成请求获取待分析产品文本信息,包括:

提取所述报告生成请求中的文本信息路径;

根据所述文本信息路径在预设的产品文本数据库中获得待分析产品文本信息。

3. 如权利要求1所述的产品报告发布方法,其特征在于,所述从所述待分析产品文本信息中抽取预设字段,得到目标字段,包括:

对所述待分析产品文本信息中的每个字符进行坐标点位置标记,得到字符信息;

利用预设字段抽取模型对所述字符信息进行字段抽取,得到所述预设字段的开始字符及中间字符;

将所述预设字段的开始字符及中间字符按照序列顺序进行组合,得到所述的目标字段。

4. 如权利要求3所述的产品报告发布方法,其特征在于,所述利用预设字段抽取模型对所述字符信息进行字段抽取之前,还包括:

构建初始模型;

获取历史文本信息集,对所述历史文本信息集进行预设字段标记,得到训练集;

利用所述训练集对所述初始模型进行迭代训练,直至所述初始模型收敛,得到所述字段抽取模型。

5. 如权利要求1所述的产品报告发布方法,其特征在于,所述根据所述历史报告差异率构建校验区间,包括:

选取所有所述历史报告差异率中的中位数,得到第一差异率;

选取所有所述历史报告差异率中的最大值,得到所述第二差异率;

将所述第一差异率及所述第二差异率作为区间端点进行区间构建,得到所述校验区间。

6. 如权利要求5所述的产品报告发布方法,其特征在于,所述根据所述评估结果对所述目标报告进行发布,包括:

判断所述目标校验差异率是否在所述校验区间内;

若所述目标校验差异率在所述校验区间内,则评估结果为报告合格,发布所述目标报告,并发送预警信息至预设第一终端设备;

若所述目标校验差异率不在所述校验区间内；

当所述目标校验差异率小于所述与第一差异率，发布所述目标报告；

当所述目标校验差异率大于所述第二差异率，则将所述目标报告发送至预设第二终端设备进行审批。

7. 如权利要求1至6中任意一项所述的产品报告发布方法，其特征在于，所述获取预设反馈时间区间内的历史报告差异率之前，还包括：

获取所述报告生成请求的请求时间；

将所述请求时间作为区间右端点；

将预设时间周期作为区间长度；

根据所述区间右端点及所述区间长度进行区间构建，得到所述反馈时间区间。

8. 一种产品报告发布装置，其特征在于，包括：

报告生成模块，用于根据产品报告生成请求获取待分析产品文本信息，从所述待分析产品文本信息中抽取预设字段，得到目标字段；提取所述报告生成请求中的产品类型，根据所述产品类型在预设的报告模板数据库中进行筛选，得到目标报告模板；将所述目标字段与所述目标报告模板进行关联匹配，得到目标报告；

报告检验模块，用于对所述目标报告进行报告差异校验，得到目标校验差异率；获取预设反馈时间区间内的历史报告差异率，根据所述历史报告差异率构建校验区间；

报告发布模块，用于根据所述校验区间及所述目标校验差异率对所述目标报告进行发布评估，得到评估结果，根据所述评估结果对所述目标报告进行发布。

9. 一种电子设备，其特征在于，所述电子设备包括：

至少一个处理器；以及，

与所述至少一个处理器通信连接的存储器；其中，

所述存储器存储有可被所述至少一个处理器执行的计算机程序，所述计算机程序被所述至少一个处理器执行，以使所述至少一个处理器能够执行如权利要求1至7中任一项所述的产品报告发布方法。

10. 一种计算机可读存储介质，存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求1至7中任一项所述的产品报告发布方法。

## 产品报告发布方法、装置、电子设备及可读存储介质

### 技术领域

[0001] 本发明涉及智能决策领域,尤其涉及一种产品报告发布方法、装置、电子设备及可读存储介质。

### 背景技术

[0002] 随着经济社会的发展,为了保障人们的购物体验,一些重要的或特殊的商品在发售前都需要发布对应产品报告,已确保用户购买对应商品时充分了解产品购买的风险,例如金融机构发行的金融产品,需要按照金融机构按监管机构要求,向投资者,全面、充分、及时和准确地披露产品运营信息,因此,需要发布对应的产品报告。

[0003] 目前的产品报告发布方法,需要利用不同的提取规则对不同的产品信息分别进行提取,利用提取的信息生成产品报告,报告生成的效率低,同时生成的产品报告需要人工审批才能发布,综合导致产品发布的效率较低。

### 发明内容

[0004] 本发明提供一种产品报告发布方法、装置、电子设备及计算机可读存储介质,其主要目的在于提高产品报告发布的效率。

[0005] 为实现上述目的,本发明提供的一种产品报告发布方法,包括:

[0006] 根据产品报告生成请求获取待分析产品文本信息,从所述待分析产品文本信息中抽取预设字段,得到目标字段;

[0007] 提取所述报告生成请求中的产品类型,根据所述产品类型在预设的报告模板数据库中进行筛选,得到目标报告模板;

[0008] 将所述目标字段与所述目标报告模板进行关联匹配,得到目标报告;

[0009] 对所述目标报告进行报告差异校验,得到目标校验差异率;

[0010] 获取预设反馈时间区间内的历史报告差异率,根据所述历史报告差异率构建校验区间;

[0011] 根据所述校验区间及所述目标校验差异率对所述目标报告进行发布评估,得到评估结果,根据所述评估结果对所述目标报告进行发布。

[0012] 可选地,所述根据产品报告生成请求获取待分析产品文本信息,包括:

[0013] 提取所述报告生成请求中的文本信息路径;

[0014] 根据所述文本信息路径在预设的产品文本数据库中获取待分析产品文本信息。

[0015] 可选地,所述从所述待分析产品文本信息中抽取预设字段,得到目标字段,包括:

[0016] 对所述待分析产品文本信息中的每个字符进行坐标点位置标记,得到字符信息;

[0017] 利用预设字段抽取模型对所述字符信息进行字段抽取,得到所述预设字段的开始字符及中间字符;

[0018] 将所述预设字段的开始字符及中间字符按照序列顺序进行组合,得到所述的目标字段。

- [0019] 可选地,所述利用预设字段抽取模型对所述字符信息进行字段抽取之前,还包括:
- [0020] 构建初始模型;
- [0021] 获取历史文本信息集,对所述历史文本信息集进行预设字段标记,得到训练集;
- [0022] 利用所述训练集对所述初始模型进行迭代训练,直至所述初始模型收敛,得到所述字段抽取模型。
- [0023] 可选地,所述根据所述历史报告差异率构建校验区间,包括:
- [0024] 选取所有所述历史报告差异率中的中位数,得到第一差异率;
- [0025] 选取所有所述历史报告差异率中的最大值,得到所述第二差异率;
- [0026] 将所述第一差异率及所述第二差异率作为区间端点进行区间构建,得到所述校验区间。
- [0027] 可选地,所述根据所述评估结果对所述目标报告进行发布,包括:
- [0028] 判断所述目标校验差异率是否在所述校验区间内;
- [0029] 若所述目标校验差异率在所述校验区间内,则评估结果为报告合格,发布所述目标报告,并发送预警信息至预设第一终端设备;
- [0030] 若所述目标校验差异率不在所述校验区间内;
- [0031] 当所述目标校验差异率小于所述与第一差异率,发布所述目标报告;
- [0032] 当所述目标校验差异率大于所述第二差异率,则将所述目标报告发送至预设第二终端设备进行审批。
- [0033] 可选地,所述获取预设反馈时间区间内的历史报告差异率之前,还包括:
- [0034] 获取所述报告生成请求的请求时间;
- [0035] 将所述请求时间作为区间右端点;
- [0036] 将预设时间周期作为区间长度;
- [0037] 根据所述区间右端点及所述区间长度进行区间构建,得到所述反馈时间区间。
- [0038] 为了解决上述问题,本发明还提供一种产品报告发布装置,所述装置包括:
- [0039] 报告生成模块,用于根据产品报告生成请求获取待分析产品文本信息,从所述待分析产品文本信息中抽取预设字段,得到目标字段;提取所述报告生成请求中的产品类型,根据所述产品类型在预设的报告模板数据库中进行筛选,得到目标报告模板;将所述目标字段与所述目标报告模板进行关联匹配,得到目标报告;
- [0040] 报告检验模块,用于对所述目标报告进行报告差异校验,得到目标校验差异率;获取预设反馈时间区间内的历史报告差异率,根据所述历史报告差异率构建校验区间;
- [0041] 报告发布模块,用于根据所述校验区间及所述目标校验差异率对所述目标报告进行发布评估,得到评估结果,根据所述评估结果对所述目标报告进行发布。
- [0042] 为了解决上述问题,本发明还提供一种电子设备,所述电子设备包括:
- [0043] 存储器,存储至少一个计算机程序;及
- [0044] 处理器,执行所述存储器中存储的计算机程序以实现上述所述的产品报告发布方法。
- [0045] 为了解决上述问题,本发明还提供一种计算机可读存储介质,所述计算机可读存储介质中存储有至少一个计算机程序,所述至少一个计算机程序被电子设备中的处理器执行以实现上述所述的产品报告发布方法。

[0046] 本发明实施例根据产品报告生成请求获取待分析产品文本信息,从所述待分析产品文本信息中抽取预设字段,得到目标字段,利用模型提取字段,不需要单独构建提取规则,字段提取更方便,提取的效率更高;提取所述报告生成请求中的产品类型;将所述目标字段与所述目标报告模板进行关联匹配,得到目标报告,根据所述产品类型在预设的报告模板数据库中进行筛选,得到目标报告模板,根据产品类型自动对提取字段进行筛选及报告模板匹配,提高了报告的生成速度;对所述目标报告进行报告差异校验,得到目标校验差异率;获取预设反馈时间区间内的历史报告差异率,根据所述历史报告差异率构建校验区间;根据所述校验区间及所述目标校验差异率对所述目标报告进行发布评估,得到评估结果,根据所述评估结果对所述目标报告进行发布,构建评估规则直接对目标报告进行评估发布,降低了人工审批的工作量,提高了报告发布的效率。因此本发明实施例提出的产品报告发布方法、装置、电子设备及可读存储介质提高了产品报告发布的效率。

### 附图说明

[0047] 图1为本发明一实施例提供的产品报告发布方法的流程示意图;

[0048] 图2为本发明一实施例提供的产品报告发布装置的模块示意图;

[0049] 图3为本发明一实施例提供的实现产品报告发布方法的电子设备的内部结构示意图;

[0050] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

### 具体实施方式

[0051] 应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0052] 本发明实施例提供一种产品报告发布方法。所述产品报告发布方法的执行主体包括但不限于服务端、终端等能够被配置为执行本申请实施例提供的该方法的电子设备中的至少一种。换言之,所述产品报告发布方法可以由安装在终端设备或服务端设备的软件或硬件来执行,所述软件可以是区块链平台。所述服务端包括但不限于:单台服务器、服务器集群、云端服务器或云端服务器集群等。

[0053] 参照图1所示的本发明一实施例提供的产品报告发布方法的流程示意图,在本发明实施例中,所述产品报告发布方法包括:

[0054] S1、根据产品报告生成请求获取待分析产品文本信息,从所述待分析产品文本信息中抽取预设字段,得到目标字段;

[0055] 详细地,本发明实施例中,所述报告生成请求包括:待分析产品文本信息对应的文本信息路径及产品类型。

[0056] 进一步地,本发明实施例中,提取所述报告生成请求中的文本信息路径,根据所述文本信息路径在预设的产品文本数据库中获得待分析产品文本信息,其中,所述产品文本数据库为包含不同产品文本信息的数据库,所述文本信息路径为所述产品文本信息在所述产品文本数据库的存储路径。

[0057] 由于所述产品文本信息中的字符数量庞大,为了提取我们需要的信息,因此,需要对所述产品文本信息进行预设字段的提取。其中,所述预设字段为需要提取的字符组合的类型。例如,本发明其中一个实施例中所述预设字段包括:姓名、日期、金额;例如:所述产品

文本信息中字符包括“张”、“某三”“汽”、“车”，当所述预设字段为姓名时，提取所述产品文本信息中姓名对应的字符组合“张三”，得到目标字段。

[0058] 可选地，本发明实施例从所述待分析产品文本信息中抽取预设字段，得到所述目标字段，包括：对所述待分析产品文本信息中的每个字符进行坐标点位置标记，得到字符信息，可选地，本发明实施例中利用预设的编程语言脚本获取所述待分析产品文本信息中的每个字符的坐标点位置，并将所述坐标点位置标记对应的字符，得到所述字符信息，其中，所述编程语言脚本可以为Python脚本；进一步地，本发明实施例利用预设字段抽取模型对所述字符信息进行字段抽取，得到所述预设字段的开始字符及中间字符；将所述预设字段的开始字符及中间字符按照序列顺序进行组合，得到所述的目标字段。例如：预设字段为姓名字段，得到的姓名字段的开始字符为“张”、中间字符为“某三”，那么按照开始字符、中间字符的顺序，将字符“张”、“某三”组合得到目标字符“张三”。

[0059] 进一步地，本发明实施例利用预设的字段抽取模型对所述待分析产品文本信息进行字段抽取之前，还包括：

[0060] 步骤I、构建初始模型；

[0061] 详细地，本发明实施例利用预设的深度学习网络模型构建基础模型；较佳地，利用 Bert base网络模型作为基础模型，在所述基础模型后面接一层全连接网络和一层序列化标注算法网络，得到所述字段抽取模型，即在所述基础模型中增加全连接网络，用于计算输入到所述基础模型中的每个字符属于不同标签的概率，根据所述概率得到预设字段对应的字符组合；及在所述全连接网络之后加入序列化标注算法网络，以对所述全连接网络得到的字符组合的顺序进行约束，得到所述字段抽取模型。例如：利用所述全连接网络计算字符“张”属于姓名字段的开始字符概率最高、计算字符“三”属于姓名字段中间字符的概率最高，因此，通过全连接层得到的姓名字段为“张三”或“三张”，由于全连接层不能确定字符组合的顺序，因此，本发明实施例通过序列化标注算法网络确定姓名字段开始字符应该在姓名中间字符前面，因此，得到姓名字段对应的最终字符组合为“张三”。

[0062] 步骤II、获取历史文本信息集，对所述历史文本信息集进行预设字段标记，得到训练集；

[0063] 本发明实施例中，所述历史文本信息集为不同的历史文本信息的集合，所述历史文本信息为产品文本信息，如：金融产品文本信息。

[0064] 详细地，本发明实施例使用BIO标记法对所述历史文本信息集进行预设字段标记，得到训练集。

[0065] 详细地，所述对所述历史文本信息集进行预设字段标记，得到训练集，包括：根据预设字段构建包含非预设字段字符标签、预设字段开始字符标签、预设字段中间字符标签的标签集合，利用所述标签集合中对应的标签对所述历史文本信息集中的每个字符进行标记，得到训练集。例如：历史文本信息集包含的文本信息为“张三向银行借款”，预设字段为姓名字段，标签集合包括：非姓名字段字符、姓名字段开始字符、姓名字段中间字符，那么利用标签集合对文本信息“张三向银行借款”进行标记，“张”字符利用姓名字段开始字符标签标记为姓名字段开始字符、“某三”字符利用姓名字段中间字符标签标记为姓名字段中间字符，“向”字符利用非姓名字段字符标签标记为非姓名字段字符，“银行”字符利用非姓名字段字符标签标记为非姓名字段字符，“借款”字符利用非姓名字段字符标签标记为

非姓名字段字符。

[0066] 步骤III、利用所述训练集对所述初始模型进行迭代训练,直至所述初始模型收敛,得到所述字段抽取模型。详细地,本发明实施例中,利用所述训练集对所述字段抽取模型进行迭代训练,直至所述字段抽取模型收敛,得到训练完成的字段抽取模型,包括:

[0067] 步骤A:利用所述字段抽取模型中的基础模型将所述训练集中的每个字符进行向量化处理,得到字符向量集;

[0068] 步骤B:利用所述字段抽取模型中的全连接网络计算字符向量集中的每个字符向量的标签分类真实值,利用所述字段抽取模型中的序列化标注算法网络计算字符向量集中每个字符向量的标签顺序真实值,根据所述字符向量集中的对应的字符的标签确定字符向量集中每个字符向量的标签分类预测值及标签顺序预测值;

[0069] 步骤C:根据所述标签分类真实值及标签分类预测值,利用预设的第一损失函数进行计算,得到第一损失值;根据所述标签顺序真实值及标签顺序预测值,利用预设的第二损失函数进行计算,得到第二损失值;将所述第一损失值及所述第二损失值进行求和计算得到目标损失值,当所述目标损失值大于或等于预设阈值时,更新所述字段抽取模型的模型参数,并返回所述步骤A;当所述目标损失值小于预设阈值时,停止训练,得到所述训练完成的字段抽取模型。

[0070] 可选地,本发明实施例中,所述第一损失函数可用如下公式进行计算:

$$[0071] \quad L_{ce} = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)]$$

[0072] 其中, $L_{ce}$ 表示所述第一损失值, $N$ 为所述字符向量集的数据数目, $i$ 为正整数, $y_i$ 为所述标签分类真实值, $p_i$ 为所述标签分类预测值。

[0073] 可选地,所述第二损失函数包括:交叉熵损失函数、绝对值损失函数、平方损失函数等。

[0074] S2、提取所述报告生成请求中的产品类型,根据所述产品类型在预设的报告模板数据库中进行筛选,得到目标报告模板;

[0075] 本发明实施例中,不同产品类型对应的报告模板是不同的,如:甲类型的产品文本信息对应的报告模板为A;乙类型的产品文本信息对应的报告模板为B。因此根据所述产品类型在预设的报告模板数据库中进行筛选,筛选所述报告模板数据库中所述产品类型对应的报告模板,得到所述目标报告模板。

[0076] S3、将所述目标字段与所述目标报告模板进行关联匹配,得到目标报告;

[0077] 详细地,本发明实施例中,将所述目标字段与所述目标报告模板进行关联匹配,得到所述目标报告;详细地,本发明实施例中将目标字段替换所述目标报告模板中对应字段类型的空白字段,得到所述目标报告。

[0078] 本发明另一实施例中,所述目标报告模板还可以存储在区块链节点中,利用区块链高吞吐的特性,提高数据的存取效率。

[0079] S4、对所述目标报告进行报告差异校验,得到目标校验差异率;

[0080] 本发明实施例中为了保证发布的产品报告更加的准确,还需要对所述目标报告进行校验后,才能发布。

[0081] 可选地,本发明实施例中计算目标报告中缺失字段的比例、目标字段的字号差异比例、字符数量差异比例;根据所述缺失字段的比例、目标字段的字号差异比例、字符数量差异比例;例如:所述目标报告中空白字段为两个、目标字段为4个,那么缺失字段比例为 $2/(4+2)=0.33$ ,所述目标报告中目标字段某字段的字号为10号字体,校验规则中该字段对应的标准字号为5号字体,那么该字段的字号差异比例为 $(10-5)/5=1$ ,所述目标字段的字号差异比例为目标字段中所有字段的字号差异比例的平均值;所述目标报告中目标字段某字段的字符数量为1个字符,校验规则中该字段对应的标准字符数量为2个字符,那么该字段的字符数量差异比例为 $(2-1)/2=0.5$ ,所述目标字段的字符数量差异比例为目标字段中所有字段的字符数量差异比例的平均值。

[0082] 可选地,本发明实施例利用如下公式计算所述目标校验差异率:

$$[0083] \quad l = a * \alpha + b * \beta + c * \gamma$$

[0084] 其中,所述a为缺失字段比例、b为字号差异比例、c为字符数量差异比例; $\alpha$ 、 $\beta$ 、 $\gamma$ 为人为设置的权重参数。

[0085] S5、获取预设反馈时间区间内的历史报告差异率,根据所述历史报告差异率构建校验区间;

[0086] 详细地,本发明实施例中所述历史报告差异率为所述反馈时间区间内发送成功的报告的对应的目标校验差异率。

[0087] 进一步地,本发明实施例中所述获取预设反馈时间区间内的历史报告差异率之前,还包括:获取所述报告生成请求的请求时间;将所述请求时间作为区间右端点,将预设时间周期作为区间长度;根据所述区间右端点及所述区间长度进行区间构建,得到所述反馈时间区间。例如:预设时间周期为3天,所述请求时间为5/13,那么所述反馈时间区间为 $[5/10, 5/13]$ 。可选地,本发明实施例中所述时间周期为365天。

[0088] 进一步地,本发明实施例中根据所述历史报告差异率构建校验区间,包括:选取所有所述历史报告差异率中的中位数,得到第一差异率;选取所有所述历史报告差异率中的最大值,得到所述第二差异率;将所述第一差异率及所述第二差异率作为区间端点进行区间构建,得到所述校验区间。例如:第一差异率为0.3,第二差异率为0.5,那么所述校验区间为 $[0.3, 0.5]$ 。

[0089] S6、根据所述校验区间及所述目标校验差异率对所述目标报告进行发布评估,得到评估结果,根据所述评估结果对所述目标报告进行发布。

[0090] 详细地,本发明实施例中根据所述校验区间及所述目标校验差异率对所述目标报告进行发布评估,包括:判断所述目标校验差异率是否在所述校验区间内,若所述目标校验差异率在所述校验区间内,则评估结果为报告合格,发布所述目标报告,并发送预警信息至预设第一终端设备,若所述所述目标校验差异率不在所述校验区间内,判断所述目标校验差异率小于所述第一差异率或大于所述第二差异率,若所述目标校验差异率小于所述与第一差异率,那么发布所述目标报告;若所述目标校验差异率大于所述第二差异率,则将所述目标报告发送至预设第二终端设备进行审批。

[0091] 如图2所示,是本发明产品报告发布装置的功能模块图。

[0092] 本发明所述产品报告发布装置100可以安装于电子设备中。根据实现的功能,所述产品报告发布装置可以包括报告生成模块101、报告检验模块102、报告发布模块103,本发

所述模块也可以称之为单元,是指一种能够被电子设备处理器所执行,并且能够完成固定功能的一系列计算机程序段,其存储在电子设备的存储器中。

[0093] 在本实施例中,关于各模块/单元的功能如下:

[0094] 所述报告生成模块101,用于根据产品报告生成请求获取待分析产品文本信息,从所述待分析产品文本信息中抽取预设字段,得到目标字段;提取所述报告生成请求中的产品类型,根据所述产品类型在预设的报告模板数据库中进行筛选,得到目标报告模板;将所述目标字段与所述目标报告模板进行关联匹配,得到目标报告;

[0095] 详细地,本发明实施例中,所述报告生成请求包括:待分析产品文本信息对应的文本信息路径及产品类型。

[0096] 进一步地,本发明实施例中,所述报告生成模块101提取所述报告生成请求中的文本信息路径,根据所述文本信息路径在预设的产品文本数据库中获取待分析产品文本信息,其中,所述产品文本数据库为包含不同产品文本信息的数据库,所述文本信息路径为所述产品文本信息在所述产品文本数据库的存储路径。

[0097] 由于所述产品文本信息中的字符数量庞大,为了提取我们需要的信息,因此,所述报告生成模块101需要对所述产品文本信息进行预设字段的提取。其中,所述预设字段为需要提取的字符组合的类型。例如,本发明其中一个实施例中所述预设字段包括:姓名、日期、金额;例如:所述产品文本信息中字符包括“张”、“某三”“汽”、“车”,当所述预设字段为姓名时,提取所述产品文本信息中姓名对应的字符组合“张三”,得到目标字段。

[0098] 可选地,本发明实施例所述报告生成模块101从所述待分析产品文本信息中抽取预设字段,得到所述目标字段,包括:对所述待分析产品文本信息中的每个字符进行坐标点位置标记,得到字符信息,可选地,本发明实施例中利用预设的编程语言脚本获取所述待分析产品文本信息中的每个字符的坐标点位置,并将所述坐标点位置标记对应的字符,得到所述字符信息,其中,所述编程语言脚本可以为Python脚本;进一步地,本发明实施例利用预设字段抽取模型对所述字符信息进行字段抽取,得到所述预设字段的开始字符及中间字符;将所述预设字段的开始字符及中间字符按照序列顺序进行组合,得到所述的目标字段。例如:预设字段为姓名字段,得到的姓名字段的开始字符为“张”、中间字符为“某三”,那么按照开始字符、中间字符的顺序,将字符“张”、“某三”组合得到目标字符“张三”。

[0099] 进一步地,本发明实施例所述报告生成模块101利用预设的字段抽取模型对所述待分析产品文本信息进行字段抽取之前,还包括:

[0100] 步骤I、构建初始模型;

[0101] 详细地,本发明实施例利用预设的深度学习网络模型构建基础模型;较佳地,利用 Bert base网络模型作为基础模型,在所述基础模型后面接一层全连接网络和一层序列化标注算法网络,得到所述字段抽取模型,即在所述基础模型中增加全连接网络,用于计算输入到所述基础模型中的每个字符属于不同标签的概率,根据所述概率得到预设字段对应的字符组合;及在所述全连接网络之后加入序列化标注算法网络,以对所述全连接网络得到的字符组合的顺序进行约束,得到所述字段抽取模型。例如:利用所述全连接网络计算字符“张”属于姓名字段的开始字符概率最高、计算字符“三”属于姓名字段中间字符的概率最高,因此,通过全连接层得到的姓名字段为“张三”或“三张”,由于全连接层不能确定字符组合的顺序,因此,本发明实施例通过序列化标注算法网络确定姓名字段开始字符应该在姓

名中间字符前面,因此,得到姓名字段对应的最终字符组合为“张三”。

[0102] 步骤II、获取历史文本信息集,对所述历史文本信息集进行预设字段标记,得到训练集;

[0103] 本发明实施例中,所述历史文本信息集为不同的历史文本信息的集合,所述历史文本信息为产品文本信息,如:金融产品文本信息。

[0104] 详细地,本发明实施例所述报告生成模块101使用BIO标记法对所述历史文本信息集进行预设字段标记,得到训练集。

[0105] 详细地,所述报告生成模块101对所述历史文本信息集进行预设字段标记,得到训练集,包括:根据预设字段构建包含非预设字段字符标签、预设字段开始字符标签、预设字段中间字符标签的标签集合,利用所述标签集合中对应的标签对所述历史文本信息集中的每个字符进行标记,得到训练集。例如:历史文本信息集包含的文本信息为“张三向银行借款”,预设字段为姓名字段,标签集合包括:非姓名字段字符、姓名字段开始字符、姓名字段中间字符,那么利用标签集合对文本信息“张三向银行借款”进行标记,“张”字符利用姓名字段开始字符标签标记为姓名字段开始字符、“某三”字符利用姓名字段中间字符标签标记为姓名字段中间字符,“向”字符利用非姓名字段字符标签标记为非姓名字段字符,“银行”字符利用非姓名字段字符标签标记为非姓名字段字符,“借款”字符利用非姓名字段字符标签标记为非姓名字段字符。

[0106] 步骤III、利用所述训练集对所述初始模型进行迭代训练,直至所述初始模型收敛,得到所述字段抽取模型。详细地,本发明实施例中,所述报告生成模块101利用所述训练集对所述字段抽取模型进行迭代训练,直至所述字段抽取模型收敛,得到训练完成的字段抽取模型,包括:

[0107] 步骤A:利用所述字段抽取模型中的基础模型将所述训练集中的每个字符进行向量化处理,得到字符向量集;

[0108] 步骤B:利用所述字段抽取模型中的全连接网络计算字符向量集中的每个字符向量的标签分类真实值,利用所述字段抽取模型中的序列化标注算法网络计算字符向量集中每个字符向量的标签顺序真实值,根据所述字符向量集中的对应的字符的标签确定字符向量集中每个字符向量的标签分类预测值及标签顺序预测值;

[0109] 步骤C:根据所述标签分类真实值及标签分类预测值,利用预设的第一损失函数进行计算,得到第一损失值;根据所述标签顺序真实值及标签顺序预测值,利用预设的第二损失函数进行计算,得到第二损失值;将所述第一损失值及所述第二损失值进行求和计算得到目标损失值,当所述目标损失值大于或等于预设阈值时,更新所述字段抽取模型的模型参数,并返回所述步骤A;当所述目标损失值小于预设阈值时,停止训练,得到所述训练完成的字段抽取模型。

[0110] 可选地,本发明实施例中,所述第一损失函数可用如下公式进行计算:

$$[0111] \quad L_{ce} = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)]$$

[0112] 其中, $L_{ce}$ 表示所述第一损失值, $N$ 为所述字符向量集的数据数目, $i$ 为正整数, $y_i$ 为所述标签分类真实值, $p_i$ 为所述标签分类预测值。

[0113] 可选地,所述第二损失函数包括:交叉熵损失函数、绝对值损失函数、平方损失函数等。

[0114] 本发明实施例中,不同产品类型对应的报告模板是不同的,如:甲类型的产品文本信息对应的报告模板为A;乙类型的产品文本信息对应的报告模板为B。因此所述报告生成模块101根据所述产品类型在预设的报告模板数据库中进行筛选,筛选所述报告模板数据库中所述产品类型对应的报告模板,得到所述目标报告模板。

[0115] 详细地,本发明实施例中,所述报告生成模块101将所述目标字段与所述目标报告模板进行关联匹配,得到所述目标报告;详细地,本发明实施例中将目标字段替换所述目标报告模板中对应字段类型的空白字段,得到所述目标报告。

[0116] 本发明另一实施例中,所述目标报告模板还可以存储在区块链节点中,利用区块链高吞吐的特性,提高数据的存取效率。

[0117] 所述报告检验模块102,用于对所述目标报告进行报告差异校验,得到目标校验差异率;获取预设反馈时间区间内的历史报告差异率,根据所述历史报告差异率构建校验区间;

[0118] 本发明实施例中,为了进一步地筛选所述第一产品集中更适合用户的产品,所述产品筛选模块102根据所述用户向量及所述产品向量进行相关度计算,得到相关度分数。

[0119] 本发明实施例中为了保证发布的产品报告更加的准确,还需要对所述目标报告进行校验后,才能发布。

[0120] 可选地,本发明实施例中所述报告检验模块102计算目标报告中缺失字段的比例、目标字段的字号差异比例、字符数量差异比例;根据所述缺失字段的比例、目标字段的字号差异比例、字符数量差异比例;例如:所述目标报告中空白字段为两个、目标字段为4个,那么缺失字段比例为 $2/(4+2)=0.33$ ,所述目标报告中目标字段某字段的字号为10号字体,校验规则中该字段对应的标准字号为5号字体,那么该字段的字号差异比例为 $(10-5)/5=1$ ,所述目标字段的字号差异比例为目标字段中所有字段的字号差异比例的平均值;所述目标报告中目标字段某字段的字符数量为1个字符,校验规则中该字段对应的标准字符数量为2个字符,那么该字段的字符数量差异比例为 $(2-1)/2=0.5$ ,所述目标字段的字符数量差异比例为目标字段中所有字段的字符数量差异比例的平均值。

[0121] 可选地,本发明实施例所述报告检验模块102利用如下公式计算所述目标校验差异率:

$$l = a * \alpha + b * \beta + c * \gamma$$

[0123] 其中,所述a为缺失字段比例、b为字号差异比例、c为字符数量差异比例; $\alpha$ 、 $\beta$ 、 $\gamma$ 为人为设置的权重参数。

[0124] 详细地,本发明实施例中所述历史报告差异率为所述反馈时间区间内发送成功的报告的对应的目标校验差异率。

[0125] 进一步地,本发明实施例中所述报告检验模块102获取预设反馈时间区间内的历史报告差异率之前,还包括:获取所述报告生成请求的请求时间;将所述请求时间作为区间右端点,将预设时间周期作为区间长度;根据所述区间右端点及所述区间长度进行区间构建,得到所述反馈时间区间。例如:预设时间周期为3天,所述请求时间为5/13,那么所述反馈时间区间为 $[5/10, 5/13]$ 。可选地,本发明实施例中所述时间周期为365天。

[0126] 进一步地,本发明实施例中所述报告检验模块102根据所述历史报告差异率构建校验区间,包括:选取所有所述历史报告差异率中的中位数,得到第一差异率;选取所有所述历史报告差异率中的最大值,得到所述第二差异率;将所述第一差异率及所述第二差异率作为区间端点进行区间构建,得到所述校验区间。例如:第一差异率为0.3,第二差异率为0.5,那么所述校验区间为[0.3,0.5]。

[0127] 所述报告发布模块103,用于根据所述校验区间及所述目标校验差异率对所述目标报告进行发布评估,得到评估结果,根据所述评估结果对所述目标报告进行发布。

[0128] 详细地,本发明实施例中所述报告发布模块103根据所述校验区间及所述目标校验差异率对所述目标报告进行发布评估,包括:判断所述目标校验差异率是否在所述校验区间内,若所述目标校验差异率在所述校验区间内,则评估结果为报告合格,发布所述目标报告,并发送预警信息至预设第一终端设备,若所述目标校验差异率不在所述校验区间内,判断所述目标校验差异率小于所述第一差异率或大于所述第二差异率,若所述目标校验差异率小于所述与第一差异率,那么发布所述目标报告;若所述目标校验差异率大于所述第二差异率,则将所述目标报告发送至预设第二终端设备进行审批。

[0129] 如图3所示,是本发明实现产品报告发布方法的电子设备的结构示意图。

[0130] 所述电子设备可以包括处理器10、存储器11、通信总线12和通信接口13,还可以包括存储在所述存储器11中并可在所述处理器10上运行的计算机程序,如产品报告发布程序。

[0131] 其中,所述存储器11至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、移动硬盘、多媒体卡、卡型存储器(例如:SD或DX存储器等)、磁性存储器、磁盘、光盘等。所述存储器11在一些实施例中可以是电子设备的内部存储单元,例如该电子设备的移动硬盘。所述存储器11在另一些实施例中也可以是电子设备的外部存储设备,例如电子设备上配备的插接式移动硬盘、智能存储卡(Smart Media Card,SMC)、安全数字(Secure Digital,SD)卡、闪存卡(Flash Card)等。进一步地,所述存储器11还可以既包括电子设备的内部存储单元也包括外部存储设备。所述存储器11不仅可以用于存储安装于电子设备的应用软件及各类数据,例如产品报告发布程序的代码等,还可以用于暂时地存储已经输出或者将要输出的数据。

[0132] 所述处理器10在一些实施例中可以由集成电路组成,例如可以由单个封装的集成电路所组成,也可以是由多个相同功能或不同功能封装的集成电路所组成,包括一个或者多个中央处理器(Central Processing unit,CPU)、微处理器、数字处理芯片、图形处理器及各种控制芯片的组合等。所述处理器10是所述电子设备的控制核心(Control Unit),利用各种接口和线路连接整个电子设备的各个部件,通过运行或执行存储在所述存储器11内的程序或者模块(例如产品报告发布程序等),以及调用存储在所述存储器11内的数据,以执行电子设备的各种功能和处理数据。

[0133] 所述通信总线12可以是外设部件互连标准(peripheral component interconnect,简称PCI)总线或扩展工业标准结构(extended industry standard architecture,简称EISA)总线等。该总线可以分为地址总线、数据总线、控制总线等。所述通信总线12总线被设置为实现所述存储器11以及至少一个处理器10等之间的连接通信。为便于表示,图中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0134] 图3仅示出了具有部件的电子设备,本领域技术人员可以理解的是,图3示出的结构并不构成对所述电子设备的限定,可以包括比图示更少或者更多的部件,或者组合某些部件,或者不同的部件布置。

[0135] 例如,尽管未示出,所述电子设备还可以包括给各个部件供电的电源(比如电池),优选地,电源可以通过电源管理装置与所述至少一个处理器10逻辑相连,从而通过电源管理装置实现充电管理、放电管理、以及功耗管理等功能。电源还可以包括一个或一个以上的直流或交流电源、再充电装置、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。所述电子设备还可以包括多种传感器、蓝牙模块、Wi-Fi模块等,在此不再赘述。

[0136] 可选地,所述通信接口13可以包括有线接口和/或无线接口(如WI-FI接口、蓝牙接口等),通常用于在该电子设备与其他电子设备之间建立通信连接。

[0137] 可选地,所述通信接口13还可以包括用户接口,用户接口可以是显示器(Display)、输入单元(比如键盘(Keyboard)),可选地,用户接口还可以是标准的有线接口、无线接口。可选地,在一些实施例中,显示器可以是LED显示器、液晶显示器、触控式液晶显示器以及OLED(Organic Light-Emitting Diode,有机发光二极管)触摸器等。其中,显示器也可以适当的称为显示屏或显示单元,用于显示在电子设备中处理的信息以及用于显示可视化的用户界面。

[0138] 应该了解,所述实施例仅为说明之用,在专利申请范围上并不受此结构的限制。

[0139] 所述电子设备中的所述存储器11存储的产品报告发布程序是多个计算机程序的组合,在所述处理器10中运行时,可以实现:

[0140] 根据产品报告生成请求获取待分析产品文本信息,从所述待分析产品文本信息中抽取预设字段,得到目标字段;

[0141] 提取所述报告生成请求中的产品类型,根据所述产品类型在预设的报告模板数据库中进行筛选,得到目标报告模板;

[0142] 将所述目标字段与所述目标报告模板进行关联匹配,得到目标报告;

[0143] 对所述目标报告进行报告差异校验,得到目标校验差异率;

[0144] 获取预设反馈时间区间内的历史报告差异率,根据所述历史报告差异率构建校验区间;

[0145] 根据所述校验区间及所述目标校验差异率对所述目标报告进行发布评估,得到评估结果,根据所述评估结果对所述目标报告进行发布。

[0146] 具体地,所述处理器10对上述计算机程序的具体实现方法可参考图1对应实施例中相关步骤的描述,在此不赘述。

[0147] 进一步地,所述电子设备集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。所述计算机可读取介质可以是非易失性的,也可以是易失性的。所述计算机可读取介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)。

[0148] 本发明实施例还可以提供一种计算机可读取存储介质,所述可读存储介质存储有计算机程序,所述计算机程序在被电子设备的处理器所执行时,可以实现:

[0149] 根据产品报告生成请求获取待分析产品文本信息,从所述待分析产品文本信息中抽取预设字段,得到目标字段;

[0150] 提取所述报告生成请求中的产品类型,根据所述产品类型在预设的报告模板数据库中进行筛选,得到目标报告模板;

[0151] 将所述目标字段与所述目标报告模板进行关联匹配,得到目标报告;

[0152] 对所述目标报告进行报告差异校验,得到目标校验差异率;

[0153] 获取预设反馈时间区间内的历史报告差异率,根据所述历史报告差异率构建校验区间;

[0154] 根据所述校验区间及所述目标校验差异率对所述目标报告进行发布评估,得到评估结果,根据所述评估结果对所述目标报告进行发布。

[0155] 进一步地,所述计算机可用存储介质可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序等;存储数据区可存储根据区块链节点的使用所创建的数据等。

[0156] 在本发明所提供的几个实施例中,应该理解到,所揭露的设备,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

[0157] 所述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单元,即可以位于一个地方,或者也可以分布到多个网络单元上。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。

[0158] 另外,在本发明各个实施例中的各功能模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能模块的形式实现。

[0159] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。

[0160] 因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附关联图标记视为限制所涉及的权利要求。

[0161] 本发明所指区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain),本质上是一个去中心化的数据库,是一串使用密码学方法相关联产生的数据块,每一个数据块中包含了一批网络交易的信息,用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层等。

[0162] 此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。系统权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第二等词语用来表示名称,而并不表示任何特定的顺序。

[0163] 最后应说明的是,以上实施例仅用以说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的

技术方案进行修改或等同替换,而不脱离本发明技术方案的精神和范围。

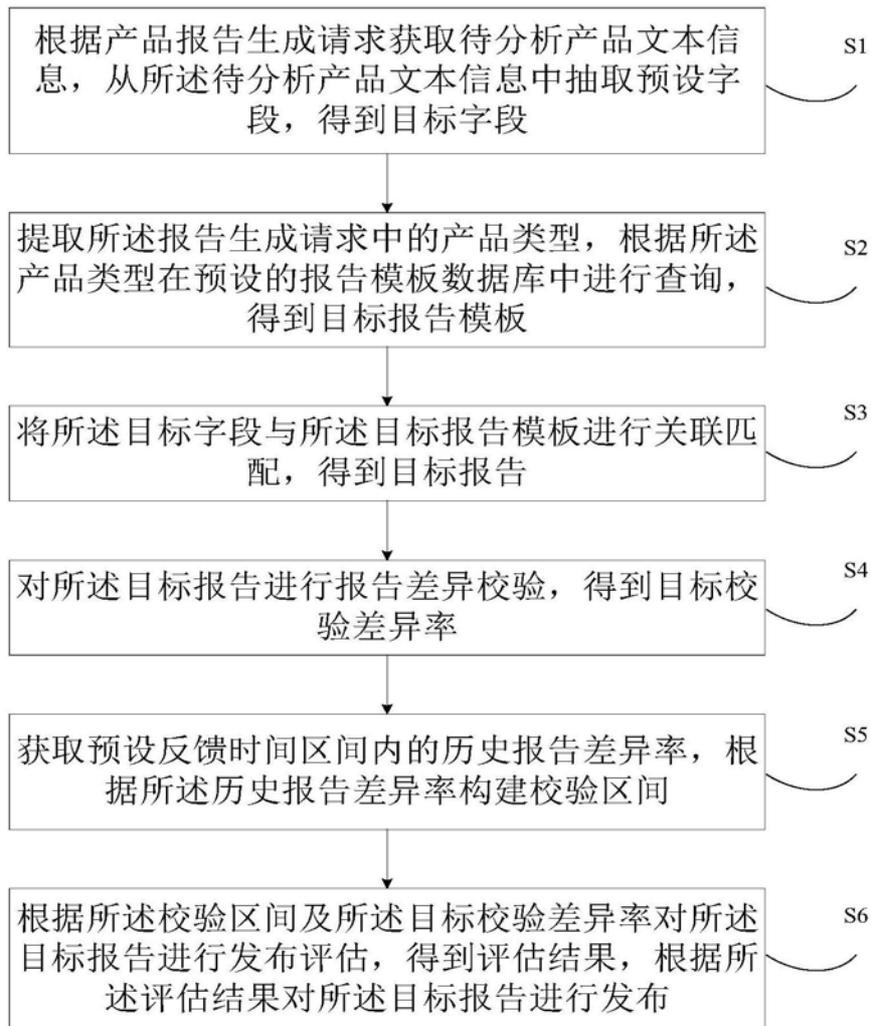


图1

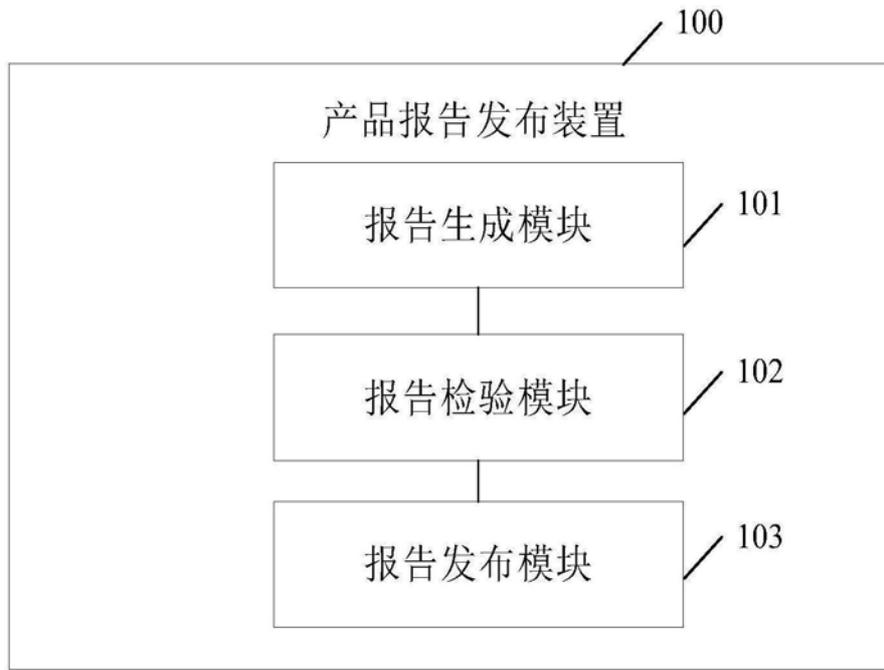


图2

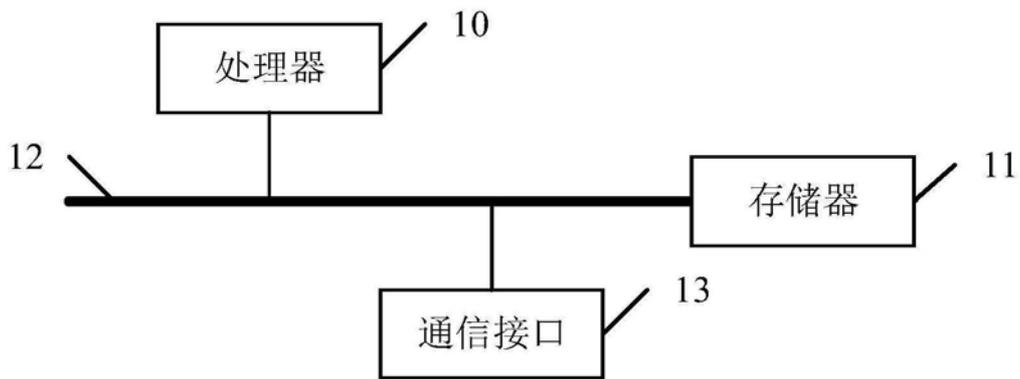


图3