

一种特征关系推荐方法及装置、一种计算设备及存储介质

技术领域

[01]本申请涉及计算机安全技术领域，特别涉及一种特征关系推荐方法及装置、一种计算设备及存储介质。

5 背景技术

[02]随着支付宝业务的发展，支付宝的安全环境时刻在变，传统的依赖于人工制作特征关系越来越显得捉襟见肘。

[03]目前，现有特征关系推荐通常采用以下方式：

[04]暴力搜索以及人工挑选判断，但是这种方案依赖于人力，且暴力搜索太耗资源；

- 10 [05]借助启发式搜索算法搜索，然后通过一定的评价模型挑选，但是该方案没有引入业务知识，搜索空间仍然很大，难以搜索出有效的特征关系。

发明内容

[06]有鉴于此，本申请实施例提供了一种特征关系推荐方法及装置、一种计算设备及存储介质，以解决现有技术中存在的技术缺陷。

- 15 [07]本申请实施例公开了一种特征关系推荐方法，所述方法包括：

[08]基于业务数据确定参数实体，其中，所述参数实体包括事件实体以及节点实体，并且所述节点实体通过所述事件实体进行关联；

[09]基于所述参数实体和预设算子实体之间的特征关系构建知识图谱；

[10]基于推理规则从所述知识图谱抽取特征关系，获得新增特征关系并推荐。

- 20 [11]另一方面，本申请实施例公开了一种特征关系推荐装置，所述装置包括：

[12]确定模块，被配置为基于业务数据确定参数实体，其中，所述参数实体包括事件实体以及节点实体，并且所述节点实体通过所述事件实体进行关联；

[13]构建模块，被配置为基于所述参数实体和预设算子实体之间的特征关系构建知识图谱；

[14]抽取模块，被配置为基于推理规则从所述知识图谱抽取特征关系，获得新增特征关系并推荐。

[15]另一方面，本申请还提供了一种计算设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机指令，所述处理器执行时实现所述特征关系推荐方法的步骤。

[16]另一方面，本申请还提供了一种计算机可读存储介质，其存储有计算机指令，该程序被处理器执行时实现所述特征关系推荐方法的步骤。

[17]本申请提供的一种特征关系推荐方法及装置、一种计算设备及存储介质，所述方法包括基于业务数据确定参数实体，其中，所述参数实体包括事件实体以及节点实体，并且所述节点实体通过所述事件实体进行关联；基于所述参数实体和预设算子实体之间的特征关系构建知识图谱；基于推理规则从所述知识图谱抽取特征关系，获得新增特征关系并推荐。

附图说明

[18]图 1 为本说明书一实施例提供的一种计算设备的结构框图；

[19]图 2 为本说明书一实施例提供的一种特征关系推荐方法的流程图；

[20]图 3 为本说明书一实施例提供的一种参数实体示意图；

[21]图 4 为本说明书一实施例提供的一种特征关系推荐方法的流程图；

[22]图 5 为本说明书一实施例提供的一种特征关系推荐方法的流程图；

[23]图 6 为本说明书一实施例提供的一种特征关系推荐方法的流程图；

[24]图 7 为本说明书一实施例提供的一种特征关系推荐装置的结构示意图。

具体实施方式

[25]在下面的描述中阐述了很多具体细节以便于充分理解本申请。但是本申请能够以很多不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本申请内涵的情况下做类似推广，因此本申请不受下面公开的具体实施的限制。

[26]在本说明书一个或多个实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本说明书一个或多个实施例。在本说明书一个或多个实施例和所附权利要求

书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文中清楚地表示其他含义。还应当理解，本说明书一个或多个实施例中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

[27]应当理解，尽管在本说明书一个或多个实施例中可能采用术语第一、第二、第三等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本说明书一个或多个实施例范围的情况下，第一也可以被称为第二，类似地，第二也可以被称为第一。取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[28]首先对本说明书一个或多个实施例中使用到的技术术语进行解释。

10 [29]知识图谱：（Knowledge Graph/Vault）又称为科学知识图谱，知识工程中以图的形式组织的知识集群，其由不同类型的实体作为节点、关系作为连接节点的边所构成的。

[30]知识推理：在知识图谱中，根据现有知识，通过一定的逻辑推断，补充和发现新的知识。

15 [31]在本说明书一个或多个实施例中，提供了一种特征关系推荐方法及装置、计算设备和计算机存储介质，在下面的实施例中逐一进行详细说明。

[32]参见图 1，本说明书一个实施例提供了一种计算设备 100 的结构框图。该计算设备 100 的部件包括但不限于存储器 110、处理器 120 和通信接口 140。所述处理器 120 与所述存储器 110 通过总线 130 相连接，数据库 150 用于存储访问请求的日志数据。

20 [33]所述通信接口 140 使得计算设备 100 能够经由一个或多个网络通信。这些网络的示例包括局域网（LAN）、广域网（WAN）、个域网（PAN）或诸如因特网的通信网络的组合。网络接口可以包括有线或无线的任何类型的网络接口（例如，网络接口卡（NIC））中的一个或多个，诸如 IEEE802.11 无线局域网（WLAN）无线接口、全球微波互联接入（Wi-MAX）接口、以太网接口、通用串行总线（USB）接口、蜂窝网络接口、蓝牙接口、近场通信（NFC）接口，等等。

25 [34]所述存储器 110，被配置为存储通信接口 140 通过总线 130 发送的业务数据以及存储在存储器 110 上并可在处理器 120 上运行的计算机指令。

[35]所述处理器 120，被配置为获取存储在存储器 110 的业务数据后，执行存储在存储器 110 上的计算机指令，实现对所述待识别对象的验证。

[36]所述计算设备 100 可以是任何类型的静止或移动计算设备，包括移动计算机或移动计算设备（例如，平板电脑、个人数字助理、膝上型计算机、笔记本电脑、上网本等）、移动电话（例如，智能手机）、可佩戴的计算设备（例如，智能手表、智能眼镜等）或其他类型的移动设备，或者诸如台式计算机或 PC 的静止计算设备。

- 5 [37]其中，处理器 120 可以执行图 2 所示方法中的步骤。图 2 是示出了本说明书一个实施例提供的特征关系推荐方法的示意性流程图，包括步骤 202 至步骤 206。

[38]步骤 202：基于业务数据确定参数实体，其中，所述参数实体包括事件实体以及节点实体，并且所述节点实体通过所述事件实体进行关联。

- 10 [39]本说明书一个或多个实施例中，所述业务数据包括但不限于交易数据、登陆数据或转账数据等。

[40]所述参数实体根据所述业务数据中的参数本体确定，所述参数本体为概念，对应的所述参数实体则是具体概念的实例。

- 15 [41]例如，参见图 3，所述参数本体包括各个概念本体：事件、账户、设备或者位置，对应的所述参数实体在具体的应用场景中则为具体概念的实例：中间实体为某个支付宝事件 eventid: 4a8167a*****、三号和四号实体为某个支付宝账号 id: 208821*****和 208890*****、二号实体为某个手机设备位置服务信息的 geohash 值: wm6n；一号实体为某个手机设备的 IP: 117.136.**.*。

[42]实际应用中所述参数实体包括但不限于事件、账户、身份证、设备、手机号、位置服务（LBS，Location Based Services）、IP 等具体的实体。

- 20 [43]其中，所述节点实体包括但不限于设备实体、账户实体和/或交易实体。

[44]本说明书一个或多个实施例中，所述参数实体包括节点实体和事件实体，各个所述节点实体之间通过所述事件实体关联，例如所述节点实体包括账户 1、账户 2，所述事件实体为转账 3 块钱，则可以所述节点实体就可以通过所述事件实体关联，表示为账户 1 向账户 2 转账 3 块钱，实际使用中可以表示支付宝事件交易或者登陆等关系。

- 25 [45]步骤 204：基于所述参数实体和预设算子实体之间的特征关系构建知识图谱。

[46]本说明书一个或多个实施例中，所述预设算子实体可以通过现有特征获取，例如，若所述现有特征包括“某时间段内支付宝账户 id 的最大移动速度”，则对应的算子实体包括梯度算子；若现有特征包括“某个支付宝手机使用设备的历史位置离散度”，则对

应的算子实体包括熵算子；若现有特征包括“某个支付宝账户 3 天内累计红包金额”，则对应的算子实体包括求和算子。

[47] 实际应用中，首先根据现有特征，对特征中算子进行分类，将不同类的算子，命名为不同的算子本体，例如 velocity 算子、sequence 算子、topology 算子等，然后对不同算子本体具体实例化后，即为各自对应的算子实体，例如累计求和、求熵、求梯度等。

[48] 基于所述参数实体和预设算子实体之间的特征关系构建知识图谱，所述算子实体可以根据所述知识图谱完成基本的查询、标注、计算、推理和预测等。

[49] 步骤 206: 基于推理规则从所述知识图谱抽取特征关系，获得新增特征关系并推荐。

[50] 本说明书一个或多个实施例中，所述推理规则包括路径排列算法（Path Ranking），采用所述 Path Ranking 算法基于所述知识图谱进行新的特征关系的抽取，获得新增特征关系并推荐。

[51] 参见图 4，基于推理规则从所述知识图谱抽取特征关系包括步骤 402 至步骤 406。

[52] 步骤 402: 基于推理规则从原有的所述节点实体与事件实体中搜索出原有的所述节点实体间的新增事件实体。

[53] 参见图 5，本说明书一个或多个实施例中，以所述节点实体包括账户 A、账户 B 和账户 C，所述事件实体包括转账 3 块钱为例，三个所述节点实体通过所述事件实体进行关联分别为：账户 A 向账户 B 转账 3 块钱，账户 B 向账户 C 转账 3 块钱，基于所述 Path Ranking 算法可以得出原有的所述节点实体间的新增事件实体：账户 A 向账户 C 转账 3 块钱。

[54] 步骤 404: 为所述新增事件实体、原有的所述节点实体与事件实体匹配符合第一预设条件的算子实体，确定所述新增事件实体、原有的所述节点实体与事件实体以及所述算子实体之间的组合特征关系。

[55] 本申请说明书一个或多个实施例中，所述第一预设条件包括但不限于实际需求特征，例如，以图 5 为例，若所述第一预设条件包括“账户 A 两天内累计转账次数”，则匹配的算子实体则为求和算子。

[56] 以新增事件实体包括账户 A 向账户 C 转账 3 块钱，原有的所述节点实体与事件实体之间的关联包括：账户 A 向账户 B 转账 3 块钱，账户 B 向账户 C 转账 3 块钱，账户 A 向账户 D 转账 1 块钱，所述算子实体包括求和算子为例，确定所述新增事件实体、原有

的所述节点实体与事件实体与所述算子实体之间的组合特征关系，所述组合特征关系包括“账户 A 累计转账次数”以及“账户 A 累计转账金额”等。

[57] 步骤 406：所述组合特征关系表征从所述知识图谱抽取的第一特征关系。

[58] 本申请说明书一个或多个实施例中，所述第一特征关系即为所述组合特征关系。

- 5 [59] 实际使用中，首先通过推理规则推理补充节点实体与事件实体关系，以推理出哪些节点实体之间可以组成抽象的 key-collection 关系；然后根据 key_collection 关系、原有的所述节点实体与事件实体的关系，匹配符合条件的算子实体，确定组合特征关系。

[60] 例如节点实体包括身份证、卡号、手机号、支付宝账号、ip、手机设备号等，然后以所述节点实体为 key 主键，对其他事件实体和节点实体包括交易金额、登陆 lbs 位置、

- 10 交易时间、登陆城市、身份证、卡号、手机号等进行 group 统计，统计身份证为 key 主键分别与交易金额、登陆 lbs 位置、交易时间、登陆城市、其他身份证、卡号、手机号等组成的 key-collection 关系，以及卡号为 key 主键分别与交易金额、登陆 lbs 位置、交易时间、登陆城市、身份证、其他卡号、手机号等组成的 key-collection 关系等。

[61] 有了 key-collection 关系，再配合算子实体，就可以组成特征关系了。以支付宝账号

- 15 为 key，对该账号下出现过所有身份证进行 group，结合 count 算子，可以组成“某个支付宝账号下经过过去重的身份证个数”组合特征关系；以身份证号为 key，对该身份证发生的交易金额 group，结合 sum 算子，组成“某个支付宝账号下累计消费金额”组合特征关系；以手机设备账号为 key，对该手机设备账号下的 lbs group，结合 maxDistance 算子，组成“该手机设备最大移动距离”组合特征关系。

- 20 [62] 本申请说明书一个或多个实施例中，所述装置还包括：第二预设条件，

[63] 基于所述第二预设条件对所述组合特征关系进行过滤；

[64] 基于过滤后的所述组合特征关系表征所述知识图谱抽取第二特征关系。

[65] 本申请说明书一个或多个实施例中，所述第二预设条件包括但不限于在所述组合特征关系中新增的知识图谱中的节点实体。

- 25 [66] 例如所述组合特征关系包括“该手机设备最大移动距离”，所述第二预设条件包括新增的知识图谱中的节点实体“时间窗口 3 天”，即可得到“该手机设备 3 天内最大移动距离”的组合特征关系，该组合特征关系即为第二特征关系。

[67] 实际使用中，所述 Path Ranking 算法将每种不同的关系路径作为一维特征，通过在

知识图谱中统计大量的关系路径构建关系分类的特征向量，建立关系分类器进行新的特征关系抽取。

- [68]本说明书一个或多个实施例中，所述推理规都是在一阶逻辑推理的框架下，采用霍恩子句的形式实现的，可以根据人工知识写成，也可以采用机器学习的方式，基于数据学习出规则，例如采用人工神经网络的学习训练实现规则，采用所述推理规则输入的是各参数实体、算子实体以及各参数实体和算子实体之间的关系，输出的是抽取出的新的特征关系，实现新增特征关系的推荐。

- [69]本说明书一个或多个实施例中，所述方法将业务数据通过参数实体、算子实体以及参数实体和算子实体之间的特征关系在知识图谱中实现，使用推理规则基于原有的特征关系对新的特征关系搜索进行指导约束，避免了大量无意义特征关系的搜索和推荐。

[70]参见图 6，本说明书一个或多个实施例提供了一种特征关系推荐方法的示意性流程图，包括步骤 602 至步骤 612。

[71]步骤 602：基于业务数据确定参数实体，其中，所述参数实体包括事件实体以及节点实体，并且所述节点实体通过所述事件实体进行关联。

- 15 [72]步骤 604：基于所述参数实体和预设算子实体之间的特征关系构建知识图谱。

[73]步骤 606：基于推理规则从所述知识图谱抽取特征关系，获得新增特征关系并推荐。

[74]步骤 608：基于预先训练的分类模型确定所述新增特征关系对应的标签，其中，所述标签包括黑标签和白标签。

[75]本说明书一个或多个实施例中，对所述分类模型进行预先训练包括如下步骤：

- 20 [76]获取训练样本数据集，其中，所述训练样本数据集中包括参数实体样本和算子实体样本之间的特征关系样本以及所述特征关系样本对应的样本标签，所述样本标签包括白样本标签和黑样本标签；

[77]基于树类模型训练所述分类模型，所述分类模型使得所述特征关系样本与所述样本标签相关联。

- 25 [78]本说明书一个或多个实施例中，通过所述分类模型得到每个与白样本标签对应的特征关系样本的重要程度。

[79]本说明书一个或多个实施例中，所述树类模型包括 XGBoost 模型。

[80]本说明书一个或多个实施例中，若将所述分类模型应用在金融行业反欺诈中，则所

述白标签可以为无欺诈风险的标签，所述黑标签可以为有欺诈风险的标签。

[81]基于预先训练的分类模型确定所述新增特征关系对应的标签包括：根据预先训练的分类模型可以确定出哪些新增特征关系为无欺诈风险的白标签，哪些新增特征关系为有欺诈风险的黑标签。

5 [82]步骤 610：基于预先训练的分类模型确定与白标签对应的新增特征关系的重要程度。

[83]本说明书一个或多个实施例中，确定出所述新增特征关系对应的标签之后，确定出与白标签对应的新增特征关系的重要程度。而对于与黑标签对应的新增特征关系的重要程度则不用获取，降低工作量提高分类模型的工作效率。

[84]步骤 612：若所述重要程度大于预设阈值，则将所述与白标签对应的新增特征关系
10 添加至知识图谱。

[85]本说明书一个或多个实施例中，所述预设阈值根据实际应用进行设定，可以设置为80%、95%或100%等，本申请对此不作任何限定。

[86]若所述预设阈值为80%，某个为白标签的新增特征关系的重要程度为85%，所述新增特征关系的重要程度大于预设阈值，则将所述新增特征关系添加至知识图谱。

15 [87]本说明书一个或多个实施例中，例如若新增特征关系为：“支付宝账户的6小时内最大移动速度”，将该特征输入到XGBoost模型中，评判该特征的重要程度为89%，预设阈值为85%，则在知识图谱中构建“支付宝账户实体、lbs实体、梯度算子实体”三者之间特征关系。

[88]本说明书一个或多个实施例中，所述方法根据分类模型实现对新增特征关系的分类
20 以及对白标签对应的新增特征关系进行重要程度评价，将重要程度达到设定要求的新增特征关系添加进构建的知识图谱中，丰富所述知识图谱，使得知识图谱的扩展性能越来越好。

[89]参见图7，本说明书一个或多个实施例提供了一种特征关系推荐装置，包括：

[90]确定模块702，被配置为基于业务数据确定参数实体，其中，所述参数实体包括事
25 件实体以及节点实体，并且所述节点实体通过所述事件实体进行关联；

[91]构建模块704，被配置为基于所述参数实体和预设算子实体之间的特征关系构建知识图谱；

[92]抽取模块706，被配置为基于推理规则从所述知识图谱抽取特征关系，获得新增特

征关系并推荐。

[93]可选地，所述装置还包括：

[94]标签选择模块，被配置为基于预先训练的分类模型确定所述新增特征关系对应的标签，其中，所述标签包括黑标签和白标签。

5 [95]可选地，所述装置还包括：

[96]重要程度确定模块，被配置为基于预先训练的分类模型确定与白标签对应的新增特征关系的重要程度；

[97]添加模块，被配置为若所述重要程度大于预设阈值，则将所述与白标签对应的新增特征关系添加至知识图谱。

10 [98]可选地，所述分类模型通过如下步骤训练：

[99]获取训练样本数据集，其中，所述训练样本数据集中包括参数实体样本和算子实体样本之间的特征关系样本以及所述特征关系样本对应的样本标签，所述样本标签包括白样本标签和黑样本标签；

[100] 基于树类模型训练所述分类模型，所述分类模型使得所述特征关系样本与所述

15 样本标签相关联。

[101] 可选地，所述装置还包括：

[102] 获取模块，被配置为通过所述分类模型得到每个与白样本标签对应的特征关系样本的重要程度。

[103] 可选地，所述抽取模块包括：

20 [104] 新增子模块，被配置为基于推理规则从原有的所述节点实体与事件实体中搜索出原有的所述节点实体间的新增事件实体；

[105] 组合特征关系确定子模块，被配置为为所述新增事件实体、原有的所述节点实体与事件实体匹配符合第一预设条件的算子实体，确定所述新增事件实体、原有的所述节点实体与事件实体与所述算子实体之间的组合特征关系；

25 [106] 特征关系抽取子模块，被配置为基于所述组合特征关系表征从所述知识图谱抽取的第一特征关系。

[107] 可选地，所述装置还包括：第二预设条件，

- [108] 基于所述第二预设条件对所述组合特征关系进行过滤;
- [109] 基于过滤后的所述组合特征关系表征从所述知识图谱抽取的第二特征关系。
- [110] 可选地, 所述树类模型包括 XGBoost 模型。
- [111] 可选地, 所述推理规则包括路径排列算法。
- 5 [112] 可选地, 所述算子实体包括求和算子、熵算子或梯度算子。
- [113] 本说明书一个或多个实施例中, 所述装置将业务数据通过参数实体、算子实体以及参数实体和算子实体之间的特征关系在知识图谱中实现, 使用推理规则基于原有的特征关系对新的特征关系搜索进行指导约束, 避免了大量无意义特征关系的搜索和推荐。
- [114] 本说明书一个或多个实施例中的各个实施例均采用递进的方式描述, 各个实施
- 10 例之间相同相似的部分互相参见即可, 每个实施例重点说明的都是与其他实施例的不同之处。尤其, 对于装置实施例而言, 由于其基本相似于方法实施例, 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。
- [115] 本申请一实施例还提供一种计算机可读存储介质, 其存储有计算机指令, 该指令被处理器执行时实现所述特征关系推荐方法的步骤。
- 15 [116] 上述为本实施例的一种计算机可读存储介质的示意性方案。需要说明的是, 该存储介质的技术方案与上述的特征关系推荐方法的技术方案属于同一构思, 存储介质的技术方案未详细描述的细节内容, 均可以参见上述特征关系推荐方法的技术方案的描述。
- [117] 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的 范围内。在一些情况下, 在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序
- 20 来执行并且仍然可以实现期望的结果。另外, 在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中, 多任务处理和并行处理也是可以的或者可能是有利的。
- [118] 本申请实施例中所述支付涉及的技术载体, 例如可以包括近场通信 (Near Field Communication, NFC)、WIFI、3G/4G/5G、POS 机刷卡技术、二维码扫码 技术、条
- 25 形码扫码技术、蓝牙、红外、短消息 (Short Message Service, SMS)、多媒体消息 (Multimedia Message Service, MMS) 等。
- [119] 所述计算机指令包括计算机指令代码, 所述计算机指令代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括: 能够携

带所述计算机指令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)、随机存取存储器 (RAM, Random Access Memory)、电载波信号、电信信号以及软件分发介质等。需要说明的是, 所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减, 例如在某些司法管辖区, 根据立法和专利实践, 计算机可读介质不包括电载波信号和电信信号。

[120] 需要说明的是, 对于前述的各方法实施例, 为了简便描述, 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, 本申请并不受所描述的动作顺序的限制, 因为依据本申请, 某些步骤可以采用其它顺序或者同时进行。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, 所涉及的动作和模块并不一定都是本申请所必须的。

[121] 在上述实施例中, 对各个实施例的描述都各有侧重, 某个实施例中未详述的部分, 可以参见其它实施例的相关描述。

[122] 以上公开的本申请优选实施例只是用于帮助阐述本申请。可选实施例并没有详尽叙述所有的细节, 也不限制该发明仅为所述的具体实施方式。显然, 根据本说明书的内容, 可作很多的修改和变化。本说明书选取并具体描述这些实施例, 是为了更好地解释本申请的原理和实际应用, 从而使所属技术领域技术人员能很好地理解和利用本申请。本申请仅受权利要求书及其全部范围和等效物的限制。