(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 112818257 A
(43)申请公布日 2021.05.18

(21)申请号 202110195120.1

(22)申请日 2021.02.19

(71)申请人 北京邮电大学
地址 100876 北京市海淀区西土城路10号

(72)发明人 徐国爱 郭燕慧 徐国胜 张淼 李思怡

(74)专利代理机构 北京风雅颂专利代理有限公司 11403
代理人 刘志永

(51)Int.Cl.
G06F 16/9536(2019.01)
G06F 16/951(2019.01)
H04L 12/58(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

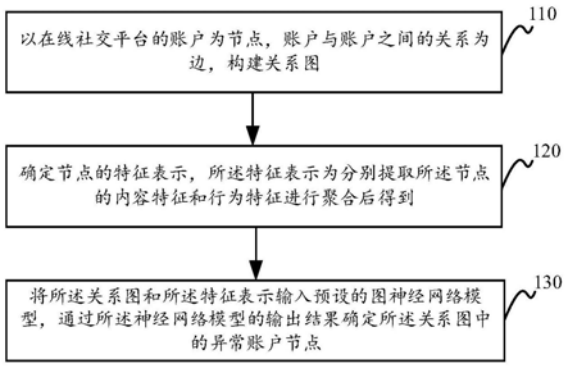
权利要求书2页 说明书13页 附图3页

(54)发明名称

基于图神经网络的账户检测方法、装置和设备

(57)摘要

本发明示例性实施例提供一种基于图神经网络的账户检测方法，包括：以在线社交平台的账户为节点，账户与账户之间的关系为边，构建关系图；确定节点的特征表示，所述特征表示为分别提取所述节点的内容特征和行为特征进行聚合后得到；将所述关系图和所述特征表示输入预设的图神经网络模型，通过所述神经网络模型的输出结果确定所述关系图中的异常账户节点；所述预设的神经网络模型为：将所述关系图的节点和边转码为能够识别的向量数据，对所述向量数据进行权重参数的训练后，输出带有标签的节点预测值，根据所述节点预测值确定节点是否为异常账号。本发明使得账户检测的稳定性、效率和准确率均得到了大幅提升。



CN 112818257 A

1.一种基于图神经网络的账户检测方法,其特征在于,包括:

以在线社交平台的账户为节点,账户与账户之间的关系为边,构建关系图;

确定节点的特征表示,所述特征表示为分别提取所述节点的内容特征和行为特征进行聚合后得到;

将所述关系图和所述特征表示输入预设的图神经网络模型,通过所述神经网络模型的输出结果确定所述关系图中的异常账户节点;

其中,所述预设的神经网络模型为:将所述关系图的节点和边转码为能够识别的向量数据,对所述向量数据进行权重参数的训练后,输出带有标签的节点预测值,根据所述节点预测值确定节点是否为异常账号。

2.根据权利要求1所述的基于图神经网络的账户检测方法,其特征在于,所述方法还包括:

预建立账户判断范围表,根据所述预设的图神经网络模型输出的节点预测值在所述账户判断范围表的匹配结果,确定节点是否为异常账号;

所述账户判断范围表包括第一数值范围和第二数值范围,当所述节点预测值处于第一范围时,确定所述节点预测值对应的节点为正常账户;当所述节点预测值处于第二范围时,确定所述节点预测值对应的节点为异常账户。

3.根据权利要求1所述的基于图神经网络的账户检测方法,其特征在于,所述方法还包括:

确定源账号节点后,根据所述源账号节点的第一关系确定一阶关系图;

根据所述一阶关系图的各一阶账号节点的第二关系确定二阶关系图;

根据所述二阶关系图的各二阶账号节点的第三关系确定三阶关系图;

重复上述过程,直至获取所述源账号节点的第N关系确定的N阶关系图,其中,所述关系为节点关注与被关注的其它节点之间存在的关系。

4.根据权利要求1所述的基于图神经网络的账户检测方法,其特征在于,所述确定节点的特征表示,所述特征表示为分别提取所述节点的内容特征和行为特征进行聚合后得到,包括:

分别获取节点的账户信息和行为信息,所述账户信息包括简介信息和发布的文本信息,所述行为信息包括账户的活动信息;

分别对所述账户信息和行为信息进行聚合后编码,确定所述节点的向量数据。

5.根据权利要求1所述的基于图神经网络的账户检测方法,其特征在于,所述方法还包括:对于每一种关系的邻居节点引入不同的权重参数,分别对属于同一种关系类型的邻居节点进行聚合后再进行不同类型关系之间的聚合;

包括:

通过以下公式进行节点之间的边关系的双重聚合:

$$h_i^{l+1} = \sigma\left(\sum_{r\in R}\sum_{v_j\in N_{v_i}^r}\frac{1}{c_{i,r}}W_r^l h_j^l + W_0^l h_i^l\right)$$

其中,R表示所述关系图中所有边的集合,$N_{v_i}^r$表示与v_i节点具有r关系的邻居集合,$c_{i,r}$为归一化参数,一般取$c_{i,r} = |N_{v_i}^r|$,W_r是具有r关系的邻居对应的权重参数,W_r是节点自身对应

的权重参数；

softmax函数与对应的Loss计算函数为：

$$S_i = \frac{e^i}{\sum_i e^i}$$

$$Loss = -\sum_{i \in y} t_i \ln h_i^L$$

其中y为有标签的节点的集合，h_i^L表示输出层有标签的第i个节点的预测值。

6.根据权利要求5所述的基于图神经网络的账户检测方法，其特征在于，所述方法还包括：

所述预设的图神经网络模型进行迭代训练至Loss计算函数变化幅度小于预定值时，确定此时所述关系图的各所述节点输出的概率值为所述预测值。

7.根据权利要求1所述的基于图神经网络的账户检测方法，其特征在于，所述预设的图神经网络模型通过R-GCN算法实现聚合邻居的策略，以对复杂无向的关系图时加快训练和计算速度。

8.一种基于图神经网络的账户检测装置，其特征在于，包括：

图网络构建模块，用于以在线社交平台的账户为节点，账户与账户之间的关系为边，构建关系图；

特征确定模块，用于确定节点的特征表示，所述特征表示为分别提取所述节点的内容特征和行为特征进行聚合后得到；

判断模块，用于将所述关系图和所述特征表示输入预设的图神经网络模型，通过所述神经网络模型的输出结果确定所述关系图中的异常账户节点；

其中，所述预设的神经网络模型为：将所述关系图的节点和边转码为能够识别的向量数据，对所述向量数据进行权重参数的训练后，输出带有标签的节点预测值，根据所述节点预测值确定节点是否为异常账号。

9.根据权利要求8所述的基于图神经网络的账户检测装置，其特征在于，所述装置还包括：

确定模块，用于预建立账户判断范围表，根据所述预设的图神经网络模型输出的节点预测值在所述账户判断范围表的匹配结果，确定节点是否为异常账号；

所述账户判断范围表包括第一数值范围和第二数值范围，当所述节点预测值处于第一范围时，确定所述节点预测值对应的节点为正常账户；当所述节点预测值处于第二范围时，确定所述节点预测值对应的节点为异常账户。

10.一种电子设备，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程序，其特征在于，所述处理器执行所述程序时实现如权利要求1至7任意一项所述的基于图神经网络的账户检测方法。

基于图神经网络的账户检测方法、装置和设备

技术领域

[0001] 本发明示例性实施例涉及信息安全技术领域，尤其涉及一种基于图神经网络的账户检测方法、装置和设备。

背景技术

[0002] 随着互联网的发展与移动终端设备的普及，在线社交平台愈发受网络用户青睐。OSN(OSN,Online Social Network,在线社交平台)是依托于在线社交平台上的一种信息网络，OSN由用户与信息构成，基于OSN的网络关系图由用户构成的节点与用户之间关系构成的边组成。保证以在线社交网络为载体所传递信息的安全是确保网络空间安全的重要组成部分。OSN的便利性为许多不法份子提供了犯罪机会，一种典型的攻击手段就是Sybil Attack，不法份子创建多个机器人账号(即异常账号)来传播不良信息、操纵舆论，是一种典型的危害网络信息安全的行为。

[0003] 目前在线社交网络平台主要基于举报与半自动化的方式检测网络中存在的异常账号，缺乏一个公开的自动化检测工具来检测OSN中的异常账号。

[0004] 有公司如Twitter所属公司采用自动化与人工相结合的方式检测异常账号，具体而言，公司采用一定技术监控大量账号的行为并发掘其中用到的行为模式，一旦发现账号的行为模式无法与正常行为模式相匹配时，需要人工对这些可疑账号展开调查以确认这些账号是否为异常账号。

[0005] 另外有公司如新浪微博平台则使用基于先验知识的方法，设置某些特征规则比如关注量与粉丝比例、是否通过实名认证、是否有详细个人信息等等来判断某账号是否为可疑账号，同时，被用户多次举报的账号也会被列入可疑账号名单中。最终通过人工审核的方式确认风险账号是否为异常账号。

[0006] 现有方法无法匹配日益复杂的异常账户检测，且目前尚无能够解决该问题的方法或装置。

发明内容

[0007] 有鉴于此，本发明示例性实施例的目的在于提出一种基于图神经网络的账户检测方法、装置和设备，以解决传统方式中账户检测效率和准确率低下的问题。

[0008] 基于上述目的，本发明示例性实施例提供了一种基于图神经网络的账户检测方法，包括：

[0009] 以在线社交平台的账户为节点，账户与账户之间的关系为边，构建关系图；

[0010] 确定节点的特征表示，所述特征表示为分别提取所述节点的内容特征和行为特征进行聚合后得到；

[0011] 将所述关系图和所述特征表示输入预设的图神经网络模型，通过所述神经网络模型的输出结果确定所述关系图中的异常账户节点；

[0012] 其中，所述预设的神经网络模型为：将所述关系图的节点和边转码为能够识别的

向量数据，对所述向量数据进行权重参数的训练后，输出带有标签的节点预测值，根据所述节点预测值确定节点是否为异常账号。

[0013] 结合上述说明，在本发明实施例另一种可能的实施方式中，所述方法还包括：

[0014] 预建立账户判断范围表，根据所述预设的图神经网络模型输出的节点预测值在所述账户判断范围表的匹配结果，确定节点是否为异常账号；

[0015] 所述账户判断范围表包括第一数值范围和第二数值范围，当所述节点预测值处于第一范围时，确定所述节点预测值对应的节点为正常账户；当所述节点预测值处于第二范围时，确定所述节点预测值对应的节点为异常账户。

[0016] 结合上述说明，在本发明实施例另一种可能的实施方式中，所述方法还包括：

[0017] 确定源账号节点后，根据所述源账号节点的第一关系确定一阶关系图；

[0018] 根据所述一阶关系图的各一阶账号节点的第二关系确定二阶关系图；

[0019] 根据所述二阶关系图的各二阶账号节点的第三关系确定三阶关系图；

[0020] 重复上述过程，直至获取所述源账号节点的第N关系确定的N阶关系图，其中，所述关系为节点关注与被关注的其它节点之间存在的关系。

[0021] 结合上述说明，在本发明实施例另一种可能的实施方式中，所述确定节点的特征表示，所述特征表示为分别提取所述节点的内容特征和行为特征进行聚合后得到，包括：

[0022] 分别获取节点的账户信息和行为信息，所述账户信息包括简介信息和发布的文本信息，包括行为信息包括账户的活动信息；

[0023] 分别对所述账户信息和行为信息进行聚合后编码，确定所述节点的向量数据。

[0024] 结合上述说明，在本发明实施例另一种可能的实施方式中，所述方法还包括：对于每一种关系的邻居节点引入不同的权重参数，分别对属于同一种关系类型的邻居节点进行聚合后再进行不同类型关系之间的聚合；

[0025] 通过以下公式进行所述节点之间的边关系的双重聚合：

[0026] $$h_i^{l+1} = \sigma\left(\sum_{r\in R}\sum_{v_j\in N_{v_i}^r}\frac{1}{c_{i,r}}W_r^l h_j^l + W_0^l h_i^l\right)$$

[0027] 其中，R表示所述关系图中所有边的集合，$N_{v_i}^r$表示与v_i节点具有r关系的邻居集合，$c_{i,r}$为归一化参数，一般取$c_{i,r} = |N_{v_i}^r|$，W_r是具有r关系的邻居对应的权重参数，W_r是节点自身对应的权重参数；

[0028] softmax函数与对应的Loss计算函数为：

[0029] $$S_i = \frac{e^i}{\sum_i e^i}$$

[0030] $$Loss = -\sum_{i\in y} t_i \ln h_i^L$$

[0031] 其中y为有标签的节点的集合，h_i^L表示输出层有标签的第i个节点的预测值。

[0032] 结合上述说明，在本发明实施例另一种可能的实施方式中，所述方法还包括：

[0033] 所述预设的图神经网络模型进行迭代训练至Loss计算函数变化幅度小于预定值时，确定此时所述关系图的各所述节点输出的概率值为所述预测值。

[0034] 结合上述说明，在本发明实施例另一种可能的实施方式中，所述预设的图神经网络模型通过R-GCN算法实现聚合邻居的策略，以对复杂无向的关系图时加快训练和计算速度。

[0035] 第二方面，本发明示例性实施例还提供了一种基于图神经网络的账户检测装置，包括：

[0036] 图网络构建模块，用于以在线社交平台的账户为节点，账户与账户之间的关系为边，构建关系图；

[0037] 特征确定模块，用于确定节点的特征表示，所述特征表示为分别提取所述节点的内容特征和行为特征进行聚合后得到；

[0038] 判断模块，用于将所述关系图和所述特征表示输入预设的图神经网络模型，通过所述神经网络模型的输出结果确定所述关系图中的异常账户节点；

[0039] 其中，所述预设的神经网络模型为：将所述关系图的节点和边转码为能够识别的向量数据，对所述向量数据进行权重参数的训练后，输出带有标签的节点预测值，根据所述节点预测值确定节点是否为异常账号。

[0040] 第三方面，本发明还提供了一种电子设备，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程序，所述处理器执行所述程序时实现所述的基于图神经网络的账户检测方法。

[0041] 从上面所述可以看出，本发明示例性实施例提供的基于图神经网络的账户检测方法、装置和设备，提供了一种准确、高效的基于图神经网络的在线社交网络异常账号检测方法及系统，通过图神经网络技术，同时考虑在线社交网络信息流动图中单个账号的独立特征与多个账号之间的关系特征，解决传统检测方法低准确率、低鲁棒性问题的同时，大幅提升检测系统的运行效率。

附图说明

[0042] 为了更清楚地说明本发明示例性实施例或现有技术中的技术方案，下面将对实施例或现有技术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示例性实施例，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。

[0043] 图1为本发明示例性实施例基于图神经网络的账户检测方法基本流程示意图；

[0044] 图2为本发明示例性实施例的关系图网络示意图；

[0045] 图3为本发明示例性实施例的特征聚合过程示意图；

[0046] 图4为本发明示例性实施例的算法聚合示意图；

[0047] 图5为本发明示例性实施例的基于图神经网络的账户检测装置示意图；

[0048] 图6为本发明示例性实施例的电子设备结构示意图。

具体实施方式

[0049] 为使本公开的目的、技术方案和优点更加清楚明白，以下结合具体实施例，并参照附图，对本公开进一步详细说明。

[0050] 需要说明的是，除非另外定义，本发明示例性实施例使用的技术术语或者科学术

语应当为本公开所属领域内具有一般技能的人士所理解的通常意义。本发明示例性实施例中使用的“第一”、“第二”以及类似的词语并不表示任何顺序、数量或者重要性，而只是用来区分不同的组成部分。

[0051] 随着互联网的发展与移动终端设备的普及，在线社交平台愈发受网络用户青睐。OSN是依托于在线社交平台上的一种信息网络，OSN由用户与信息构成，基于OSN的网络关系图由用户构成的节点与用户之间关系构成的边组成。保证以在线社交网络为载体所传递信息的安全是确保网络空间安全的重要组成部分。OSN的便利性为许多不法份子提供了犯罪机会，一种典型的攻击手段就是Sybil Attack，不法份子创建多个机器人账号(在本发明中写作异常账号)来传播不良信息、操纵舆论，是一种典型的危害网络信息安全的行为。

[0052] 目前在线社交网络平台主要基于举报与半自动化的方式检测网络中存在的异常账号，缺乏一个公开的自动化检测工具来检测OSN中的异常账号。

[0053] Twitter所属公司采用自动化与人工相结合的方式检测异常账号，具体而言，公司采用一定技术监控大量账号的行为并发掘其中用到的行为模式，一旦发现账号的行为模式无法与正常行为模式相匹配时，需要人工对这些可疑账号展开调查以确认这些账号是否为异常账号。

[0054] 新浪微博平台则使用基于先验知识的方法，设置某些特征规则比如关注量与粉丝比例、是否通过实名认证、是否有详细个人信息等等来判断某账号是否为可疑账号，同时，被用户多次举报的账号也会被列入可疑账号名单中。最终通过人工审核的方式确认风险账号是否为异常账号。

[0055] 检测在线社交网络中异常账号的方法可分为人工检测方法、机器自动化检测方法以及半自动化方法三种，其中机器自动化检测在线社交网络中的异常账号相对人工方法有速度快、耗费小、准确率高的优势。机器自动化检测方法可分为三类:基于行为特征的检测方法、基于内容特征的检测方法和基于图的检测方法。

[0056] 基于行为特征的检测方法

[0057] 行为指的是账号在OSN上进行活动的集合，比如发布与转发信息，活动行为特征则是从此集合中提取的抽象数据，一般以数组或矩阵的形式表示。

[0058] 基于行为特征的检测方法利用异常账号与正常账号在网络中通讯的行为模式的不同来判断账号是否异常。首先，需要确定待检测账号并收集账号在一定时间范围内进行的网络活动(在社交网络中通常表现为转发、评论等行为)。然后，编码收集得到的活动数据，通过一定的先验知识将活动转化为方便机器处理的形式(特征)。最后，将特征输入分类器模型中得到分类结果判定此账户是否为异常账号，分类器可以是人工定义的判别模型比如二叉决策树，也可以是一些有监督机器学习比如SVM、RNN等。

[0059] 基于内容特征的检测方法

[0060] 内容指的是账号在OSN中发布/转发消息的内容以及账户个人信息的集合，内容特征则是从此集合中提取的抽象数据，一般以数组或矩阵的形式表示。

[0061] 基于内容特征的检测方法也需要先确定待检测的行号，但与基于行为特征的检测方法不同，基于内容特征的检测方法将重点放在静态信息的内容上。具体而言，此方法首先需要收集用户在一定时间内发布的信息内容(多为文本形式)，然后对消息文本分词将消息内容表示为单词词组，利用word2vec技术将词组转换为方便机器处理的向量或矩阵，最后

以此矩阵为特征采用有监督机器学习方法训练分类器来区分恶意账号与正常账号,其中主流的机器学习方法是以长短期记忆网络(LSTM,Long Short-Term Memory)为代表的循环神经网络算法(RNN,Recurrent Neural Network)算法。

[0062] 基于图的检测方法

[0063] 与基于行为特征和基于内容特征的检测方法针对单一账号不同,基于图的检测方法需要依据账号之间的相互关注情况构造一个账号关系图。账号关系图以账号作为节点,账号之间的联系作为边。当某个节点与待检测节点之间有信息的流动(关注/被关注、转发/被转发、评论/被评论等)时,这两个节点之间有边连接,否则没有边连接。

[0064] 在线社交网络中的账号关系图中通常存在三种关系模式:异常账号与正常账号之间的联系、正常账号与正常账号之间的联系、异常账号与异常账号之间的联系。基于图的检测方法认为这三种关系模式存在明显的差异,比如异常账号倾向于与异常账号相互协作,而正常账号通常很少主动连接至异常账号。此类方法通过区别图中异常账号与正常账号关系网的不同结构和连接方式来检查图中某个节点是否为异常节点。比如对于某个不确定节点,可以通过指定最长深度的遍历搜集其周围节点信息,以搜集到的节点中异常节点的比例来判断此节点是否为异常节点。

[0065] OSN中存在大量账号,人工判断一个账号是否为异常账号需要耗费大量时间与精力,而在线社交网络中账号创建方式较为简单,攻击者可以在短时间内创建大量账号,人工检测的速度难以跟上恶意账号产生速度。

[0066] 自动化检测方法中,基于内容特征的检测理论上是最高效的,这类方法能够在账号向社交网络中发布信息时对信息进行审核以确定此信息是否为不良信息并将发布不良信息的账号判定为可疑账号。但基于内容特征与基于行为特征的检测都存在鲁棒性差的问题,攻击者可以通过伪造账号信息绕过大部分基于规则的异常检测,而针对机器学习方法,攻击者也可以通过分析检测模型或试错方式还原检测模型,并控制账号发送的信息内容以及行为活动以避免出现某些易被检测为恶意账号的特征来绕过机器学习模型检测,同时新注册的账号因为行为特征与内容特征的缺失而难以被识别并可能导致模型的误识别。

[0067] 基于图的检测模型考虑多个用户之间的通讯关系,虽然在鲁棒性上强于上述两类模型,但需要构建复杂的社交网络关系图,且检测的准确率依托于图建模水平,要达到准确的检测效果需要较强的专业知识。此外,前文介绍的三类方法只能在同一时间对一个账号进行检测,不能同时判断多个账号是否异常,一定程度上限制了模型检测的性能,因此目前亟需解决的两个问题如下:

[0068] 如何设计出一个鲁棒性强的检测模型,鲁棒性强的检测模型应当能够辨别攻击者为了绕过检测刻意伪造的信息并给出正确的检测结果。

[0069] 如何提升模型的检测效率以满足在社交网络图中能同时检测多个节点是否异常的需求。

[0070] 本发明的方法能够解决上述问题,且本发明的方法所使用的源代码支持c++/python/java等主流开发语言,适用于twitter、facebook、微博等多个在线社交网络平台,并可部署在网络用户终端或在线社交网络平台服务器上。

[0071] 本发明涉及一种基于图神经网络的账户检测方法、装置和设备,其主要运用于场景中,其基本思想是:构建OSN关系图,图节点为网络中的账号,图中边表示账号之间的关

系，定义的关系为{关注、被关注}；从一个已知节点(即源节点，能够确定为正常或异常)出发，得到其一阶、二阶与三阶邻居节点并依照关注或被关注关系构造网络关系图；针对OSN关系图我们需要确定节点的特征表示，本发明设计了一种OSN关系图节点特征编码方案能够自动化地收集信息并将节点表示为向量特征以便后续操作；在完成OSN图的构建后，以图为输入设计半监督的异构社交网络，图中节点部分标注；深度学习的任务是将图中每一个节点进行标注以检测异常账号；本发明设计的一个基于图神经网络的OSN异常账号检测方法，能够在不良消息传播时自动检测网络中操控消息传播的异常账号，维护在线社交网络安全。

[0072] 本实施例可适用于带有卷积神经网络模块的智能型终端中以进行账户检测的情况中，该方法可以由智能学习的装置来执行，其中该装置可以由软件和/或硬件来实现，一般地可集成于移动终端中，或者终端中的中心控制模块来控制，如图1所示，为本发明示例性实施例的基于图神经网络的账户检测方法的基本流程示意图，所述方法具体包括如下步骤：

[0073] 在步骤110中，以在线社交平台的账户为节点，账户与账户之间的关系为边，构建关系图；

[0074] 在步骤120中，确定节点的特征表示，所述特征表示为分别提取所述节点的内容特征和行为特征进行聚合后得到；

[0075] 在步骤130中，将所述关系图和所述特征表示输入预设的图神经网络模型，通过所述神经网络模型的输出结果确定所述关系图中的异常账户节点；

[0076] 其中，所述预设的神经网络模型为：将所述关系图的节点和边转码为能够识别的向量数据，对所述向量数据进行权重参数的训练后，输出带有标签的节点预测值，根据所述节点预测值确定节点是否为异常账号。

[0077] 结合图2所示的关系图，本发明示例性实施例中基于图神经网络的在线社交网络异常账号检测系统包括关系信息收集、账号特征收集与异常账号检测三部分，其中方法的核心和重心放在异常账号检测部分。

[0078] 在基于图的检测方法中，本发明示例性实施例的实施方式中需要使用爬虫自动化地爬取用户数据构造OSN中的账号关系图。假设已确定某源用户为正常账号或异常账号，需要分以下几个步骤从单一账号开始获取有限深度的账号关系图：

[0079] 将源账号依据已知情况标记为正常账号或异常账号，获取此账号关注的所有账号并标记关系为R1，获取此账号的所有粉丝(关注了此账号的其它账户)并标记关系为R2，此时生成了源账号的一阶关系图。

[0080] 对一阶关系图中的每一个节点重复1)中的操作，可以将一阶关系图扩充为二阶关系图。

[0081] 对节点进行持续扩充，得到更高阶的OSN关系图，重复上述过程，能够得到第N阶关系图，然而根据实际情况，也为了加快后续图计算的收敛、保证检测精度，一般将关系图深度设置为3阶，即N＝3。

[0082] 生成的社交网络关系图中总共有两种关系：若节点a关注节点b，则a、b之间存在一条由a指向b的边；若节点a为节点b的粉丝，则a、b之间存在一条由b指向a的边。如图1所示为一个二阶关系图，颜色较深的黑色节点表示已标记的异常账号，颜色较浅的灰色节点表示

未标记账号。

[0083] 本发明示例性实施例的一种实施方式中，还包括：确定预建立账户判断范围表，根据所述预设的图神经网络模型输出的节点预测值在所述账户判断范围表的匹配结果，确定节点是否为异常账号；

[0084] 所述账户判断范围表包括第一数值范围和第二数值范围，当所述节点预测值处于第一范围时，确定所述节点预测值对应的节点为正常账户；当所述节点预测值处于第二范围时，确定所述节点预测值对应的节点为异常账户。

[0085] 所述节点预测值可为(0,1)之间的数值，所述判断范围表可包括(0～0.2，正常账户)(0.2～0.8，异常账户)，则当节点预测值为0.1时，根据匹配确定该节点为正常账户，当节点为0.5时，确定该节点为异常账户。

[0086] 所述预设的卷积神经网络模型，包括：

[0087] 生成OSN关系图后，需要使用一定的数据结构表示关系图网络中的各个节点，本发明综合使用了内容特征与行为特征表示网络中的每一个节点。首先需要对每一个账号的必要信息进行爬取，以微博为例，具体考虑的信息如下：

[0088] 账号个人简介：账号主页通常会包含一段简短的个人简介，普通账号的个人简介通常涵盖了持有账号用户的喜好以及所属领域，而异常账号的个人简介通常为空或意义不明的一段文字。

[0089] 账号一段时间内发布的信息：为了提升特征提取的高效性，本发明只考虑文本类型数据。账号发布的信息通常可以分为原创信息与转发信息，原创信息指由账号自行发布的原创文本消息，转发信息指账号转发其他账号发布所发布的消息并附上自身的评论性言论的文本消息。

[0090] 结合图3所示的特征表示，爬取数据完毕后需要对收集得到的信息进行编码使其变成机器能够识别的向量数据。在这里需要分为针对文本信息的编码与针对行为信息的编码。

[0091] 针对文本信息包括账号个人简介与发布微博文本内容，目的是使用基于自然语言处理(NLP)的编码技术将文本转化为特征向量。

[0092] 首先，使用文本分词工具JieBa将文本句划分为多个中文单词。JieBa工具支持三种分词模式：精确模式试图将句子最精确地切开，适合文本分析；全模式把句子中所有的可以成词的词语都扫描出来，速度非常快，但是不能解决歧义；搜索引擎模式会在精确模式的基础上，对长词再次切分，提高召回率，适合用于搜索引擎分词。此处为了提高系统总体效率本发明使用JieBa全模式进行分词工作。

[0093] 分词之后需要将单词映射至向量空间中，为了减少向量空间的维度，本发明基于word2vec技术中的神经网络分布式词向量表示，使用预训练的词向量模型将每一个单词映射为固定维度的词向量，然后将文本中所有词向量取平均操作得到整个文本的向量表示。

[0094] 针对账号原创文本信息，只需生成一个文本向量表示；对于账号转发的信息，需要分别生成所转发信息的文本向量表示以及账号转发附带文字的文本向量表示。最后需要将所有文本信息的向量表示聚合为一个向量。

[0095] 针对行为信息包括账号一段时间内在社交网络中所进行的活动，由于新浪微博平台的隐私保护措施，非本账号只能在账号主页中爬取账号发送的公开微博信息。此部分针

对账号的每条微博需要记录的信息有:发送时间、发送地点、发送类型(原创/转发)、评论数、点赞数、转发数、内容类型(图片/文字/视频)。收集一段时间内的活动数据后使用人工指定的特征编码规则对信息进行编码,比如使用3个二进制位表示微博内容类型:001表示包含文字,010表示包含图片,100表示包含视频。

[0096] 最后,根据文本信息向量与行为信息向量得到单个节点的最终向量表示作为所述特征表示。

[0097] 经过前面的两步得到了一个有向在线社交网络关系图,图中每个节点具备一个向量特征表示,图中每条边表示两节点之间的关注与被关注关系,图中除了源节点外其他节点均为未标注数据。在此使用异构的图卷积神经网络算法R-GCN提取图特征,基于半监督学习方法可以将源节点的标签信息扩展到图中所有节点,给图中每一个节点进行标注以检测其他异常节点(异常账号)。

[0098] 图神经网络(GNN,Graph Neural Network)在图信号处理上存在的天然优势,GNN算法分为spectral-based与spatial-based两类。图卷积神经网络(GCN,Graph Convolutional Network)是一类典型的spectral-based图神经网络算法。GraphSAGE算法是一个基于GCN进行改进的spatial-based算法,它通过聚合邻居的策略实现卷积操作可以加快GNN在较大无向图上的计算训练速度。

[0099] 结合图4所示的算法中邻居节点聚合过程示例图,在OSN关系图中,两节点的边分为关注与被关注两种结构,边是异构的,R-GCN是GraphSAGE的一个变体,它能够处理异构边图模型的spatial-based GNN算法。R-GCN对于每一种关系的邻居节点引入不同的权重参数,分别对属于同一种关系类型的邻居节点进行聚合后再进行不同类型关系之间的聚合。从计算角度来看,R-GCN在GraphSAGE聚合邻居操作的基础上又增加了一个聚合关系的维度,使得节点的聚合操作变成一个双重聚合的过程,其核心公式如下:

[0100] 通过以下公式进行双重聚合:

[0101] $$h_i^{l+1} = \sigma\left(\sum_{r\in R}\sum_{v_j\in N_{v_i}^r}\frac{1}{c_{i,r}}W_r^l h_j^l + W_0^l h_i^l\right)$$

[0102] 其中,R表示所述关系图中所有边的集合,$N_{v_i}^r$表示与v_i节点具有r关系的邻居集合,$c_{i,r}$为归一化参数,一般取$c_{i,r} = |N_{v_i}^r|$,W_r是具有r关系的邻居对应的权重参数,W_r是节点自身对应的权重参数;

[0103] 通过将多个如上图所示的GCN聚合层堆叠得到一个GCN中间网络,在GCN最后一层后叠加一个softmax层计算分类输出,softmax函数可以将多个神经元的输出映射到(0,1)区间内,此时输出即可表示为表示节点属于某一类别的概率函数,通过概率函数与图中以标记的一些节点数据并使用交叉熵计算Loss值,最后将Loss反向传播至图神经网络中进行权重参数的调整(网络训练)。

[0104] softmax函数与对应的Loss计算函数为:

[0105] $$S_i = \frac{e^i}{\sum_t ej}$$

[0106] $$Loss = -\sum_{iey} t_i lnh_i^L$$

[0107] 其中y为有标签的节点的集合，h_i^L表示输出层有标签的第i个节点的预测值。

[0108] 迭代进行此过程直到网络收敛(Loss变化幅度小于预定值)，就可以依据每个节点输出的概率值确定该节点对应账号是否为异常账号。

[0109] 本发明的方法有益效果包括：

[0110] 在处理图数据的过程中，使用机器学习模型代替传统基于先验知识的模型，减少了对先验知识的需求。同时深度学习方法相对于普通机器学习方法能够自动化地从复杂数据中提取特征，削减了特征工程时的人工投入。

[0111] 由于有监督的深度学习需要大量人工标注数据支持训练，本发明的方法使用半监督图神经网络，使模型能够在只存在少量标记数据的数据集中精准识别节点特征，将图中的少量标记扩展到全图标记满足分类需求。

[0112] 提升了检测准确率：

[0113] 基于特征的检测只考虑单个节点的活动特征，基于图的检测只考虑节点之间的关系建模，当节点特征或关系特征数据不足时，模型检测准确率会严重降低。而本发明中提出的图卷积神经网络可以同时考虑节点自身的特征与节点之间的关系，使模型依赖于更多的数据特征以避免当某几个特征缺失时模型失效的情况，可以显著提升检测准确率。

[0114] 提升模型鲁棒性：

[0115] 向OSN中投放不良信息的攻击者往往会使用一些伪装技术修饰其创建的异常账号，比如复制正常账号的个人简介等信息绕过基于内容特征的检测，或者模拟正常账号的行为活动以绕过基于行为特征的检测。本发明继承了基于图的检测技术的优点，不止提取节点个体特征还提取节点周围关系特征，能够有效抵抗在单个节点上进行的伪装。而攻击者想要继续绕过检测需要同时修改网络中的多个节点特征及其连接关系，这不仅难度较大还可能会使攻击者无法正常进行攻击。

[0116] 提高检测效率：

[0117] 大部分基于机器学习的在线社交网络异常账号检测方法都属于有监督学习，这需要大量的人工标注数据，尽管存在一些人工标注的数据集，但在线社交网络中异常账号的数量与特征都在不断增多，在老旧数据集上训练的模型无法在新数据中得到较高的准确率，而人工标注效率较低，这在一定程度上限制了适用于新特征模型的训练。本发明中提出的半监督图神经网络可以在少量有标注的数据中进行训练并将标注扩展到图中的所有节点，一定程度上提升了模型训练效率。

[0118] 除此之外，不论是基于特征的方法还是基于图的方法，都只能同时对一个节点进行检测，而基于图神经网络的方法可以同时输出多个节点的分类结果，提升了检测效率。

[0119] 使用了半监督的图卷积神经网络来预测社交网络关系图中未标注节点的特征以检测在线社交网络中的异常账号，不仅提升了检测准确率与鲁棒性，还提升了系统检测性能。

[0120] 可以理解，该方法可以通过任何具有计算、处理能力的装置、设备、平台、设备集群来执行。

[0121] 基于同一发明构思，图5为本发明实施例提供的一种基于图神经网络的账户检测装置的结构示意图，该装置可由软件和/或硬件实现，一般地集成于智能终端中，可通过基于图神经网络的账户检测方法来实现。如图所示，本实施例以上述任意的基于图神经网络

的账户检测方法实施例相对应地，提供了一种基于图神经网络的账户检测装置，其主要包括了图网络构建模块510、特征确定模块520、判断模块530和算法模块540。

[0122] 其中的图网络构建模块510，用于以在线社交平台的账户为节点，账户与账户之间的关系为边，构建关系图；

[0123] 其中的特征确定模块520，用于确定节点的特征表示，所述特征表示为分别提取所述节点的内容特征和行为特征进行聚合后得到；

[0124] 其中的判断模块530，用于将所述关系图和所述特征表示输入预设的图神经网络模型，通过所述神经网络模型的输出结果确定所述关系图中的异常账户节点；

[0125] 算法模块540，用于建立所述预设的神经网络模型：将所述关系图的节点和边转码为能够识别的向量数据，对所述向量数据进行权重参数的训练后，输出带有标签的节点预测值，根据所述节点预测值确定节点是否为异常账号。

[0126] 本发明示例性实施例的另一种实施方式中，所述装置还包括判断模块，用于：预建立账户判断范围表，根据所述预设的图神经网络模型输出的节点预测值在所述账户判断范围表的匹配结果，确定节点是否为异常账号；

[0127] 所述账户判断范围表包括第一数值范围和第二数值范围，当所述节点预测值处于第一范围时，确定所述节点预测值对应的节点为正常账户；当所述节点预测值处于第二范围时，确定所述节点预测值对应的节点为异常账户。

[0128] 本发明示例性实施例的另一种实施方式中，所述装置还包括图阶确定模块，用于：确定源账号节点后，根据所述源账号节点的第一关系确定一阶关系图；

[0129] 根据所述一阶关系图的各一阶账号节点的第二关系确定二阶关系图；

[0130] 根据所述二阶关系图的各二阶账号节点的第三关系确定三阶关系图；

[0131] 重复上述过程，直至获取所述源账号节点的第N关系确定的N阶关系图，其中，所述关系为节点关注与被关注的其它节点之间存在的关系。

[0132] 本发明示例性实施例的另一种实施方式中，所述算法模块还用于：

[0133] 分别获取节点的账户信息和行为信息，所述账户信息包括简介信息和发布的文本信息，包括行为信息包括账户的活动信息；

[0134] 分别对所述账户信息和行为信息进行聚合后编码，确定所述节点的向量数据。

[0135] 本发明示例性实施例的另一种实施方式中，所述装置还包括聚合模块，用于：对于每一种关系的邻居节点引入不同的权重参数，分别对属于同一种关系类型的邻居节点进行聚合后再进行不同类型关系之间的聚合；

[0136] 通过以下公式进行所述节点之间的边关系的双重聚合，聚合的是节点与节点之间的关系特征，即关注与被关注关系，本发明示例性实施例的一种实施方式中，可指将内容特征与行为特征所表示的向量进行连接操作：

[0137] $$h_i^{l+1} = \sigma\left(\sum_{r\in R}\sum_{v_j\in N_{v_i}^r}\frac{1}{c_{i,r}}W_r^l h_j^l + W_0^l h_i^l\right)$$

[0138] 其中，R表示所述关系图中所有边的集合，$N_{v_i}^r$表示与v_i节点具有r关系的邻居集合，$c_{i,r}$为归一化参数，一般取$c_{i,r} = |N_{v_i}^r|$，W_r是具有r关系的邻居对应的权重参数，W_r是节点自身对应的权重参数；

[0139] softmax函数与对应的Loss计算函数为：

[0140] $$S_i = \frac{e^i}{\sum_i e^j}$$

[0141] $$Loss = -\sum_{i \in y} t_i \ln h_i^L$$

[0142] 其中y为有标签的节点的集合，h_i^L表示输出层有标签的第i个节点的预测值。

[0143] 本发明示例性实施例的另一种实施方式中，所述装置还包括确定模块，用于所述预设的图神经网络模型进行迭代训练至Loss计算函数变化幅度小于预定值时，确定此时所述关系图的各所述节点输出的概率值为所述预测值。

[0144] 本发明示例性实施例的另一种实施方式中，所述装置还包括策略模块，用于：所述预设的图神经网络模型通过R-GCN算法实现聚合邻居的策略，以对复杂无向的关系图时加快训练和计算速度。

[0145] 为了描述的方便，上述装置分别以功能分为各类模块进行分别描述，当然，本发明示例性实施例实施时可以将各模块的功能在同一个软件和/或硬件实现，且，上述实施例中提供的基于图神经网络的账户检测装置可执行本发明中任意实施例中所提供的基于图神经网络的账户检测方法，具备执行该方法相应的功能模块和有益效果，未在上述实施例中详细描述的技术细节，可参见本发明任意实施例中所提供的基于图神经网络的账户检测方法，在此不再赘述。

[0146] 在线社交网络中存在大量的异常账号，这些账号通过协作转发不良信息达到操纵舆论的目的，严重危害在线社交网络信息安全，及时发现并屏蔽异常账号是在线社交平台需要迫切解决的问题，此发明期望能够以少量的先验知识以及人工参与达到快速准确检测OSN中异常账号的目的。

[0147] 我们引入异构图的概念以及深度学习技术到在线社交网络安全领域，提出一种新型的基于图神经网络的在线社交网络中异常用户检测算法，设计了一个准确、高效的在线社交网络异常账号检测系统。发明基于图神经网络的在线社交网络异常账号检测系统目的如下：

[0148] 鲁棒性

[0149] 需要一个强鲁棒性的模型防止攻击者通过伪造个人信息或故意构造消息内容来绕过异常账号检测。许多攻击者会复制在线社交网络中某真实用户的信息与部分行为以伪装为真实用户，这种方法能有效绕过基于行为特征的检测与基于内容特征的检测。在线社交网络异常账号检测系统需要具备分辨此类伪造信息的能力。本发明提出的GNN方法是一种半监督的深度学习方法，它以OSN用户关系图作为输入，同时考虑图中的节点特征与节点关系特征，能够有效消除单一节点信息易被伪造的缺点。

[0150] 准确性

[0151] 基于特征的检测往往会因为关键特征缺失影响到检测结果的准确率，当攻击者使用新注册的异常账号实施攻击时，检测模型由于无法捕捉到关键特征而无法区分正常注册账号与异常注册账号，将导致大量的误识别。我们的系统需要能够避免此类误识别，将特征范围进一步扩充，使用基于账号的特征以及其在OSN中的关系特征判断账号是否异常。

[0152] 高效性

[0153] 一般的检测方法一次只能对一个账号进行检测，而在OSN中，正常账号数量远远大于异常账号数量，对选定账号检测的方法无法有效检测出网络中的异常账号，因此，高效的异常账号检测系统需要能够同时检测OSN中多个账号的异常情况。

[0154] 基于同一发明构思，与上述任意实施例方法相对应的，本说明书一个或多个实施例还提供了一种电子设备，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程序，所述处理器执行所述程序时实现上任意一实施例所述的基于图神经网络的账户检测方法。

[0155] 需要说明的是，本发明示例性实施例的方法可以由单个设备执行，例如一台计算机或服务器等。本实施例的方法也可以应用于分布式场景下，由多台设备相互配合来完成。在这种分布式场景的情况下，这多台设备中的一台设备可以只执行本发明示例性实施例的方法中的某一个或多个步骤，这多台设备相互之间会进行交互以完成所述的基于图神经网络的账户检测方法。

[0156] 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。

[0157] 为了描述的方便，描述以上装置时以功能分为各种模块分别描述。当然，在实施本发明示例性实施例时可以把各模块的功能在同一个或多个软件和/或硬件中实现。

[0158] 图6示出了本实施例所提供的一种更为具体的电子设备硬件结构示意图，该设备可以包括：处理器1010、存储器1020、输入/输出接口1030、通信接口1040和总线1050。其中处理器1010、存储器1020、输入/输出接口1030和通信接口1040通过总线1050实现彼此之间在设备内部的通信连接。

[0159] 处理器1010可以采用通用的CPU(Central Processing Unit，中央处理器)、微处理器、应用专用集成电路(Application Specific Integrated Circuit，ASIC)、或者一个或多个集成电路等方式实现，用于执行相关程序，以实现本说明书实施例所提供的技术方案。

[0160] 存储器1020可以采用ROM(Read Only Memory，只读存储器)、RAM(Random Access Memory，随机存取存储器)、静态存储设备，动态存储设备等形式实现。存储器1020可以存储操作系统和其他应用程序，在通过软件或者固件来实现本说明书实施例所提供的技术方案时，相关的程序代码保存在存储器1020中，并由处理器1010来调用执行本发明实施例的基于图神经网络的账户检测方法。

[0161] 输入/输出接口1030用于连接输入/输出模块，以实现信息输入及输出。输入输出/模块可以作为组件配置在设备中(图中未示出)，也可以外接于设备以提供相应功能。其中输入设备可以包括键盘、鼠标、触摸屏、麦克风、各类传感器等，输出设备可以包括显示器、扬声器、振动器、指示灯等。

[0162] 通信接口1040用于连接通信模块(图中未示出)，以实现本设备与其他设备的通信交互。其中通信模块可以通过有线方式(例如USB、网线等)实现通信，也可以通过无线方式(例如移动网络、WIFI、蓝牙等)实现通信。

[0163] 总线1050包括一通路，在设备的各个组件(例如处理器1010、存储器1020、输入/输出接口1030和通信接口1040)之间传输信息。

[0164] 需要说明的是，尽管上述设备仅示出了处理器1010、存储器1020、输入/输出接口1030、通信接口1040以及总线1050，但是在具体实施过程中，该设备还可以包括实现正常运行所必需的其他组件。此外，本领域的技术人员可以理解的是，上述设备中也可以仅包含实现本说明书实施例方案所必需的组件，而不必包含图中所示的全部组件。

[0165] 上述实施例的电子设备用于实现前述任一实施例中相应的基于图神经网络的账户检测方法，并且具有相应的方法实施例的有益效果，在此不再赘述。

[0166] 所属领域的普通技术人员应当理解：以上任何实施例的讨论仅为示例性的，并非旨在暗示本公开的范围(包括权利要求)被限于这些例子；在本公开的思路下，以上实施例或者不同实施例中的技术特征之间也可以进行组合，步骤可以以任意顺序实现，并存在如上所述的本发明示例性实施例的不同方面的许多其它变化，为了简明它们没有在细节中提供。

[0167] 另外，为简化说明和讨论，并且为了不会使本发明示例性实施例难以理解，在所提供的附图中可以示出或可以不示出与集成电路(IC)芯片和其它部件的公知的电源/接地连接。此外，可以以框图的形式示出装置，以便避免使本发明示例性实施例难以理解，并且这也考虑了以下事实，即关于这些框图装置的实施方式的细节是高度取决于将要实施本发明示例性实施例的平台的(即，这些细节应当完全处于本领域技术人员的理解范围内)。在阐述了具体细节(例如，电路)以描述本公开的示例性实施例的情况下，对本领域技术人员来说显而易见的是，可以在没有这些具体细节的情况下或者这些具体细节有变化的情况下实施本发明示例性实施例。因此，这些描述应被认为是说明性的而不是限制性的。

[0168] 尽管已经结合了本公开的具体实施例对本公开进行了描述，但是根据前面的描述，这些实施例的很多替换、修改和变型对本领域普通技术人员来说将是显而易见的。例如，其它存储器架构(例如，动态RAM(DRAM))可以使用所讨论的实施例。

[0169] 本发明示例性实施例旨在涵盖落入所附权利要求的宽泛范围之内的所有这样的替换、修改和变型。因此，凡在本发明示例性实施例的精神和原则之内，所做的任何省略、修改、等同替换、改进等，均应包含在本公开的保护范围之内。

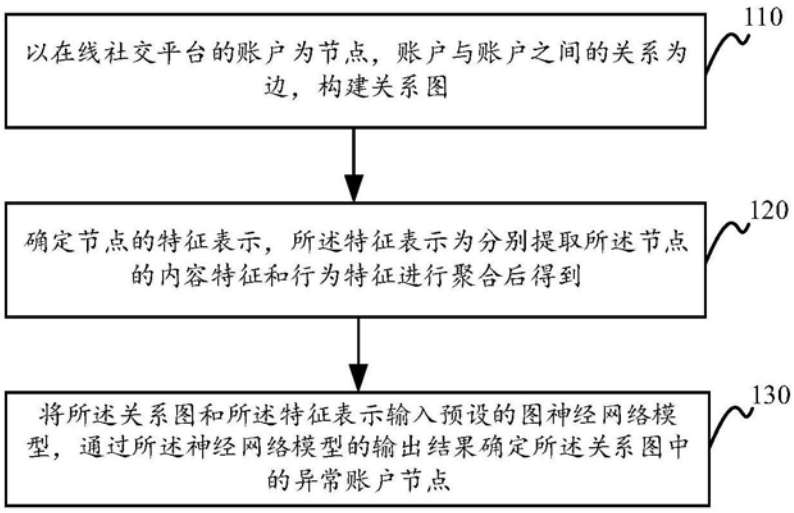

以在线社交平台的账户为节点，账户与账户之间的关系为边，构建关系图
110
确定节点的特征表示，所述特征表示为分别提取所述节点的内容特征和行为特征进行聚合后得到
120
将所述关系图和所述特征表示输入预设的图神经网络模型，通过所述神经网络模型的输出结果确定所述关系图中的异常账户节点
130

图1

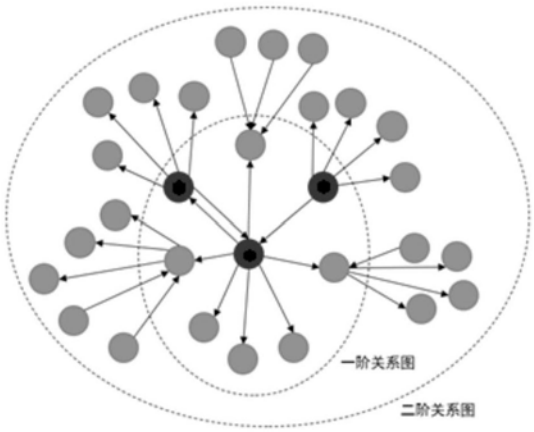

一阶关系图
二阶关系图

图2

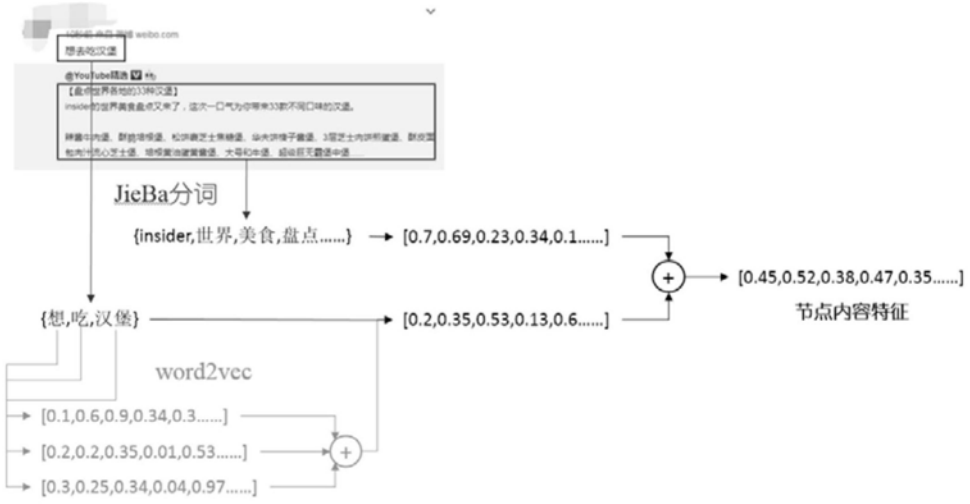

JieBa分词
{insider,世界,美食,盘点……}
[0.7,0.69,0.23,0.34,0.1……]
[0.45,0.52,0.38,0.47,0.35……]
节点内容特征
{想,吃,汉堡}
[0.2,0.35,0.53,0.13,0.6……]
word2vec
[0.1,0.6,0.9,0.34,0.3……]
[0.2,0.2,0.35,0.01,0.53……]
[0.3,0.25,0.34,0.04,0.97……]

图3

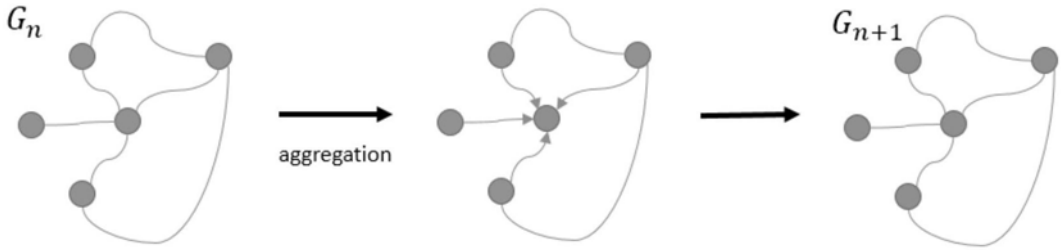

G_n
aggregation
G_{n+1}

图4

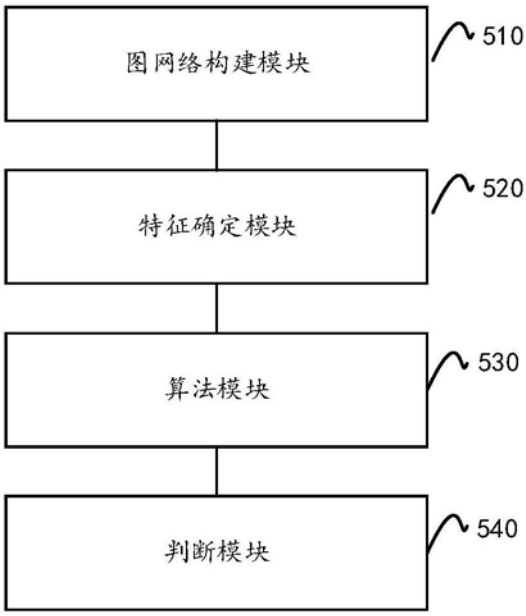

图网络构建模块
510
特征确定模块
520
算法模块
530
判断模块
540

图5

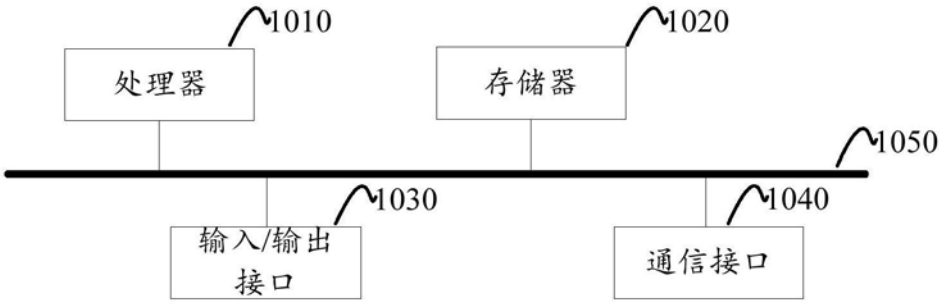


图6