(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 112732919 A
(43)申请公布日 2021.04.30

(21)申请号 202110052560.1

(22)申请日 2021.01.15

(71)申请人 中国科学院地理科学与资源研究所
地址 100101 北京市朝阳区大屯路甲11号

(72)发明人 郭启全 江东

(74)专利代理机构 北京中和立达知识产权代理事务所(普通合伙) 11756
代理人 杨磊

(51)Int.Cl.
G06F 16/35(2019.01)
G06F 16/951(2019.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)
H04L 29/06(2006.01)

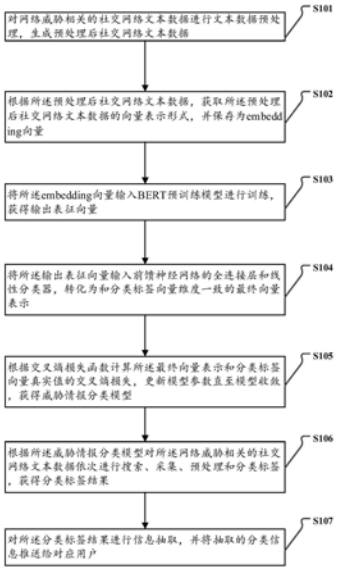
权利要求书3页 说明书11页 附图10页

(54)发明名称

一种面向网络安全威胁情报的智能分类标签方法及系统

(57)摘要

本发明提供了一种面向网络安全威胁情报的智能分类标签方法及系统。该方案包括对网络威胁相关的社交网络文本数据进行文本数据预处理,获取向量表示形式,并输入BERT预训练模型进行训练;进而利用前馈神经网络的全连接层和线性分类器,转化为和分类标签向量维度一致向量;根据交叉熵损失函数计算更新模型参数直至模型收敛,获得威胁情报分类模型;根据所述威胁情报分类模型对社交网络文本数据依次进行搜索、采集、预处理和分类标签,获得分类标签结果。本方案采用预训练模型学习文本数据的上下文语义信息和句子间关系,获取语义表征,生成的威胁情报分类模型准确率高,可提高训练效率,缩短直接模型训练时间。



CN 112732919 A

1.一种面向网络安全威胁情报的智能分类标签方法,其特征在于,包括:

对网络威胁相关的社交网络文本数据进行文本数据预处理,生成预处理后社交网络文本数据;

根据所述预处理后社交网络文本数据,获取所述预处理后社交网络文本数据的向量表示形式,并保存为embedding向量;

将所述embedding向量输入BERT预训练模型进行训练,获得输出表征向量;

将所述输出表征向量输入前馈神经网络的全连接层和线性分类器,转化为和分类标签向量维度一致的最终向量表示;

根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失,更新模型参数直至模型收敛,获得威胁情报分类模型;

根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签,获得分类标签结果;

对所述分类标签结果进行信息抽取,并将抽取的分类信息推送给对应用户。

2.如权利要求1所述的一种面向网络安全威胁情报的智能分类标签方法,其特征在于,所述对网络威胁相关的社交网络文本数据进行文本数据预处理,生成预处理后社交网络文本数据,具体包括:

将所述网络威胁相关的社交网络文本数据划分成若干个顺序固定的单个句子;

删除所有的所述单个句子中的特殊符号,保存为删除特殊符号的单个句子;

对所有的所述删除特殊符号的单个句子利用WordPiece分词算法进行单词拆分,生成预处理后社交文本数据。

3.如权利要求1所述的一种面向网络安全威胁情报的智能分类标签方法,其特征在于,所述embedding向量包括token embedding向量、position embedding向量和segment embedding向量。

4.如权利要求3所述的一种面向网络安全威胁情报的智能分类标签方法,其特征在于,所述根据所述预处理后社交网络文本数据,获取所述预处理后社交网络文本数据的向量表示形式,并保存为embedding向量,具体包括:

获取所述预处理后社交文本数据中的所述token embedding向量;

在所述token embedding向量中插入特殊分离符,用于分割所述预处理后社交文本数据;

在所述token embedding向量中插入特殊分隔符,用于分割所述预处理后社交文本数据中的不同句子;

利用所述segment embedding向量进行所述预处理后社交文本数据中的相邻2个句子的向量表示;

利用所述position embedding向量进行所述预处理后社交文本数据中的序列位置信息的表示。

5.如权利要求1所述的一种面向网络安全威胁情报的智能分类标签方法,其特征在于,所述将所述embedding向量输入BERT预训练模型进行训练,获得输出表征向量,具体包括:

将所述embedding向量输入所述BERT预训练模型;

利用所述BERT预训练模型进行第一步训练,所述第一步训练为通过随机遮蔽掉一个句

子中的词，利用上下文进行预测，获得一个双向深度上下文语义信息；

利用所述BERT预训练模型进行第二步训练，所述第二步训练为预测一个句子的下一个句子，获得一个句子与句子之间的关系；

反复进行预先设定次数的所述第一步训练和所述第二步训练，根据所有的所述双向深度上下文语义信息和所有的所述句子与句子之间的关系生成所述输出表征向量。

6.如权利要求1所述的一种面向网络安全威胁情报的智能分类标签方法，其特征在于，所述将所述输出表征向量输入前馈神经网络的全连接层和线性分类器，转化为和分类标签向量维度一致的最终向量表示，具体包括：

获取所述输出表征向量，输入到所述前馈神经网络的全连接层；

通过所述前馈神经网络的全连接层生成第一中间向量；

将所述第一中间向量输入到一个线性映射的分类器，生成和标签向量数量维度一致的向量；

将所述和标签向量数量维度一致的向量保存为所述最终向量表示。

7.如权利要求1所述的一种面向网络安全威胁情报的智能分类标签方法，其特征在于，所述根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失，更新模型参数直至模型收敛，获得威胁情报分类模型，具体包括：

根据交叉熵损失函数随机计算获取一个所述最终向量表示和分类标签向量真实值的交叉熵损失，作为初始损失值；

进行新损失值运算，所述新损失值运算为根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失，作为新的损失值；

反复进行新损失值运算，直到模型收敛或达到预设运算次数时，输出所述威胁情报分类模型，其中，所述模型收敛为在所述新损失值运算过程中连续K次所述新的损失值不大于所述初始损失值的情况，K为用户在建模前预先设定的收敛次数。

8.如权利要求1所述的一种面向网络安全威胁情报的智能分类标签方法，其特征在于，所述模型参数具体包括：所述线性分类器的参数、所述全连接层的参数和所述BERT预训练模型的参数。

9.如权利要求1所述的一种面向网络安全威胁情报的智能分类标签方法，其特征在于，所述根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签，获得分类标签结果，具体包括：

设置重点关注网络安全列表；

根据所述重点关注网络安全列表利用搜素引擎定时对社交网络中安全主体的相关信息进行搜索和采集，获取待分类社交网络文本数据；

对所述待分类社交网络文本数据进行预处理，获得获得输入文本的向量表示；

利用所述威胁情报分类模型根据所述输入文本的向量表示进行分类结果计算，生成所述分类标签结果。

10.一种面向网络安全威胁情报的智能分类标签系统，其特征在于，该系统包括：

预处理模块，用于对网络威胁相关的社交网络文本数据进行文本数据预处理，生成预处理后社交网络文本数据；

第一向量生成模块，用于根据所述预处理后社交网络文本数据，获取所述预处理后社

交网络文本数据的向量表示形式，并保存为embedding向量；

第二向量生成模块，用于将所述embedding向量输入BERT预训练模型进行训练，获得输出表征向量；

第三向量生成模块，用于将所述输出表征向量输入前馈神经网络的全连接层和线性分类器，转化为和分类标签向量维度一致的最终向量表示；

模型训练模块，用于根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失，更新模型参数直至模型收敛，获得威胁情报分类模型；

分类标签模块，用于根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签，获得分类标签结果；

分类结果推送模块，用于对所述分类标签结果进行信息抽取，并将抽取的分类信息推送给对应用户。

一种面向网络安全威胁情报的智能分类标签方法及系统

技术领域

[0001] 本发明涉及网络安全技术领域，更具体地，涉及一种面向网络安全威胁情报的智能分类标签方法及系统。

背景技术

[0002] 网络安全威胁情报是与网络空间安全威胁相关的线索和证据。网络安全威胁情报可作为网络安全基础知识使用，包括网络安全的主体、主体涉及场景、机制、指标、影响和可执行的建议等信息。掌握网络安全威胁情报可以及时、有效提高针对网络攻击威胁的监测发现与应急响应能力。

[0003] 在本发明之前，现有技术中主要采用神经网络对网络空间安全威胁的数据处理。循环神经网络可以产生记忆效应，适合处理自然语言类的序列数据。但是，由于梯度消失和梯度爆炸，当一个遥远序列传递到当前，梯度变得很小时，无法建立长期记忆，导致循环神经网络存在长程依赖的问题。长短时记忆网络是一种特殊的循环神经网络结构，包含输入门、遗忘门和输出门。长短时记忆网络通过门控机制可以建立较长的长距离时序依赖关系，有助于解决自然语言序列中的长程依赖问题。但如果直接使用神经网络对网络威胁相关的自然语言文本进行分类模型学习，通常需要大量的训练数据，由于网络威胁相关数据与不相关数据的不平衡问题，导致用于模型训练的数据量较为匮乏，容易导致学习过程的过拟合。因此，直接使用传统的模型，尚不能很好解决机器学习在威胁情报领域使用时有效训练样本的数据量小，且训练样本分布不均衡而导致的过拟合、模型准确度低的问题。

发明内容

[0004] 鉴于上述问题，本发明提出了一种面向网络安全威胁情报的智能分类标签方法及系统，在预训练模型的基础上进行分类器训练，能够解决传统机器学习方法在应用到威胁情报分类领域时，由于有效训练样本的数据量小和训练样本分布不均衡而导致的过拟合、模型准确度低的问题。

[0005] 在发明实施例的第一方面，提供可一种面向网络安全威胁情报的智能分类标签方法包括：

[0006] 对网络威胁相关的社交网络文本数据进行文本数据预处理，生成预处理后社交网络文本数据；

[0007] 根据所述预处理后社交网络文本数据，获取所述预处理后社交网络文本数据的向量表示形式，并保存为embedding向量；

[0008] 将所述embedding向量输入BERT预训练模型进行训练，获得输出表征向量；

[0009] 将所述输出表征向量输入前馈神经网络的全连接层和线性分类器，转化为和分类标签向量维度一致的最终向量表示；

[0010] 根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失，更新模型参数直至模型收敛，获得威胁情报分类模型；

[0011] 根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签，获得分类标签结果；

[0012] 对所述分类标签结果进行信息抽取，并将抽取的分类信息推送给对应用户。

[0013] 在一个或多个实施例中，优选地，所述对网络威胁相关的社交网络文本数据进行文本数据预处理，生成预处理后社交网络文本数据，具体包括：

[0014] 将所述网络威胁相关的社交网络文本数据划分成若干个顺序固定的单个句子；

[0015] 删除所有的所述单个句子中的特殊符号，保存为删除特殊符号的单个句子；

[0016] 对所有的所述删除特殊符号的单个句子利用WordPiece分词算法进行单词拆分，生成预处理后社交文本数据。

[0017] 在一个或多个实施例中，优选地，所述embedding向量包括token embedding向量、position embedding向量和segment embedding向量。

[0018] 在一个或多个实施例中，优选地，所述根据所述预处理后社交网络文本数据，获取所述预处理后社交网络文本数据的向量表示形式，并保存为embedding向量，具体包括：

[0019] 获取所述预处理后社交文本数据中的所述token embedding向量；

[0020] 在所述token embedding向量中插入特殊分离符，用于分割所述预处理后社交文本数据；

[0021] 在所述token embedding向量中插入特殊分隔符，用于分割所述预处理后社交文本数据中的不同句子；

[0022] 利用所述segment embedding向量进行所述预处理后社交文本数据中的相邻2个句子的向量表示；

[0023] 利用所述position embedding向量进行所述预处理后社交文本数据中的序列位置信息的表示。

[0024] 在一个或多个实施例中，优选地，所述将所述embedding向量输入BERT预训练模型进行训练，获得输出表征向量，具体包括：

[0025] 将所述embedding向量输入所述BERT预训练模型；

[0026] 利用所述BERT预训练模型进行第一步训练，所述第一步训练为通过随机遮蔽掉一个句子中的词，利用上下文进行预测，获得一个双向深度上下文语义信息；

[0027] 利用所述BERT预训练模型进行第二步训练，所述第二步训练为预测一个句子的下一个句子，获得一个句子与句子之间的关系；

[0028] 反复进行预先设定次数的所述第一步训练和所述第二步训练，根据所有的所述双向深度上下文语义信息和所有的所述句子与句子之间的关系生成所述输出表征向量。

[0029] 在一个或多个实施例中，优选地，所述将所述输出表征向量输入前馈神经网络的全连接层和线性分类器，转化为和分类标签向量维度一致的最终向量表示，具体包括：

[0030] 获取所述输出表征向量，输入到所述前馈神经网络的全连接层；

[0031] 通过所述前馈神经网络的全连接层生成第一中间向量；

[0032] 将所述第一中间向量输入到一个线性映射的分类器，生成和标签向量数量维度一致的向量；

[0033] 将所述和标签向量数量维度一致的向量保存为所述最终向量表示。

[0034] 在一个或多个实施例中，优选地，所述根据交叉熵损失函数计算所述最终向量表

示和分类标签向量真实值的交叉熵损失，更新模型参数直至模型收敛，获得威胁情报分类模型，具体包括：

[0035] 根据交叉熵损失函数随机计算获取一个所述最终向量表示和分类标签向量真实值的交叉熵损失，作为初始损失值；

[0036] 进行新损失值运算，所述新损失值运算为根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失，作为新的损失值；

[0037] 反复进行新损失值运算，直到模型收敛或达到预设运算次数时，输出所述威胁情报分类模型，其中，所述模型收敛为在所述新损失值运算过程中连续K次所述新的损失值不大于所述初始损失值的情况，K为用户在建模前预先设定的收敛次数。

[0038] 在一个或多个实施例中，优选地，所述模型参数具体包括：所述线性分类器的参数、所述全连接层的参数和所述BERT预训练模型的参数。

[0039] 在一个或多个实施例中，优选地，所述根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签，获得分类标签结果，具体包括：

[0040] 设置重点关注网络安全列表；

[0041] 根据所述重点关注网络安全列表利用搜索引擎定时对社交网络中安全主体的相关信息进行搜索和采集，获取待分类社交网络文本数据；

[0042] 对所述待分类社交网络文本数据进行预处理，获得获得输入文本的向量表示；

[0043] 利用所述威胁情报分类模型根据所述输入文本的向量表示进行分类结果计算，生成所述分类标签结果。

[0044] 在发明实施例的第二方面，提供了一种面向网络安全威胁情报的智能分类标签系统包括：

[0045] 预处理模块，用于对网络威胁相关的社交网络文本数据进行文本数据预处理，生成预处理后社交网络文本数据；

[0046] 第一向量生成模块，用于根据所述预处理后社交网络文本数据，获取所述预处理后社交网络文本数据的向量表示形式，并保存为embedding向量；

[0047] 第二向量生成模块，用于将所述embedding向量输入BERT预训练模型进行训练，获得输出表征向量；

[0048] 第三向量生成模块，用于将所述输出表征向量输入前馈神经网络的全连接层和线性分类器，转化为和分类标签向量维度一致的最终向量表示；

[0049] 模型训练模块，用于根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失，更新模型参数直至模型收敛，获得威胁情报分类模型；

[0050] 分类标签模块，用于根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签，获得分类标签结果；

[0051] 分类结果推送模块，用于对所述分类标签结果进行信息抽取，并将抽取的分类信息推送给对应用户。

[0052] 本发明的实施例提供的技术方案可以包括以下有益效果：

[0053] 1)本发明采用的预训练模型可以学习到文本数据的双向深度上下文语义信息，以及句子与句子之间的关系，从而获得更高层次的语义表征，因此，所生成的威胁情报分类标

签模型具有更高的准确率。

[0054] 2)本发明使用BERT模型特征表示作为任务的词嵌入特征,来对社交网络中的文本信息进行特征表示,筛选威胁情报相关信息,在BERT的基础上训练分类任务模型,有助于解决传统神经网络在面对实际训练数据不足和训练样本分布不均匀的情况下存在的梯度消失和过拟合问题,同时可以提高训练效率,缩短直接训练时间。

[0055] 本发明的其它特征和优点将在随后的说明书中阐述,并且,部分地从说明书中变得显而易见,或者通过实施本发明而了解。本发明的目的和其他优点可通过在所写的说明书、权利要求书、以及附图中所特别指出的结构来实现和获得。

[0056] 下面通过附图和实施例,对本发明的技术方案做进一步的详细描述。

附图说明

[0057] 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。

[0058] 图1是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法的流程图。

[0059] 图2是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的对网络威胁相关的社交网络文本数据进行文本数据预处理的流程图。

[0060] 图3是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的根据所述预处理后社交网络文本数据获取所述预处理后社交网络文本数据的向量表示形式,并保存为embedding向量的流程图。

[0061] 图4是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的将所述embedding向量输入BERT预训练模型进行训练获得输出表征向量的流程图。

[0062] 图5是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的将所述输出表征向量输入前馈神经网络的全连接层和线性分类器,转化为和分类标签向量维度一致的最终向量表示的流程图。

[0063] 图6是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失,更新模型参数直至模型收敛,威胁情报分类模型的流程图。

[0064] 图7是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的威胁情报分类模型的数据处理关系示意图。

[0065] 图8是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签的流程图。

[0066] 图9是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的威胁情报获取推送的流程图。

[0067] 图10是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签系统的结构图。

具体实施方式

[0068] 在本发明的说明书和权利要求书及上述附图中的描述的一些流程中，包含了按照特定顺序出现的多个操作，但是应该清楚了解，这些操作可以不按照其在本文中出现的顺序来执行或并行执行，操作的序号如101、102等，仅仅是用于区分开各个不同的操作，序号本身不代表任何的执行顺序。另外，这些流程可以包括更多或更少的操作，并且这些操作可以按顺序执行或并行执行。需要说明的是，本文中的“第一”、“第二”等描述，是用于区分不同的消息、设备、模块等，不代表先后顺序，也不限定“第一”和“第二”是不同的类型。

[0069] 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。

[0070] 网络安全威胁情报是与网络空间安全威胁相关的线索和证据。网络安全威胁情报可作为网络安全基础知识使用，包括网络安全的主体、主体涉及场景、机制、指标、影响和可执行的建议等信息。网络安全威胁情报可以及时、有效提高针对网络攻击威胁的监测发现与应急响应能力。

[0071] 社交网络如微博和知识共享网站等公共数据共享平台聚集大量信息，如何从中高效地提取网络安全威胁相关信息，是开展网络安全工作的重要途径。利用机器学习进行文本分类，为网络安全威胁信息的提取提供了可能。

[0072] 自然语言作为一种天生的序列数据，词序对语义的表达具有重要的意义。传统的统计学习方法如支持向量机、朴素贝叶斯、随机森林等，通常会忽略自然语言文本数据中的自然顺序结构以及上下文信息，因此无法学习获得文本的语义信息。深度学习方则法避免了人工提取特征，而是将文本表达为具有相关语义信息的张量形式，以便实现预测、分类和信息提取。

[0073] 在本发明之前，现有技术中主要采用神经网络对网络空间安全威胁的处理。循环神经网络可以产生记忆效应，适合处理自然语言类的序列数据。但是，由于梯度消失和梯度爆炸，当一个遥远序列传递到当前，梯度变得很小时，无法建立长期记忆，导致循环神经网络存在长程依赖的问题。长短时记忆网络是一种特殊的循环神经网络结构，包含输入门、遗忘门和输出门。长短时记忆网络通过门控机制可以建立较长的长距离时序依赖关系，有助于解决自然语言序列中的长程依赖问题。但如果直接使用神经网络对网络威胁相关的自然语言文本进行分类模型学习，通常需要大量的训练数据，由于网络威胁相关数据与不相关数据的不平衡问题，导致用于模型训练的数据量较为匮乏，容易导致学习过程的过拟合。因此，直接使用传统的模型，尚不能很好解决机器学习在威胁情报领域使用时有效训练样本的数据量小，且训练样本分布不均衡而导致的过拟合、模型准确度低的问题。

[0074] 本发明实施例中，提供了一种面向网络安全威胁情报的智能分类标签方法及系统。通过该方案采用预训练模型学习到文本数据的双向深度上下文语义信息，以及句子与句子之间的关系，从而获得更高层次的语义表征，因此所生成的威胁情报分类标签模型具有更高的准确率，有助于解决传统神经网络在面对实际训练数据不足和训练样本分布不均匀的情况下存在的梯度消失和过拟合问题，同时也可以提高训练效率，缩短直接训练时间。

[0075] 在发明实施例的第一方面，提供可一种面向网络安全威胁情报的智能分类标签方

法。

[0076] 图1是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法的流程图。

[0077] 如图1所示,所述面向网络安全威胁情报的智能分类标签方法包括:

[0078] S101、对网络威胁相关的社交网络文本数据进行文本数据预处理,生成预处理后社交网络文本数据;

[0079] S102、根据所述预处理后社交网络文本数据,获取所述预处理后社交网络文本数据的向量表示形式,并保存为embedding向量;

[0080] S103、将所述embedding向量输入BERT预训练模型进行训练,获得输出表征向量;

[0081] 其中,embedding表示嵌入层向量,BERT的全称是Bidirectional Encoder Representation from Transformers,中文意思是基于Transformer的双向编码器表示模型,所述Transformer是一种利用自我注意来计算其输入和输出的表示的模型。

[0082] S104、将所述输出表征向量输入前馈神经网络的全连接层和线性分类器,转化为和分类标签向量维度一致的最终向量表示;

[0083] S105、根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失,更新模型参数直至模型收敛,获得威胁情报分类模型;

[0084] S106、根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签,获得分类标签结果;

[0085] S107、对所述分类标签结果进行信息抽取,并将抽取的分类信息推送给对应用户。

[0086] 本发明实施例中,对传统的神经网络学习模型在威胁情报分类任务中存在的问题,提供一种基于预训练模型迁移学习语义特征来进行社交网络中威胁情报分类标签的方法和装置。具体的,采用BERT预训练模型为基础,训练威胁情报分类器,在训练数据量不足且样本分布不均匀的实际情况下,有助于解决传统学习模型过拟合的问题,提高分类模型的训练效率和准确性。

[0087] 图2是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的对网络威胁相关的社交网络文本数据进行文本数据预处理的流程图。

[0088] 如图2所示,在一个或多个实施例中,优选地,所述对网络威胁相关的社交网络文本数据进行文本数据预处理,生成预处理后社交网络文本数据,具体包括:

[0089] S201、将所述网络威胁相关的社交网络文本数据划分成若干个顺序固定的单个句子;

[0090] S202、删除所有的所述单个句子中的特殊符号,保存为删除特殊符号的单个句子;

[0091] S203、对所有的所述删除特殊符号的单个句子利用WordPiece分词算法进行单词拆分,生成预处理后社交文本数据。

[0092] 具体的,所述WordPiece分词算法是一种基于概率生成的单词拆分方法,所述WordPiece分词算法将所有的所述删除特殊符号的单个句子拆分成预先设定的词表大小,或者或概率增量低于某一阈值。

[0093] 本发明实施例中,首先将文本数据中的每一个句子进行特殊符号的删除以及分词处理。采用WordPiece分词算法拆分单词,拆分粒度位于字符和单词之间,可以更好地表征词根、词缀之间的联系,也可以更好地处理网络空间安全领域新出现的名称或词语。

[0094] 在一个或多个实施例中,优选地,所述embedding向量包括token embedding向量、position embedding向量和segment embedding向量。

[0095] 其中,所述token embedding为对所述预处理后社交文本数据的词汇编码;其中,segment embedding为对所述预处理后社交文本数据中的词语出现的段进行编码;position embedding为对所述预处理后社交文本数据中的词语出现的位置进行编码。

[0096] 图3是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的根据所述预处理后社交网络文本数据获取所述预处理后社交网络文本数据的向量表示形式,并保存为embedding向量的流程图。

[0097] 如图3所示,在一个或多个实施例中,优选地,所述根据所述预处理后社交网络文本数据,获取所述预处理后社交网络文本数据的向量表示形式,并保存为embedding向量,具体包括:

[0098] S301、获取所述预处理后社交文本数据中的所述token embedding向量;

[0099] S302、在所述token embedding向量中插入特殊分离符,用于分割所述预处理后社交文本数据;

[0100] S303、在所述token embedding向量中插入特殊分隔符,用于分割所述预处理后社交文本数据中的不同句子;

[0101] S304、利用所述segment embedding向量进行所述预处理后社交文本数据中的相邻2个句子的向量表示;

[0102] S305、利用所述position embedding进行所述预处理后社交文本数据中的序列位置信息的表示。

[0103] 本发明实施例中,获取到的WordPiece分词生成的token embedding向量,加入两个特殊的符号,具体为特殊分离符和特殊分隔符,用于之后的分类任务。预训练的另一项任务是下一句子预测任务,用来学习句子与句子之间的语义信息。两个句子被简单拼接在一起后送入到模型中。Segment Embedding向量辅助进行两个句子的向量表示,Position Embedding向量则让预训练模型可以学习到序列的位置信息。

[0104] 图4是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的将所述embedding向量输入BERT预训练模型进行训练获得输出表征向量的流程图。

[0105] 如图4所示,在一个或多个实施例中,优选地,所述将所述embedding向量输入BERT预训练模型进行训练,获得输出表征向量,具体包括:

[0106] S401、将所述embedding向量输入所述BERT预训练模型;

[0107] S402、利用所述BERT预训练模型进行第一步训练,所述第一步训练为通过随机遮蔽掉一个句子中的词,利用上下文进行预测,获得一个双向深度上下文语义信息;

[0108] S403、利用所述BERT预训练模型进行第二步训练,所述第二步训练为预测一个句子的下一个句子,获得一个句子与句子之间的关系;

[0109] S404、反复进行预先设定次数的所述第一步训练和所述第二步训练,根据所有的所述双向深度上下文语义信息和所有的所述句子与句子之间的关系生成所述输出表征向量。

[0110] 本发明实施例中,获取到的token embedding向量输入BERT模型,以获得更高层次的语义表征信息。BERT模型的学习过程包括两项任务,一是随机遮蔽掉一个句子中的词,利

用上下文进行预测;二是预测此句子的下一句。通过BERT模型的Transformer网络,可以学习到文本数据的双向深度上下文语义信息,以及句子与句子之间的关系,从而获得更高层次的语义表征。

[0111] 图5是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的将所述输出表征向量输入前馈神经网络的全连接层和线性分类器,转化为和分类标签向量维度一致的最终向量表示的流程图。

[0112] 如图5所示,在一个或多个实施例中,优选地,所述将所述输出表征向量输入前馈神经网络的全连接层和线性分类器,转化为和分类标签向量维度一致的最终向量表示,具体包括:

[0113] S501、获取所述输出表征向量,输入到所述前馈神经网络的全连接层;

[0114] S502、通过所述前馈神经网络的全连接层生成第一中间向量;

[0115] S503、将所述第一中间向量输入到一个线性映射的分类器,生成和标签向量数量维度一致的向量;

[0116] S504、将所述和标签向量数量维度一致的向量保存为所述最终向量表示。

[0117] 本发明实施例中,先经过一个全连接前馈神经网络,之后经过一个线性分类器,该分类器通过一个线性映射,将其转换为和标签数量维度一致的向量,作为后续数据分析的基础。

[0118] 图6是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失,更新模型参数直至模型收敛,威胁情报分类模型的流程图。

[0119] 如图6所示,在一个或多个实施例中,优选地,所述根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失,更新模型参数直至模型收敛,获得威胁情报分类模型,具体包括:

[0120] S601、根据交叉熵损失函数随机计算获取一个所述最终向量表示和分类标签向量真实值的交叉熵损失,作为初始损失值;

[0121] S602、进行新损失值运算,所述新损失值运算为根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失,作为新的损失值;

[0122] S603、反复进行新损失值运算,直到模型收敛或达到预设运算次数时,输出所述威胁情报分类模型,其中,所述模型收敛为在所述新损失值运算过程中连续K次所述新的损失值不大于所述初始损失值的情况,K为用户在建模前预先设定的收敛次数。

[0123] 本发明实施例中,在获得最终的向量表示后,先计算对应的交叉熵损失,进而更新线性分类器、全连接前馈神经网络和BERT预训练模型的参数,获取更低的损失值,实现模型收敛,得到最佳分类效果。

[0124] 图7是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的威胁情报分类模型的数据处理关系示意图。

[0125] 如图7所示,在一个或多个实施例中,优选地,所述模型参数具体包括:所述线性分类器的参数、所述全连接层的参数和所述BERT预训练模型的参数。

[0126] 本发明实施例中,在语义表征向量的基础上,使用前馈神经网络和线性分类器,组合形成一个完整的分类模型,最后得到与分类标签向量维度一样的表征向量。根据分类结

果计算交叉熵损失，通过反向传播修正分类模型的各项参数，其中所述各项参数包括预训练模型参数、前馈神经网络参数和线性分类器参数，从而获得最佳的分类效果。

[0127] 图8是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签的流程图。

[0128] 如图8所示，在一个或多个实施例中，优选地，所述根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签，获得分类标签结果，具体包括：

[0129] S801、设置重点关注网络安全列表；

[0130] S802、根据所述重点关注网络安全列表利用搜索引擎定时对社交网络中安全主体的相关信息进行搜索和采集，获取待分类社交网络文本数据；

[0131] S803、对所述待分类社交网络文本数据进行预处理，获得获得输入文本的向量表示；

[0132] S804、利用所述威胁情报分类模型根据所述输入文本的向量表示进行分类结果计算，生成所述分类标签结果。

[0133] 本发明实施例中，为了帮助用户获取到其重点关注的网络安全主体相关威胁情报，本发明中采取搜索加分类的方法，首先根据用户设定的网络安全主体关键词，定时从社交网络中搜索和采集网络安全主体相关的社交网络文本数据，缩小分类范围；进而，将采集的社交网络文本数据，通过本发明实现的基于BERT预训练模型迁移得到的威胁情报分类模型进行分类标签；最终，将威胁情报相关的社交网络文本数据进行信息抽取，根据标签结果推送给相关用户，帮助用户及时了解其关注的网络安全主体状况。

[0134] 图9是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签方法中的威胁情报获取推送的流程图。

[0135] 本发明实施例中，所述BERT预处理模型是基于Transformer的双向编码器表示模型，BERT模型算法采用Transformer框架，可更好地学习表征单词的双向语义特征。Transformer框架并未采用传统的神经网络网络架构。BERT模型本质上是在海量语料数据的基础上进行无监督学习得到表征单词的嵌入向量，提供了一个为其他学习任务进行迁移学习的基本模型。因此，使用BERT的特征表示作为该任务的词嵌入特征，来对社交网络中的文本信息进行特征表示，筛选威胁情报相关信息，在BERT的基础上训练分类任务模型。提出在此预训练模型的基础上进行分类器训练，能够解决传统机器学习方法在应用到威胁情报分类领域时，由于有效训练样本的数据量小，且训练样本分布不均衡而导致的过拟合、模型准确度低等问题。

[0136] 在发明实施例的第二方面，提供了一种面向网络安全威胁情报的智能分类标签系统。

[0137] 图10是本发明一个实施例的一种面向网络安全威胁情报的智能分类标签系统的结构图。

[0138] 如图10所示，所述面向网络安全威胁情报的智能分类标签系统包括：

[0139] 预处理模块，用于对网络威胁相关的社交网络文本数据进行文本数据预处理，生成预处理后社交网络文本数据；

[0140] 第一向量生成模块1001，用于根据所述预处理后社交网络文本数据，获取所述预处理后社交网络文本数据的向量表示形式，并保存为embedding向量；

[0141] 第二向量生成模块1002，用于将所述embedding向量输入BERT预训练模型进行训练，获得输出表征向量；

[0142] 第三向量生成模块1003，用于将所述输出表征向量输入前馈神经网络的全连接层和线性分类器，转化为和分类标签向量维度一致的最终向量表示；

[0143] 模型训练模块1004，用于根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失，更新模型参数直至模型收敛，获得威胁情报分类模型；

[0144] 分类标签模块1005，用于根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签，获得分类标签结果；

[0145] 分类结果推送模块1006，用于对所述分类标签结果进行信息抽取，并将抽取的分类信息推送给对应用户。

[0146] 本发明实施例中，提供一种基于预训练模型迁移学习语义特征来进行社交网络中威胁情报分类标签的方法。在此基础上，提供了一种面向网络安全威胁情报的智能分类标签系统，该系统采用的预训练模型可以学习到文本数据的双向深度上下文语义信息，以及句子与句子之间的关系，从而获得更高层次的语义表征，因此所生成的威胁情报分类标签模型具有更高的准确率，有助于解决传统神经网络在面对实际训练数据不足和训练样本分布不均匀的情况下存在的梯度消失和过拟合问题，同时，可提高训练效率，缩短直接训练时间。

[0147] 本发明的实施例提供的技术方案可以包括以下有益效果：

[0148] 1)本发明采用的预训练模型可以学习到文本数据的双向深度上下文语义信息，以及句子与句子之间的关系，从而获得更高层次的语义表征，因此，所生成的威胁情报分类标签模型具有更高的准确率。

[0149] 2)本发明使用BERT模型特征表示作为任务的词嵌入特征，来对社交网络中的文本信息进行特征表示，筛选威胁情报相关信息，在BERT的基础上训练分类任务模型，有助于解决传统神经网络在面对实际训练数据不足和训练样本分布不均匀的情况下存在的梯度消失和过拟合问题，同时可以提高训练效率，缩短直接训练时间。

[0150] 本领域内的技术人员应明白，本发明的实施例可提供为方法、系统、或计算机程序产品。因此，本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器和光学存储器等)上实施的计算机程序产品的形式。

[0151] 本发明是参照根据本发明实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0152] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特

定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0153] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0154] 显然，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样，倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

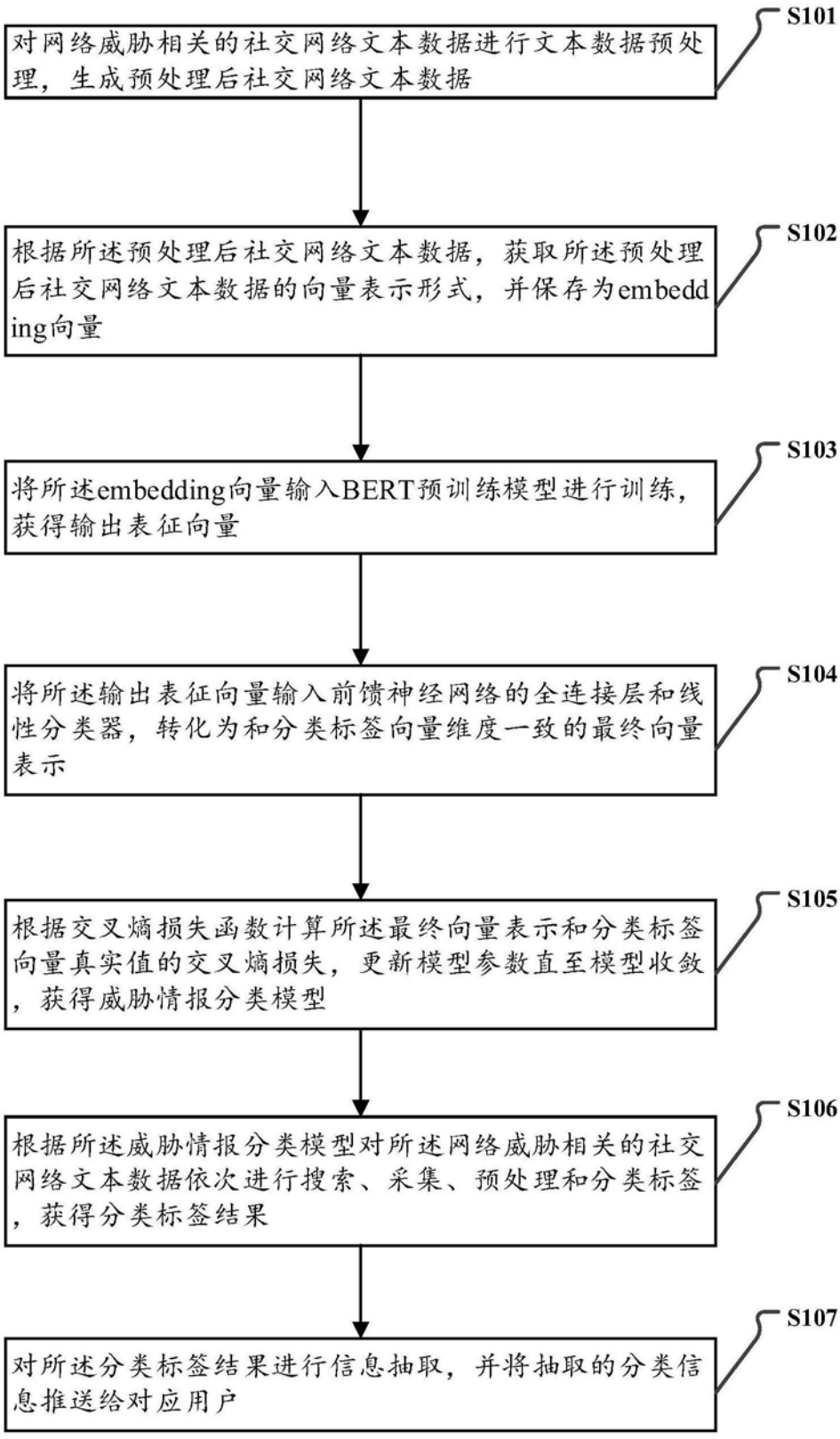

S101
对网络威胁相关的社交网络文本数据进行文本数据预处理，生成预处理后社交网络文本数据
S102
根据所述预处理后社交网络文本数据，获取所述预处理后社交网络文本数据的向量表示形式，并保存为embedding向量
S103
将所述embedding向量输入BERT预训练模型进行训练，获得输出表征向量
S104
将所述输出表征向量输入前馈神经网络的全连接层和线性分类器，转化为和分类标签向量维度一致的最终向量表示
S105
根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失，更新模型参数直至模型收敛，获得威胁情报分类模型
S106
根据所述威胁情报分类模型对所述网络威胁相关的社交网络文本数据依次进行搜索、采集、预处理和分类标签，获得分类标签结果
S107
对所述分类标签结果进行信息抽取，并将抽取的分类信息推送给对应用户

图1

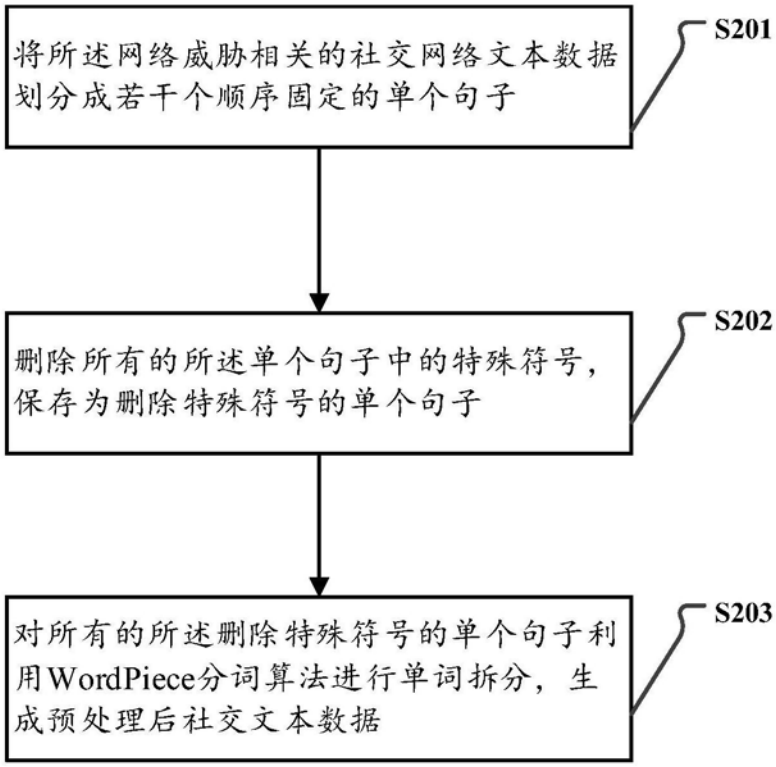

将所述网络威胁相关的社交网络文本数据划分成若干个顺序固定的单个句子
S201
删除所有的所述单个句子中的特殊符号，保存为删除特殊符号的单个句子
S202
对所有的所述删除特殊符号的单个句子利用WordPiece分词算法进行单词拆分，生成预处理后社交文本数据
S203

图2

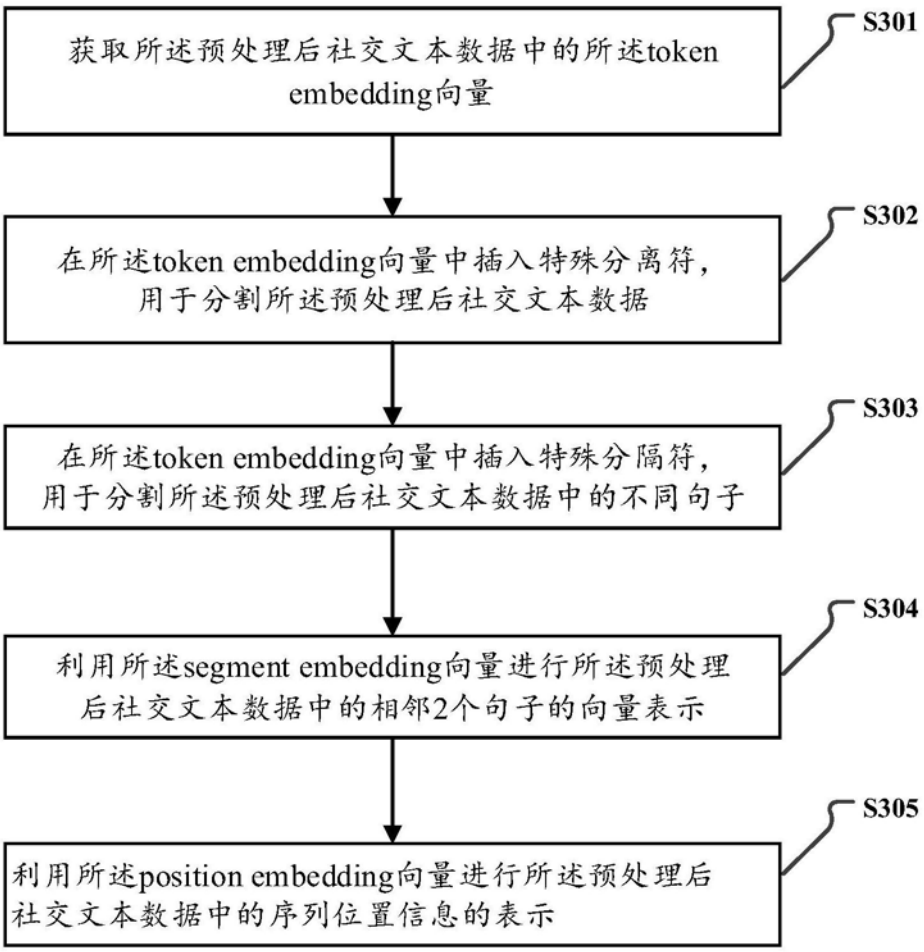

获取所述预处理后社交文本数据中的所述token embedding向量
S301
在所述token embedding向量中插入特殊分离符，用于分割所述预处理后社交文本数据
S302
在所述token embedding向量中插入特殊分隔符，用于分割所述预处理后社交文本数据中的不同句子
S303
利用所述segment embedding向量进行所述预处理后社交文本数据中的相邻2个句子的向量表示
S304
利用所述position embedding向量进行所述预处理后社交文本数据中的序列位置信息的表示
S305

图3

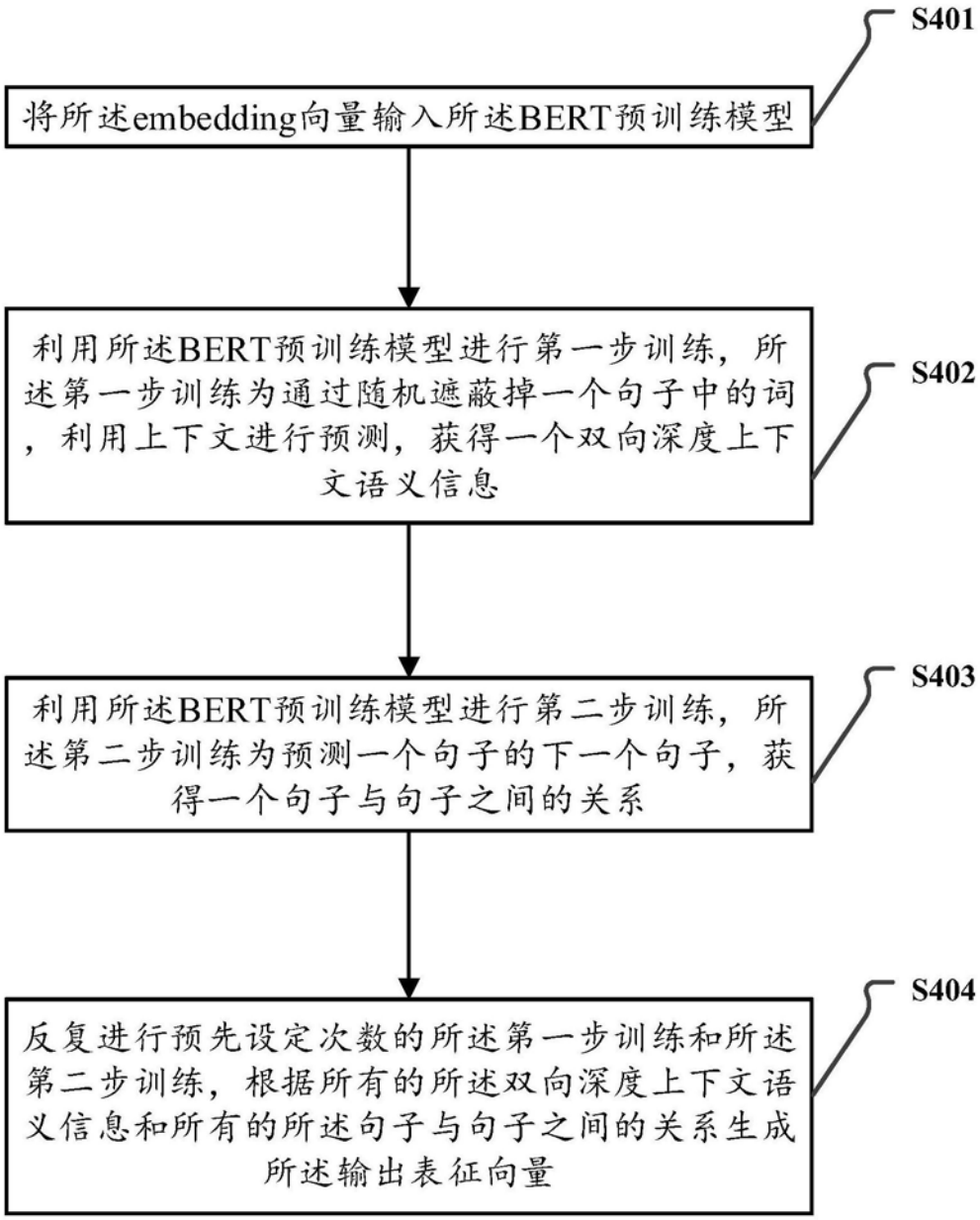

S401
将所述embedding向量输入所述BERT预训练模型
利用所述BERT预训练模型进行第一步训练，所述第一步训练为通过随机遮蔽掉一个句子中的词，利用上下文进行预测，获得一个双向深度上下文语义信息
S402
利用所述BERT预训练模型进行第二步训练，所述第二步训练为预测一个句子的下一个句子，获得一个句子与句子之间的关系
S403
反复进行预先设定次数的所述第一步训练和所述第二步训练，根据所有的所述双向深度上下文语义信息和所有的所述句子与句子之间的关系生成所述输出表征向量
S404

图4

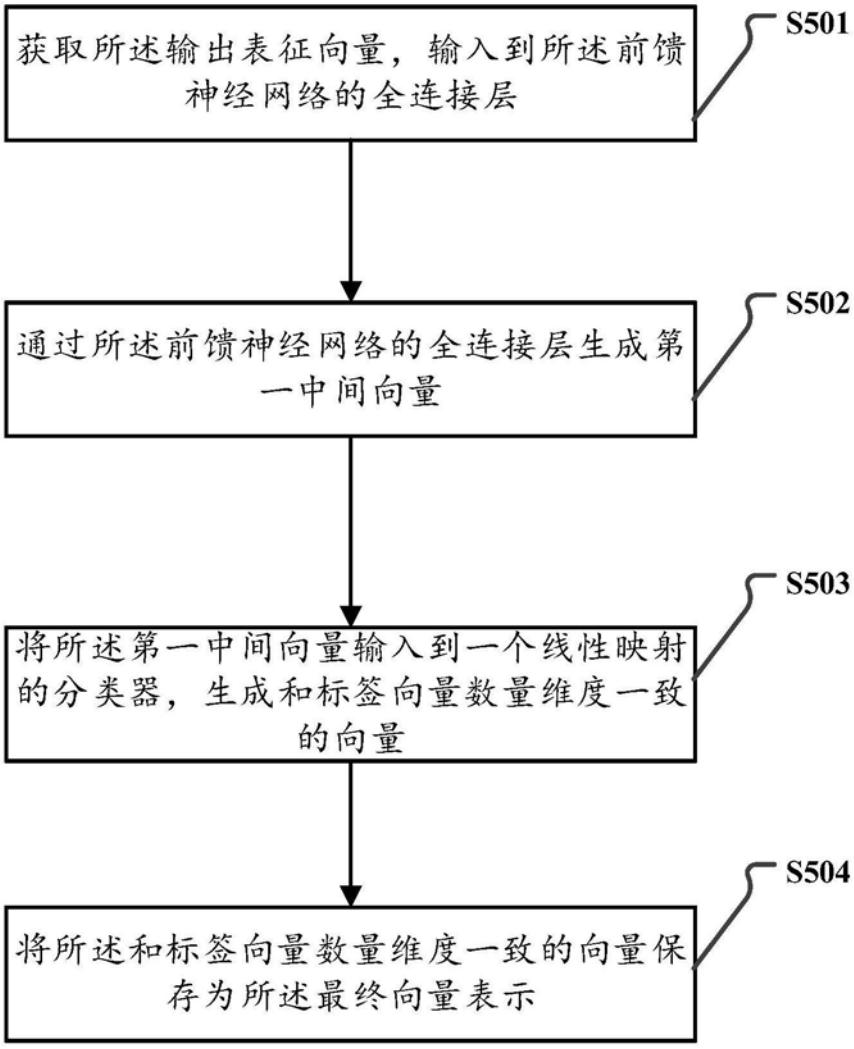

获取所述输出表征向量，输入到所述前馈神经网络的全连接层
S501
通过所述前馈神经网络的全连接层生成第一中间向量
S502
将所述第一中间向量输入到一个线性映射的分类器，生成和标签向量数量维度一致的向量
S503
将所述和标签向量数量维度一致的向量保存为所述最终向量表示
S504

图5

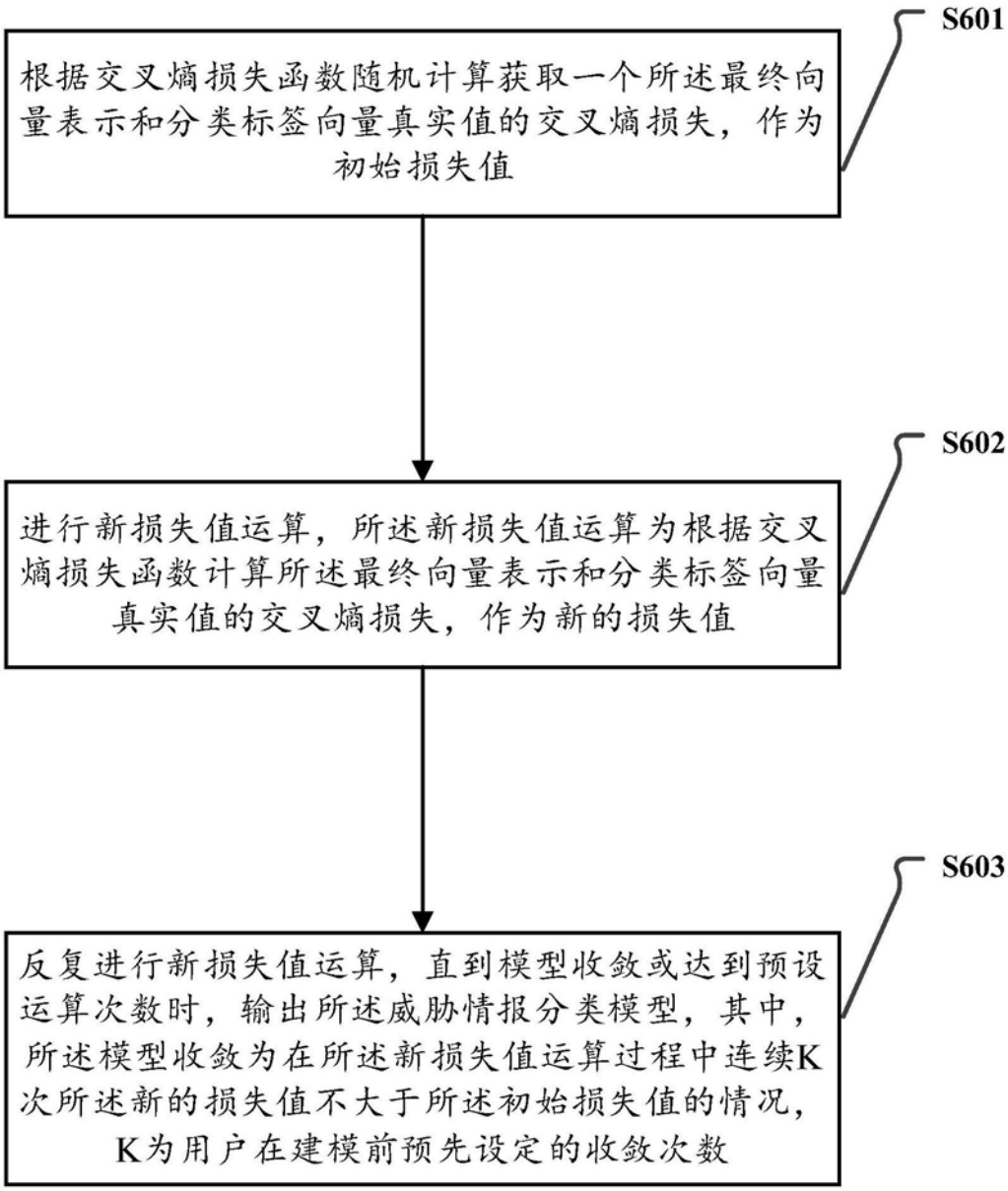

S601
根据交叉熵损失函数随机计算获取一个所述最终向量表示和分类标签向量真实值的交叉熵损失，作为初始损失值
S602
进行新损失值运算，所述新损失值运算为根据交叉熵损失函数计算所述最终向量表示和分类标签向量真实值的交叉熵损失，作为新的损失值
S603
反复进行新损失值运算，直到模型收敛或达到预设运算次数时，输出所述威胁情报分类模型，其中，所述模型收敛为在所述新损失值运算过程中连续K次所述新的损失值不大于所述初始损失值的情况，K为用户在建模前预先设定的收敛次数

图6

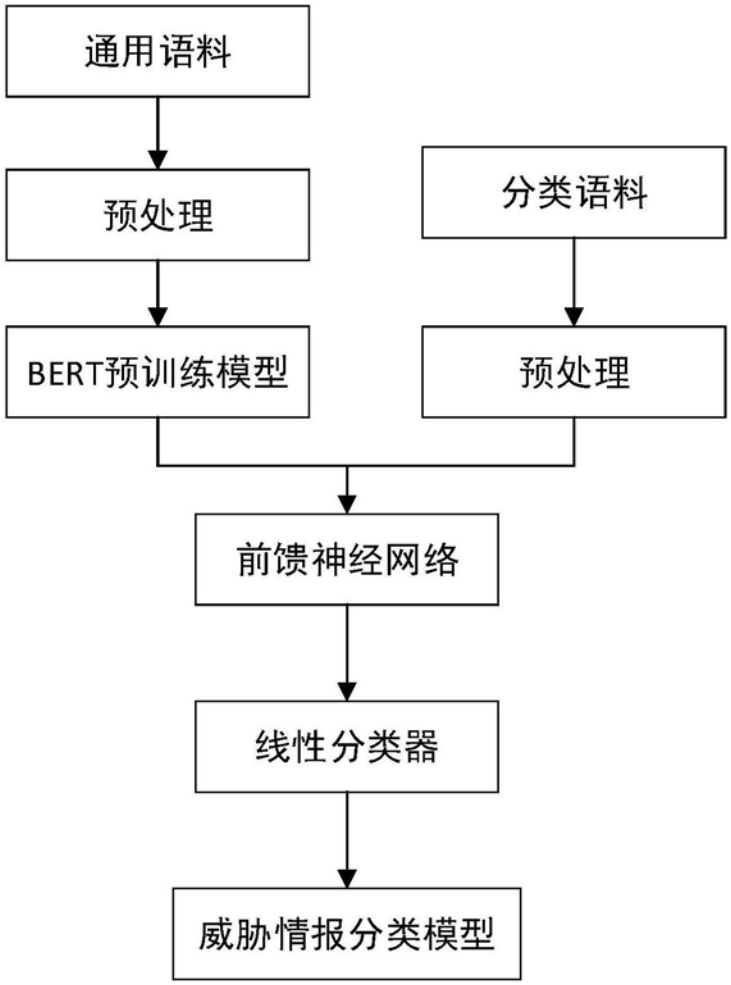


图7

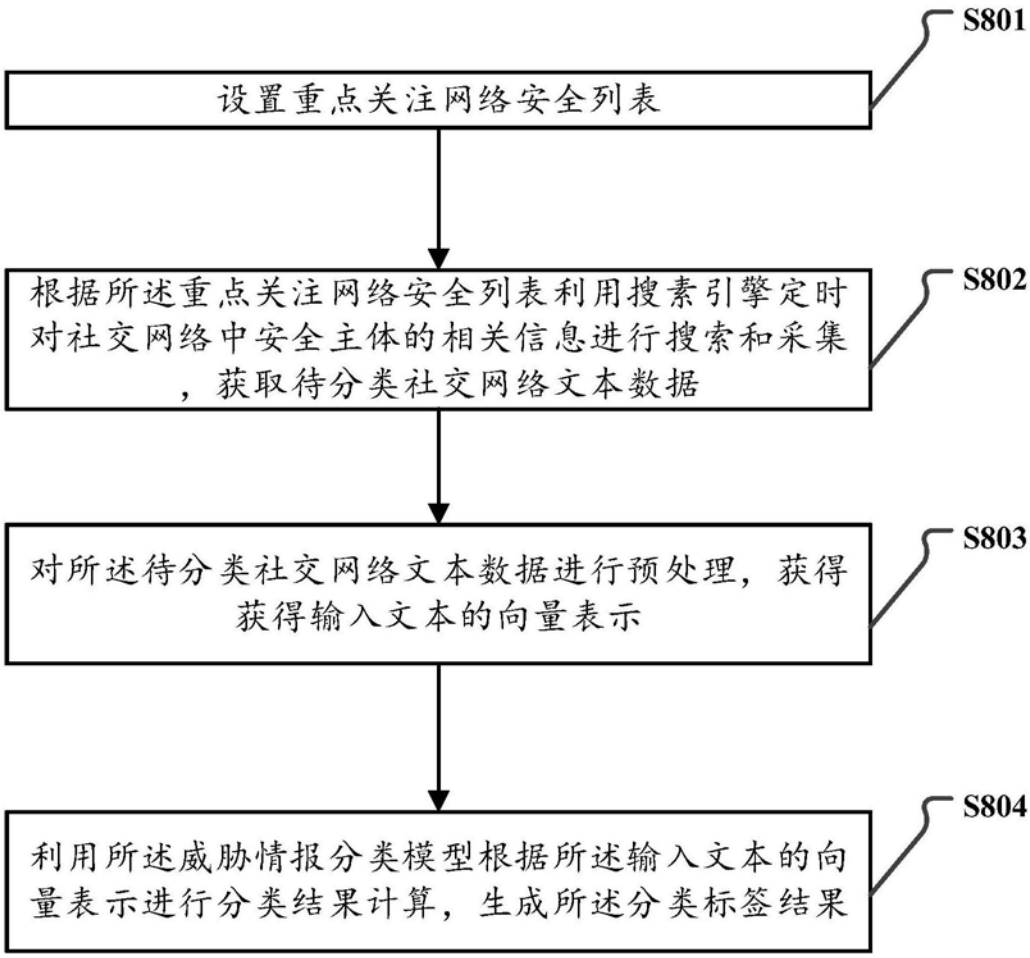

S801
设置重点关注网络安全列表
S802
根据所述重点关注网络安全列表利用搜索引擎定时对社交网络中安全主体的相关信息进行搜索和采集，获取待分类社交网络文本数据
S803
对所述待分类社交网络文本数据进行预处理，获得获得输入文本的向量表示
S804
利用所述威胁情报分类模型根据所述输入文本的向量表示进行分类结果计算，生成所述分类标签结果

图8

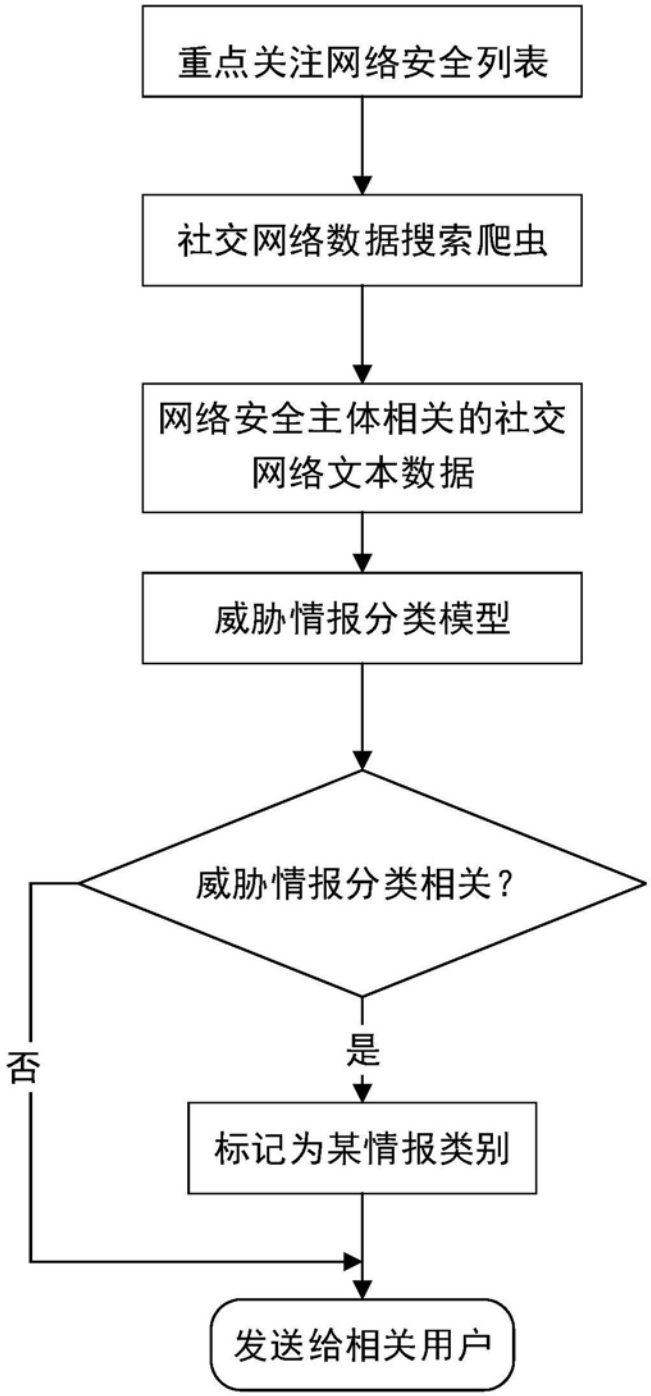

重点关注网络安全列表
社交网络数据搜索爬虫
网络安全主体相关的社交网络文本数据
威胁情报分类模型
威胁情报分类相关？
否
是
标记为某情报类别
发送给相关用户

图9

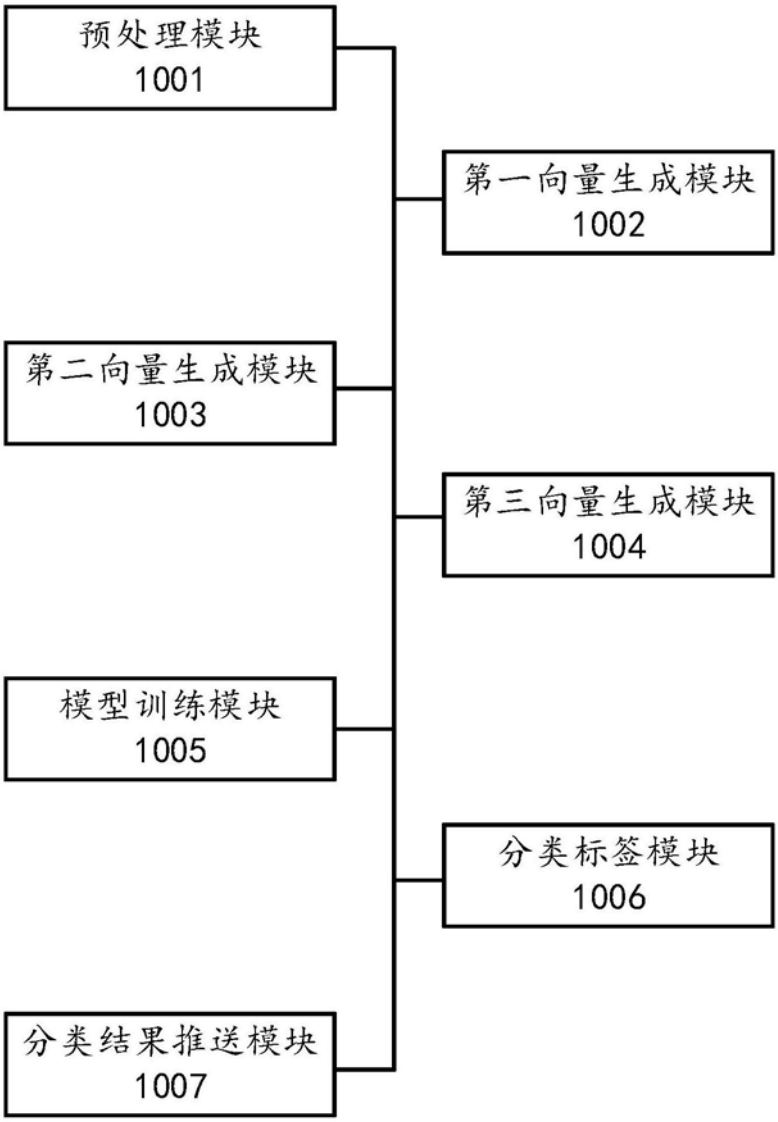

预处理模块
1001
第一向量生成模块
1002
第二向量生成模块
1003
第三向量生成模块
1004
模型训练模块
1005
分类标签模块
1006
分类结果推送模块
1007

图10