



(12)发明专利申请

(10)申请公布号 CN 108182279 A

(43)申请公布日 2018.06.19

(21)申请号 201810077890.4

(22)申请日 2018.01.26

(71)申请人 有米科技股份有限公司

地址 510000 广东省广州市番禺区小谷围街外环东路232号13栋B232

(72)发明人 王秋文 李百川 陈第

(74)专利代理机构 广州华进联合专利商标代理有限公司 44224

代理人 余永文

(51) Int. Cl.

G06F 17/30(2006.01)

G06K 9/62(2006.01)

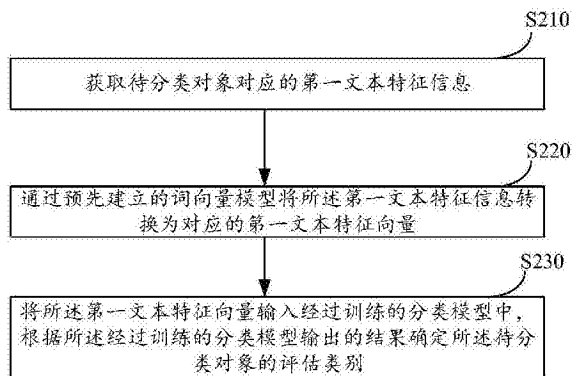
权利要求书2页 说明书14页 附图4页

(54)发明名称

基于文本特征的对象分类方法、装置和计算机设备

(57)摘要

本发明涉及基于文本特征的对象分类方法、装置和计算机设备,属于网络技术领域。所述方法包括:获取待分类对象对应的第一文本特征信息;通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量;将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。上述技术方案,解决了对文本对象分析时分类模型不够准确的问题,能对文本对象进行准确归类。



1. 一种基于文本特征的对象分类方法,其特征在于,包括以下步骤:

获取待分类对象对应的第一文本特征信息;

通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量;

将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。

2. 根据权利要求1所述的基于文本特征的对象分类方法,其特征在于,所述将所述第一文本特征向量输入经过训练的分类模型中的步骤之前,还包括:

获取多个参考对象对应的第二文本特征向量;分别对所述参考对象的实际类别进行标注;

通过各个参考对象对应的第二文本特征向量以及实际类别训练预先建立的分类模型,得到经过训练的分类模型。

3. 根据权利要求2所述的基于文本特征的对象分类方法,其特征在于,所述分类模型包括至少一个二分类子模型,每个二分类子模型分别对应一个评估类别;

所述通过各个参考对象对应的第二文本特征向量以及实际类别训练预先建立的分类模型的步骤,包括:

将某一所述第二文本特征向量分别输入各个二分类子模型中,分别得到所述第二文本特征向量与对应的评估类别的匹配度;

根据所述匹配度确定参考对象的评估类别;

将参考对象的评估类别与对应的实际类别进行比对,根据比对结果调整所述分类模型。

4. 根据权利要求3所述的基于文本特征的对象分类方法,其特征在于,所述根据所述匹配度确定参考对象的评估类别的步骤,包括:

确定所述匹配度中的最高匹配度值,获取与所述最高匹配度值对应的评估类别,作为对应的待分类对象的评估类别。

5. 根据权利要求1至4任一所述的基于文本特征的对象分类方法,其特征在于,所述通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量的步骤之前,还包括:

从预设的文本信息库中确定特征词的上下文信息,通过one hot工具确定所述特征词的词向量;

根据所述词向量确定所述上下文信息出现的条件概率;

根据所述条件概率和所述上下文信息建立词向量模型。

6. 根据权利要求5所述的基于文本特征的对象分类方法,其特征在于,所述第一文本特征信息中包括至少一个特征词;

所述通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量的步骤,包括:

通过预先建立的词向量模型将所述第一文本特征信息中的各个特征词转换为对应的特征词向量,根据各个所述特征词向量确定所述待分类对象对应的第一文本特征向量。

7. 根据权利要求1、2、3、4或6所述的基于文本特征的对象分类方法,其特征在于,所述

获取待分类对象对应的第一文本特征信息的步骤,包括:

通过网络爬虫工具获取待分类对象对应的ID、昵称、简介、经营范围、帐号主体和/或推送消息,从中获取待分类对象对应的第一文本特征信息。

8. 一种基于文本特征的对象分类装置,其特征在于,包括:

信息获取模块,用于获取待分类对象对应的第一文本特征信息;

向量转换模块,用于通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量;

以及,分类模块,用于将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。

9. 一种计算机设备,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计算机程序时实现权利要求1至7任一项所述的方法的步骤。

10. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至7任一项所述的方法的步骤。

基于文本特征的对象分类方法、装置和计算机设备

技术领域

[0001] 本发明涉及网络技术领域,特别是涉及基于文本特征的对象分类方法、装置、计算机设备和存储介质。

背景技术

[0002] 分类是一种重要的数据挖掘技术。分类的目的是根据数据集的特点把未知类别的样本映射到给定类别中的某一个。现有的对文本进行分类的方法主要有人工分类法和模型文本法,人工分类法利用人的自有知识对信息进行分类,而模型分类法通过相似度模型、概率模型、线性模型、非线性模型以及组合模型等模型对信息分类。在实现本发明过程中,发明人发现现有技术中至少存在如下问题:人工文本分类尽管利用人工分类,基于已有知识、常识分类,能够保证准确性,但对于微信公众号等类别众多的文本,分类效率低下,后期分类容易产生偏差以及误判;而对于模型分类法,模型各有利弊,针对不同领域有不同效果。因此,有必要找到一种合适的能对文本对象进行准确归类的方法。

发明内容

[0003] 基于此,本发明提供了基于文本特征的对象分类方法、装置、计算机设备和存储介质,能对文本对象进行准确归类。

[0004] 本发明实施例的内容如下:

[0005] 一种基于文本特征的对象分类方法,包括以下步骤:获取待分类对象对应的第一文本特征信息;通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量;将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。

[0006] 在其中一个实施例中,所述将所述第一文本特征向量输入经过训练的分类模型中的步骤之前,还包括:获取多个参考对象对应的第二文本特征向量;分别对所述参考对象的实际类别进行标注;通过各个参考对象对应的第二文本特征向量以及实际类别训练预先建立的分类模型,得到经过训练的分类模型。

[0007] 在其中一个实施例中,所述分类模型包括至少一个二分类子模型,每个二分类子模型分别对应一个评估类别;所述通过各个参考对象对应的第二文本特征向量以及实际类别训练预先建立的分类模型的步骤,包括:将某一所述第二文本特征向量分别输入各个二分类子模型中,分别得到所述第二文本特征向量与对应的评估类别的匹配度;根据所述匹配度确定参考对象的评估类别;将参考对象的评估类别与对应的实际类别进行比对,根据比对结果调整所述分类模型。

[0008] 在其中一个实施例中,所述根据所述匹配度确定参考对象的评估类别的步骤,包括:确定所述匹配度中的最高匹配度值,获取与所述最高匹配度值对应的评估类别,作为对应的待分类对象的评估类别。

[0009] 在其中一个实施例中,所述通过预先建立的词向量模型将所述第一文本特征信息

转换为对应的第一文本特征向量的步骤之前,还包括:从预设的文本信息库中确定特征词的上下文信息,通过one hot工具确定所述特征词的词向量;根据所述词向量确定所述上下文信息出现的条件概率;根据所述条件概率和所述上下文信息建立词向量模型。

[0010] 在其中一个实施例中,所述第一文本特征信息中包括至少一个特征词;所述通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量的步骤,包括:通过预先建立的词向量模型将所述第一文本特征信息中的各个特征词转换为对应的特征词向量,根据各个所述特征词向量确定所述待分类对象对应的第一文本特征向量。

[0011] 在其中一个实施例中,所述获取待分类对象对应的第一文本特征信息的步骤,包括:通过网络爬虫工具获取待分类对象对应的ID、昵称、简介、经营范围、帐号主体和/或推送消息,从中获取待分类对象对应的第一文本特征信息。

[0012] 相应的,本发明实施例提供一种基于文本特征的对象分类装置,包括:信息获取模块,用于获取待分类对象对应的第一文本特征信息;向量转换模块,用于通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量;以及,分类模块,用于将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。

[0013] 上述基于文本特征的对象分类方法和装置,首先获取待分类对象对应的第一文本特征信息;通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量;将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。能根据预先训练好的模型对待分类对象进行准确的分类,进而根据得到的分类信息对该待分类对象进行针对性的操作,可以有效防止对各种类别的对象均进行操作而造成资源的浪费。

[0014] 一种计算机设备,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计算机程序时实现以下步骤:获取待分类对象对应的第一文本特征信息;通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量;将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。

[0015] 上述计算机设备,能根据预先训练好的模型对待分类对象进行准确的分类,进而根据得到的分类信息对该待分类对象进行针对性的操作,可以有效防止对各种类别的对象均进行操作而造成资源的浪费。

[0016] 一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现以下步骤:获取待分类对象对应的第一文本特征信息;通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量;将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。

[0017] 上述计算机可读存储介质,能根据预先训练好的模型对待分类对象进行准确的分类,进而根据得到的分类信息对该待分类对象进行针对性的操作,可以有效防止对各种类别的对象均进行操作而造成资源的浪费。

附图说明

- [0018] 图1为一个实施例中基于文本特征的对象分类方法的应用环境图；
- [0019] 图2为一个实施例中基于文本特征的对象分类方法的流程示意图；
- [0020] 图3为另一个实施例中基于文本特征的对象分类方法的流程示意图；
- [0021] 图4为一个实施例中基于文本特征的对象分类方法的具体应用实例图；
- [0022] 图5为一个实施例中基于文本特征的对象分类装置的结构框图；
- [0023] 图6为一个实施例中计算机设备的内部结构。

具体实施方式

[0024] 为了使本发明的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本发明进行进一步详细说明。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。

[0025] 本发明实施例以微信公众号为例进行描述，但是本发明实施例的基于文本特征的对象分类方法还可以应用于其他的需要对对象进行分类的应用场景中。

[0026] 微信平台提供了公众号服务，将受众范围确定为整个微信用户群体，极大地扩大了宣传的范围，给广告主提供了新的广告推广渠道。但是公众号数量庞大且领域广泛，对合适的公众号的筛选是营销活动中最重要、最繁重的部分。广告主选择的依据是日常获取的信息以及规则搜索，分类信息就成了筛选过程的一个重要成分。

[0027] 目前，对公众号进行分类的方法主要有手工文本分类和模型文本分类等。手工文本分类是利用人的自有知识来对公众号分类。基于已有知识、常识分类的人工文本分类准确度有保障，但由于公众号众多，容易受主观意识以及精神体力的影响，其分类效率低下，且后期分类可能会存在偏差以及误判。而模型文本分类是根据相似度模型、概率模型、线性模型、非线性模型和组合模型等对文本进行分类的方法。但是各种模型各有利弊，针对不同领域不同模型有不同的效果；很多模型不适用于对公众号进行分类。如基于LDA主题聚类对公众号进行分类的方法，该方法是通过LDA提取主题后再聚类，该方法存在着几个弊端：对离群点敏感、局部最优非全局最优导致结果不稳定、可解释性不强、对相似度较高的类区分能力不足等。因此，本发明实施例提供一种基于文本特征的对象分类方法，能通过合适的模型对文本对象进行准确归类。

[0028] 本申请实施例提供的基于文本特征的对象分类方法，可以应用于如图1所示的应用环境中。其中，服务器110之间通过网络进行通信，某一服务器调用某一对待分类对象对应的服务器的接口，从中获取待分类对象对应的信息，进而实现对该待分类对象的分类。服务器110可以用独立的服务器或者是多个服务器组成的服务器集群来实现。服务器110还可以替换为各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备等终端，服务器对终端的某些相关信息进行分析并对该信息对应的对象进行分类。

[0029] 如图2所示，本发明实施例提供一种基于文本特征的对象分类方法，包括以下步骤：

[0030] S210、获取待分类对象对应的第一文本特征信息。

[0031] 其中，待分类对象指的是需要进行分类的对象，可以是在精准营销过程中的营销对象，如：公众号、网站、应用等等。本发明实施例对待分类对象的具体形式不做限制，该待分类对象中包含有文本并能通过该文本实现分类即可。

[0032] 此外,第一文本特征信息为待分类对象提供的文本(可以为某个词、语料或者由字符组成的文段等)以及与该文本相关的信息,如某一微信公众号的简介、推送消息等。第一文本特征信息还可以为对待分类对象提供的信息进行处理后得到的具有代表性的文本以及与这些文本相关的信息。通过该第一文本特征信息可以确定该待分类对象的相关信息,进而可以确定该待分类对象的所属类别。

[0033] S220、通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量。

[0034] 在本步骤中借助词向量模型将文本特征信息进行量化,将其转化为第一文本特征向量。

[0035] 其中,词向量模型是用于对第一文本特征信息进行处理使其能够符合某种规则的模型。

[0036] 本发明实施例对文本特征向量中的数值位数以及向量的维度不做限制。

[0037] S230、将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。

[0038] 其中,评估类别指的是待分类对象可能的类别,如:某一微信公众号的评估类别可以是“美食”“搞笑”“影视”“阅读”等。本发明实施例对评估类别的数量不做限制,该评估类别的数量还可以根据实际情况进行调整。

[0039] 其中,分类模型可以是Logistic分类器、softmax分类器、SVM支持向量机等,也可以为其他的分类模型。

[0040] 本步骤通过训练好的分类模型对第一文本特征向量进行分析并得到分类结果,进而确定待分类对象的评估类别。

[0041] 本实施例能根据预先训练好的模型对待分类对象进行准确的分类,进而根据得到的分类信息对该待分类对象进行针对性的操作,可以有效防止对各种类别的对象均进行操作而造成资源的浪费。

[0042] 在一个实施例中,所述将所述第一文本特征向量输入经过训练的分类模型中的步骤之前,还包括:获取多个参考对象对应的第二文本特征向量;分别对所述参考对象的实际类别进行标注;通过各个参考对象对应的第二文本特征向量以及实际类别训练预先建立的分类模型,得到经过训练的分类模型。

[0043] 其中,参考对象是待分类对象进行参考的对象,也即用于对分类模型进行训练的对象。参考对象和待分类对象可以是同样的形式,如两者都是微信公众号;也可以是不同的形式,如第一文本特征向量是微信公众号,而第二文本特征向量是与该微信公众号的账号主体对应的网站。根据该参考对象的第二文本特征向量可以对分类模型进行训练,经过训练的分类模型能实现对待分类对象的分类。

[0044] 其中,第二文本特征向量与第一文本特征向量的格式对应一致,是对分类模型进行训练时用到的向量。

[0045] 其中,实际类别可以是人工对参考对象进行分析以后得出的分类结果,也可以是结合一定算法得到的分类结果。这些实际类别可以作为模型训练过程的参考。

[0046] 本实施例通过多个参考对象对应的特征向量以及实际类别来训练分类模型,这些参考对象可以有效地表征待分类对象的信息,通过经过训练的分类模型能实现对待分类对

象的准确分类。

[0047] 在一个实施例中,所述分类模型包括至少一个二分类子模型,每个二分类子模型分别对应一个评估类别;所述通过各个参考对象对应的第二文本特征向量以及实际类别训练预先建立的分类模型的步骤,包括:将某一所述第二文本特征向量分别输入各个二分类子模型中,分别得到所述第二文本特征向量与对应的评估类别的匹配度;根据所述匹配度确定参考对象的评估类别;将参考对象的评估类别与对应的实际类别进行比对,根据比对结果调整所述分类模型。

[0048] 可选地,二分类子模型可以是一个、两个或者多个。本发明实施例对二分类子模型的个数不做限制。

[0049] 可选地,本实施例的具体过程可以为:分类模型 $F(x)$ 中包含有三个二分类子模型 z_1 、 z_2 和 z_3 , z_1 、 z_2 和 z_3 分别是“搞笑”“影视”和“美食”对应的二分类器。当将某一第二文本特征向量分别输入到 z_1 、 z_2 和 z_3 中时,这些二分类子模型分别计算该第二文本特征向量与“搞笑”“影视”和“美食”这些类别的匹配度,得到匹配度结果为 $[0.2, 0.3, 0.9]$ 。根据该匹配度结果确定参考对象的评估类别,如“美食”;将参考对象的评估类别与对应的实际类别进行比对,根据比对结果调整所述分类模型:若第二文本特征向量对应的参考对象的实际类别为“影视”,则分类模型得到的分类结果不准确,对分类模型进行调整;若第二文本特征向量对应的参考对象的实际类别为“美食”,则分类模型得到的分类结果准确。

[0050] 可选地,根据比对结果调整所述分类模型的步骤还可以是:确定各个比对结果的准确率,当准确率低于某一阈值时,对分类模型进行调整;若准确率高于一阈值时,完成分类模型的训练过程。

[0051] 可选地,分类模型为SVM支持向量机模型,建立二分类子模型的过程可以为:

[0052] 对于任意公众号 i ,其评估类别表示为: y_i ,文本特征向量表示为: \vec{x}_i ,则有总体大小为 n 的训练集: $(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n)$,模型计算过程如下:

[0053] 首先,假设数据线性可分,则存在可以区分两类数据的超平面,超平面由方程簇表示: $\vec{w} \cdot \vec{x} - b = 1$ 或 $\vec{w} \cdot \vec{x} - b = -1$,其中, \vec{w} 为法向量; b 为截距。

[0054] 两个超平面的距离是: $\frac{2}{\|\vec{w}\|}$,要使两平面间的距离最大,即最小化 $\|\vec{w}\|$ 。

[0055] 为了使样本点都在超平面的间隔区外,对于所有的 i ,需满足以下其中一个条件:

[0056] $\vec{w} \cdot \vec{x}_i - b \geq 1$, if $y_i = 1$;

[0057] 或

[0058] $\vec{w} \cdot \vec{x}_i - b \leq -1$, if $y_i = -1$;

[0059] 上述两式子可合并为: $y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1$, for all $1 \leq i \leq n$

[0060] 因此,距离优化问题可以转化为:对于 $i = 1, \dots, n$,在 $y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1$ 的条件

下,求最小化 $\|\vec{w}\|$ 。

[0061] 其次,考虑数据线性不可分,引入铰链损失函数: $\max(0,1-y_i(\vec{w}\cdot\vec{x}_i-b))$

[0062] 则距离优化问题可以转化为:

$$[0063] \left[\frac{1}{n} \sum_{i=1}^n \max(0,1-y_i(\vec{w}\cdot\vec{x}_i-b)) \right] + \lambda \|\vec{w}\|^2$$

[0064] 引入变量: $\zeta_i = \max(0,1-y_i(\vec{w}\cdot\vec{x}_i-b))$,因此上式可以改写为目标函数可微的约束优化问题:

$$[0065] \text{ minimize } \frac{1}{n} \sum_{i=1}^n \zeta_i + \lambda \|\vec{w}\|^2$$

[0066] 其中, λ 是调节间隔大小, $\lambda \|\vec{w}\|$ 可以给模型增加“软间隔”(soft margin),这样可以允许部分训练集出错(正负样本区重叠);对于所有i的取值, $y_i(\vec{w}\cdot\vec{x}_i-b) \geq 1-\zeta_i$, $\zeta_i \geq 0$ 。

[0067] 拉格朗日对偶简化后,得到:

$$[0068] \text{ maximize } f(c_1, \dots, c_n) = \sum_{i=1}^n c_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i c_i (x_i \cdot x_j) y_j c_j$$

[0069] 其中, c_i 为拉格朗日乘子;

$$[0070] \text{ 对于所有i的取值, } \sum_{i=1}^n c_i y_i = 0, \quad 0 \leq c_i \leq \frac{1}{2n\lambda}。$$

[0071] 根据上式可得 \vec{w} 、b:

$$[0072] \vec{w} = \sum_{i=1}^n c_i y_i \vec{x}_i$$

$$[0073] b = \vec{w} \cdot \vec{x}_i - y_i$$

[0074] 假设变换后的数据点为 $\varphi(\vec{x}_i)$,存在一核函数 $k: k(\vec{x}_i, \vec{x}_j) = \varphi(\vec{x}_i) \cdot \varphi(\vec{x}_j)$,则 \vec{w} 满足:

$$[0075] \vec{w} = \sum_{i=1}^n c_i y_i \varphi(\vec{x}_i)$$

[0076] 优化问题求解 c_i :

$$[0077] \text{ maximize } f(c_1, \dots, c_n) = \sum_{i=1}^n c_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i c_i k(\vec{x}_i, \vec{x}_j) y_j c_j$$

$$[0078] \text{ 其中,对于所有i的取值, } \sum_{i=1}^n c_i y_i = 0, \quad 0 \leq c_i \leq \frac{1}{2n\lambda};$$

[0079] 求解b可得:

$$[0080] b = y_i - \vec{w} \cdot \varphi(\vec{x}_i) = y_i - \left[\sum_{k=1}^n c_k y_k k(\vec{x}_k, \vec{x}_i) \right]$$

[0081] 则分类函数 Z ;

$$[0082] \quad \vec{z} = \text{sgn}(\vec{w} \cdot \varphi(\vec{z}) - b) = \text{sgn}\left(\left[\sum_{i=1}^n c_i y_i k(x_i, \vec{z})\right] - b\right)$$

[0083] 该分类函数 \vec{z} 即为二分类子模型,多个二分类子模型即构成分类模型。

[0084] 本实施例中的分类模型包括至少一个二分类子模型,利用“一对多的方法”构建多分类模型 $F(x)$,多分类模型 $F(x)$ 的分类结果是根据各个二分类子模型的分类结果得到的。这样的方式能有效降低模型的复杂度,进而可以提高分类效率。

[0085] 在一个实施例中,所述根据所述匹配度确定参考对象的评估类别的步骤,包括:确定所述匹配度中的最高匹配度值,获取与所述最高匹配度值对应的评估类别,作为对应的待分类对象的评估类别。

[0086] 具体过程举例如下:从匹配度结果[0.2、0.3、0.9](该匹配度结果中的三个维度分别对应三个二分类子模型)中确定其中的最高匹配度值为0.9,若0.9对应的评估类别为“美食”,则确定对应的待分类对象的评估类别为“美食”。

[0087] 本实施例根据最高匹配度值确定待分类对象的评估类别,能方便直接地根据二分类子模型的结果确定待分类对象的评估类别。

[0088] 在一个实施例中,所述通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量的步骤之前,还包括:从预设的文本信息库中确定特征词的上下文信息,通过one hot工具确定所述特征词的词向量;根据所述词向量确定所述上下文信息出现的条件概率;根据所述条件概率和所述上下文信息建立词向量模型。

[0089] 可选地,预设的文本信息库可以为万维网、百科网页、新闻、文档、微信等网页对应的文本信息。

[0090] 具体地,本发明实施例从微信公众号中获取相关的文本信息,这些文本信息中包括多个特征词以及这些特征词对应的上下信息。

[0091] 其中,特征词可以是将预设的文本信息库中的文本进行分词、去除停用词等处理后得到的能代表文本信息库的特征的词。对每个文本信息库的信息进行筛选,将对分类没有意义或对分类贡献值较小的信息过滤,减小处理维度。可选地,特征词可以为一个,也可以是两个或多个。

[0092] 其中,上下文信息指的是特征词周围的词的集合。上下文信息的篇幅可长可短,本发明实施例对上下文信息的篇幅不做限制。

[0093] 可选地,确定特征词时还可以对全半角、大小写等进行区别。

[0094] 可选地,经过词向量模型输出的各个第一文本特征向量的维度一致,也可以不一致,该维度可以根据具体情况变化。

[0095] 可选地,利用word2vec训练得到若干维的词向量模型,确定特征词的词向量的过程可以是:假设有一系列的文档:document1、document2、document3…。其中document1为:我去球场,分词之后得到特征词:[我,去,球场]。经过类似的处理,得到所有的document的特征词为:[我,去,球场,学校,飞机,广工,…。按所有的特征词的顺序来定义所有词的词向量,通过one hot工具将特征词用词向量表示,则:“我”=[1,0,0,0,。],“去”=[0,1,0,0,。],通过这样的方式就将文本信息转换为了数值信息;通过数值信息能更加方便地进行数值计算以及模型建立。

[0096] 可选地,词向量表征的是特征词的位置信息,并不能将特征词与预设的文本信息库向结合,即无法表示特征词的特征信息。

[0097] 可选地,建立词向量模型的具体过程可以为:

[0098] 采用基于Hierarchical Softmax的Skip-gram模型训练词向量,假设特征词w的上下文是Context(w) (由特征词w前后c个词构成),优化的目标函数为:

$$[0099] \quad L = \sum_{w \in C} \log p(\text{Context}(w) | w) \quad (1)$$

[0100] 其中C表示预料 (Corpus);

[0101] 条件概率函数 $p(\text{Context}(w) | w)$ 可转化为:

$$[0102] \quad p(\text{Context}(w) | w) = \prod_{u \in \text{Context}(w)} p(u | w)$$

[0103] 其中,u为特征词w的上下文信息包含的词数。

[0104] 根据Hierarchical Softmax和逻辑回归可知,一个节点被分为正类(目标类别)的概率为:

$$[0105] \quad \delta(v(w)^T \theta_j^u) = \frac{1}{1 + e^{-v(w)^T \theta_j^u}}$$

[0106] 其中,v(w)为特征词w的词向量, $v(w) \in R^m$,m为词向量的长度; p^w 为从根节点出发到达w对应叶子结点的路径; θ_j^u 为路径 p^w 第j个中非叶子结点对应的向量,也就是节点的概率值。

[0107] 根据Hierarchical Softmax将条件概率函数 $p(\text{Context}(w) | w)$ 转化为:

$$[0108] \quad p(u | w) = \prod_{j=2}^{l^w} p(d_j^u | v(w), \theta_{j-1}^u)$$

[0109] 其中,

$$[0110] \quad p(d_j^u | v(w), \theta_{j-1}^u) = [\delta(v(w)^T \theta_{j-1}^u)]^{1-d_j^u} \cdot [1 - \delta(v(w)^T \theta_{j-1}^u)]^{d_j^u} \quad (2)$$

[0111] 其中, l^w 为路径 p^w 中包含的节点个数; d_j^u 为w的Huffman编码, l^w-1 位编码,表示路径 p^w 中第j个节点的编码;

[0112] 将式(2)代入式(1),可得对数似然函数的表达式为:

$$[0113] \quad L = \sum_{w \in C} \sum_{u \in \text{Context}(w)} \sum_{j=2}^{l^w} \{ (1-d_j^u) \cdot \log[\delta(v(w)^T \theta_{j-1}^u)] + d_j^u \cdot \log[1 - \delta(v(w)^T \theta_{j-1}^u)] \}$$

[0114] 该对数似然函数即为Skip-gram的目标函数,采用随机梯度上升优化,从而训练出词向量模型。

[0115] 本实施例从文本信息库中提取特征词以及该特征词对应的上下信息,这些上下文信息能有效地表征该特征词的相关特征,根据这些相关特征建立的词向量模型也能很好地表征特征词的特征。

[0116] 在一个实施例中,所述第一文本特征信息中包括至少一个特征词;所述通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量的步骤,包括:通过预先建立的词向量模型将所述第一文本特征信息中的各个特征词转换为对应的特征词向量,根据各个所述特征词向量确定所述待分类对象对应的第一文本特征向量。

[0117] 其中,根据各个所述特征词向量确定所述待分类对象对应的第一文本特征向量的步骤可以是各个所述特征词向量经过一定的算法得到第一文本特征向量,该算法可以是各个特征词向量直接相加,也可以是添加对应的权值再相加,还可以是其他的算法。

[0118] 可选地,本实施例的实现过程可以为:第一文本特征信息中的特征词为[陈翔、爆笑、笑点],将陈翔、爆笑、笑点这三个特征词输入预先建立的词向量模型中,得到特征词对应的特征词向量:陈翔=[0.1、0.1、0.3]、爆笑=[0.2、0.1、0.5]、笑点=[0.2、0.4、0.7]。将这些特征词向量相加得到待分类对象对应的第一文本特征向量=[0.5、0.6、1.5],该第一文本特征向量可以表征待分类对象的特征。

[0119] 本实施例通过预先建立的词向量模型实现特征词与特征词向量的转化,计算过程简单,之后再根据这些特征词向量得到与待分类对象对应的第一文本特征向量,待分类对象与第一文本特征向量一一对应。

[0120] 在一个实施例中,所述获取待分类对象对应的第一文本特征信息的步骤,包括:通过网络爬虫工具获取待分类对象对应的ID、昵称、简介、经营范围、帐号主体和/或推送消息,从中获取待分类对象对应的第一文本特征信息。

[0121] 可选地,获取第一文本特征信息之后,需要对第一文本特征信息进行分词、去除停用词等处理,从中提取具有代表性的特征词。第一文本特征信息也可以指提取得到的特征词的集合。

[0122] 可选地,可以通过jieba工具等对每个微信公众号的文本特征信息分词后根据TF-IDF提取前N个(N可以为任意正整数)特征词,根据这些特征词构建公众号的特征词列表。这些特征词包括但不限于名词、动词等可以用于将该公众号与其他网页内容相区别的词。

[0123] 本实施例通过网络爬虫工具调用待分类对象的API并获取待分类对象的相关信息,根据这些信息得到待分类对象对应的第一文本特征信息。

[0124] 可选地,如图3所示,图3为基于文本特征的对象分类方法的示意性流程图,所述基于文本特征的对象分类方法包括以下步骤:

[0125] S310、获取多个参考对象对应的第二文本特征向量;分别对所述参考对象的实际类别进行标注。

[0126] S320、通过各个参考对象对应的第二文本特征向量以及实际类别训练预先建立的分类模型,得到经过训练的分类模型。

[0127] S330、从预设的文本信息库中确定特征词的上下文信息,通过one hot工具确定所述特征词的词向量。

[0128] S340、根据所述词向量确定所述上下文信息出现的条件概率。

[0129] S350、根据所述条件概率和所述上下文信息建立词向量模型。

[0130] S360、获取待分类对象对应的第一文本特征信息。

[0131] S370、通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量。

[0132] S380、将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。

[0133] 可选地,S310至S350为离线计算,S360至380为在线计算,对每个待分类公众号进行分类时可以实时进行,提高微信公众号分类的效率。

[0134] 为了更好地理解上述方法,以下详细阐述一个本发明基于文本特征的对象分类方法的应用实例,如图4所示,图4为基于文本特征的对象分类方法的具体应用实例图。其中,以类别为“阅读”“美食”和“搞笑”这三个类别为例。

[0135] 现有两个微信公众号的数据:

[0136] 公众号1:陈翔六点半,简介:《陈翔六点半》是全网第一创意爆笑迷你剧。有灵活的场景和固定的时长,家庭幽默录像式的小情节短剧。无固定演员固定角色,具有鲜明的网络特点,每集均有至少一个笑点,时长均不超过一分钟。由一到两个情节组成,其目的就是让观众用最短的时间和通过最方便的移动互联网平台、解压、放松、快乐。

[0137] 公众号2:大胃王密子君,简介:跟我一起做一个快乐的吃货吧。

[0138] 基于文本特征的对象分类方法的具体过程为:

[0139] 1) 分别对公众号1和公众号2的实际类别进行标注,即公众号1的实际类别为“搞笑”,公众号2的实际类别为“美食”。

[0140] 2) 分别对公众号1和公众号2进行分词和去除停用词处理,得到各个公众号的特征词,公众号1的特征词为陈翔、爆笑和笑点,公众号2的特征词为大胃王和吃货。

[0141] 3) 将上述特征词输入预先建立并经过训练的词向量模型中,得到特征词对应的特征词向量:陈翔=[0.1、0.1、0.3]、爆笑=[0.2、0.1、0.5]、笑点=[0.2、0.4、0.7];大胃王=[0.7、0.1、0.05]、吃货=[0.6、0.2、0.05]。将这些特征词向量相加得到公众号1对应的第二文本特征向量=[0.5、0.6、1.5];公众号2对应的第二文本特征向量=[1.3、0.3、0.1]。

[0142] 4) SVM分类模型(支持向量机模型)包括三个二分类子模型,这三个二分类子模型分别与类别“阅读”“美食”和“搞笑”对应。将这两个第二文本特征向量分别输入SVM分类模型的各个二分类子模型中。与“阅读”对应的二分类子模型对第二文本特征向量[0.5、0.6、1.5]得到的结果为0.1,对第二文本特征向量[1.3、0.3、0.1]得到的匹配度结果为0.9;与“美食”对应的二分类子模型对第二文本特征向量[0.5、0.6、1.5]得到的匹配度结果为0.1,对第二文本特征向量[1.3、0.3、0.1]得到的结果为0.2;与“搞笑”对应的二分类子模型对第二文本特征向量[0.5、0.6、1.5]得到的结果为0.8,对第二文本特征向量[1.3、0.3、0.1]得到的匹配度结果为0.2。

[0143] 根据各个二分类子模型的分类结果得到公众号1对应的匹配度[0.1、0.1、0.8],最高匹配度为0.8,与0.8对应的评估类别为“搞笑”,将其与公众号1的实际类别“搞笑”进行比对,发现分类模型得到的分类结果正确。

[0144] 根据各个二分类子模型的分类结果得到公众号2对应的匹配度[0.9、0.2、0.2],最高匹配度为0.9,与0.9对应的评估类别为“阅读”,将其与公众号1的实际类别“美食”进行比对,发现分类模型得到的分类结果错误。

[0145] 5) 根据以上分类结果得到分类模型的分类准确率为50%,低于预设的阈值99%,对该分类模型进行调整,直到准确率高于该阈值。优选地,支持向量机模型 $F(x)$ 的各个参数为惩罚松弛系数为1,分类决策采用“One-vs-Rest”模式,核函数采用“poly”函数,“poly”核函数的维度取1、系数为1/33、c值为1。

[0146] 6) 获取待分类公众号的信息:大胃王mini,简介:大胃王mini的美食频道。

[0147] 7) 对待分类公众号进行分词、去除停用词处理,得到特征词:大胃王、美食,将这些特征词输入到词向量模型中得到对应的特征词向量:大胃王=[0.7、0.1、0.05]、美食=

[0.7、0.2、0.1],将这两个特征词向量相加得到该待分类公众号对应的第一文本特征向量 = [1.4、0.3、0.15]。

[0148] 8) 将该第一文本特征向量 = [1.4、0.3、0.15]输入分类模型中的各个二分类子模型中,并得到该待分类公众号的匹配度[0.1、0.9、0.2],最高匹配度值为0.9,与该0.9对应的二分类子模型的评估类别为“美食”,则输出该待分类公众号的评估类别为“美食”。

[0149] 将本发明实施例的基于文本特征的对象分类方法应用在万象平台微信公众号分类上,测试集(多个待分类对象)的表现为precision(正确率):0.76,recall(召回率):0.71,f1-score(f1值):0.73。相比人工分类,该技术在保证准确率的前提下分类速度大幅领先。此外,调高阈值可降低召回率提升准确率,证明了该方法的有效性。

[0150] 需要说明的是,对于前述的各方法实施例,为了简便描述,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,本发明并不受所描述的动作顺序的限制,因为依据本发明,某些步骤可以采用其它顺序或者同时进行。

[0151] 基于与上述实施例中的基于文本特征的对象分类方法相同的思想,本发明还提供基于文本特征的对象分类装置,该装置可用于执行上述基于文本特征的对象分类方法。为了便于说明,基于文本特征的对象分类装置实施例的结构示意图中,仅仅示出了与本发明实施例相关的部分,本领域技术人员可以理解,图示结构并不构成对装置的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。

[0152] 本发明实施例提供一种基于文本特征的对象分类装置,如图5所示,所述基于文本特征的对象分类装置包括:信息获取模块510,用于获取待分类对象对应的第一文本特征信息;向量转换模块520,用于通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量;以及,分类模块530,用于将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。

[0153] 本实施例能根据预先训练好的模型对待分类对象进行准确的分类,进而根据得到的分类信息对该待分类对象进行针对性的操作,可以有效防止对各种类别的对象均进行操作而造成资源的浪费。

[0154] 在一个实施例中,所述的基于文本特征的对象分类装置,还包括:类别标注模块,用于获取多个参考对象对应的第二文本特征向量;分别对所述参考对象的实际类别进行标注;以及,模型训练模块,用于通过各个参考对象对应的第二文本特征向量以及实际类别训练预先建立的分类模型,得到经过训练的分类模型。

[0155] 在一个实施例中,所述分类模型包括至少一个二分类子模型,每个二分类子模型分别对应一个评估类别;所述模型训练模块,包括:匹配度获取子模块,用于将某一所述第二文本特征向量分别输入各个二分类子模型中,分别得到所述第二文本特征向量与对应的评估类别的匹配度;类别确定子模块,用于根据所述匹配度确定参考对象的评估类别;以及,模型调整子模块,用于将参考对象的评估类别与对应的实际类别进行比对,根据比对结果调整所述分类模型。

[0156] 在一个实施例中,所述类别确定子模块,还用于确定所述匹配度中的最高匹配度值,获取与所述最高匹配度值对应的评估类别,作为对应的待分类对象的评估类别。

[0157] 在一个实施例中,所述的基于文本特征的对象分类装置,还包括:词向量确定模

块,用于从预设的文本信息库中确定特征词的上下文信息,通过one hot工具确定所述特征词的词向量;条件概率计算模块,用于根据所述词向量确定所述上下文信息出现的条件概率;以及,词向量模型建立模块,用于根据所述条件概率和所述上下文信息建立词向量模型。

[0158] 在一个实施例中,所述第一文本特征信息中包括至少一个特征词;所述向量转换模块,还用于通过预先建立的词向量模型将所述第一文本特征信息中的各个特征词转换为对应的特征词向量,根据各个所述特征词向量确定所述待分类对象对应的第一文本特征向量。

[0159] 在一个实施例中,所述信息获取模块510,还用于通过网络爬虫工具获取待分类对象对应的ID、昵称、简介、经营范围、帐号主体和/或推送消息,从中获取待分类对象对应的第一文本特征信息。

[0160] 需要说明的是,本发明的基于文本特征的对象分类装置与本发明的基于文本特征的对象分类方法一一对应,在上述基于文本特征的对象分类方法的实施例阐述的技术特征及其有益效果均适用于基于文本特征的对象分类装置的实施例中,具体内容可参见本发明方法实施例中的叙述,此处不再赘述,特此声明。

[0161] 此外,上述示例的基于文本特征的对象分类装置的实施方式中,各程序模块的逻辑划分仅是举例说明,实际应用中可以根据需要,例如出于相应硬件的配置要求或者软件的实现的便利考虑,将上述功能分配由不同的程序模块完成,即将所述基于文本特征的对象分类装置的内部结构划分成不同的程序模块,以完成以上描述的全部或者部分功能。

[0162] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是服务器,其内部结构图可以如图6所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储分类数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种基于文本特征的对象分类方法。

[0163] 本领域技术人员可以理解,图6中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0164] 在一个实施例中,提供了一种计算机设备,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,处理器执行计算机程序时实现以下步骤:获取待分类对象对应的第一文本特征信息;通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量;将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。

[0165] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:获取多个参考对象对应的第二文本特征向量;分别对所述参考对象的实际类别进行标注;通过各个参考对象对应的第二文本特征向量以及实际类别训练预先建立的分类模型,得到经过训练的分类模型。

[0166] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:将某一所述第二文

本特征向量分别输入各个二分类子模型中,分别得到所述第二文本特征向量与对应的评估类别的匹配度;根据所述匹配度确定参考对象的评估类别;将参考对象的评估类别与对应的实际类别进行比对,根据比对结果调整所述分类模型。

[0167] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:确定所述匹配度中的最高匹配度值,获取与所述最高匹配度值对应的评估类别,作为对应的待分类对象的评估类别。

[0168] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:从预设的文本信息库中确定特征词的上下文信息,通过one hot工具确定所述特征词的词向量;根据所述词向量确定所述上下文信息出现的条件概率;根据所述条件概率和所述上下文信息建立词向量模型。

[0169] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:通过预先建立的词向量模型将所述第一文本特征信息中的各个特征词转换为对应的特征词向量,根据各个所述特征词向量确定所述待分类对象对应的第一文本特征向量。

[0170] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:通过网络爬虫工具获取待分类对象对应的ID、昵称、简介、经营范围、帐号主体和/或推送消息,从中获取待分类对象对应的第一文本特征信息。

[0171] 在一个实施例中,提供了一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现以下步骤:获取待分类对象对应的第一文本特征信息;通过预先建立的词向量模型将所述第一文本特征信息转换为对应的第一文本特征向量;将所述第一文本特征向量输入经过训练的分类模型中,根据所述经过训练的分类模型输出的结果确定所述待分类对象的评估类别。

[0172] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:获取多个参考对象对应的第二文本特征向量;分别对所述参考对象的实际类别进行标注;通过各个参考对象对应的第二文本特征向量以及实际类别训练预先建立的分类模型,得到经过训练的分类模型。

[0173] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:根据所述将某一所述第二文本特征向量分别输入各个二分类子模型中,分别得到所述第二文本特征向量与对应的评估类别的匹配度;根据所述匹配度确定参考对象的评估类别;将参考对象的评估类别与对应的实际类别进行比对,根据比对结果调整所述分类模型。

[0174] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:确定所述匹配度中的最高匹配度值,获取与所述最高匹配度值对应的评估类别,作为对应的待分类对象的评估类别。

[0175] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:从预设的文本信息库中确定特征词的上下文信息,通过one hot工具确定所述特征词的词向量;根据所述词向量确定所述上下文信息出现的条件概率;根据所述条件概率和所述上下文信息建立词向量模型。

[0176] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:通过预先建立的词向量模型将所述第一文本特征信息中的各个特征词转换为对应的特征词向量,根据各个所述特征词向量确定所述待分类对象对应的第一文本特征向量。

[0177] 在一个实施例中, 计算机程序被处理器执行时还实现以下步骤: 通过网络爬虫工具获取待分类对象对应的ID、昵称、简介、经营范围、帐号主体和/或推送消息, 从中获取待分类对象对应的第一文本特征信息。

[0178] 本领域普通技术人员可以理解, 实现上述实施例方法中的全部或部分流程, 是可以通过计算机程序来指令相关的硬件来完成, 所述的程序可存储于一计算机可读取存储介质中, 作为独立的产品销售或使用。计算机可读介质的更具体的示例(非穷尽性列表)包括以下: 具有一个或多个布线的电连接部(电子装置), 便携式计算机盘盒(磁装置), 随机存取存储器(RAM), 只读存储器(ROM), 可擦除可编程只读存储器(EPROM或闪速存储器), 光纤装置, 以及便携式光盘只读存储器(CDROM)。另外, 计算机可读介质甚至可以是可在其上打印所述程序的纸或其他合适的介质, 因为可以例如通过对纸或其他介质进行光学扫描, 接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得所述程序, 然后将其存储在计算机存储器中。

[0179] 应当理解, 本发明的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中, 多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如, 如果用硬件来实现, 和在另一实施方式中一样, 可用本领域公知的下列技术中的任一项或他们的组合来实现: 具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路, 具有合适的组合逻辑门电路的专用集成电路, 可编程门阵列(PGA), 现场可编程门阵列(FPGA)等。

[0180] 本发明实施例的术语“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(模块)单元的过程、方法、系统、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它步骤或单元。

[0181] 以上所述实施例的各技术特征可以进行任意的组合, 为使描述简洁, 未对上述实施例中的各个技术特征所有可能的组合都进行描述, 然而, 只要这些技术特征的组合不存在矛盾, 都应当认为是本说明书记载的范围。

[0182] 以上所述实施例仅表达了本发明的几种实施方式, 不能理解为对本发明专利范围的限制。应当指出的是, 对于本领域的普通技术人员来说, 在不脱离本发明构思的前提下, 还可以做出若干变形和改进, 这些都属于本发明的保护范围。因此, 发明专利的保护范围应以所附权利要求为准。

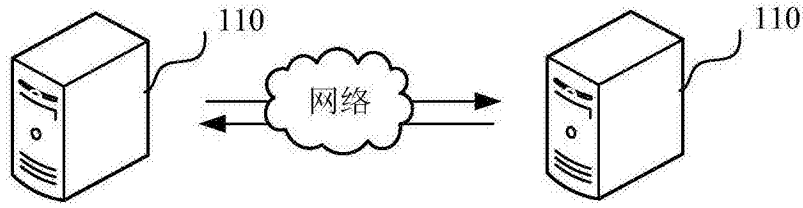


图1

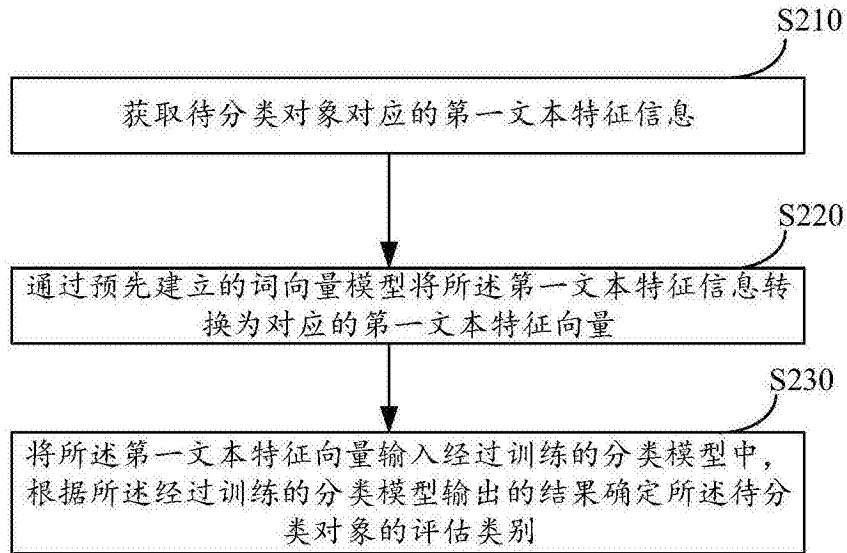


图2

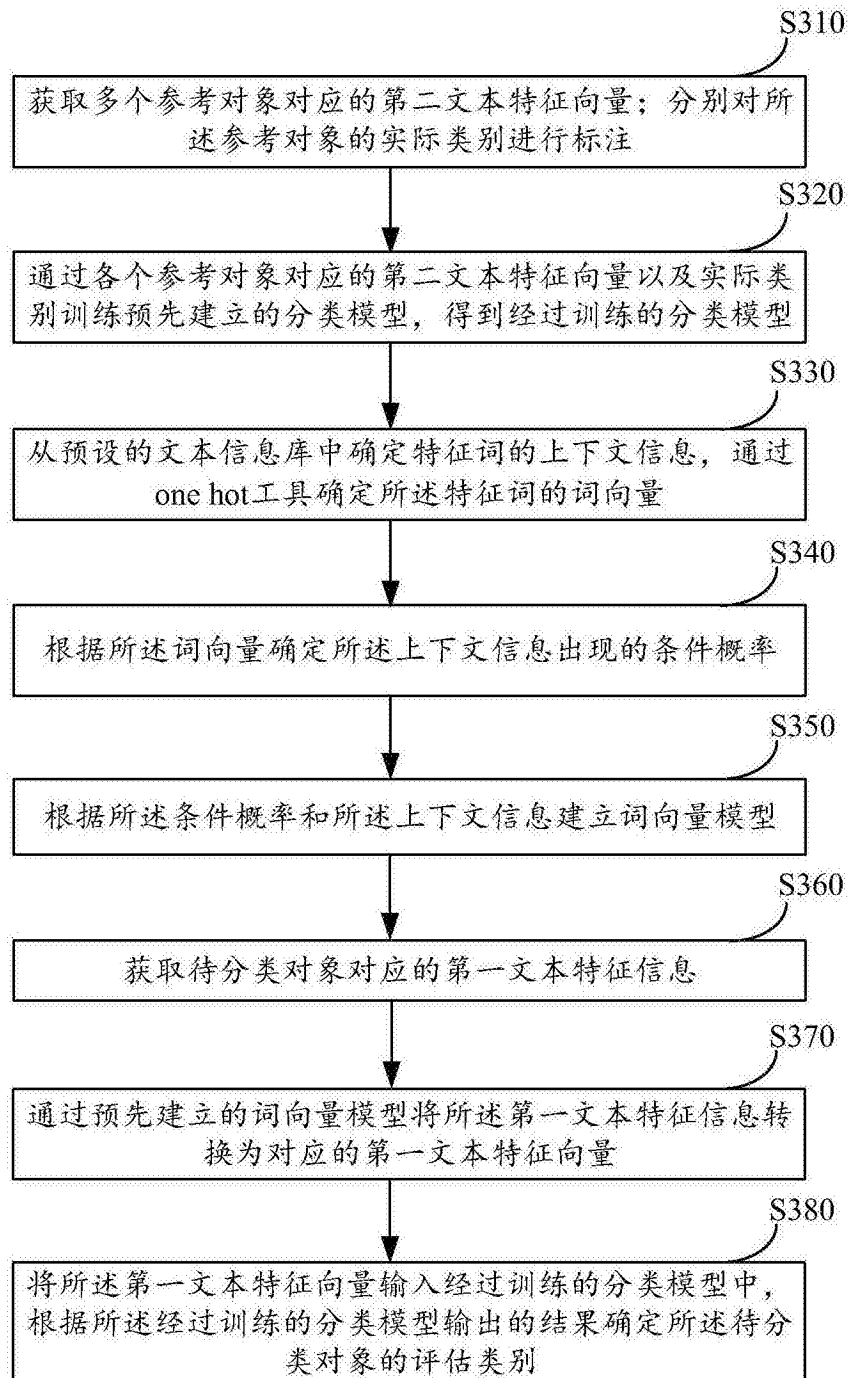


图3

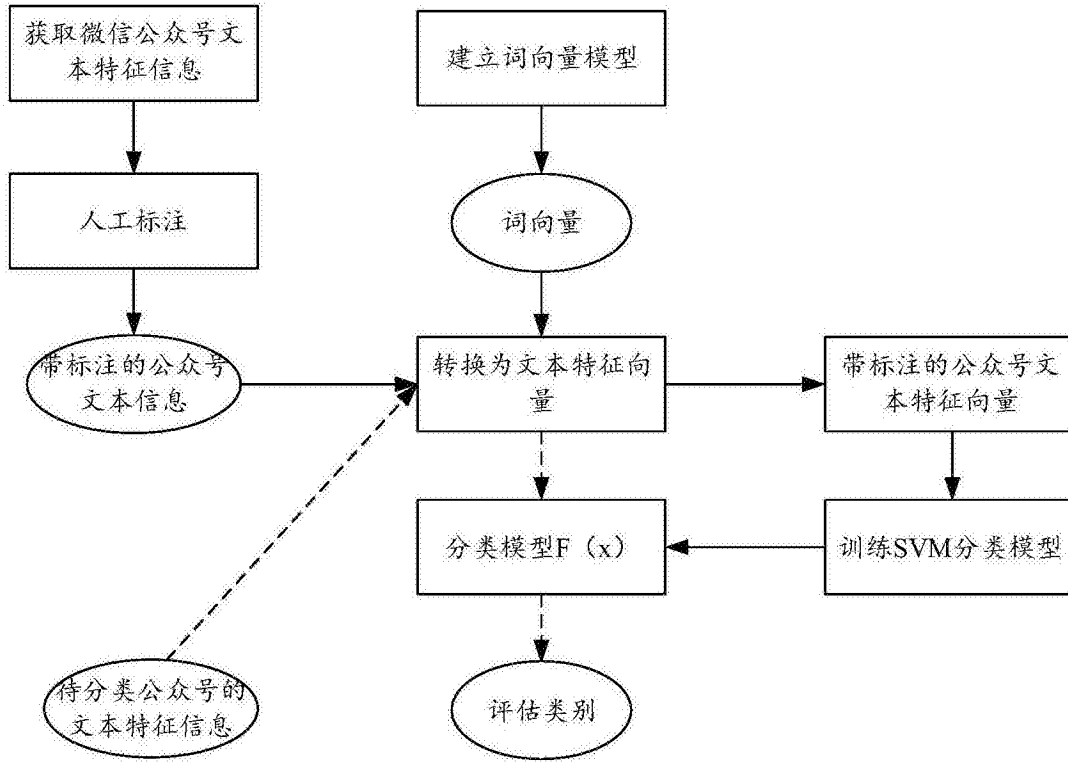


图4

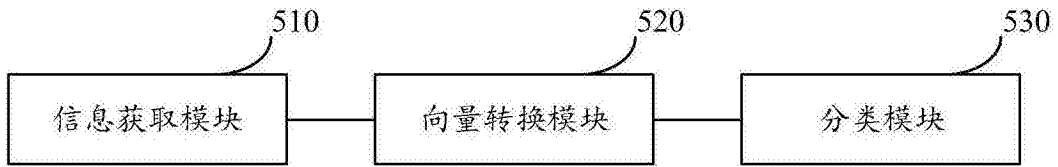


图5

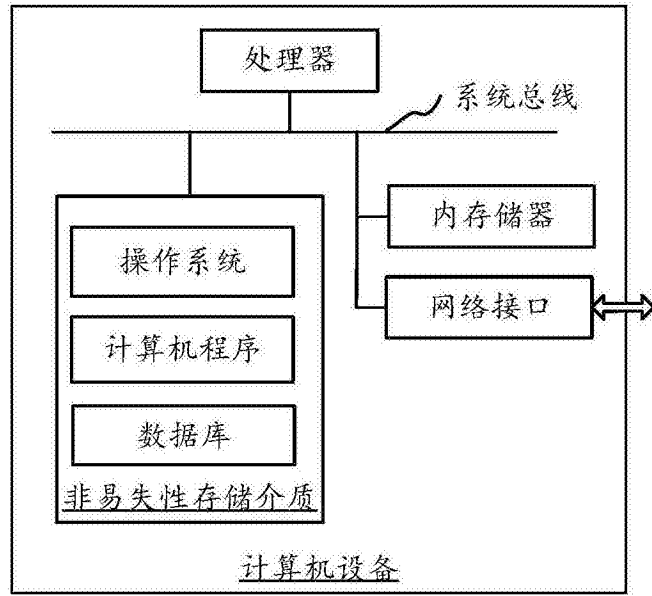


图6