(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 111259169 A
(43)申请公布日 2020.06.09

(21)申请号 202010080507.8

(22)申请日 2020.02.05

(71)申请人 四川无声信息技术有限公司
地址 610000 四川省成都市高新区芳草东街76号

(72)发明人 吕振远 许春阳 程芃森 陈航 张冬 崔凯铜

(74)专利代理机构 北京超凡宏宇专利代理事务所(特殊普通合伙) 11463
代理人 安卫静

(51)Int.Cl.
G06F 16/38(2019.01)
G06F 40/279(2020.01)
G06K 9/62(2006.01)

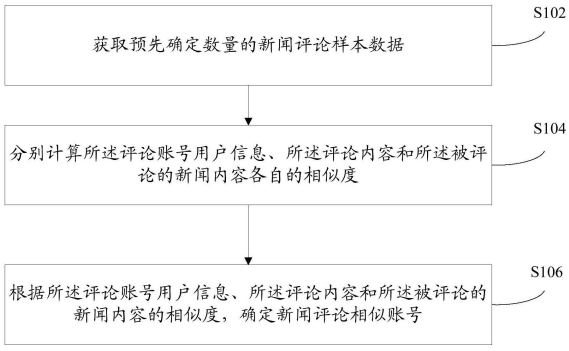
权利要求书2页 说明书7页 附图2页

(54)发明名称

新闻评论相似账号确定方法及装置

(57)摘要

本发明提供了一种新闻评论相似账号确定方法及装置,涉及相似账号确定技术领域。该方法包括:获取预先确定数量的新闻评论样本数据;分别计算所述评论账号用户信息、所述评论内容和所述被评论的新闻内容各自的相似度;根据所述评论账号用户信息、所述评论内容和所述被评论的新闻内容的相似度,确定新闻评论相似账号。本发明实施例的新闻评论相似账号确定方法及装置通过分别计算评论账号用户信息、评论内容和被评论的新闻内容各自的相似度,获取新闻评论相似账号,从而达到了方便追踪特定账号的技术效果。



CN 111259169 A

1.一种新闻评论相似账号确定方法,其特征在于,所述方法包括以下步骤:

获取预先确定数量的新闻评论样本数据;其中,所述新闻评论样本数据包括评论账号用户信息、评论内容和被评论的新闻内容;

分别计算所述评论账号用户信息、所述评论内容和所述被评论的新闻内容各自的相似度;

根据所述评论账号用户信息、所述评论内容和所述被评论的新闻内容的相似度,确定新闻评论相似账号。

2.根据权利要求1所述的方法,其特征在于,所述计算评论账号用户信息、评论内容和被评论的新闻内容的相似度的步骤,包括:

对评论账号用户信息、评论内容和被评论的新闻内容添加停用词,去除其中的符号和表情,获取经预处理的评论账号用户信息、评论内容和被评论的新闻内容;

计算所述经预处理的评论账号用户信息、评论内容和被评论的新闻内容的相似度。

3.根据权利要求1所述的方法,其特征在于,所述计算所述评论账号用户信息的相似度的步骤,包括

提取任两条所述评论账号用户信息中的用户的年龄信息、性别信息、所在地信息;

根据所述年龄信息、所述性别信息、所述所在地信息计算所述评论账号用户信息的相似度。

4.根据权利要求1所述的方法,其特征在于,所述计算所述评论内容的相似度的步骤,包括:

对任两条评论内容中的每个词的词向量进行累加并求平均值,获取所述任两条评论内容之间的词向量;

根据所述任两条评论内容之间的词向量的余弦值,计算所述评论内容的相似度;其中,所述余弦值越接近1,所述评论内容的相似度越高。

5.根据权利要求1所述的方法,其特征在于,所述计算所述被评论的新闻内容的相似度的步骤,包括:

对任两条所述被评论的新闻内容中的每个词的词向量进行累加并求平均值,获取所述任两条所述被评论的新闻内容之间的词向量;

根据所述任两条所述被评论的新闻内容之间的词向量的余弦值,计算所述被评论的新闻内容的相似度;其中,所述余弦值越接近1,所述被评论的新闻内容的相似度越高。

6.一种新闻评论相似账号确定装置,其特征在于,所述装置包括:

获取模块,用于获取预先确定数量的新闻评论样本数据;其中,所述新闻评论样本数据包括评论账号用户信息、评论内容和被评论的新闻内容;

计算模块,用于分别计算所述评论账号用户信息、所述评论内容和所述被评论的新闻内容各自的相似度;

确定模块,用于根据所述评论账号用户信息、所述评论内容和所述被评论的新闻内容的相似度,确定新闻评论相似账号。

7.根据权利要求6所述的装置,其特征在于,所述计算模块用于:

对评论账号用户信息、评论内容和被评论的新闻内容添加停用词,去除其中的符号和表情,获取经预处理的评论账号用户信息、评论内容和被评论的新闻内容;

计算所述经预处理的评论账号用户信息、评论内容和被评论的新闻内容的相似度。

8.根据权利要求6所述的装置，其特征在于，所述计算模块还用于：

提取任两条所述评论账号用户信息中的用户的年龄信息、性别信息、所在地信息；

根据所述年龄信息、所述性别信息、所述所在地信息计算所述评论账号用户信息的相似度。

9.一种服务器，其特征在于，包括处理器和存储器，所述存储器存储有能够被所述处理器执行的计算机可执行指令，所述处理器执行所述计算机可执行指令以实现权利要求1至5任一项所述的方法。

10.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有计算机可执行指令，所述计算机可执行指令在被处理器调用和执行时，所述计算机可执行指令促使处理器实现权利要求1至5任一项所述的方法。

新闻评论相似账号确定方法及装置

技术领域

[0001] 本发明相似账号确定技术领域，尤其是涉及一种新闻评论相似账号确定方法及装置。

背景技术

[0002] 目前，对于社会上的各种各样的新闻，网民通过网络等方式对这些新闻表达自己的观点并且反映自身诉求，这种近似观点性、影响力、倾向性的评论，如果不加以有效的舆论引导和管控，容易造成舆情隐患，甚至会扰乱社会秩序。

发明内容

[0003] 有鉴于此，本发明的目的在于提供一种新闻评论相似账号确定方法及装置，以改善各种评论容易造成舆情隐患甚至会扰乱社会秩序的技术问题。

[0004] 第一方面，本发明实施例提供了一种新闻评论相似账号确定方法，所述方法包括以下步骤：

[0005] 获取预先确定数量的新闻评论样本数据；其中，所述新闻评论样本数据包括评论账号用户信息、评论内容和被评论的新闻内容；

[0006] 分别计算所述评论账号用户信息、所述评论内容和所述被评论的新闻内容各自的相似度；

[0007] 根据所述评论账号用户信息、所述评论内容和所述被评论的新闻内容的相似度，确定新闻评论相似账号。

[0008] 结合第一方面，本发明实施例提供了第一方面的第一种可能的实施方式，其中，所述计算评论账号用户信息、评论内容和被评论的新闻内容的相似度的步骤，包括：

[0009] 对评论账号用户信息、评论内容和被评论的新闻内容添加停用词，去除其中的符号和表情，获取经预处理的评论账号用户信息、评论内容和被评论的新闻内容；

[0010] 计算所述经预处理的评论账号用户信息、评论内容和被评论的新闻内容的相似度。

[0011] 结合第一方面，本发明实施例提供了第一方面的第二种可能的实施方式，其中，所述计算所述评论账号用户信息的相似度的步骤，包括：

[0012] 提取任两条所述评论账号用户信息中的用户的年龄信息、性别信息、所在地信息；

[0013] 根据所述年龄信息、所述性别信息、所述所在地信息计算所述评论账号用户信息的相似度。

[0014] 结合第一方面，本发明实施例提供了第一方面的第三种可能的实施方式，其中，所述计算所述评论内容的相似度的步骤，包括：

[0015] 对任两条评论内容中的每个词的词向量进行累加并求平均值，获取所述任两条评论内容之间的词向量；

[0016] 根据所述任两条评论内容之间的词向量的余弦值，计算所述评论内容的相似度；

其中，所述余弦值越接近1，所述评论内容的相似度越高。

[0017] 结合第一方面，本发明实施例提供了第一方面的第四种可能的实施方式，所述计算所述被评论的新闻内容的相似度的步骤，包括：

[0018] 对任两条所述被评论的新闻内容中的每个词的词向量进行累加并求平均值，获取所述任两条所述被评论的新闻内容之间的词向量；

[0019] 根据所述任两条所述被评论的新闻内容之间的词向量的余弦值，计算所述被评论的新闻内容的相似度；其中，所述余弦值越接近1，所述被评论的新闻内容的相似度越高。

[0020] 第二方面，本发明实施例还提供一种新闻评论相似账号确定装置，所述装置包括：

[0021] 获取模块，用于获取预先确定数量的新闻评论样本数据；其中，所述新闻评论样本数据包括评论账号用户信息、评论内容和被评论的新闻内容；

[0022] 计算模块，用于分别计算所述评论账号用户信息、所述评论内容和所述被评论的新闻内容各自的相似度；

[0023] 确定模块，用于根据所述评论账号用户信息、所述评论内容和所述被评论的新闻内容的相似度，确定新闻评论相似账号。

[0024] 结合第二方面，本发明实施例提供了第二方面的第一种可能的实施方式，其中，所述计算模块用于：

[0025] 对评论账号用户信息、评论内容和被评论的新闻内容添加停用词，去除其中的符号和表情，获取经预处理的评论账号用户信息、评论内容和被评论的新闻内容；

[0026] 计算所述经预处理的评论账号用户信息、评论内容和被评论的新闻内容的相似度。

[0027] 结合第二方面，本发明实施例提供了第二方面的第二种可能的实施方式，所述计算模块还用于：

[0028] 提取任两条所述评论账号用户信息中的用户的年龄信息、性别信息、所在地信息；

[0029] 根据所述年龄信息、所述性别信息、所述所在地信息计算所述评论账号用户信息的相似度。

[0030] 第三方面，本发明实施例还提供一种服务器，所述服务器包括：处理器和存储器，所述存储器存储有能够被所述处理器执行的计算机可执行指令，所述处理器执行所述计算机可执行指令以实现上文所述的方法。

[0031] 第四方面，本发明实施例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存储有计算机可执行指令，所述计算机可执行指令在被处理器调用和执行时，所述计算机可执行指令促使处理器实现上文所述的方法。

[0032] 本发明实施例带来了以下有益效果：本发明实施例提供的一种新闻评论相似账号确定方法及装置，获取新闻评论样本数据中的评论账号用户信息、评论内容和被评论的新闻内容，分别计算其各自的相似度，并根据该相似度确定新闻评论相似账号。本发明实施例的新闻评论相似账号确定方法及装置通过分别计算评论账号用户信息、评论内容和被评论的新闻内容各自的相似度，获取新闻评论相似账号，从而达到了方便追踪特定账号的技术效果。

[0033] 本发明的其他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在说明书、权利要求书以

及附图中所特别指出的结构中实现和获得。

[0034] 为使本发明的上述目的、特征和优点能够更加明显易懂，下文特举优选实施例，并配合所附附图，作详细说明如下。

附图说明

[0035] 为了更清楚地说明本发明的具体实施方式或现有技术中的技术方案，下面将对具体实施方式或现有技术描述中所需要使用的附图进行简单的介绍，显而易见地，下面描述中的附图是本发明的一些实施方式，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。

[0036] 图1为本发明实施例提供的一种新闻评论相似账号确定方法的流程图；

[0037] 图2为本发明实施例提供的另一种新闻评论相似账号确定方法的流程图；

[0038] 图3为本发明实施例提供的一种新闻评论相似账号确定装置的结构框图；

[0039] 图4为本发明实施例提供的一种服务器的结构示意图。

具体实施方式

[0040] 为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合附图对本发明的技术方案进行清楚、完整的描述，显然，所描述的实施例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做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。

[0041] 目前，目前，对于社会上的各种各样的新闻，网民通过网络等方式对这些新闻表达自己的观点并且反映自身诉求，这种近似观点性、影响力、倾向性的评论，如果不加以有效的舆论引导和管控，容易造成舆情隐患，甚至会扰乱社会秩序。基于此，本发明实施例提供了一种新闻评论相似账号确定方法及装置，以缓解上述问题。

[0042] 为了便于对本实施例进行理解，首先对本发明实施例所公开的一种新闻评论相似账号确定方法进行详细介绍。

[0043] 在一种可能的实施方式中，本发明提供了一种新闻评论相似账号确定方法。如图1所示为本发明实施例提供的一种新闻评论相似账号确定方法的流程图，所述方法包括以下步骤：

[0044] 步骤S102：获取预先确定数量的新闻评论样本数据。

[0045] 其中，所述新闻评论样本数据包括评论账号用户信息、评论内容和被评论的新闻内容。

[0046] 需要进一步说明的是，所获取的新闻评论样本数据要足够多，要涵盖多样性的新闻评论，从而确保新闻评论样本数据的多样性。

[0047] 步骤S104：分别计算所述评论账号用户信息、所述评论内容和所述被评论的新闻内容各自的相似度。

[0048] 步骤S106：根据所述评论账号用户信息、所述评论内容和所述被评论的新闻内容的相似度，确定新闻评论相似账号。

[0049] 本发明实施例带来了以下有益效果：本发明实施例通过一种新闻评论相似账号确定方法，获取新闻评论样本数据中的评论账号用户信息、评论内容和被评论的新闻内容，分

别计算其各自的相似度，并根据该相似度确定新闻评论相似账号。本发明实施例的新闻评论相似账号确定方法及装置通过分别计算评论账号用户信息、评论内容和被评论的新闻内容各自的相似度，获取新闻评论相似账号，从而达到了方便追踪特定账号的技术效果。

[0050] 在实际使用时，为了对分别计算所述评论账号用户信息、所述评论内容和所述被评论的新闻内容各自的相似度的过程进行更加详细的描述，本发明实施例在图2中示出了本发明实施例提供的另一种新闻评论相似账号确定方法的流程图，该方法包括以下步骤：

[0051] 步骤S202：获取预先确定数量的新闻评论样本数据。

[0052] 其中，所述新闻评论样本数据包括评论账号用户信息、评论内容和被评论的新闻内容。

[0053] 步骤S204：对评论账号用户信息、评论内容和被评论的新闻内容添加停用词，去除其中的符号和表情，获取经预处理的评论账号用户信息、评论内容和被评论的新闻内容。

[0054] 其中，在对新闻评论样本数据进行预处理的过程中，训练word2vec(word2vec为计算word vector的开源工具)词向量模型需要有符合的训练语料，这需要对原始的新闻评论样本数据进行清理工作。在新闻评论中，评论者通常会添加一些表情图、各种符号等无关紧要的内容，需要去除与训练语料无关的这些内容，通常通过添加停用词将这些符号和表情去除。

[0055] 步骤S206：计算所述经预处理的评论账号用户信息、评论内容和被评论的新闻内容的相似度。

[0056] 步骤S208：根据所述评论账号用户信息、所述评论内容和所述被评论的新闻内容的相似度，确定新闻评论相似账号。

[0057] 进一步地，为了对分别计算所述评论账号用户信息、所述评论内容和所述被评论的新闻内容各自的相似度的步骤进行更进一步的描述，分别对上述三种相似度计算过程进行更加详细的描述。

[0058] 具体地，所述计算所述评论账号用户信息的相似度的步骤，包括：

[0059] 提取任两条所述评论账号用户信息中的用户的年龄信息、性别信息、所在地信息；

[0060] 根据所述年龄信息、所述性别信息、所述所在地信息计算所述评论账号用户信息的相似度。

[0061] 其中，本发明实施例基于实体识别模型对评论账号用户信息进行提取。

[0062] 具体地，所述计算所述评论内容的相似度的步骤，包括：

[0063] 对任两条评论内容中的每个词的词向量进行累加并求平均值，获取所述任两条评论内容之间的词向量；

[0064] 根据所述任两条评论内容之间的词向量的余弦值，计算所述评论内容的相似度；其中，所述余弦值越接近1，所述评论内容的相似度越高。

[0065] 具体地，所述计算所述被评论的新闻内容的相似度的步骤，包括：

[0066] 对任两条所述被评论的新闻内容中的每个词的词向量进行累加并求平均值，获取所述任两条所述被评论的新闻内容之间的词向量；

[0067] 根据所述任两条所述被评论的新闻内容之间的词向量的余弦值，计算所述被评论的新闻内容的相似度；其中，所述余弦值越接近1，所述被评论的新闻内容的相似度越高。

[0068] 综上所述，本发明的新闻评论相似账号确定方法及装置，获取新闻评论样本数据

中的评论账号用户信息、评论内容和被评论的新闻内容，分别计算其各自的相似度，并根据该相似度确定新闻评论相似账号。本发明实施例的新闻评论相似账号确定方法及装置通过分别计算评论账号用户信息、评论内容和被评论的新闻内容各自的相似度，获取新闻评论相似账号，从而达到了方便追踪特定账号的技术效果。

[0069] 在另一种可能的实施方式中，对应于上述实施方式提供的新闻评论相似账号确定方法，本发明实施例还提供了一种新闻评论相似账号确定装置，图3本发明实施例提供的一种新闻评论相似账号确定装置的结构框图。如图3所示，所述装置包括：

[0070] 获取模块301，用于获取预先确定数量的新闻评论样本数据。

[0071] 其中，所述新闻评论样本数据包括评论账号用户信息、评论内容和被评论的新闻内容；

[0072] 计算模块302，用于分别计算所述评论账号用户信息、所述评论内容和所述被评论的新闻内容各自的相似度；

[0073] 确定模块303，用于根据所述评论账号用户信息、所述评论内容和所述被评论的新闻内容的相似度，确定新闻评论相似账号。

[0074] 在实际使用时，所述计算模块302用于：

[0075] 对评论账号用户信息、评论内容和被评论的新闻内容添加停用词，去除其中的符号和表情，获取经预处理的评论账号用户信息、评论内容和被评论的新闻内容；

[0076] 计算所述经预处理的评论账号用户信息、评论内容和被评论的新闻内容的相似度。

[0077] 在实际使用时，所述计算模块还用于：

[0078] 提取任两条所述评论账号用户信息中的用户的年龄信息、性别信息、所在地信息；

[0079] 根据所述年龄信息、所述性别信息、所述所在地信息计算所述评论账号用户信息的相似度。

[0080] 在又一种可能的实施方式中，本发明实施例还提供了一种服务器，图4示出了本发明实施例提供的一种服务器的结构示意图，参见图4，该服务器包括：处理器400、存储器401、总线402和通信接口403，该处理器400、存储器401、通信接口403和通过总线402连接；处理器400用于执行存储器401中存储的可执行模块，例如计算机程序。

[0081] 其中，存储器401存储有能够被处理器400执行的计算机可执行指令，处理器400执行计算机可执行指令以实现上文所述的方法。

[0082] 进一步地，存储器401可能包含高速随机存取存储器(RAM，Random Access Memory)，也可能还包括非不稳定的存储器(non-volatile memory)，例如至少一个磁盘存储器。通过至少一个通信接口403(可以是有线或者无线)实现该系统网元与至少一个其他网元之间的通信连接，可以使用互联网，广域网，本地网，城域网等。

[0083] 总线402可以是ISA总线、PCI总线或EISA总线等。该总线可以分为地址总线、数据总线、控制总线等。为便于表示，图4中仅用一个双向箭头表示，但并不表示仅有一根总线或一种类型的总线。

[0084] 其中，存储器401用于存储程序，处理器400在接收到程序执行指令后，执行所述程序，前述本发明实施例任一实施例揭示的新闻评论相似账号确定方法可以应用于处理器400中，或者由处理器400实现。

[0085] 此外，处理器400可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器400中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器400可以是通用处理器，包括中央处理器(Central Processing Unit，简称CPU)、网络处理器(Network Processor，简称NP)等；还可以是数字信号处理器(Digital Signal Processing，简称DSP)、专用集成电路(Application Specific Integrated Circuit，简称ASIC)、现成可编程门阵列(Field-Programmable Gate Array，简称FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器，闪存、只读存储器，可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器401，处理器400读取存储器401中的信息，结合其硬件完成上述方法的步骤。

[0086] 在又一种可能的实施方式中，本发明实施例还提供了一种计算机可读存储介质，计算机可读存储介质存储有计算机可执行指令，所述计算机可执行指令在被处理器调用和执行时，计算机可执行指令促使处理器实现上文所述的方法。

[0087] 本发明实施例提供的新闻评论相似账号确定装置，与上述实施例提供的新闻评论相似账号确定方法具有相同的技术特征，所以也能解决相同的技术问题，达到相同的技术效果。

[0088] 本发明实施例所提供的新闻评论相似账号确定方法及装置的计算机程序产品，包括存储了程序代码的计算机可读存储介质，所述程序代码包括的指令可用于执行前面方法实施例中所述的方法，具体实现可参见方法实施例，在此不再赘述。

[0089] 所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的装置的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0090] 另外，在本发明实施例的描述中，除非另有明确的规定和限定，术语“安装”、“相连”、“连接”应做广义理解，例如，可以是固定连接，也可以是可拆卸连接，或一体地连接；可以是机械连接，也可以是电连接；可以是直接相连，也可以通过中间媒介间接相连，可以是两个元件内部的连通。对于本领域技术人员而言，可以具体情况理解上述术语在本发明中的具体含义。

[0091] 所述功能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时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备(可以是个人计算机，服务器，或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器(ROM，ReaD-Only Memory)、随机存取存储器(RAM，RanDom Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0092] 在本发明的描述中，需要说明的是，术语“中心”、“上”、“下”、“左”、“右”、“竖直”、“水平”、“内”、“外”等指示的方位或位置关系为基于附图所示的方位或位置关系，仅是为了便于描述本发明和简化描述，而不是指示或暗示所指的装置或元件必须具有特定的方位、

以特定的方位构造和操作，因此不能理解为对本发明的限制。此外，术语“第一”、“第二”、“第三”仅用于描述目的，而不能理解为指示或暗示相对重要性。

[0093] 最后应说明的是：以上实施例，仅为本发明的具体实施方式，用以说明本发明的技术方案，而非对其限制，本发明的保护范围并不局限于此，尽管参照前述实施例对本发明进行了详细的说明，本领域技术人员应当理解：任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，其依然可以对前述实施例所记载的技术方案进行修改或可轻易想到变化，或者对其中部分技术特征进行等同替换；而这些修改、变化或者替换，并不使相应技术方案的本质脱离本发明实施例技术方案的精神和范围，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应以权利要求的保护范围为准。

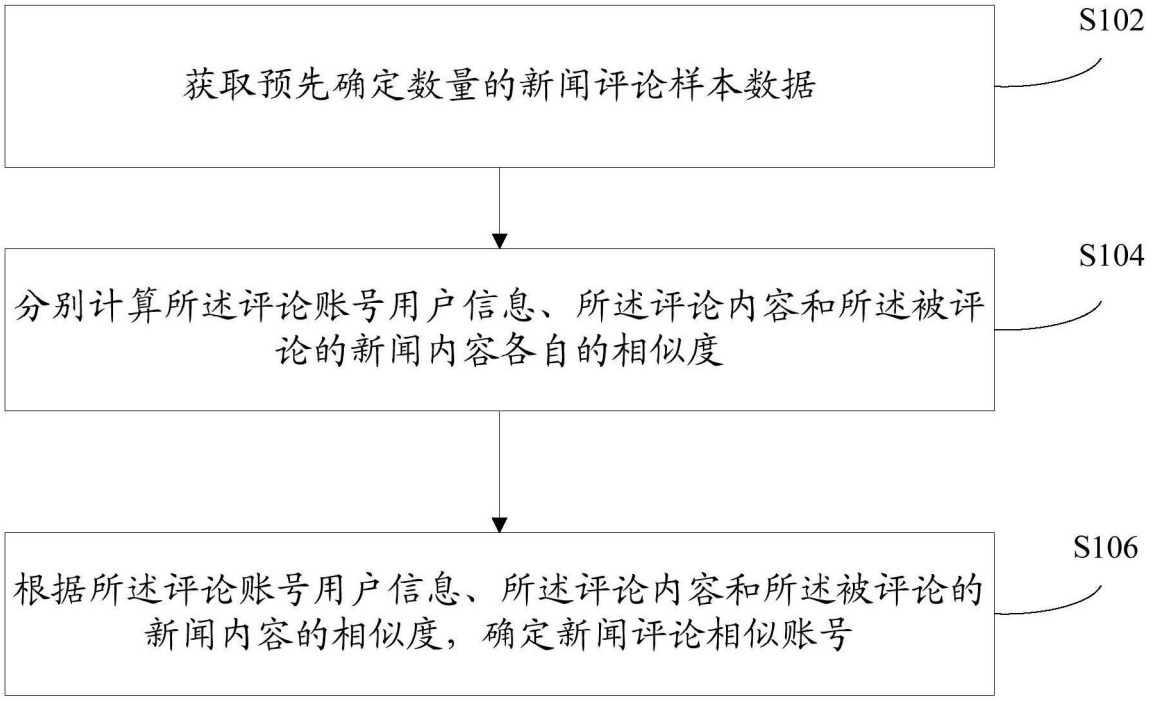

S102
获取预先确定数量的新闻评论样本数据
S104
分别计算所述评论账号用户信息、所述评论内容和所述被评论的新闻内容各自的相似度
S106
根据所述评论账号用户信息、所述评论内容和所述被评论的新闻内容的相似度，确定新闻评论相似账号

图1

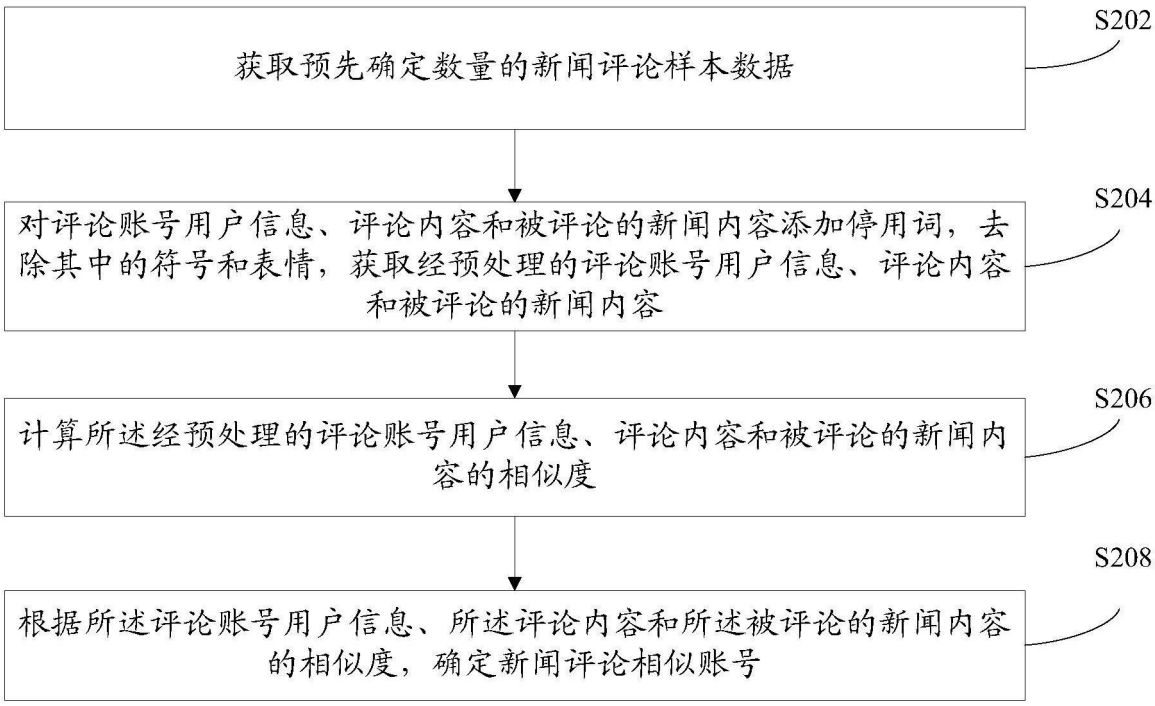

S202
获取预先确定数量的新闻评论样本数据
S204
对评论账号用户信息、评论内容和被评论的新闻内容添加停用词，去除其中的符号和表情，获取经预处理的评论账号用户信息、评论内容和被评论的新闻内容
S206
计算所述经预处理的评论账号用户信息、评论内容和被评论的新闻内容的相似度
S208
根据所述评论账号用户信息、所述评论内容和所述被评论的新闻内容的相似度，确定新闻评论相似账号

图2

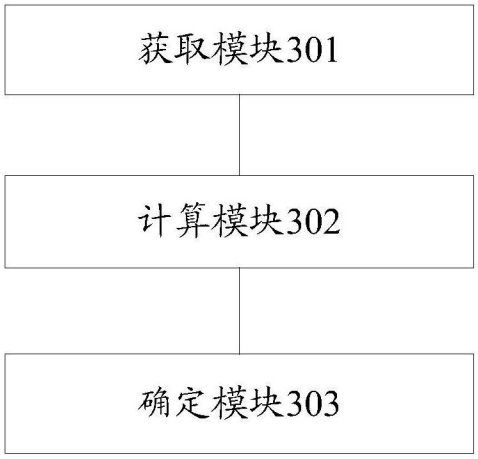


图3

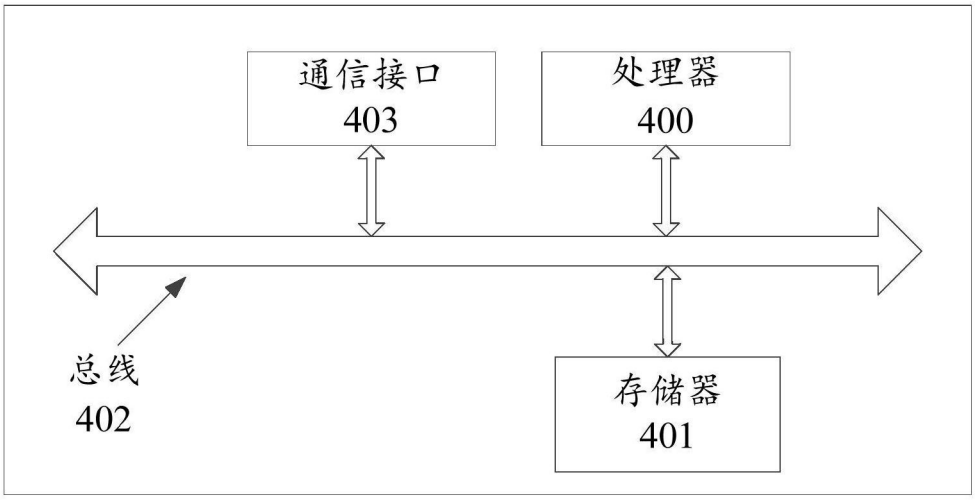


图4