(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 108021619 B
(45)授权公告日 2020.05.05

(21)申请号 201711115994.1

(22)申请日 2017.11.13

(65)同一申请的已公布的文献号
申请公布号 CN 108021619 A

(43)申请公布日 2018.05.11

(73)专利权人 星潮闪耀移动网络科技(中国)有限公司
地址 100193 北京市海淀区东北旺西路中关村软件园二期(西扩)N-1、N-2地块新浪总部科研楼5层517室

(72)发明人 杨宠 王晓栋

(74)专利代理机构 北京国昊天诚知识产权代理有限公司 11315
代理人 许志勇

(51)Int.Cl.
G06F 16/9535(2019.01)
G06F 40/30(2020.01)

(56)对比文件
CN 105488154 A,2016.04.13,
CN 103324665 A,2013.09.25,
CN 105447045 A,2016.03.30,
US 2007192300 A1,2007.08.16,

审查员 倪赛华

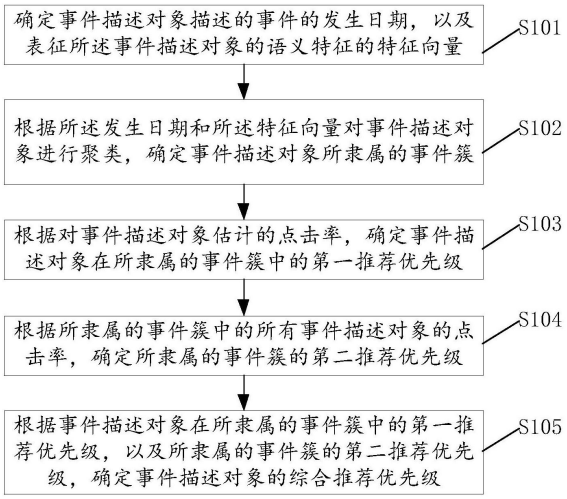
权利要求书4页 说明书15页 附图4页

(54)发明名称

一种事件描述对象推荐方法及装置

(57)摘要

本申请公开了一种事件描述对象推荐方法及装置,可以确定事件描述对象描述的事件的发生日期以及表征所述事件描述对象的语义特征的特征向量;根据发生日期和特征向量对事件描述对象进行聚类,确定事件描述对象所隶属的事件簇;根据对事件描述对象估计的点击率确定事件描述对象在所隶属的事件簇中的第一推荐优先级;根据所隶属的事件簇中的所有事件描述对象的点击率,确定所隶属的事件簇的第二推荐优先级;根据事件描述对象在所隶属的事件簇中的第一推荐优先级和所隶属的事件簇的第二推荐优先级,确定事件描述对象的综合推荐优先级。而不依赖于用户的历史浏览记录或点击数据,因此,可以感知用户的兴趣变化或隐藏性趣,满足用户的潜在需求。



CN 108021619 B

1.一种事件描述对象推荐方法,其特征在于,所述方法包括:

确定事件描述对象描述的事件的发生日期,以及表征所述事件描述对象的语义特征的特征向量;

根据所述发生日期和所述特征向量对事件描述对象进行聚类,确定事件描述对象所隶属的事件簇;

根据对事件描述对象估计的点击率,确定事件描述对象在所隶属的事件簇中的第一推荐优先级;

根据所隶属的事件簇中的所有事件描述对象的点击率,确定所隶属的事件簇的第二推荐优先级;

根据事件描述对象在所隶属的事件簇中的第一推荐优先级,以及所隶属的事件簇的第二推荐优先级,确定事件描述对象的综合推荐优先级;

所述根据对事件描述对象估计的点击率,确定事件描述对象在所隶属的事件簇中的第一推荐优先级,包括:

提取事件描述对象的多维度静态特征;

将所述多维度静态特征和所述特征向量进行合并,得到表征事件描述对象的组合向量;

根据所述组合向量和点击率预估模型,估计事件描述对象的点击率;所述点击率预估模型,是根据已推荐的事件描述对象的组合向量和真实点击率确定的、用于估计事件描述对象的点击率的模型;

根据估计出的事件描述对象的点击率的大小,确定事件描述对象在所隶属的事件簇中的第一推荐优先级的高低。

2.如权利要求1所述的方法,其特征在于,在所述根据所述发生日期和所述特征向量对事件描述对象进行聚类前,所述方法还包括:

根据事件领域模型确定事件描述对象所属的事件领域;所述事件领域模型是根据已知事件领域的事件描述对象确定的、用于确定事件描述对象属于预设事件领域的概率的模型;

根据事件描述对象所属的事件领域,对事件描述对象进行预聚类;则,

所述根据所述发生日期和所述特征向量对事件描述对象进行聚类,包括:

根据所述发生日期和所述特征向量对预聚类后的事件描述对象进行聚类。

3.如权利要求2所述的方法,其特征在于,所述根据事件领域模型确定事件描述对象所属的事件领域,包括:

确定事件描述对象的分词的词向量;

将事件描述对象的分词的词向量组成的矩阵输入事件领域模型,获得事件描述对象属于预设事件领域的概率;

根据事件描述对象属于预设事件领域的概率的大小,确定事件描述对象所属的事件领域。

4.如权利要求1所述的方法,其特征在于,所述确定表征事件描述对象的语义特征的特征向量,包括:

对事件描述对象进行分词处理获得事件描述对象的分词结果;

根据分词结果和语义特征向量模型，确定表征事件描述对象的语义特征的特征向量；所述语义特征向量模型为doc2vec模型。

5.如权利要求1所述的方法，其特征在于，所述根据所述发生日期和所述特征向量对事件描述对象进行聚类，确定事件描述对象所隶属的事件簇，包括：

根据所述发生日期确定事件描述对象描述的事件的时间属性，所述时间属性包括：未来型或当前型；

根据确定出的时间属性和所述发生日期，对事件描述对象进行预聚类；

根据所述特征向量对预聚类后的事件描述对象进行聚类，确定事件描述对象所隶属的事件簇。

6.如权利要求5所述的方法，其特征在于，所述根据所述特征向量对预聚类后的事件描述对象进行聚类，确定事件描述对象所隶属的事件簇，包括：

计算所述特征向量与目标事件簇的聚类中心的余弦相似度；所述聚类中心为表征事件簇中存储的事件描述对象的特征向量的平均向量；所述目标事件簇为与预聚类后的事件描述对象描述的事件的发生时间和时间属性相同的事件簇；

确定计算出的余弦相似度中的最大值是否大于第一阈值；

若为是，将所述最大值对应的目标事件簇确定为事件描述对象所隶属的事件簇。

7.如权利要求6所述的方法，其特征在于，若计算出的余弦相似度中的最大值不大于第一阈值，所述方法还包括：

当事件描述对象对应的时间属性为未来型时，新建事件簇作为事件描述对象所隶属的事件簇；

当事件描述对象对应的时间属性为当前型时，将事件描述对象丢弃。

8.如权利要求5所述的方法，其特征在于，当事件描述对象对应的时间属性为未来型时，所述方法还包括：

利用正则表达式提取事件描述对象中描述事件发生时间的第一时间词；

根据表征第一时间词的词向量，扩展出与所述第一时间词的语义相同或语义相近的第二时间词；

从未聚类的事件描述对象中检索出与所述第二时间词匹配的事件描述对象，并返回执行所述确定事件描述对象描述的事件的发生日期的步骤至所述确定事件描述对象的综合推荐优先级的步骤。

9.如权利要求5所述的方法，其特征在于，所述方法还包括：

在时间属性为未来型的事件发生日期之前的预设时间，按照事件描述对象的综合推荐优先级向用户推荐时间属性为未来型的第一事件描述对象；

若监测到所述用户关注所述第一事件描述对象，在所述第一事件描述对象描述的事件发生日或发生日后，按照事件描述对象的综合推荐优先级向用户推荐时间属性为当前型的第二事件描述对象；所述第二事件描述对象与所述第一事件描述对象描述的事件相关。

10.如权利要求1所述的方法，其特征在于，所述根据所隶属的事件簇中的所有事件描述对象的点击率，确定所隶属的事件簇的第二推荐优先级，包括：

对所隶属的事件簇中的所有事件描述对象的点击率进行加权求和，获得所隶属的事件簇的预估关注度；

根据所隶属的事件簇的预估关注度，确定所隶属的事件簇的第二推荐优先级。

11.一种事件描述对象推荐装置，其特征在于，所述装置包括：

第一确定模块，用于确定事件描述对象描述的事件的发生日期，以及表征所述事件描述对象的语义特征的特征向量；

第一聚类模块，用于根据所述发生日期和所述特征向量对事件描述对象进行聚类，确定事件描述对象所隶属的事件簇；

第一预估模块，用于根据对事件描述对象估计的点击率，确定事件描述对象在所隶属的事件簇中的第一推荐优先级；

第二预估模块，用于根据所隶属的事件簇中的所有事件描述对象的点击率，确定所隶属的事件簇的第二推荐优先级；

优先级确定模块，用于根据事件描述对象在所隶属的事件簇中的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级，确定事件描述对象的综合推荐优先级；

所述第一预估模块，包括：

提取单元，用于提取事件描述对象的多维度静态特征；

向量合并单元，用于将所述多维度静态特征和所述特征向量进行合并，得到表征事件描述对象的组合向量；

点击率预估单元，用于根据所述组合向量和点击率预估模型，估计事件描述对象的点击率；所述点击率预估模型是根据已推荐的事件描述对象的组合向量和真实点击率确定的、用于估计事件描述对象的点击率的模型；

优先级确定单元，用于根据估计出的事件描述对象的点击率的大小，确定事件描述对象在所隶属的事件簇中的第一推荐优先级的高低。

12.如权利要求11所述的装置，其特征在于，还包括：

领域确定模块，用于在所述根据所述发生日期和所述特征向量对事件描述对象进行聚类前，根据事件领域模型确定事件描述对象所属的事件领域；所述事件领域模型是根据已知事件领域的事件描述对象确定的、用于确定事件描述对象属于预设事件领域的概率的模型；

第二聚类模块，用于根据事件描述对象所属的事件领域，对事件描述对象进行预聚类；则，

所述第一聚类模块，具体用于根据所述发生日期和所述特征向量对预聚类后的事件描述对象进行聚类。

13.如权利要求12所述的装置，其特征在于，所述领域确定模块，具体用于确定事件描述对象的分词的词向量；将事件描述对象的分词的词向量组成的矩阵输入事件领域模型，获得事件描述对象属于预设事件领域的概率；根据事件描述对象属于预设事件领域的概率的大小，确定事件描述对象所属的事件领域。

14.如权利要求11所述的装置，其特征在于，所述第一确定模块，具体用于对事件描述对象进行分词处理获得事件描述对象的分词结果；根据分词结果和语义特征向量模型，确定表征事件描述对象的语义特征的特征向量；所述语义特征向量模型为doc2vec模型。

15.如权利要求11所述的装置，其特征在于，所述第一聚类模块包括：

时间属性确定单元，用于根据所述发生日期确定事件描述对象描述的事件的时间属

性,所述时间属性包括:未来型或当前型;

第一聚类单元,用于根据确定出的时间属性和所述发生日期,对事件描述对象进行预聚类;

第二聚类单元,用于根据所述特征向量对预聚类后的事件描述对象进行聚类,确定事件描述对象所隶属的事件簇。

16.如权利要求15所述的装置,其特征在于,所述第二聚类单元包括:

计算子单元,用于计算所述特征向量与目标事件簇的聚类中心的余弦相似度;所述聚类中心为表征事件簇中存储的事件描述对象的特征向量的平均向量;所述目标事件簇为与预聚类后的事件描述对象描述的事件的发生时间和时间属性相同的事件簇;

判断子单元,用于确定计算出的余弦相似度中的最大值是否大于第一阈值;

确定子单元,用于在所述判断子单元获得的判断结果为是时,将所述最大值对应的目标事件簇确定为事件描述对象所隶属的事件簇。

17.如权利要求16所述的装置,其特征在于,所述装置还包括:

第三聚类单元,用于在所述判断子单元获得的判断结果为否的情况下,当事件描述对象对应的时间属性为未来型时,新建事件簇作为事件描述对象所隶属的事件簇;当事件描述对象对应的时间属性为当前型时,将事件描述对象丢弃。

18.如权利要求15所述的装置,其特征在于,所述装置还包括:

时间词提取模块,用于当事件描述对象对应的时间属性为未来型时,利用正则表达式提取事件描述对象中描述事件发生时间的第一时间词;

时间词扩展模块,用于根据表征第一时间词的词向量,扩展出与所述第一时间词的语义相同或语义相近的第二时间词;

检索触发模块,用于从未聚类的事件描述对象中检索出与所述第二时间词匹配的事件描述对象,并触发所述第一确定模块至所述优先级确定模块。

19.如权利要求11所述的装置,其特征在于,所述装置还包括:

第一推荐模块,用于在时间属性为未来型的事件发生日期之前的预设时间,按照事件描述对象的综合推荐优先级向用户推荐时间属性为未来型的第一事件描述对象;

第二推荐模块,用于若监测到所述用户关注所述第一事件描述对象,在所述第一事件描述对象描述的事件发生日或发生日后,按照事件描述对象的综合推荐优先级向用户推荐时间属性为当前型的第二事件描述对象;所述第二事件描述对象与所述第一事件描述对象描述的事件相关。

20.如权利要求11所述的装置,其特征在于,所述第二预估模块,具体用于对所隶属的事件簇中的所有事件描述对象的点击率进行加权求和,获得所隶属的事件簇的预估关注度;根据所隶属的事件簇的预估关注度,确定所隶属事件簇的第二推荐优先级。

21.一种电子设备,其特征在于,包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述计算机程序被所述处理器执行时实现如权利要求1至10中任一项所述的方法的步骤。

22.一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如权利要求1至10中任一项所述的方法的步骤。

一种事件描述对象推荐方法及装置

技术领域

[0001] 本申请涉及计算机技术领域，尤其涉及一种事件描述对象推荐方法及装置。

背景技术

[0002] 当今社会正处于信息爆炸的时代，伴随着互联网技术的蓬勃发展，人们可以方便、快捷地从互联网上获得丰富多样的信息，例如用户可以通过安装在手机上的新闻客户端获得各类新闻资讯。与此同时，由于不同用户感兴趣的信息内容不尽相同，使得用户想要获得自己感兴趣的信息内容的需求越来越强烈。

[0003] 为了满足不同用户对信息内容的个性化需求，现有技术引入了个性化信息推荐技术。现有的个性化信息推荐技术，是结合用户的历史浏览记录和历史点击数据确定出用户的兴趣点，优先向用户推荐用户与该兴趣点匹配的、当前正在发生或已经发生的事件的相关信息的技术。

[0004] 由于现有的个性化信息推荐技术是基于用户实时或长期积累的兴趣进行推荐的，而用户的兴趣可能会随着未来世界的变化而发生变化，但现有的信息推荐技术无法感知用户的兴趣在未来可能发生的变化，这使得现有的信息推荐技术无法满足用户的潜在需求。

发明内容

[0005] 本申请实施例提供一种事件描述是对象方法及装置，以解决现有的信息推荐技术无法满足用户的潜在需求的技术问题。

[0006] 第一方面，本申请实施例提供一种事件描述对象推荐方法，所述方法包括：

[0007] 确定事件描述对象描述的事件的发生日期，以及表征所述事件描述对象的语义特征的特征向量；

[0008] 根据所述发生日期和所述特征向量对事件描述对象进行聚类，确定事件描述对象所隶属的事件簇；

[0009] 根据对事件描述对象估计的点击率，确定事件描述对象在所隶属的事件簇中的第一推荐优先级；

[0010] 根据所隶属的事件簇中的所有事件描述对象的点击率，确定所隶属的事件簇的第二推荐优先级；

[0011] 根据事件描述对象在所隶属的事件簇中的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级，确定事件描述对象的综合推荐优先级。

[0012] 第二方面，本申请实施例还提供一种事件描述对象推荐装置，所述装置包括：

[0013] 第一确定模块，用于确定事件描述对象描述的事件的发生日期，以及表征所述事件描述对象的语义特征的特征向量；

[0014] 第一聚类模块，用于根据所述发生日期和所述特征向量对事件描述对象进行聚类，确定事件描述对象所隶属的事件簇；

[0015] 第一预估模块，用于根据对事件描述对象估计的点击率，确定事件描述对象在所

隶属的事件簇中的第一推荐优先级；

[0016] 第二预估模块，用于根据所隶属的事件簇中的所有事件描述对象的点击率，确定所隶属事件簇的第二推荐优先级；

[0017] 优先级确定模块，用于根据事件描述对象在所隶属的事件簇中的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级，确定事件描述对象的综合推荐优先级。

[0018] 第三方面，本申请实施例还提供了一种电子设备，包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述计算机程序被所述处理器执行时实现如第一方面所述的方法的步骤。

[0019] 第四方面，本申请实施例还提供了一种计算机可读存储介质，所述计算机可读存储介质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如第一方面所述的方法的步骤。

[0020] 本申请实施例采用的上述至少一个技术方案，由于可以根据所述发生日期和所述特征向量对事件描述对象进行聚类，以确定事件描述对象所隶属的事件簇；可以估计出事件描述对象的点击率和事件描述对象在所隶属的事件簇的第一推荐优先级，和所隶属的事件簇的第二推荐优先级；然后根据事件描述对象在所隶属的事件簇中的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级，确定事件描述对象的综合推荐优先级。而不依赖于用户的历史浏览记录或点击数据确定事件描述对象的推荐优先级，因此，可以试探、挖掘或感知用户的兴趣变化或隐藏性趣，进而可以满足用户的潜在需求。

附图说明

[0021] 此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：

[0022] 图1为本申请实施例提供的一种事件描述对象推荐方法的一种具体实现方式的流程示意图；

[0023] 图2为本申请实施例提供的一种训练事件领域模型的训练过程的原理示意图；

[0024] 图3为图1所示实施例中的步骤S102的一种详细流程示意图；

[0025] 图4为本申请实施例提供的一种事件簇索引的层次结构示意图；

[0026] 图5为本申请实施例提供的一种事件描述对象推荐装置的一种具体实现方式的结构框图；

[0027] 图6为图5所示实施例中的模块502的一种详细结构框图；

[0028] 图7为本申请实施例提供的一种电子设备的结构示意图。

具体实施方式

[0029] 为使本申请的目的、技术方案和优点更加清楚，下面将结合本申请具体实施例及相应的附图对本申请技术方案进行清楚、完整地描述。显然，所描述的实施例仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通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，都属于本申请保护的范围。

[0030] 为了解决现有技术中的信息推荐技术无法感知用户的兴趣在未来可能发生的变化，进而无法满足用户的潜在需求的技术问题。例如，一个平时对体育不感兴趣的用户可能

会随着奥运会或世界杯的即将来临而对相关的新闻产生兴趣;再如,一个平时不太关注科技新闻的用户可能会对即将发布的新款手机(例如iPhone)产生兴趣,但是现有技术并不能发现用户的这些潜在兴趣。本申请实施例提供了一种事件描述对象推荐方法和装置,下面分别进行说明。

[0031] 下面先对本申请实施例提供的一种事件描述对象推荐方法进行说明。

[0032] 需要说明的是,实施本申请实施例提供的一种事件描述对象推荐方法的及装置的执行主体,可以是各事件描述对象推荐客户端的服务器,具体可以是事件描述对象推荐客户端的服务器上的一个数据分析平台或服务平台,例如具体可以是新闻客户端(如新浪新闻)的服务器上的Simba数据分析平台,或thrift服务。本申请实施例对实施上述方法及装置的具体实施主体不做限定。

[0033] 以下结合附图,详细说明本申请各实施例提供的技术方案。

[0034] 如图1所示,本申请实施例提供的一种事件描述对象推荐方法,可以包括如下步骤:

[0035] S101、确定事件描述对象描述的事件的发生日期,以及表征所述事件描述对象的语义特征的特征向量;

[0036] 事件可以是比较重大、对一定的人群会产生一定影响的事情。事件描述对象可以是能够描述事件的相关信息的载体。事件描述对象可以是文本形式的也可以是非文本形式的。其中,文本形式的事件描述对象既可以是短文本也可以是长文本,短文本例如可以是一条微博,长文本例如可以是一篇文章等;非文本形式的事件描述对象例如可以是一张图片、一段动画、一段视频或一段音频,等等。事件的发生日期可以是指事件的实际发生日期。

[0037] 在本申请实施例中,待聚类的事件描述对象的数量可以是一个也可以是多个,本申请实施例对此不做限定。

[0038] 在步骤S101中,确定事件描述对象描述的事件的发生日期的具体方式可以包括:利用正则表达式提取事件描述对象中描述事件发生时间的时间词;根据所述描述事件发生时间的时间词,确定事件描述对象描述的事件的发生日期。

[0039] 正则表达式是用于匹配预设字符串的表达式,例如,一段文本形式的事件描述对象为:金州勇士队将于11月12日在主场迎战迈阿密热火队,正则表达式的形式可以为“XX月XX日”。

[0040] 在步骤S101中,确定表征事件描述对象的语义特征的特征向量,具体可以包括:对事件描述对象进行分词处理获得事件描述对象的分词结果;根据分词结果和语义特征向量模型,确定表征事件描述对象的语义特征的特征向量。

[0041] 通常来讲,表征事件描述对象的语义特征的特征向量中的特征为事件描述对象的低维度特征,低维度特征可以理解为是可以从事件描述对象中直接获得或者对事件描述对象做少量的处理即可获得的特征。例如,若事件描述对象为一篇文章,那么这些低维度特征可以是:文章的统一资源定位符(Uniform/Universal Resource Locator,URL)、文章的标题、文章的作者、文章的发布时间、文章的分词结果等。

[0042] 相应的,需要说明的是,文章的高维度特征可以理解为是对事件描述对象的低维度特征进行分析计算得到的特征。例如,若事件描述对象为一篇文章,那么这些高维度特征可以是:文章所属的领域(体育领域还是科技领域等)、文章的主题(例如NBA比赛还是奥斯

卡颁奖典礼等)、文章的时效性(短期事件还是长期事件等)。

[0043] 在实际应用中,可以采用现有技术中的分词方式对事件描述对象进行分词处理,得到事件描述对象的分词结果;其中,语义特征向量模型可以为现有的doc2vec模型。

[0044] S102、根据所述发生日期和所述特征向量对事件描述对象进行聚类,确定事件描述对象所隶属的事件簇;

[0045] 具体来说,可以将事件描述对象归类至与该事件描述对象描述的事件的发生日期相同,并且与该事件描述对象的语义特征相似或相近的事件描述对象所隶属的事件簇中,从而确定出事件描述对象所隶属的事件簇。

[0046] 具体可以计算表征事件描述对象的特征向量与已聚类的同一发生日期对应的事件簇中的事件描述对象的特征向量的余弦相似度,来确定事件描述对象所隶属的事件簇。

[0047] 由于特征向量反映了事件描述对象的语义特征,而余弦相似度又能反映两个向量的相似程度,因此表征两个事件描述对象的特征向量的余弦相似度越大,说明两个事件描述对象的语义越相近,可以归至同一类中。

[0048] S103、根据对事件描述对象估计的点击率,确定事件描述对象在所隶属的事件簇中的第一推荐优先级;

[0049] 在一种具体实施方式中,步骤S103可以包括:

[0050] 子步骤1、提取事件描述对象的多维度静态特征;

[0051] 其中,多维静态特征可以包括:是否包含地域信息、标题、正文长度、段落数、标签数、时效性、新闻级别、人名数量、机构名数量、标题质量、版权状况、是否属于标题党、是否属于三俗文章、文本质量、文章情感类型、包含的图片数、推广信息数量、是否包含二维码、明星数量和所属媒体级别等特征中的一种或多种。

[0052] 子步骤2、将所述多维度静态特征和所述特征向量进行合并,得到表征事件描述对象的组合向量;

[0053] 子步骤3、根据所述组合向量和点击率预估模型,估计事件描述对象的点击率;所述点击率预估模型,是根据已推荐的事件描述对象的组合向量和真实点击率确定的、用于估计事件描述对象的点击率的模型;

[0054] 在一种具体实施方式中,点击率预估模型可以为根据已推荐的事件描述对象的组合向量和真实点击率训练获得的梯度提升决策树(Gradient Boosting Decision Tree,GBDT)模型,简称GBDT回归模型,下文会简要介绍GBDT模型的训练过程,详见下文。

[0055] 子步骤4、根据估计出的事件描述对象的点击率的大小,确定事件描述对象在所隶属的事件簇的中的第一推荐优先级的高低。一般而言,估计出的点击率越大,事件描述对象在所隶属的事件簇中的第一推荐优先级越高。

[0056] 可选地,为了降低存储事件簇的存储负担,当估计出的点击率小于第二阈值时,在步骤S103中,还可以将该点击率对应的事件描述对象丢弃。

[0057] 由于事件描述对象的点击率越高,说明用户对该事件描述对象描述的事件越感兴趣,反之,说明用户对该事件描述对象描述的事件的兴趣可能不高。因此,估计出未推荐的事件描述对象的点击率,也就意味着可以预估出用户对未推荐的事件描述对象描述的事件的感兴趣程度,从而可以试探或感知用户的潜在兴趣点。

[0058] 下面以事件描述对象为文章为例,对训练GBDT模型的过程进行简要地介绍。

[0059] 首先，获取已推荐的历史文章和这些历史文章对应的真实点击率构成训练集；其次，提取训练集中的各历史文章的多维静态特征，并利用doc2vec模型得到表征各历史文章的语义特征的doc2vec向量；再次，将各历史文章的多维静态特征和doc2vec向量组合在一起得到各历史文章的组合向量，并利用这些组合向量训练出多个弱分类器(训练弱分类器的过程属于现有技术，本文不再详述)；在实际训练多个弱分类器的过程中，训练文本(历史文章)的标签为真实点击率的95％威尔逊置信空间的下限，例如，假设训练文本的真实点击率为0.2，那么训练文本的标签就为0.2乘以0.95，等于0.19。最后为多个弱分类器分别分配权重得到GBDT回归模型。

[0060] 在实际应用中，将事件描述对象的多维度静态特征和doc2vec特征向量构成的组合向量输入GBDT回归模型可以得到估计出的点击率。

[0061] 上述威尔逊置信空间的思路是，虽然用户对一篇文章的点击率p(p＝u/v，其中，u为点击被推荐的文章的用户数量，v为收到被推荐的文章的用户的总数量)越大，就代表这篇文章的关注度越高，越应该优先推荐。但是，p的可信度与点击用户的数量密切相关，如果样本太小，p的可信度并不高。本领域技术人员知道，p是“二项分布”中某个事件的发生概率，因此我们可以计算出p的置信区间。

[0062] 所谓置信区间，可以理解为就某个概率而言，p会落在那个区间。比如，某个产品的好评率是80％，但是这个值不一定可信。根据统计学原理，我们只能说，有95％的把握可以确定该产品的好评率在75％到85％之间，即置信区间是[75％，85％]。这样处理的原理是，置信区间的宽窄与样本的数量有关。比如文章A被推荐给10个用户，其中8个用户对文章A进行了点击，2个用户未对文章A进行点击；文章B被推荐给100个用户，其中80个用户对文章B进行了点击，20个用户未对文章B进行点击。这两篇文章的点击率都是80％，但是文章B的置信区间(假设为[75％，85％])会比文章A的置信区间(假设为[70％，90％])窄，因此，文章B的置信区间的下限值(75％)会比文章A(70％)大，所以文章B应该排在文章A的前面。

[0063] 计算威尔逊置信度的公式如下：

[0064] $$\text{Score}=\frac{p+\frac{1}{2n}z^2_{1-\frac{\alpha}{2}}\pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}+\frac{z^2_{1-\frac{\alpha}{2}}}{4n^2}}}{1+\frac{1}{n}z^2_{1-\frac{\alpha}{2}}}$$

[0065] 其中，Score为置信度，p为一篇文章的实际点击率，n为样本的大小(接收被推荐的文章的用户的总数量)，z和α为常数，在本申请实施例中，z可以取95％。不难理解，利用上述公式就可以计算出历史文章的真实点击率的置信度。

[0066] S104、根据所隶属的事件簇中的所有事件描述对象的点击率，确定所隶属的事件簇的第二推荐优先级；

[0067] 在步骤S104中，可以对所隶属的事件簇中的所有事件描述对象的点击率进行加权求和，获得所隶属的事件簇的预估关注度；根据所隶属的事件簇的预估关注度，确定所隶属的事件簇的第二推荐优先级。通常情况下，所隶属的事件簇对应的预估关注度越高，说明该事件簇中的事件描述对象描述的事件越热门，该事件簇的第二推荐优先级也越高。

[0068] S105、根据事件描述对象在所隶属的事件簇中的第一推荐优先级，以及所隶属的

事件簇的第二推荐优先级，确定事件描述对象的综合推荐优先级。

[0069] 可以理解，事件描述对象在所隶属的事件簇中的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级均较高时，事件描述对象的综合推荐优先级也较高。

[0070] 在具体实现时，可以对各事件簇按照第二推荐优先级由高到低的顺序进行排序，同时对每一事件簇中的事件描述对象按照估计出的点击率由大到小的顺序进行排序得到事件描述对象的推荐序列，这样，可以快速地确定出综合优先级较高(排在推荐序列前面)的事件描述对象向用户进行推荐。

[0071] 还可以理解，事件描述对象的综合优先级越高，说明该事件描述对象提供的内容越优质，将这些优质的内容推荐给用户后，能更准确、快速地感知用户兴趣的变化。

[0072] 本申请实施例提供的一种事件描述对象推荐方法，由于可以根据事件描述对象描述的事件的发生日期和事件描述对象的特征向量对事件描述对象进行聚类，确定出事件描述对象所隶属的事件簇；并且，可以估计出事件描述对象的点击率和事件描述对象在所隶属的事件簇的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级；然后根据事件描述对象在所隶属的事件簇中的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级，确定事件描述对象的综合推荐优先级。而不依赖于用户的历史浏览记录或点击数据确定事件描述对象的推荐优先级，因此，可以试探、挖掘或感知用户的兴趣变化或隐藏性趣，因此可以满足用户的潜在需求。

[0073] 在另一个实施例中，在上述步骤S102之前，也即在根据所述发生日期和所述特征向量对事件描述对象进行聚类前，图1所示的一种事件描述对象推荐方法，还可以包括：

[0074] 步骤1、根据事件领域模型确定事件描述对象所属的事件领域；

[0075] 事件领域可以是常见的领域，例如，事件领域可以是：军事、科技、娱乐、体育和财经等等。在实际应用中，事件领域可以根据实际需要进行设定，本申请实施例对此不做限定。

[0076] 事件领域模型是根据已知事件领域的事件描述对象确定的、用于确定事件描述对象属于预设事件领域的概率的模型，具体可以是根据已知事件领域的事件描述对象的分词的词向量组成的矩阵训练得到的模型。

[0077] 这样，步骤1具体可以包括：确定事件描述对象的分词的词向量；将事件描述对象的分词的词向量组成的矩阵输入事件领域模型，获得事件描述对象属于预设事件领域的概率；根据事件描述对象属于预设事件领域的概率的大小，确定事件描述对象所属的事件领域。

[0078] 可选地，事件领域模型可以是利用已知事件领域的事件描述对象训练得到的卷积神经网络模型。下文会结合图2对训练卷积神经网络模型的过程进行简要说明，详见下文。

[0079] 步骤2、根据事件描述对象所属的事件领域，对事件描述对象进行预聚类。

[0080] 在此基础上，上述步骤S102具体可以包括：根据所述发生日期和所述特征向量对预聚类后的事件描述对象进行聚类。

[0081] 可以理解，对事件描述对象根据事件领域预聚类后，再根据事件的发生时间和事件描述对象的特征向量进行下一步的聚类，可以缩小对事件描述对象进行聚类的范围，从而减小对事件描述对象进行聚类的计算量、缩短对事件描述对象进行聚类的时间。

[0082] 下面以事件描述对象为文章为例，结合图2对训练卷积神经网络模型的过程进行

简要说明。

[0083] 首先，可以从互联网爬取相关网站(如新浪新闻)已推送的多篇文章，并对其中的一部分文章描述的事件的事件领域进行人工标注，得到训练集；

[0084] 其次，对训练集中的各文章进行分词处理，得到各文章的分词结果；

[0085] 最后，将文章的分词结果输入word2vec模型获得文章的分词的词向量；将文章的分词的词向量组成的矩阵作为卷积神经网络的输入，训练得到卷积神经网络模型各层的参数，从而得到可以确定事件描述对象描述的对象属于预设事件领域的概率的卷积神经网络模型。

[0086] 示例性地，如图2所示，上述训练过程可以为：步骤S1、将文章的分词的词向量组成的m×n维的矩阵11(例如图2中所示的9×6维的矩阵)输入卷积神经网络模型；步骤S2、使用卷积核对m×n维的矩阵中的特征进行组合得到m×1维的矩阵12；步骤S3、使用激活函数在池化层对m×1维的矩阵12进行降维得到矩阵j×1维的矩阵13，其中，j小于m；步骤S4、通过softmax函数确定文章描述的事件属于预设事件领域的概率，并通过梯度下降算法和向后传播算法优化卷积神经网络模型的损失函数，最终得到可用的卷积神经网络模型。

[0087] 如图3所示，在本申请的又一个实施例中，图1所示的一种事件描述对象推荐方法中的步骤S102可以包括：

[0088] S301、根据所述发生日期确定事件描述对象描述的事件的时间属性；

[0089] 其中，时间属性可以包括：未来型或当前型；具体可以根据事件的发生日期与当前日期的先后关系来确定事件的事件属性。一般来说，将发生日期晚于当前日期的事件的时间属性确定为未来型，而将发生日期为当前日期或当前日期之前的日期的事件的时间属性确定为当前型。例如，假设当前日期是2017年11月10日(星期五)，那么在2017年11月10日发布的微博“由成龙主演的新电影《新警察故事》即将在本周末上映”描述的就是时间属性为未来型的事件，而在2017年11月12日发布的微博“成龙主演的新电影《新警察故事》首映日票房破亿”描述的是时间属性为当前型的事件。

[0090] S302、根据确定出的时间属性和所述发生日期，对事件描述对象进行预聚类；

[0091] 具体可以将同一发生日期、同一时间属性的事件描述对象归为一类。

[0092] S303、根据所述特征向量对预聚类后的事件描述对象进行聚类，确定事件描述对象所隶属的事件簇。

[0093] 在一种具体实现方式中，步骤S303具体可以包括：

[0094] 子步骤1、计算所述特征向量与目标事件簇的聚类中心的余弦相似度；

[0095] 其中，聚类中心为表征事件簇中存储的事件描述对象的总体语义特征的向量，具体可以用事件簇中存储的各事件描述对象的特征向量的平均向量来表征，平均向量可以通过对事件簇中存储的各事件描述对象的特征向量求和之后再求平均得到。

[0096] 其中，目标事件簇为与预聚类后的事件描述对象描述的事件的发生时间和时间属性相同的事件簇；

[0097] 子步骤2、确定计算出的余弦相似度中的最大值是否大于第一阈值；若为是执行下述子步骤3；否则，执行下述子步骤4；

[0098] 第一阈值的取值范围通常在[0.7,1]之间，例如可以是0.8，实际应用中可以根据实际情况进行设定，此处不做限定。

[0099] 子步骤3、将所述最大值对应的目标事件簇确定为事件描述对象所隶属的事件簇；

[0100] 由于余弦相似度的大小可以反映两个向量的相似程度，因此通过上述过程可以将事件描述对象聚类至与该事件描述对象的语义特征最相近的事件簇中。

[0101] 不难理解，对事件描述对象根据事件的发生时间和时间属性预聚类后，再基于事件聚类模型进行下一步的聚类，也可以缩小对事件描述对象进行聚类的范围，从而减小对事件描述对象进行聚类的计算量、缩短对事件描述对象进行聚类的时间。

[0102] 子步骤4、当事件描述对象对应的时间属性为未来型时，新建事件簇作为事件描述对象所隶属的事件簇；当事件描述对象对应的时间属性为当前型时，将事件描述对象丢弃。

[0103] 可以理解，如果一篇文章描述的事件既不属于未来型的事件，又没有与之相似的当前类型的事件，说明书这篇文章为普通文章，不具有推荐价值，可以将该文章进行丢弃处理。

[0104] 在实际应用中，可以构建一个事件簇索引，并将聚类后的事件簇对应存储至该事件簇索引中。例如如图4所示，可以构建一个按照事件的发生日期、事件领域、事件描述对象的语义特征和事件的时间属性聚类的事件簇索引。在图4中，在同一发生日期下可以对应包括n个事件领域，每一时间领域下可以对应包括n类语义特征，每一语义特征下可以分别对应包括未来型的事件簇和当前型的事件簇。并且，当某一事件簇为当前型的事件簇时，对应的事件发生日期可以是事件描述对象的发布日期，当前型的事件簇用于描述未来事件的后续进展。事件簇索引中的事件簇会因新的事件描述对象的发布而不断更新，上文中通过步骤S101-S102对事件描述对象进行聚类的过程，也可以理解为是对事件簇索引进行更新的一个过程。

[0105] 可选地，当监测到事件簇索引被更新以后，可以向执行本申请实施例提供的一种事件描述对象推荐方法的执行主体的前端发送更新通知，该更新通知用于通知前端事件簇索引发生了更新，以供前端决定是否向用户推荐已更新的事件簇索引中的事件描述对象。

[0106] 在又一个实施例中，上述任一实施例提供的一种事件描述对象推荐方法，还可以包括：

[0107] 步骤1、在时间属性为未来型的事件发生日期之前的预设时间，按照事件描述对象的综合推荐优先级向用户推荐时间属性为未来型的第一事件描述对象；

[0108] 在本申请实施例中，可以将按照综合推荐优先级推荐给用户的、未来型的第一事件描述对象形象地称为“明日头条”。

[0109] 步骤2、若监测到所述用户关注所述第一事件描述对象，在所述第一事件描述对象描述的事件发生日或发生日后，按照事件描述对象的综合推荐优先级向用户推荐时间属性为当前型的第二事件描述对象；所述第二事件描述对象与所述第一事件描述对象描述的事件相关，或者说所述第二事件描述对象描述了所述第一时间描述对象描述的事件的后续进展。

[0110] 例如，在2017年11月10日星期五将“本周日世界杯即将开赛”的第一事件描述对象推荐给用户之后，如果用户点击了第一事件描述对象，说明用户对第一事件描述对象描述的未来型感兴趣，则可以在本周日(11月12日)当天将世界杯开赛当天的赛况信息作为第二事件描述对象推荐给用户，以便于用户了解本次世界杯的后续进展。

[0111] 可以想象，由于是在未来型的事件发生日期之前将描述未来型的第一事件描述对

象推荐给用户，因此，可以通过用户是否关注该第一事件描述对象试探、挖掘或感知用户的兴趣变化或隐藏性趣；进一步地，由于能够在感知到用户的兴趣发生变化后，在未来型的事件发生日或发生日之后继续将描述该未来事件的第二事件描述对象推荐给用户，因此可以满足用户的潜在需求。也即本申请实施例提供的一种事件描述对象推荐方法可以解决现有的信息推荐技术无法满足用户的潜在需求的技术问题。

[0112] 总而言之，本申请实施例提供的一种事件描述对象推荐方法，融合了word2vec、doc2vec、卷积神经网络模型和GBDT回归模型等多种自然语言处理算法，并结合流式计算平台，实时对事件描述对象进行聚类，以对存储的事件簇进行更新，从而挖掘出未来即将发生的热点事件的事件描述对象，并确定出这些事件描述对象的综合推荐优先级，从而将优质的事件描述对象推荐给用户，因此能够感知和试探用户隐藏的兴趣，进而满足用户的潜在需求。

[0113] 在实际推荐时，可以事先为聚类后的事件描述对象设置一个唯一标识该事件描述对象的ID，执行本申请实施例提供的一种事件描述对象推荐方法的执行主体的前端可以根据事件描述对象的ID、事件的发生日期、事件的时间属性、事件领域向用户推送信息。

[0114] 在又一个实施例中，上述任一实施例提供的一种事件描述对象推荐方法，还可以包括：

[0115] 步骤1、利用正则表达式提取事件描述对象中描述事件发生时间的第一时间词；

[0116] 时间词可以是任何能表示时间的词汇，例如：XX月XX日、七夕节、乞巧节、中国情人节、本周五、本周末等。

[0117] 如前文所述，正则表达式是用于匹配预设字符串的表达式，例如，一段文本形式的事件描述对象为：金州勇士队将于11月12日在主场迎战迈阿密热火队，正则表达式的形式可以为“XX月XX日”。

[0118] 在实际应用中，用于检索第一时间词和日期相关的正则表达式有若干个，例如：“XXX电影将于本周六上映”、“XXX电视剧将于中秋节当天与大家见面”、“11月10日晚，我们一起关注广州恒大的比赛”。

[0119] 步骤2、根据表征第一时间词的词向量，扩展出与所述第一时间词的语义相同或语义相近的第二时间词；

[0120] 具体可以使用word2vec模型进行第一时间词的扩展得到第二时间词。例如，当第一时间词为“七夕节”时，可以扩展出“乞巧节”、“中国情人节”等第二时间词。

[0121] 步骤3、从未聚类的事件描述对象中检索出与所述第二时间词匹配的事件描述对象，并返回执行上述步骤S101至S105，也即返回执行所述确定事件描述对象描述的事件的发生日期的步骤至所述确定事件描述对象的综合推荐优先级的步骤。

[0122] 还可以将检索到的与第二时间词匹配的事件描述对象中的表示事件发生日期的词汇，与当前日期进行比较，将未来预设天数内(例如未来7天内)内会发生的事件标记为未来型的事件，并对这些未来型的事件对应的事件描述对象返回执行上述步骤S101至S105。

[0123] 该实施例提供的事件描述对象推荐方法，可以检索获得尽可能多的事件描述对象，并进行聚类，从而能够最大限度地挖掘出优质事件描述对象，当将这些优质事件描述对象推荐给用户时，可以更好的感知用户兴趣的变化，能够更好的满足用户的潜在需求。

[0124] 相应于上述方法实施例，本申请实施例还提供了一种事件描述对象推荐装置，下

面进行介绍。

[0125] 如图5所示，本申请实施例提供的一种事件描述对象推荐装置，可以包括：第一确定模块501、第一聚类模块502、第一预估模块503、第二预估模块504和优先级确定模块505。

[0126] 第一确定模块501，用于确定事件描述对象描述的事件的发生日期，以及表征所述事件描述对象的语义特征的特征向量；

[0127] 事件可以是比较重大、对一定的人群会产生一定影响的事情。事件描述对象可以是能够描述事件的相关信息的载体。

[0128] 在本申请实施例中，待聚类的事件描述对象的数量可以是一个也可以是多个，本申请实施例对此不做限定。

[0129] 在第一确定模块501中，具体可以利用正则表达式提取事件描述对象中描述事件发生时间的时间词；根据所述描述事件发生时间的时间词，确定事件描述对象描述的事件的发生日期。其中，正则表达式是用于匹配预设字符串的表达式。

[0130] 在第一确定模块501中，具体可以对事件描述对象进行分词处理获得事件描述对象的分词结果；根据分词结果和语义特征向量模型，确定表征事件描述对象的语义特征的特征向量；其中，所述语义特征向量模型可以为现有的doc2vec模型。

[0131] 通常来讲，表征事件描述对象的语义特征的特征向量中的特征为事件描述对象的低维度特征，关于低维度特征的具体说明请参见方法实施例部分，此处不再重复描述。

[0132] 第一聚类模块502，用于根据所述发生日期和所述特征向量对事件描述对象进行聚类，确定事件描述对象所隶属的事件簇；

[0133] 具体来说，第一聚类模块502可以将事件描述对象归类至与该事件描述对象描述的事件的发生日期相同，并且与该事件描述对象的语义特征相似或相近的事件描述对象所隶属的事件簇中，从而确定出事件描述对象所隶属的事件簇。

[0134] 更为详细的，第一聚类模块502可以计算表征事件描述对象的特征向量与已聚类的同一发生日期对应的事件簇中的事件描述对象的特征向量的余弦相似度，来确定事件描述对象所隶属的事件簇。

[0135] 由于特征向量反映了事件描述对象的语义特征，而余弦相似度又能反映两个向量的相似程度，因此表征两个事件描述对象的特征向量的余弦相似度越大，说明两个事件描述对象的语义越相近，可以归至同一类中。

[0136] 第一预估模块503，用于根据对事件描述对象估计的点击率，确定事件描述对象在所隶属的事件簇中的第一推荐优先级；

[0137] 在一种具体实施方式中，第一预估模块503，可以包括：提取单元、向量合并单元、点击率预估单元和优先级确定单元。

[0138] 提取单元，用于提取事件描述对象的多维度静态特征；

[0139] 向量合并单元，用于将所述多维度静态特征和所述特征向量进行合并，得到表征事件描述对象的组合向量；

[0140] 点击率预估单元，用于根据所述组合向量和点击率预估模型，估计事件描述对象的点击率；所述点击率预估模型是根据已推荐的事件描述对象的组合向量和真实点击率确定的、用于估计事件描述对象的点击率的模型；

[0141] 点击率预估模型可以为根据已推荐的事件描述对象的组合向量和真实点击率训

练获得的GBDT回归模型。

[0142] 优先级确定单元，用于根据估计出的事件描述对象的点击率的大小，确定事件描述对象在所隶属的事件簇的中的第一推荐优先级的高低。一般而言，估计出的点击率越大，事件描述对象在所隶属的事件簇中的第一推荐优先级越高。

[0143] 可选地，为了降低存储事件簇的存储负担，第一预估模块503，可以包括：丢弃单元，用于当估计出的点击率小于第二阈值时，将该点击率对应的事件描述对象丢弃。

[0144] 不难理解，由于事件描述对象的点击率越高，说明用户对该事件描述对象描述的事件越感兴趣，反之，说明用户对该事件描述对象描述的事件的兴趣可能不高。因此，估计出未推荐的事件描述对象的点击率，也就意味着可以预估出用户对未推荐的事件描述对象描述的事件的感兴趣程度，从而可以试探或感知用户的潜在兴趣点。

[0145] 第二预估模块504，用于根据所隶属的事件簇中的所有事件描述对象的点击率，确定所隶属的事件簇的第二推荐优先级；

[0146] 在第二预估模块504中，可以对估计出的事件簇中的所有事件描述对象的点击率进行加权求和，获得所隶属的事件簇的预估关注度；根据事件簇的预估关注度，确定所隶属的事件簇的第二推荐优先级。通常情况下，事件簇对应的预估关注度越高，说明该事件簇中的事件描述对象描述的事件越热门，该事件簇的第二推荐优先级也越高。

[0147] 优先级确定模块505，用于根据事件描述对象在所隶属的事件簇中的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级，确定事件描述对象的综合推荐优先级。

[0148] 可以理解，事件描述对象在所隶属的事件簇中的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级均较高时，事件描述对象的综合推荐优先级也较高。

[0149] 在具体实现时，可以对各事件簇按照第二推荐优先级由高到低的顺序进行排序，同时对每一事件簇中的事件描述对象按照估计出的点击率由大到小的顺序进行排序得到事件描述对象的推荐序列，这样，可以快速地确定出综合优先级较高(排在推荐序列前面)的事件描述对象向用户进行推荐。

[0150] 还可以理解，事件描述对象的综合优先级越高，说明该事件描述对象提供的内容越优质，将这些优质的内容推荐给用户后，能更准确、快速地感知用户兴趣的变化。

[0151] 本申请实施例提供的一种事件描述对象推荐装置，由于可以根据事件描述对象描述的事件的发生日期和事件描述对象的特征向量对事件描述对象进行聚类，确定出事件描述对象所隶属的事件簇；并且，可以估计出事件描述对象的点击率和事件描述对象在所隶属的事件簇的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级；然后根据事件描述对象在所隶属的事件簇中的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级，确定事件描述对象的综合推荐优先级。而不依赖于用户的历史浏览记录或点击数据确定事件描述对象的推荐优先级，因此，可以试探、挖掘或感知用户的兴趣变化或隐藏性趣，因此可以满足用户的潜在需求。

[0152] 在另一个实施例中，图5所示的一种事件描述对象推荐装置，还可以包括：领域确定模块和第二聚类模块。

[0153] 领域确定模块，用于在所述根据所述发生日期和所述特征向量对事件描述对象进行聚类前，根据事件领域模型确定事件描述对象所属的事件领域；

[0154] 事件领域可以是常见的领域，例如，事件领域可以是：军事、科技、娱乐、体育和财

经等等。

[0155] 事件领域模型是根据已知事件领域的事件描述对象确定的、用于确定事件描述对象属于预设事件领域的概率的模型，具体可以是根据已知事件领域的事件描述对象的分词的词向量组成的矩阵训练得到的模型。

[0156] 这样，上述领域确定模块，具体可以用于确定事件描述对象的分词的词向量；将事件描述对象的分词的词向量组成的矩阵输入事件领域模型，获得事件描述对象属于预设事件领域的概率；根据事件描述对象属于预设事件领域的概率的大小，确定事件描述对象所属的事件领域。

[0157] 可选地，事件领域模型可以是利用已知事件领域的事件描述对象训练得到的卷积神经网络模型。

[0158] 第二聚类模块，用于根据事件描述对象所属的事件领域，对事件描述对象进行预聚类；

[0159] 则在此基础上，第一聚类模块502，具体用于根据所述发生日期和所述特征向量对预聚类后的事件描述对象进行聚类。

[0160] 可以理解，对事件描述对象根据事件领域预聚类后，再根据事件的发生时间和事件描述对象的特征向量进行下一步的聚类，可以缩小对事件描述对象进行聚类的范围，从而减小对事件描述对象进行聚类的计算量、缩短对事件描述对象进行聚类的时间。

[0161] 在本申请的又一个实施例中，如图6所示，图5所示的一种事件描述对象推荐装置中的第一聚类模块502具体可以包括：时间属性确定单元601、第一聚类单元602和第二聚类单元603。

[0162] 时间属性确定单元601，用于根据所述发生日期确定事件描述对象描述的事件的时间属性，所述时间属性包括：未来型或当前型；

[0163] 第一聚类单元602，用于根据确定出的时间属性和所述发生日期，对事件描述对象进行预聚类；

[0164] 第二聚类单元603，用于根据所述特征向量对预聚类后的事件描述对象进行聚类，以确定事件描述对象所隶属的事件簇。

[0165] 其中，第二聚类单元603具体可以包括：计算子单元、判断子单元、确定子单元和第三聚类单元。

[0166] 计算子单元，用于计算所述特征向量与目标事件簇的聚类中心的余弦相似度；

[0167] 聚类中心为表征事件簇中存储的事件描述对象的特征向量的平均向量。

[0168] 目标事件簇为与预聚类后的事件描述对象描述的事件的发生时间和时间属性相同的事件簇。

[0169] 判断子单元，用于确定计算出的余弦相似度中的最大值是否大于第一阈值；

[0170] 第一阈值的取值范围通常在[0.7,1]之间，例如可以是0.8，实际应用中可以根据实际情况进行设定，此处不做限定。

[0171] 确定子单元，用于在所述判断子单元获得的判断结果为是时，将所述最大值对应的目标事件簇确定为事件描述对象所隶属的事件簇。

[0172] 由于余弦相似度的大小可以反映两个向量的相似程度，因此通过上述过程可以将事件描述对象聚类至与该事件描述对象的语义特征最相近的事件簇中。

[0173] 不难理解，对事件描述对象根据事件的发生时间和时间属性预聚类后，再基于事件聚类模型进行下一步的聚类，也可以缩小对事件描述对象进行聚类的范围，从而减小对事件描述对象进行聚类的计算量、缩短对事件描述对象进行聚类的时间。

[0174] 第三聚类单元，用于在所述判断子单元获得的判断结果为否的情况下，当事件描述对象对应的时间属性为未来型时，新建事件簇作为事件描述对象所隶属的事件簇；当事件描述对象对应的时间属性为当前型时，将事件描述对象丢弃。

[0175] 可以理解，如果一篇文章描述的事件既不属于未来型的事件，又没有与之相似的当前类型的事件，说明书这篇文章为普通文章，不具有推荐价值，可以将该文章进行丢弃处理。

[0176] 在又一个实施例中，上述任一实施例提供的一种事件描述对象推荐装置还可以包括：第一推荐模块和第二推荐模块。

[0177] 第一推荐模块，用于在时间属性为未来型的事件发生日期之前的预设时间，按照事件描述对象的综合推荐优先级向用户推荐时间属性为未来型的第一事件描述对象；

[0178] 在本申请实施例中，可以将按照综合推荐优先级推荐给用户的、未来型的第一事件描述对象形象地称为“明日头条”。

[0179] 第二推荐模块，用于若监测到所述用户关注所述第一事件描述对象，在所述第一事件描述对象描述的事件发生日或发生日后，按照事件描述对象的综合推荐优先级向用户推荐时间属性为当前型的第二事件描述对象；所述第二事件描述对象与所述第一事件描述对象描述的事件相关。

[0180] 可以想象，由于是在未来型的事件发生日期之前将描述未来型的第一事件描述对象推荐给用户，因此，可以通过用户是否关注该第一事件描述对象试探、挖掘或感知用户的兴趣变化或隐藏性趣；进一步地，由于能够在感知到用户的兴趣发生变化后，在未来型的事件发生日或发生日之后继续将描述该未来事件的第二事件描述对象推荐给用户，因此可以满足用户的潜在需求。也即本申请实施例提供的一种事件描述对象推荐方法可以解决现有的信息推荐技术无法满足用户的潜在需求的技术问题。

[0181] 总而言之，本申请实施例提供的一种事件描述对象推荐装置，融合了word2vec、doc2vec、卷积神经网络模型和GBDT回归模型等多种自然语言处理算法，并结合流式计算平台，实时对事件描述对象进行聚类，以对存储的事件簇进行更新，从而挖掘出未来即将发生的热点事件的事件描述对象，并确定出这些事件描述对象的综合推荐优先级，从而将优质的事件描述对象推荐给用户，因此能够感知和试探用户隐藏的兴趣，进而满足用户的潜在需求。

[0182] 在又一个实施例中，上述任一实施例提供的一种事件描述对象推荐装置还可以包括：时间词提取模块、时间词扩展模块和检索触发模块。

[0183] 时间词提取模块，用于当事件描述对象对应的时间属性为未来型时，利用正则表达式提取事件描述对象中描述事件发生时间的第一时间词；

[0184] 时间词可以是任何能表示时间的词汇，例如：XX月XX日、七夕节、乞巧节、中国情人节、本周五、本周末等。

[0185] 时间词扩展模块，用于根据表征第一时间词的词向量，扩展出与所述第一时间词的语义相同或语义相近的第二时间词；

[0186] 具体可以使用word2vec模型进行第一时间词的扩展得到第二时间词。例如，当第一时间词为“七夕节”时，可以扩展出“乞巧节”、“中国情人节”等第二时间词。

[0187] 检索触发模块，用于从未聚类的事件描述对象中检索出与所述第二时间词匹配的事件描述对象，并触发上述第一确定模块501至优先级确定模块505。

[0188] 该实施例提供的事件描述对象推荐装置，可以检索获得尽可能多的事件描述对象，并进行聚类，从而能够最大限度地挖掘出优质事件描述对象，当将这些优质事件描述对象推荐给用户时，可以更好的感知用户兴趣的变化，能够更好的满足用户的潜在需求。

[0189] 需要说明的是，由于装置实施例执行的内容与方法实施例类似，因此，本文对装置实施例部分描述的较为简略，相关之处请参见方法实施例部分。

[0190] 图7示出了是本申请实施例提供的一种电子设备的结构示意图。请参考图7，在硬件层面，该电子设备包括处理器，可选地还包括内部总线、网络接口、存储器。其中，存储器可能包含内存，例如高速随机存取存储器(Random-Access Memory，RAM)，也可能还包括非易失性存储器(non-volatile memory)，例如至少1个磁盘存储器等。当然，该电子设备还可能包括其他业务所需要的硬件。

[0191] 处理器、网络接口和存储器可以通过内部总线相互连接，该内部总线可以是ISA(Industry Standard Architecture，工业标准体系结构)总线、PCI(Peripheral Component Interconnect，外设部件互连标准)总线或EISA(Extended Industry Standard Architecture，扩展工业标准结构)总线等。所述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7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。

[0192] 存储器，用于存放程序。具体地，程序可以包括程序代码，所述程序代码包括计算机操作指令。存储器可以包括内存和非易失性存储器，并向处理器提供指令和数据。

[0193] 处理器从非易失性存储器中读取对应的计算机程序到内存中然后运行，在逻辑层面上形成事件描述对象推荐装置。处理器，执行存储器所存放的程序，并具体用于执行本申请实施例提供的事件描述对象推荐方法。

[0194] 上述如本申请图7所示实施例揭示的事件描述对象推荐装置执行的方法可以应用于处理器中，或者由处理器实现。处理器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。上述的处理器可以是通用处理器，包括中央处理器(Central Processing Unit，CPU)、网络处理器(Network Processor，NP)等；还可以是数字信号处理器(Digital Signal Processor，DSP)、专用集成电路(Application Specific Integrated Circuit，ASIC)、现场可编程门阵列(Field－Programmable Gate Array，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申请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。结合本申请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，处理器读取存储器中的信息，结合其硬件完成上述方法的步骤。

[0195] 本申请实施例还提出了一种计算机可读存储介质，该计算机可读存储介质存储一个或多个程序，该一个或多个程序包括指令，该指令当被包括多个应用程序的电子设备执行时，能够使该电子设备执行图7所示实施例中事件描述对象推荐装置执行的方法，并具体用于执行本申请实施例提供的事件描述对象推荐方法。

[0196] 本领域内的技术人员应明白，本申请的实施例可提供为方法、系统、或计算机程序产品。因此，本申请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，本申请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、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0197] 本申请是参照根据本申请实施例的方法、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。

[0198] 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。

[0199] 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。

[0200] 需要说明的是，本申请中的各个实施例均采用相关的方式描述，各个实施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。

[0201] 还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0202] 以上仅为本申请的实施例而已，并不用于限制本申请。对于本领域技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在本申请的权利要求范围之内。

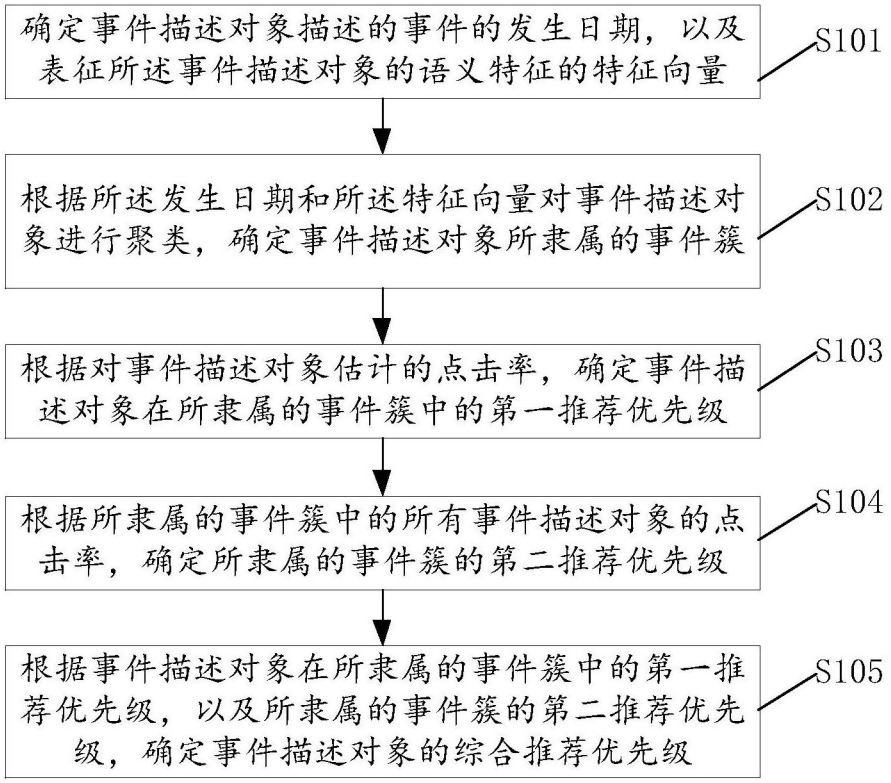

确定事件描述对象描述的事件的发生日期，以及表征所述事件描述对象的语义特征的特征向量
S101
根据所述发生日期和所述特征向量对事件描述对象进行聚类，确定事件描述对象所隶属的事件簇
S102
根据对事件描述对象估计的点击率，确定事件描述对象在所隶属的事件簇中的第一推荐优先级
S103
根据所隶属的事件簇中的所有事件描述对象的点击率，确定所隶属的事件簇的第二推荐优先级
S104
根据事件描述对象在所隶属的事件簇中的第一推荐优先级，以及所隶属的事件簇的第二推荐优先级，确定事件描述对象的综合推荐优先级
S105

图1

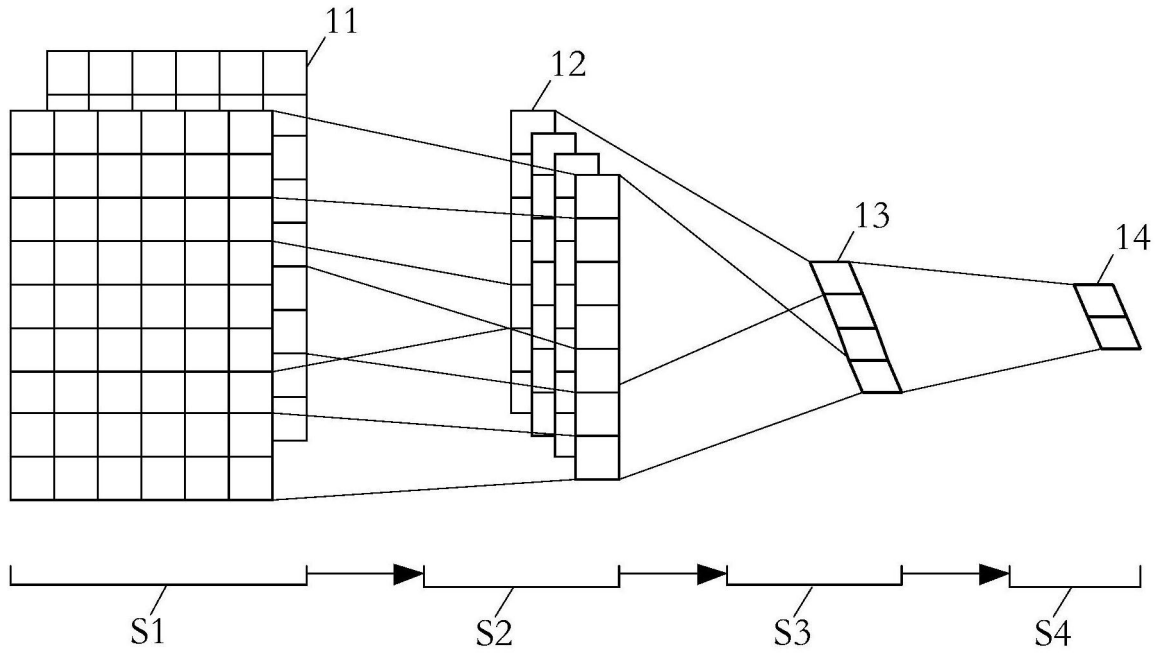

11
12
13
14
S1
S2
S3
S4

图2

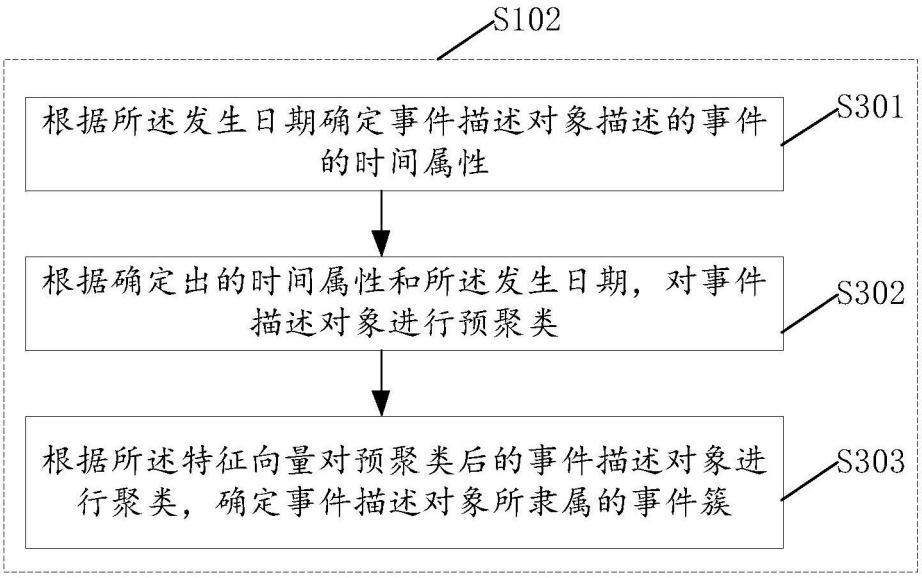

S102
根据所述发生日期确定事件描述对象描述的事件的时间属性
S301
根据确定出的时间属性和所述发生日期，对事件描述对象进行预聚类
S302
根据所述特征向量对预聚类后的事件描述对象进行聚类，确定事件描述对象所隶属的事件簇
S303

图3

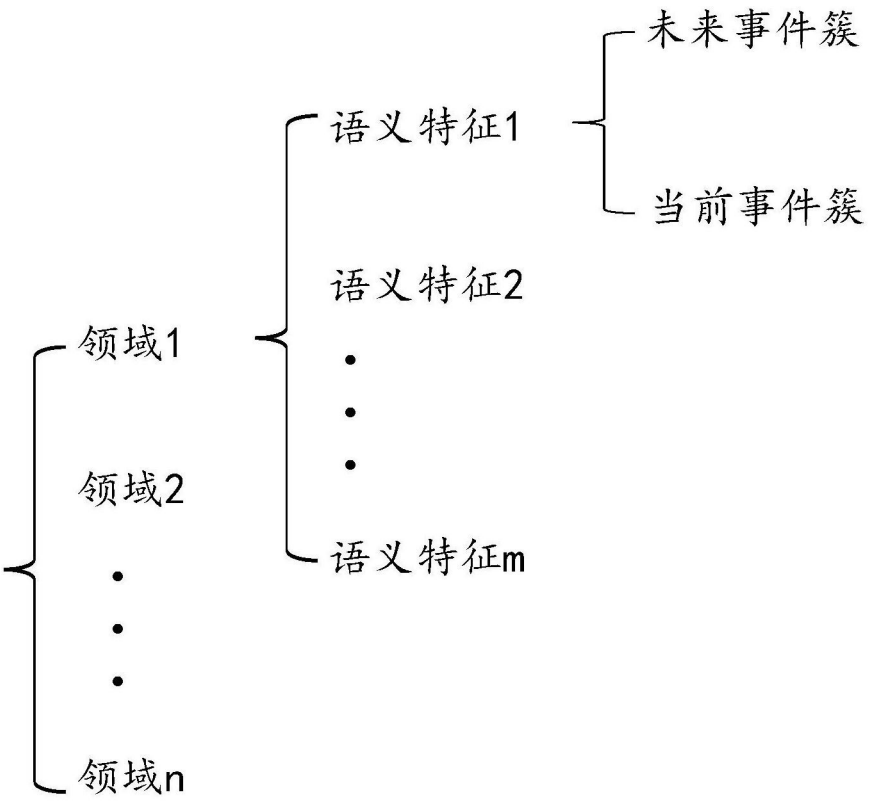

未来事件簇
语义特征1
当前事件簇
语义特征2
领域1
语义特征m
领域2
领域n

图4

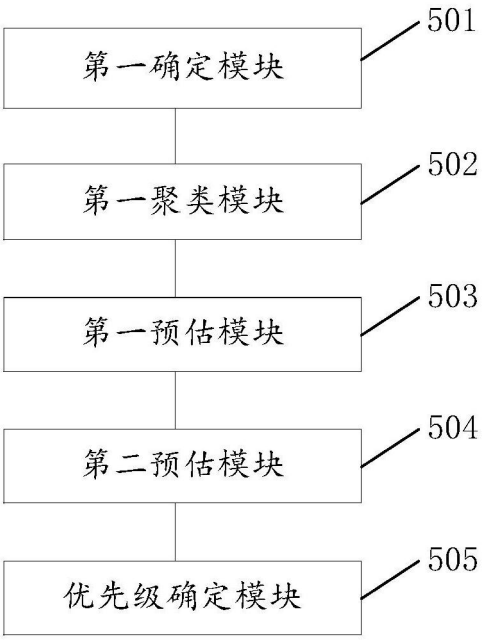

501
第一确定模块
502
第一聚类模块
503
第一预估模块
504
第二预估模块
505
优先级确定模块

图5

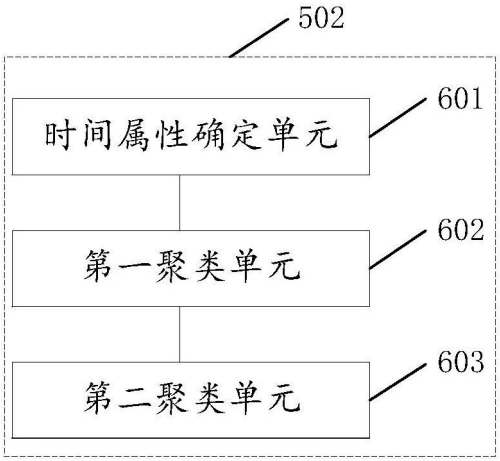

502
601
时间属性确定单元
602
第一聚类单元
603
第二聚类单元

图6

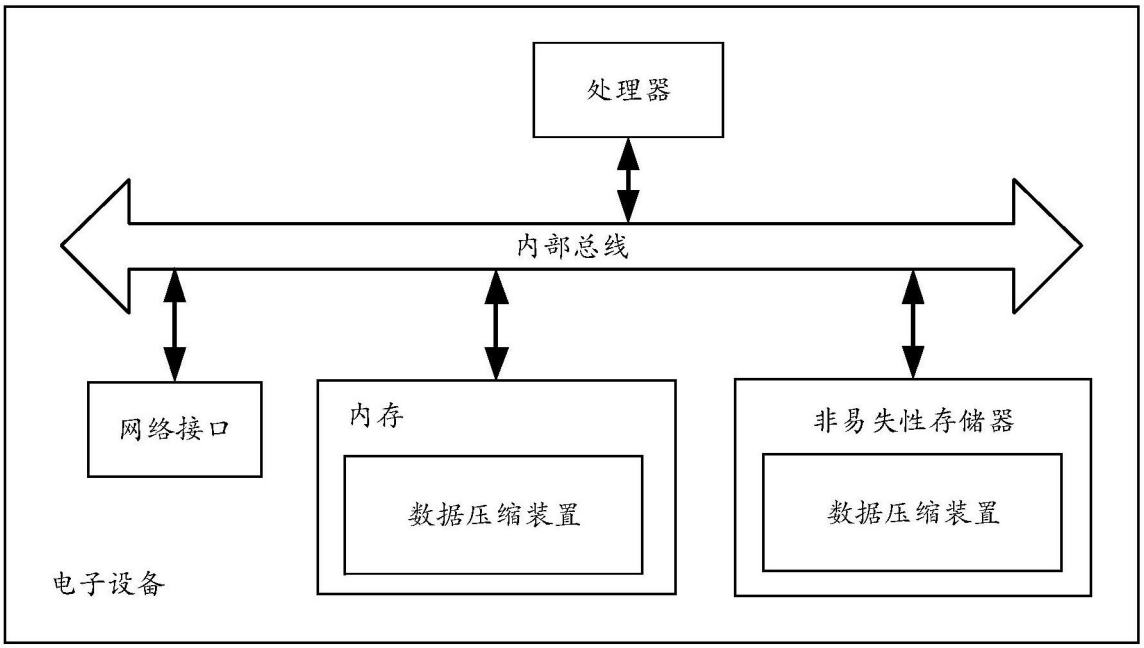


图7