(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 108170663 A
(43)申请公布日 2018.06.15

(21)申请号 201711123278.8

(22)申请日 2017.11.14

(71)申请人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱

(72)发明人 曹绍升 杨新星 周俊

(74)专利代理机构 北京晋德允升知识产权代理有限公司 11623
代理人 杨移

(51)Int.Cl.
G06F 17/27(2006.01)
G06K 9/62(2006.01)

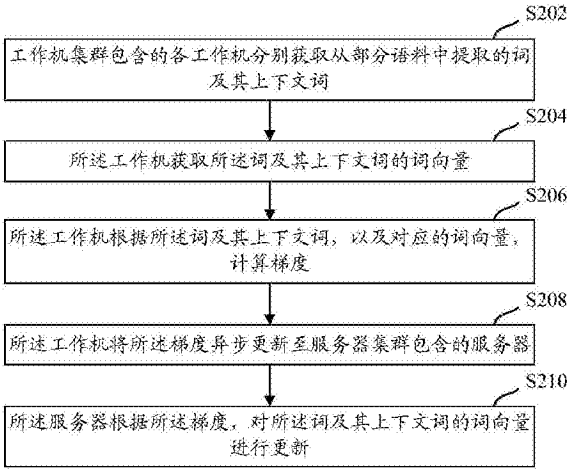
权利要求书3页 说明书10页 附图3页

(54)发明名称

基于集群的词向量处理方法、装置以及设备

(57)摘要

本说明书实施例公开了基于集群的词向量处理方法、装置以及设备，方案包括：集群包括服务器集群和工作机集群；工作机集群中的各工作机分别读取部分语料，并从读取的语料中提取词及其上下文词，从服务器集群中的服务器获取对应的词向量并计算梯度，将梯度异步更新至服务器；服务器根据梯度，对词及其上下文词的词向量进行更新。



CN 108170663 A

1.一种基于集群的词向量处理方法，所述集群包括多个工作机和服务器，所述方法包括：

各所述工作机分别执行：

获取从部分语料中提取的词及其上下文词；

获取所述词及其上下文词的词向量；

根据所述词及其上下文词，以及对应的词向量，计算梯度；

将所述梯度异步更新至所述服务器；

所述服务器根据所述梯度，对所述词及其上下文词的词向量进行更新。

2.如权利要求1所述的方法，所述获取从部分语料中提取的词及其上下文词前，所述方法还包括：

各所述工作机分布式地读取得到部分语料；

所述获取从部分语料中提取的词及其上下文词，具体包括：

根据自己所读取得到的语料，建立相应的词对，所述词对包含当前词及其上下词。

3.如权利要求2所述的方法，所述获取所述词及其上下文词的词向量，具体包括：

根据自己建立的各所述词对，提取得到当前词集合和上下文词集合；

从所述服务器获取所述当前词集合和上下文词集合包含的词的词向量。

4.如权利要求2所述的方法，所述根据所述词及其上下文词，以及对应的词向量，计算梯度，具体包括：

根据指定的损失函数、负样例词、自己建立的各所述词对，以及所述词及其上下文词的词向量，计算各词分别对应的梯度。

5.如权利要求1所述的方法，所述计算梯度，具体包括：

所述工作机上的一个或者多个线程以异步计算且不加锁更新的方式，计算梯度。

6.如权利要求1所述的方法，所述工作机将所述梯度异步更新至所述服务器，具体包括：

所述工作机计算得到所述梯度后，将所述梯度发送给所述服务器，其中，所述发送动作的执行无需等待其他工作机向所述服务器发送梯度。

7.如权利要求4所述的方法，所述服务器根据所述梯度，对所述词及其上下文词的词向量进行更新，具体包括：

按照以下公式，对所述词及其上下文词，以及所述负样例词的词向量进行迭代更新：

$$\vec{w}_{t+1}=\vec{w}_t-\alpha\cdot\nabla\cdot\vec{c}_t, w\in B_k$$

$$\vec{c}_{t+1}=\vec{c}_t-\alpha\cdot\nabla\cdot\vec{w}_t, c\in\Gamma(w)$$

其中，$\nabla=\sigma(\vec{w}\cdot\vec{c})-y$，$y=\begin{cases}1,\{w,c\}\\0,\{w,c'\}\end{cases}$，w表示当前词，c表示w的上下文词，c'表示负样例词，$\vec{w}$表示w的词向量，$\vec{c}$表示c的词向量，$\vec{w}_t$和$\vec{c}_t$表示在所述服务器上的第t次更新，B_k表示所述工作机上第k组语料，Γ(w)表示w的上下文词和负样例词的集合，α表示学习率，σ为Sigmoid函数。

8.一种基于集群的词向量处理装置，所述集群包括多个工作机和服务器，所述装置位

于所述集群，包括位于所述工作机的第一获取模块、第二获取模块、梯度计算模块、异步更新模块、位于所述服务器的词向量更新模块；

各工作机通过相应的模块分别执行：

所述第一获取模块获取从部分语料中提取的词及其上下文词；

所述第二获取模块获取所述词及其上下文词的词向量；

所述梯度计算模块根据所述词及其上下文词，以及对应的词向量，计算梯度；

所述异步更新模块将所述梯度异步更新至所述服务器；

所述服务器的所述词向量更新模块根据所述梯度，对所述词及其上下文词的词向量进行更新。

9.如权利要求8所述的装置，所述第一获取模块获取从部分语料中提取的词及其上下文词前，分布式地读取得到部分语料；

所述第一获取模块获取从部分语料中提取的词及其上下文词，具体包括：

所述第一获取模块根据自己所读取得到的语料，建立相应的词对，所述词对包含当前词及其上下词。

10.如权利要求9所述的装置，所述第二获取模块获取所述词及其上下文词的词向量，具体包括：

所述第二获取模块根据所述第一获取模块建立的各所述词对，提取得到当前词集合和上下文词集合；

从所述服务器获取所述当前词集合和上下文词集合包含的词的词向量。

11.如权利要求9所述的装置，所述梯度计算模块根据所述词及其上下文词，以及对应的词向量，计算梯度，具体包括：

所述梯度计算模块根据指定的损失函数、负样例词、自己建立的各所述词对，以及所述词及其上下文词的词向量，计算各词分别对应的梯度。

12.如权利要求8所述的装置，所述梯度计算模块计算梯度，具体包括：

所述梯度计算模块的一个或者多个线程以异步计算且不加锁更新的方式，计算梯度。

13.如权利要求8所述的装置，所述异步更新模块将所述梯度异步更新至所述服务器，具体包括：

所述异步更新模块在所述梯度计算模块计算得到所述梯度后，将所述梯度发送给所述服务器，其中，所述发送动作的执行无需等待其他工作机的异步更新模块向所述服务器发送梯度。

14.如权利要求11所述的装置，所述词向量更新模块根据所述梯度，对所述词及其上下文词的词向量进行更新，具体包括：

所述词向量更新模块按照以下公式，对所述词及其上下文词，以及所述负样例词的词向量进行迭代更新：

$$\vec{w}_{t+1}=\vec{w}_t-\alpha\cdot\nabla\cdot\vec{c}_t,w\in B_k$$

$$\vec{c}_{t+1}=\vec{c}_t-\alpha\cdot\nabla\cdot\vec{w}_t,c\in\Gamma(w)$$

其中，$\nabla=\sigma\left(\vec{w}\cdot\vec{c}\right)-y$，$y=\begin{cases}1,\{w,c\}\\0,\{w,c'\}\end{cases}$，w表示当前词，c表示w的上下文词，c'表示负样例词，$\vec{w}$表示w的词向量，$\vec{c}$表示c的词向量，$\vec{w}_t$和$\vec{c}_t$表示在所述服务器上的第t次更新，B_k表示所述工作机上第k组语料，Γ(w)表示w的上下文词和负样例词的集合，α表示学习率，σ为Sigmoid函数。

15.一种基于集群的词向量处理设备，所述设备属于所述集群，包括：

至少一个处理器；以及，

与所述至少一个处理器通信连接的存储器；其中，

所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够：

获取从部分语料中提取的词及其上下文词；

获取所述词及其上下文词的词向量；

根据所述词及其上下文词，以及对应的词向量，计算梯度；

将所述梯度异步更新；

根据异步更新的梯度，对所述词及其上下文词的词向量进行更新。

基于集群的词向量处理方法、装置以及设备

技术领域

[0001] 本说明书涉及计算机软件技术领域，尤其涉及基于集群的词向量处理方法、装置以及设备。

背景技术

[0002] 如今的自然语言处理的解决方案，大都采用基于神经网络的架构，而在这种架构下一个重要的基础技术就是词向量。词向量是将词映射到一个固定维度的向量，该向量表征了该词的语义信息。

[0003] 在现有技术中，常见的用于生成词向量的算法比如包括谷歌公司的单词向量算法、微软公司的深度神经网络算法等，往往在单机上运行。

[0004] 基于现有技术，需要高效的大规模词向量训练方案。

发明内容

[0005] 本说明书实施例提供基于集群的词向量处理方法、装置以及设备，用以解决如下技术问题：需要高效的大规模词向量训练方案。

[0006] 为解决上述技术问题，本说明书实施例是这样实现的：

[0007] 本说明书实施例提供的一种基于集群的词向量处理方法，所述集群包括多个工作机和服务器，所述方法包括：

[0008] 各所述工作机分别执行：

[0009] 获取从部分语料中提取的词及其上下文词；

[0010] 获取所述词及其上下文词的词向量；

[0011] 根据所述词及其上下文词，以及对应的词向量，计算梯度；

[0012] 将所述梯度异步更新至所述服务器；

[0013] 所述服务器根据所述梯度，对所述词及其上下文词的词向量进行更新。

[0014] 本说明书实施例提供的一种基于集群的词向量处理装置，所述集群包括多个工作机和服务器，所述装置位于所述集群，包括位于所述工作机的第一获取模块、第二获取模块、梯度计算模块、异步更新模块、位于所述服务器的词向量更新模块；

[0015] 各工作机通过相应的模块分别执行：

[0016] 所述第一获取模块获取从部分语料中提取的词及其上下文词；

[0017] 所述第二获取模块获取所述词及其上下文词的词向量；

[0018] 所述梯度计算模块根据所述词及其上下文词，以及对应的词向量，计算梯度；

[0019] 所述异步更新模块将所述梯度异步更新至所述服务器；

[0020] 所述服务器的所述词向量更新模块根据所述梯度，对所述词及其上下文词的词向量进行更新。

[0021] 本说明书实施例提供的一种基于集群的词向量处理设备，所述设备属于所述集群，包括：

[0022] 至少一个处理器;以及,

[0023] 与所述至少一个处理器通信连接的存储器;其中,

[0024] 所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够:

[0025] 获取从部分语料中提取的词及其上下文词;

[0026] 获取所述词及其上下文词的词向量;

[0027] 根据所述词及其上下文词,以及对应的词向量,计算梯度;

[0028] 将所述梯度异步更新;

[0029] 根据异步更新的梯度,对所述词及其上下文词的词向量进行更新。

[0030] 本说明书实施例采用的上述至少一个技术方案能够达到以下有益效果:在训练过程中,各工作机而无需相互等待,向服务器异步更新针对各词计算出的梯度,进而由服务器根据梯度更新各词的词向量,因此,有利于提高词向量训练收敛速度,再加上集群的分布式处理能力,使得该方案能够适用于大规模词向量训练且效率较高。

附图说明

[0031] 为了更清楚地说明本说明书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说明书中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0032] 图1为本说明书的方案在一种实际应用场景下涉及的一种整体架构示意图;

[0033] 图2为本说明书实施例提供的一种基于集群的词向量处理方法的流程示意图;

[0034] 图3为本说明书实施例提供的一种实际应用场景下,基于集群的词向量处理方法的原理示意图;

[0035] 图4为本说明书实施例提供的对应于图3的一种基于集群的词向量处理方法的详细流程示意图;

[0036] 图5为本说明书实施例提供的对应于图2的一种基于集群的词向量处理装置的结构示意图。

具体实施方式

[0037] 本说明书实施例提供基于集群的词向量处理方法、装置以及设备。

[0038] 为了使本技术领域的人员更好地理解本说明书中的技术方案,下面将结合本说明书实施例中的附图,对本说明书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申请一部分实施例,而不是全部的实施例。基于本说明书实施例,本领域普通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,都应当属于本申请保护的范围。

[0039] 本说明书的方案适用于集群,在集群下对于大规模词向量的处理效率更高,具体地:可以拆分训练语料,集群中的多个工作机分布式地分别根据拆分的部分语料,配合一个或者多个服务器训练所述部分语料对应的词向量,在训练过程中,各工作机负责计算各词对应的梯度,并异步更新至服务器,服务器负责根据梯度更新词向量。

[0040] 方案涉及的集群可以有一个或者多个，以图1为例，涉及了两个集群。

[0041] 图1为本说明书的方案在一种实际应用场景下涉及的一种整体架构示意图。该整体架构中，主要涉及三部分：包含多个服务器的服务器集群、包含多个工作机的工作机集群、数据库。数据库保存有用于训练的语料，供工作机集群读取，服务器集群保存原始的词向量，工作机集群与服务器集群进行配合，通过异步更新梯度，实现对词向量的训练。

[0042] 图1中的架构是示例性的，并非唯一。比如，方案也可以只涉及一个集群，该集群中包含至少一个调度机和多个工作机，由调度机完成上述服务器集群的工作；再比如，方案也可以涉及一个工作机集群和一个服务器；等等。

[0043] 下面基于图1中的架构，对本说明书的方案进行详细说明。

[0044] 图2为本说明书实施例提供的一种基于集群的词向量处理方法的流程示意图，所述集群包括工作机集群和服务器集群。图2中各步骤由集群中的至少一个机器(或者机器上的程序)执行，不同步骤的执行主体可以不同，图2中的流程可以执行多轮，每轮可以使用不同组的语料，语料用于训练词向量。

[0045] 图2中的流程包括以下步骤：

[0046] S202：工作机集群包含的各工作机分别获取从部分语料中提取的词及其上下文词。

[0047] S204：所述工作机获取所述词及其上下文词的词向量。

[0048] S206：所述工作机根据所述词及其上下文词，以及对应的词向量，计算梯度。

[0049] S208：所述工作机将所述梯度异步更新至服务器集群包含的服务器。

[0050] S210：所述服务器根据所述梯度，对所述词及其上下文词的词向量进行更新。

[0051] 在本说明书实施例中，各工作机可以分布式地执行步骤S202～S208，各工作机对应的部分语料通常是不同的，如此能够高效利用大规模的训练语料，也能够提高词向量的训练效率。比如，对于当前用于训练词向量的语料，可以将语料拆分为多份，各工作机可以分别读取一部分，进而基于自己读取的部分语料执行步骤S202～S208。

[0052] 为了便于描述，对于步骤S202～S208，以下各实施例主要从某一个工作机的角度进行说明。

[0053] 在本说明书实施例中，若本轮流程是第一轮流程，步骤S204中获取的词向量可以是初始化得到的。比如，可以采用随机初始化的方式或者按照指定概率分布初始化的方式，初始化各词的词向量，以及各词的上下文词的词向量，指定概率分布比如是0-1分布等。而若本轮流程并非第一轮流程，则步骤S204中获取的词向量可以是上轮流程执行完毕后更新并保存的词向量。

[0054] 在本说明书实施例中，训练词向量的过程主要包括计算梯度以及根据梯度更新向量，分别由工作机集群和服务器集群执行。在训练过程中，工作机计算完成后，需要将结果同步到服务器，通常有两种模式：同步更新与异步更新。同步更新是指：各工作机采用某种方式进行模型平均后再更新至服务器(一般地，不同的平均策略会造成不同的结果，模型平均的策略设计是同步更新重要的一环)。而异步更新是指任一个工作机计算完成就立即向服务器更新数据，而不等待其他工作机更不用进行模型平均。从最终效果上讲，异步更新由于不需要等待其他工作机计算完成，因此训练收敛速度往往更快，本说明书的方案主要基于异步更新的方式进行说明，具体异步更新的数据包括由工作机计算的各词对应的梯度。

[0055] 在本说明书实施例中，步骤S208由服务器集群执行，更新后的词向量也保存于服务器集群，以便下轮流程使用。当然，在图1以外的其他架构中，步骤S208也可以由与工作机属于同一集群的调度机或服务器执行。

[0056] 以此类推，进行多轮流程直至所有组的训练语料全部使用完毕后，服务器集群可以将最终更新得到的词向量写出到数据库，以便用于需求词向量的各种场景。

[0057] 通过图2的方法，在训练过程中，各工作机而无需相互等待，向服务器异步更新针对各词计算出的梯度，进而由服务器根据梯度更新各词的词向量，因此，有利于提高词向量训练收敛速度，再加上集群的分布式处理能力，使得该方案能够适用于大规模词向量训练且效率较高。

[0058] 基于图2的方法，本说明书实施例还提供了该方法的一些具体实施方案，以及扩展方案，继续基于图1中的架构进行说明。

[0059] 在本说明书实施例中，从语料中提取词及其上下文词可以由工作机执行，也可以由其他设备预先执行。以前一种方式为例，则对于步骤S202，所述获取从部分语料中提取的词及其上下文词前，还可以执行：各所述工作机分布式地读取得到部分语料。语料若保存于数据库，则可以从数据库读取。

[0060] 在本说明书实施例中，所述获取从部分语料中提取的词及其上下文词，具体可以包括：根据自己所读取得到的语料，建立相应的词对，所述词对包含当前词及其上下词。比如，可以扫描自己所读取得到的语料中的词，当前扫描的词为当前词记作w，根据设定的滑窗距离确定包含w的一个滑窗，将该滑窗内的其他每个词分别作为w的一个上下文词，记作c，如此构成词对{w,c}。

[0061] 进一步地，假定词向量保存于服务器集群包含的多个服务器上。则对于步骤S204，所述获取所述词及其上下文词的词向量，具体可以包括：根据自己建立的各所述词对，提取得到当前词集合和上下文词集合；从所述服务器获取所述当前词集合和上下文词集合包含的词的词向量。当然，这并非唯一实施方式，比如，也可以在扫描语料时，同步地从服务器获取当前扫描到的词的词向量而未必要依赖于建立的词对，等等。

[0062] 在本说明书实施例中，可以根据指定的损失函数，自己建立的各所述词对，以及所述词及其上下文词的词向量，计算各词分别对应的梯度。

[0063] 为了获得更好的训练效果以及更快地收敛，还可以引入指定的负样例词作为上下文词的对照计算梯度，负样例词被视为：相比于上下文词，与对应的当前词相关性相对低的词，一般可以在全部词中随机选择若干个。在这种情况下，对于步骤S206，所述根据所述词及其上下文词，以及对应的词向量，计算梯度，具体可以包括：根据指定的损失函数、负样例词、自己建立的各所述词对，以及所述词及其上下文词的词向量，计算各词分别对应的梯度。

[0064] 当前词及其每个负样例词也可以构成一个词对(称为负样例词对)，用c'表示负样例词，负样例词对记作{w,c'}，假定有λ个负样例词，相应的λ个负样例词对可以记作{w,c'$_1$}、{w,c'$_2$}、…、{w,c'$_\lambda$}，为了便于描述将负样例词对和上面的上下文词对(当前词及其上下文词构成的词对)统一记作{w,c}，并用y来区分，对于上下文词对，y＝1，对于负样例词对，y＝0。

[0065] 在本发明实施例中，上述的损失函数可以有多种形式，一般包含至少两项，一项反

映当前词与其上下文之间的相似度，另一项反映当前词与其负样例词之间的相似度，其中，可以用向量点乘度量相似度，也可以采用其他方式度量相似度。以一种实际应用场景为例，比如利用以下公式计算当前词对应的梯度▽：

[0066] $$\nabla=\sigma\left(\vec{w}\cdot\vec{c}\right)-y;\ (公式一)$$

[0067] 其中，$\vec{w}$表示w的词向量，$\vec{c}$表示c的词向量，σ是激活函数，假定为Sigmoid函数，则$\sigma=\frac{1}{1+e^{-x}}$。

[0068] 进一步地，每个工作机上的一个或者多个线程可以以异步计算且不加锁更新的方式，计算梯度。从而，工作机内各线程也可以并行计算梯度且不会相互妨碍，能够进一步地提高计算效率。

[0069] 在本说明书实施例中，对于步骤S208，所述工作机将所述梯度异步更新至所述服务器，具体可以包括：所述工作机计算得到所述梯度后，将所述梯度发送给所述服务器，其中，所述发送动作的执行无需等待其他工作机向所述服务器发送梯度。

[0070] 在本说明书实施例中，服务器获得工作机异步更新的梯度后，可以利用该梯度更新对应的当前词的词向量。不仅如此，服务器还可以利用该梯度，更新当前词的上下文词以及负样例词的词向量，具体的更新方式可以参照梯度下降法进行。

[0071] 例如，对于步骤S210，所述服务器根据所述梯度，对所述词及其上下文词的词向量进行更新，具体可以包括：

[0072] 按照以下公式，对所述词及其上下文词，以及所述负样例词的词向量进行迭代更新：

[0073] $$\vec{w}_{t+1}=\vec{w}_t-\alpha\cdot\nabla\cdot\vec{c}_t,w\in B_k;\ (公式二)$$

[0074] $$\vec{c}_{t+1}=\vec{c}_t-\alpha\cdot\nabla\cdot\vec{w}_t,c\in\Gamma(w);\ (公式三)$$

[0075] 其中，$\nabla=\sigma\left(\vec{w}\cdot\vec{c}\right)-y$，$y=\begin{cases}1,\{w,c\}\\0,\{w,c'\}\end{cases}$，w表示当前词，c表示w的上下文词，c'表示负样例词，$\vec{w}$表示w的词向量，$\vec{c}$表示c的词向量，$\vec{w}_t$和$\vec{c}_t$表示在所述服务器上的第t次更新，B_k表示所述工作机上第k组语料，Γ(w)表示w的上下文词和负样例词的集合，α表示学习率，σ比如为Sigmoid函数。

[0076] 根据上面的说明，本说明书实施例还提供了一种实际应用场景下，基于集群的词向量处理方法的原理示意图，如图3所示，进一步地，本说明书实施例还提供了对应于图3的一种基于集群的词向量处理方法的详细流程示意图，如图4所示。

[0077] 在图3中，示例性地示出了工作机0～2、服务器0～2，主要针对工作机0进行说明，而工作机1和2简略地进行了表示，工作方式与工作机0是一致的。“wid”、“cid”为标识，分别表示当前词和上下文词，“wid list”、“cid list”是标识列表，分别表示当前词集合和上下文词集合。图3中的简略工作流程包括：各工作机分布式地读取语料，建立词对；各工作机从服务器集群获取相应的词向量；各工作机利用读取的语料计算梯度并异步更新至服务器集群；服务器集群根据梯度更新词向量。

[0078] 图4中示出了更详细的流程，主要包括以下步骤：

[0079] S402：各工作机分布式地读取部分语料，建立词对{w,c}，从词对中提取wid list和cid list，如图4中的工作机0所示。

[0080] S404：工作机根据wid list和cid list从服务器拉取对应的词向量，服务器发送对应的词向量给工作机。

[0081] S406：工作机根据词对和对应的词向量，计算梯度，具体采用上述的公式一进行计算。

[0082] S408：工作机的每个线程均以异步计算且不加锁更新的方式，计算梯度，完成梯度计算后，不等待其他工作机，直接将计算出的该工作机上所有词对应的梯度传给服务器。

[0083] S410：服务器集群根据梯度更新词向量，具体采用上述的公式二和公式三进行计算。

[0084] 基于同样的思路，本说明书实施例还提供了上述方法的对应装置，如图5所示。

[0085] 图5为本说明书实施例提供的对应于图2的一种基于集群的词向量处理装置的结构示意图，所述集群包括多个工作机和服务器，所述装置位于所述集群，包括位于所述工作机的第一获取模块501、第二获取模块502、梯度计算模块503、异步更新模块504、位于所述服务器的词向量更新模块505；

[0086] 各工作机通过相应的模块分别执行：

[0087] 所述第一获取模块501获取从部分语料中提取的词及其上下文词；

[0088] 所述第二获取模块502获取所述词及其上下文词的词向量；

[0089] 所述梯度计算模块503根据所述词及其上下文词，以及对应的词向量，计算梯度；

[0090] 所述异步更新模块504将所述梯度异步更新至所述服务器；

[0091] 所述服务器的所述词向量更新模块505根据所述梯度，对所述词及其上下文词的词向量进行更新。

[0092] 可选地，所述第一获取模块501获取从部分语料中提取的词及其上下文词前，分布式地读取得到部分语料；

[0093] 所述第一获取模块501获取从部分语料中提取的词及其上下文词，具体包括：

[0094] 所述第一获取模块501根据自己所读取得到的语料，建立相应的词对，所述词对包含当前词及其上下词。

[0095] 可选地，所述第二获取模块502获取所述词及其上下文词的词向量，具体包括：

[0096] 所述第二获取模块502根据所述第一获取模块501建立的各所述词对，提取得到当前词集合和上下文词集合；

[0097] 从所述服务器获取所述当前词集合和上下文词集合包含的词的词向量。

[0098] 可选地，所述梯度计算模块503根据所述词及其上下文词，以及对应的词向量，计算梯度，具体包括：

[0099] 所述梯度计算模块503根据指定的损失函数、负样例词、自己建立的各所述词对，以及所述词及其上下文词的词向量，计算各词分别对应的梯度。

[0100] 可选地，所述梯度计算模块503计算梯度，具体包括：

[0101] 所述梯度计算模块503的一个或者多个线程以异步计算且不加锁更新的方式，计算梯度。

[0102] 可选地，所述异步更新模块504将所述梯度异步更新至所述服务器，具体包括：

[0103] 所述异步更新模块504在所述梯度计算模块503计算得到所述梯度后，将所述梯度发送给所述服务器，其中，所述发送动作的执行无需等待其他工作机的异步更新模块504向所述服务器发送梯度。

[0104] 可选地，所述词向量更新模块505根据所述梯度，对所述词及其上下文词的词向量进行更新，具体包括：

[0105] 所述词向量更新模块505按照以下公式，对所述词及其上下文词，以及所述负样例词的词向量进行迭代更新：

[0106] $$\vec{w}_{t+1}=\vec{w}_t-\alpha\cdot\nabla\cdot\vec{c}_t, w\in B_k$$

[0107] $$\vec{c}_{t+1}=\vec{c}_t-\alpha\cdot\nabla\cdot\vec{w}_t, c\in\Gamma(w)$$

[0108] 其中，$\nabla=\sigma(\vec{w}\cdot\vec{c})-y$， $y=\begin{cases}1,\{w,c\}\\0,\{w,c'\}\end{cases}$，w表示当前词，c表示w的上下文

[0109] 词，c'表示负样例词，$\vec{w}$表示w的词向量，$\vec{c}$表示c的词向量，$\vec{w}_t$和$\vec{c}_t$表示

[0110] 在所述服务器上的第t次更新，B_k表示所述工作机上第k组语料，Γ(w)表

[0111] 示w的上下文词和负样例词的集合，α表示学习率，σ为Sigmoid函数。

[0112] 基于同样的思路，本说明书实施例还提供了对应于图2的一种基于集群的词向量处理设备，该设备属于所述集群，包括：

[0113] 至少一个处理器；以及，

[0114] 与所述至少一个处理器通信连接的存储器；其中，

[0115] 所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够：

[0116] 获取从部分语料中提取的词及其上下文词；

[0117] 获取所述词及其上下文词的词向量；

[0118] 根据所述词及其上下文词，以及对应的词向量，计算梯度；

[0119] 将所述梯度异步更新；

[0120] 根据异步更新的梯度，对所述词及其上下文词的词向量进行更新。

[0121] 基于同样的思路，本说明书实施例还提供了对应于图2的一种非易失性计算机存储介质，存储有计算机可执行指令，所述计算机可执行指令设置为：

[0122] 获取从部分语料中提取的词及其上下文词；

[0123] 获取所述词及其上下文词的词向量；

[0124] 根据所述词及其上下文词，以及对应的词向量，计算梯度；

[0125] 将所述梯度异步更新；

[0126] 根据异步更新的梯度，对所述词及其上下文词的词向量进行更新。

[0127] 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。

[0128] 本说明书中的各个实施例均采用递进的方式描述，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、设备、非易失性计算机存储介质实施例而言，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。

[0129] 本说明书实施例提供的装置、设备、非易失性计算机存储介质与方法是对应的，因此，装置、设备、非易失性计算机存储介质也具有与对应方法类似的有益技术效果，由于上面已经对方法的有益技术效果进行了详细说明，因此，这里不再赘述对应装置、设备、非易失性计算机存储介质的有益技术效果。

[0130] 在20世纪90年代，对于一个技术的改进可以很明显地区分是硬件上的改进(例如，对二极管、晶体管、开关等电路结构的改进)还是软件上的改进(对于方法流程的改进)。然而，随着技术的发展，当今的很多方法流程的改进已经可以视为硬件电路结构的直接改进。设计人员几乎都通过将改进的方法流程编程到硬件电路中来得到相应的硬件电路结构。因此，不能说一个方法流程的改进就不能用硬件实体模块来实现。例如，可编程逻辑器件(Programmable Logic Device,PLD)(例如现场可编程门阵列(Field Programmable Gate Array,FPGA))就是这样一种集成电路，其逻辑功能由用户对器件编程来确定。由设计人员自行编程来把一个数字系统“集成”在一片PLD上，而不需要请芯片制造厂商来设计和制作专用的集成电路芯片。而且，如今，取代手工地制作集成电路芯片，这种编程也多半改用“逻辑编译器(logic compiler)”软件来实现，它与程序开发撰写时所用的软件编译器相类似，而要编译之前的原始代码也得用特定的编程语言来撰写，此称之为硬件描述语言(Hardware Description Language,HDL)，而HDL也并非仅有一种，而是有许多种，如ABEL(Advanced Boolean Expression Language)、AHDL(Altera Hardware Description Language)、Confluence、CUPL(Cornell University Programming Language)、HDCal、JHDL(Java Hardware Description Language)、Lava、Lola、MyHDL、PALASM、RHDL(Ruby Hardware Description Language)等，目前最普遍使用的是VHDL(Very-High-Speed Integrated Circuit Hardware Description Language)与Verilog。本领域技术人员也应该清楚，只需要将方法流程用上述几种硬件描述语言稍作逻辑编程并编程到集成电路中，就可以很容易得到实现该逻辑方法流程的硬件电路。

[0131] 控制器可以按任何适当的方式实现，例如，控制器可以采取例如微处理器或处理器以及存储可由该(微)处理器执行的计算机可读程序代码(例如软件或固件)的计算机可读介质、逻辑门、开关、专用集成电路(Application Specific Integrated Circuit,ASIC)、可编程逻辑控制器和嵌入微控制器的形式，控制器的例子包括但不限于以下微控制器：ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20以及Silicone Labs C8051F320，存储器控制器还可以被实现为存储器的控制逻辑的一部分。本领域技术人员也知道，除了以纯计算机可读程序代码方式实现控制器以外，完全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部件，而对其内包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至，可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

[0132] 上述实施例阐明的系统、装置、模块或单元，具体可以由计算机芯片或实体实现，或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的，计算机例如可以为个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

[0133] 为了描述的方便，描述以上装置时以功能分为各种单元分别描述。当然，在实施本说明书时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[0134] 本领域内的技术人员应明白，本说明书实施例可提供为方法、系统、或计算机程序产品。因此，本说明书实施例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，本说明书实施例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。

[0135] 本说明书是参照根据本说明书实施例的方法、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。

[0136] 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。

[0137] 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。

[0138] 在一个典型的配置中，计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0139] 内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM)和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0140] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算

机可读介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0141] 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,并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0142] 本说明书可以在由计算机执行的计算机可执行指令的一般上下文中描述,例如程序模块。一般地,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本说明书,在这些分布式计算环境中,由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中,程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0143] 本说明书中的各个实施例均采用递进的方式描述,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,对于系统实施例而言,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。

[0144] 以上所述仅为本说明书实施例而已,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。

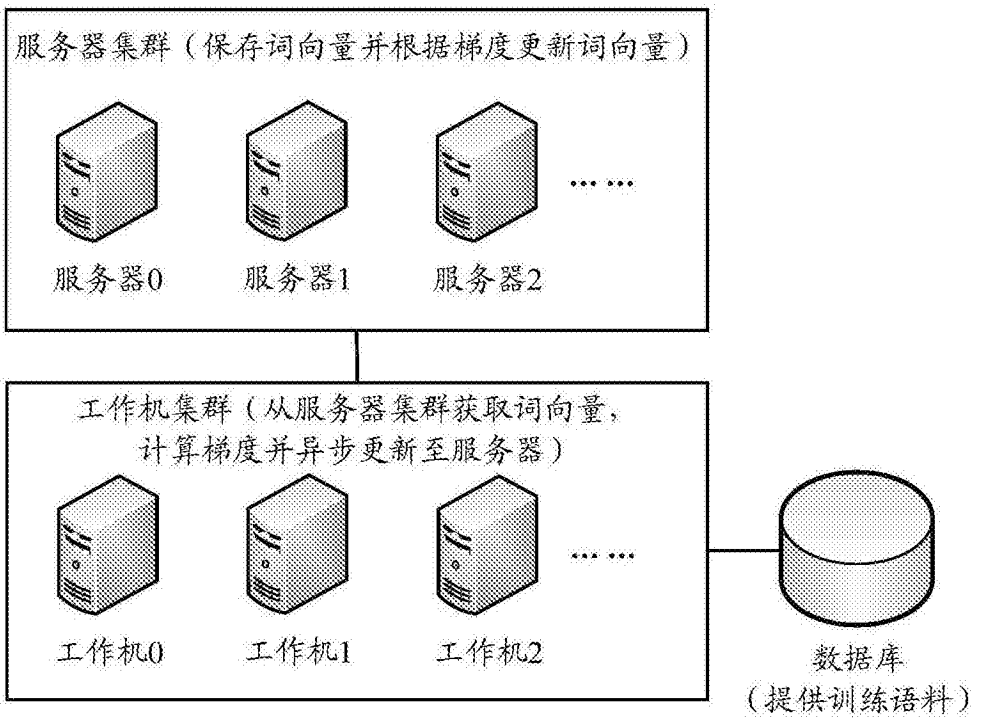

服务器集群（保存词向量并根据梯度更新词向量）
服务器0
服务器1
服务器2
……
工作机集群（从服务器集群获取词向量，
计算梯度并异步更新至服务器）
工作机0
工作机1
工作机2
……
数据库
（提供训练语料）

图1

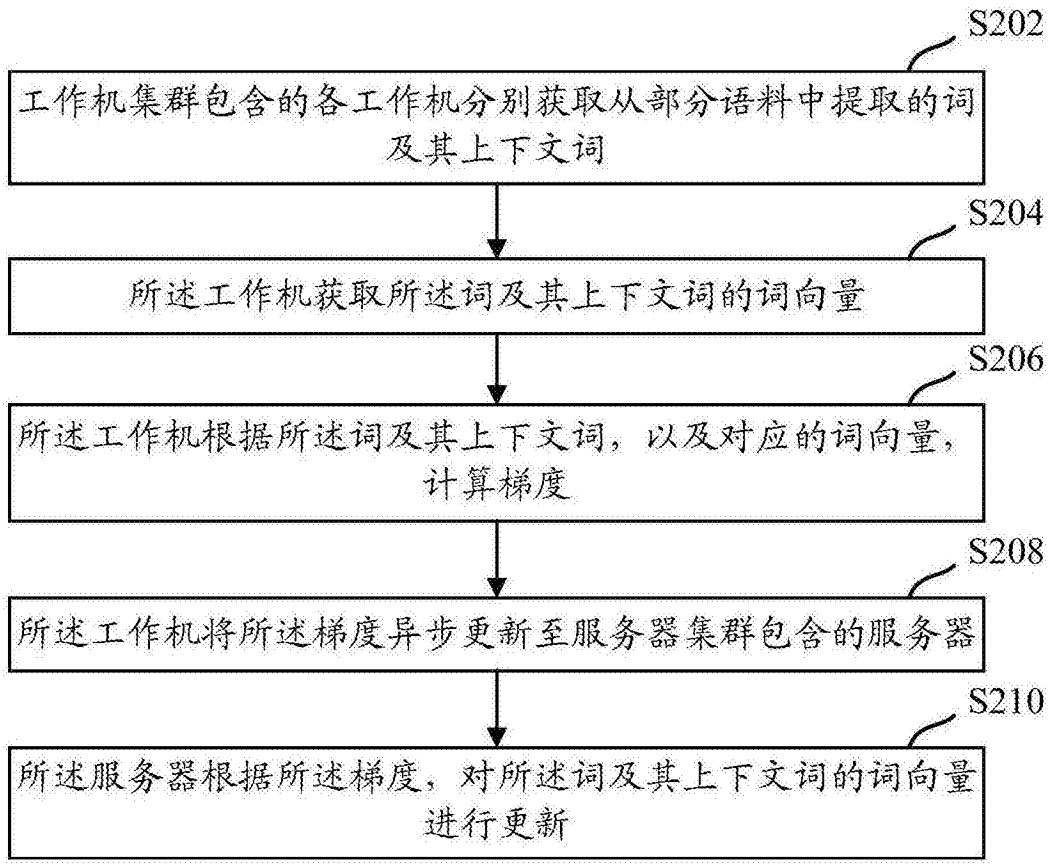

S202
工作机集群包含的各工作机分别获取从部分语料中提取的词及其上下文词
S204
所述工作机获取所述词及其上下文词的词向量
S206
所述工作机根据所述词及其上下文词，以及对应的词向量，计算梯度
S208
所述工作机将所述梯度异步更新至服务器集群包含的服务器
S210
所述服务器根据所述梯度，对所述词及其上下文词的词向量进行更新

图2

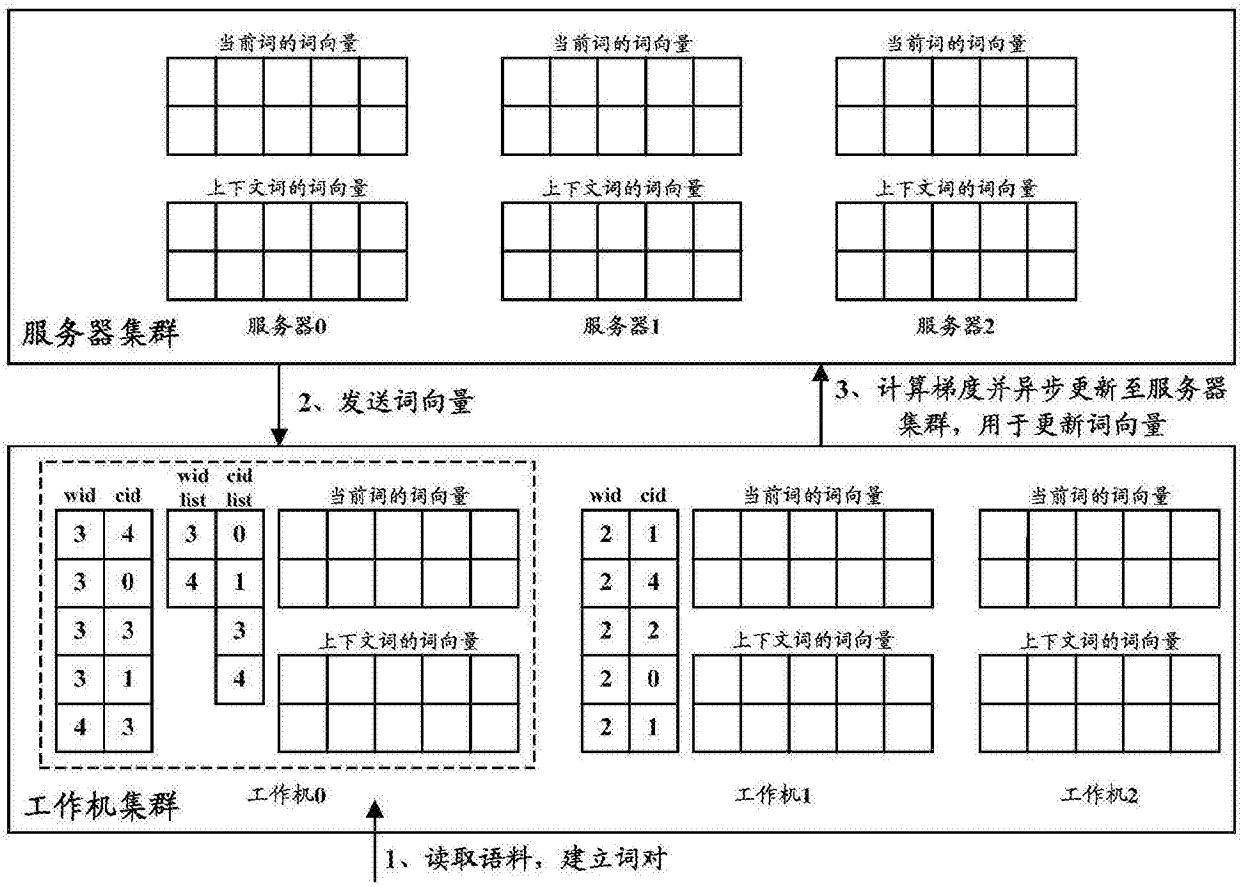

当前词的词向量
当前词的词向量
当前词的词向量
上下文词的词向量
上下文词的词向量
上下文词的词向量
服务器集群
服务器0
服务器1
服务器2
2、发送词向量
3、计算梯度并异步更新至服务器集群，用于更新词向量
wid cid
wid list
cid list
当前词的词向量
上下文词的词向量
工作机集群
工作机0
工作机1
工作机2
1、读取语料，建立词对

图3

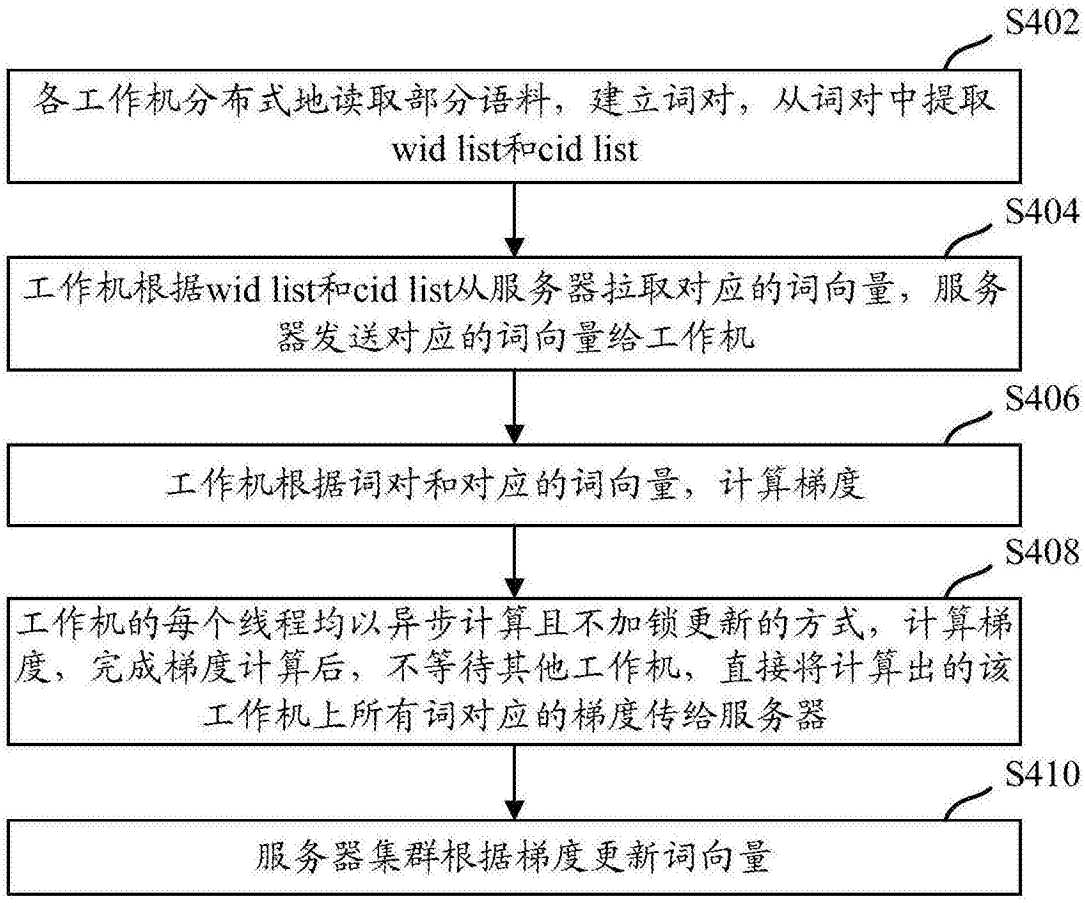

S402
各工作机分布式地读取部分语料，建立词对，从词对中提取wid list和cid list
S404
工作机根据wid list和cid list从服务器拉取对应的词向量，服务器发送对应的词向量给工作机
S406
工作机根据词对和对应的词向量，计算梯度
S408
工作机的每个线程均以异步计算且不加锁更新的方式，计算梯度，完成梯度计算后，不等待其他工作机，直接将计算出的该工作机上所有词对应的梯度传给服务器
S410
服务器集群根据梯度更新词向量

图4

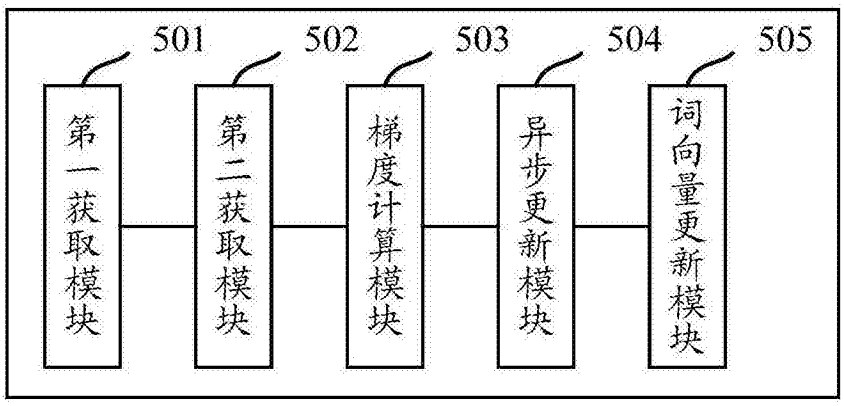

501
502
503
504
505
第一获取模块
第二获取模块
梯度计算模块
异步更新模块
词向量更新模块

图5