

权利要求书

- 1、一种分布式图嵌入方法，图数据结构至少分布于两个计算节点，所述方法包括：
在利用预设采样算法对图数据结构进行路径采样时，将当前采样顶点的标识和当前嵌入向量加入预定义的消息数据结构，所述消息数据结构中记录有采样过程中采样顶点的标识和当前嵌入向量、且所述消息数据结构由上一采样顶点所在计算节点传递而获得；
在消息数据结构的数据长度达到词向量算法所定义的窗口长度时，利用所述词向量算法计算更新后的消息数据结构中各采样顶点的嵌入向量的更新梯度；
基于更新后的消息数据结构中的标识，将各更新梯度分别发送至相应顶点，并利用所述更新梯度对归属当前计算节点的采样顶点的当前嵌入向量进行更新；
达到预设收敛条件后，将图数据结构中各顶点的最终嵌入向量，构成图嵌入矩阵。
- 2、根据权利要求 1 所述的方法，图数据结构中顶点的当前嵌入向量存储于顶点所在计算节点的图顶点数据中。
- 3、根据权利要求 1 所述的方法，所述方法还包括：
在消息数据结构的数据长度未达到词向量算法所定义的窗口长度时，利用预设采样算法对与所述当前采样顶点相邻且连接的顶点进行采样，将更新后的消息数据结构发送至所确定的采样顶点所在计算节点。
- 4、根据权利要求 1 所述的方法，所述消息数据结构的数据长度上限值为窗口长度，所述方法还包括：
在消息数据结构的数据长度达到或大于词向量算法所定义的窗口长度时，删除消息数据结构中加入时间最早的采样顶点的标识和当前嵌入向量。
- 5、根据权利要求 1 至 4 任一项所述的方法，所述方法还包括：
基于预设任务利用图嵌入矩阵构建训练样本，并利用所构建的训练样本对预设模型进行训练，获得用于处理预设任务的模型。
- 6、一种分布式图嵌入装置，图数据结构至少分布于两个计算节点，所述装置设于其中一个计算节点，包括：
消息更新模块，用于在利用预设采样算法对图数据结构进行路径采样时，将当前采样顶点的标识和当前嵌入向量加入预定义的消息数据结构，所述消息数据结构中记录有采样过程中采样顶点的标识和当前嵌入向量、且所述消息数据结构由上一采样顶点所在计算节点传递而获得；
梯度计算模块，用于在消息数据结构的数据长度达到词向量算法所定义的窗口长度时，利用所述词向量算法计算更新后的消息数据结构中各采样顶点的嵌入向量的更新梯

度;

向量更新模块,用于基于更新后的消息数据结构中的标识,将各更新梯度分别发送至相应顶点,并利用所述更新梯度对归属当前计算节点的采样顶点的当前嵌入向量进行更新;

- 5 矩阵生成模块,用于达到预设收敛条件后,将图数据结构中各顶点的最终嵌入向量,构成图嵌入矩阵。

7、根据权利要求6所述的装置,图数据结构中顶点的当前嵌入向量存储于顶点所在计算节点的图顶点数据中。

8、根据权利要求6所述的装置,所述装置还包括:

- 10 顶点采样模块,用于在消息数据结构的数据长度未达到词向量算法所定义的窗口长度时,利用预设采样算法对与所述当前采样顶点相邻且连接的顶点进行采样;

信息传输模块,用于将更新后的消息数据结构发送至所确定的采样顶点所在计算节点。

9、根据权利要求6所述的装置,所述消息数据结构的数据长度上限值为窗口长度,

- 15 所述消息更新模块还用于:

在消息数据结构的数据长度达到或大于词向量算法所定义的窗口长度时,删除消息数据结构中加入时间最早的采样顶点的标识和当前嵌入向量。

10、一种分布式图嵌入系统,所述系统至少包括两个计算节点,图数据结构分布式存储在所述系统中,任意一个计算节点执行以下方法:

- 20 在利用预设采样算法对图数据结构进行路径采样时,将当前采样顶点的标识和当前嵌入向量加入预定义的消息数据结构,所述消息数据结构中记录有采样过程中采样顶点的标识和当前嵌入向量、且所述消息数据结构由上一采样顶点所在计算节点传递而获得;

在消息数据结构的数据长度达到词向量算法所定义的窗口长度时,利用所述词向量算法计算更新后的消息数据结构中各采样顶点的嵌入向量的更新梯度;

- 25 基于更新后的消息数据结构中的标识,将各更新梯度分别发送至相应顶点,并利用所述更新梯度对归属当前计算节点的采样顶点的当前嵌入向量进行更新;

达到预设收敛条件后,将图数据结构中各顶点的最终嵌入向量,构成图嵌入矩阵。

11、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,其中,图数据结构至少分布于两个计算节点,所述处理器执行所述程序

- 30 时实现如下方法:

在利用预设采样算法对图数据结构进行路径采样时,将当前采样顶点的标识和当前

嵌入向量加入预定义的消息数据结构，所述消息数据结构中记录有采样过程中采样顶点的标识和当前嵌入向量、且所述消息数据结构由上一采样顶点所在计算节点传递而获得；

在消息数据结构的数据长度达到词向量算法所定义的窗口长度时，利用所述词向量算法计算更新后的消息数据结构中各采样顶点的嵌入向量的更新梯度；

- 5 基于更新后的消息数据结构中的标识，将各更新梯度分别发送至相应顶点，并利用所述更新梯度对归属当前计算节点的采样顶点的当前嵌入向量进行更新；

达到预设收敛条件后，将图数据结构中各顶点的最终嵌入向量，构成图嵌入矩阵。