



## LINE:Large-scale Information Network Embedding阅读笔记



v1ncent....  
计算机研究生

关注他

25 人赞了该文章

论文：LINE：Large-scale Information Network Embedding

发表时间：2015

作者：Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, Qiaozhu Mei

单位：Microsoft Research Asia;Peking University,China;University of Michigan

**前言：大规模信息网络(large-scale information Network)**无论在存取性，使用性上比起普通的信息处理方式更加复杂，更加多变，例如航空公司网络，出版物网络，社会和通信网络以及万维网。LINE模型致力于将这种大型的信息网络嵌入到低维的向量空间中，且该模型适用于任何类型(有向、无向亦或是有权重)的信息网络。并提出了一种解决经典随机梯度下降限制的边缘采样算法，提高了推理的有效性和效率。可用于可视化，节点分类以及关系预测等方面。

**关键词：**信息网络嵌入；可扩展性；特征学习；降维。

### 一、写作动机

当前大多数图形嵌入方法不能对包含数百万个节点的真实信息网络进行扩展，分析大型信息网络在学术界和行业中一直受到越来越多的关注。而现在的大多数嵌入方法在小型网络中适用性非常不错，但当网络包含数百万乃至数亿节点时，看起来并不那么有效，其时间复杂度至少是节点数的二次方。最重要的是，它们着重于关注节点之间的一阶相似性，及两点之间是否直接相连，而忽略了其二阶相似性(即拥有许多共同的邻节点)。因此LINE模型就是为了在信息网络嵌入至低维空间时保留其一阶相似以及二阶相似。

### 二、LINE模型构建

在模型构建之前，我们首先要对4个重要的名词做定义，以便读者方便阅读和理解LINE模型。

#### 1、信息网络的定义

信息网络定义为  $G = (V, E)$ ，其中  $V$  是顶点集合，顶点表示数据对象， $E$  是顶点之间的边缘的集合，每条边表示两个数据对象之间的关系

与权重  $W_{uv} > 0$  相关联，权重表示关系的强

25

20 条评论

分享

收藏

...



$W_{uv} \equiv W_{vu}$ ; 如果  $G$  是指向的, 我们有  $(u, v) \neq (v, u)$  和  $W_{uv} \neq W_{vu}$ 。关于边权重问题, 本次论文中只研究非负权重的情况。

## 2、一阶相似性的定义

网络中的一阶相似性是两个顶点之间的局部点对的邻近度。对于由边缘  $(u, v)$  链接的每对顶点, 该边缘的权重  $W_{uv}$  表示  $u$  和  $v$  之间的一阶相似性, 如果在  $u$  和  $v$  之间没有观察到边缘, 它们的一阶相似性为0。

## 3、二阶相似度的定义

二阶相似度指的是一对顶点之间的接近程度  $(u, v)$  在网络中是其邻域网络结构之间的相似性。数学上, 让  $p_u = (w_{u,1}, \dots, w_{u,|V|})$  表示一阶附近  $u$  与所有其他的顶点, 那么  $u$  和  $v$  之间的二阶相似性由  $p_u$  和  $p_v$  之间的相似性来决定。如果没有一个顶点同时和  $u$  与  $v$  链接, 那么  $u$  和  $v$  的二阶相似性是0。

## 4、大规模信息网络嵌入的定义

给定大网络  $G = (V, E)$ , 大规模信息网络嵌入是将每个顶点  $v \in V$  表示为低维空间  $R^d$  中的向量, 学习一个函数  $f_G: V \rightarrow R^d$ , 其中  $d \ll |V|$ 。在空间  $R^d$  中, 顶点之间的一阶相似性和二阶相似性都被保留。

## 5、LINE:大规模信息网络嵌入

前文已经提到了信息网络的两大相似性, LINE模型也是由此展开的。

①LINE模型的一阶相似性: 对每个无向边  $(i, j)$ , 定义顶点  $v_i$  和  $v_j$  的联合概率分布为:

$$p_1(v_i, v_j) = \frac{1}{1 + \exp(-\vec{u}_i^T \cdot \vec{u}_j)}, \quad (1)$$

$u_i \in R^d$  是顶点  $v_i$  的低维向量表示, 为保持其一阶相似性, 一种直接方法是最小化以下的目标函数:

$$O_1 = d(\hat{p}_1(\cdot, \cdot), p_1(\cdot, \cdot)),$$

$d(\cdot, \cdot)$  为两种分布之间的距离,  $p(\cdot)$  为空间  $V \times V$  上的一个分布:  $\frac{w_{ij}}{W}$ ,  $W$  为  $i, j$  两点间边权重总和。我们选择尽量减少两个概率分布的KL 散度。将  $d(\cdot)$  替换为 KL 散度并省略一些常数, 我们得到:

$$O_1 = - \sum_{(i,j) \in E} w_{ij} \log p_1(v_i, v_j),$$

需要注意的是, 一阶相似度仅适用于无向图, 而不适用于有向图。

②LINE模型的二阶相似性: 二阶相似性假定与其他顶点共享邻居顶点的两个点彼此相似(无向有向均可), 一个向量  $u$  和  $u'$  分别表示顶点本身和其他顶点的特定“上下文”, 意为二阶相似。对于每个有向边  $(i, j)$ , 我们首先定义由生成“上下文”的概率:

$$p_2(v_j | v_i) = \frac{\exp(\vec{u}_j'^T \cdot \vec{u}_i)}{\sum_{k=1}^{|V|} \exp(\vec{u}_k'^T \cdot \vec{u}_i)}, \quad (4)$$

可以看到, 上式其实是一个条件分布, 我们取  $i$  为研究对象,  $p(\cdot, v_i)$ , 在降维之后使其接近与经验分布  $p_2^\uparrow$ 。因此最小化以下目标函数:



$$O_2 = \sum_{i \in V} \lambda_i d(\hat{p}_2(\cdot | v_i), p_2(\cdot | v_i)),$$

$d(\cdot)$ 上文已经说明， $\lambda_i$  来表示网络中顶点 $i$ 的声望,本文中即是顶点 $i$ 的度数，因此二阶相似性的计算公式为：

$$O_2 = - \sum_{(i,j) \in E} w_{ij} \log p_2(v_j | v_i).$$

将上述的  $O_1$  以及  $O_2$  进行训练，再进行综合。

③模型优化： $O_2$  的计算代价十分的昂贵，因此优化时使用了负采样方法(论文参考[13])，为每条边指定了一个目标函数：

$$\log \sigma(\vec{u}_j^T \cdot \vec{u}_i) + \sum_{i=1}^K E_{v_n \sim P_n(v)} [\log \sigma(-\vec{u}_n^T \cdot \vec{u}_i)],$$

而上述函数又可通过采用异步随机梯度下降算法 (ASGD) 来优化。每一步中，ASGD算法对小批量边缘进行抽样，然后更新模型参数。但是这也带来一个问题，如果我们根据小权重的边缘选择较大的学习率，那么大权重的边上的梯度就会爆炸式的过大，如果我们根据具有较大权重的边选择学习小的速率，那么小权重上的边的梯度将变得太小。因此边缘采样同样要优化。从起始边缘采样并将采样的边缘作为二进制边缘，其中采样概率与原始边缘的权重成比例。

### 三、实验

我们将该方法应用于不同类型的几个大型现实世界网络，包括语言网络，两个社交网络和两个引用网络。

数据集：

(1)语言网络：整个英文维基百科页面构建了一个词共同网络

(2)社交网络：Flickr和Youtube

(3)引用网络：作者引文网和论文引文网

以上网络的详细统计数据总结在表1中，代表一系列信息网络(有向或无向、加权与否)；每个网络包含至少50万个节点和数百万个边。

算法比较：将LINE模型与一些常见的图嵌入算法进行比较，包括GF算法，Deep Walk算法，LINE-SGD，LINE算法等。

数据集表一：

Table 1: Statistics of the real-world information networks.

|             | Language Network    | Social Network    |                   | Citation Network     |                      |
|-------------|---------------------|-------------------|-------------------|----------------------|----------------------|
| Name        | WIKIPEDIA           | FLICKR            | YOUTUBE           | DBLP(AUTHORCITATION) | DBLP(PAPER CITATION) |
| Type        | undirected,weighted | undirected,binary | undirected,binary | directed,weighted    | directed,binary      |
| V           | 1,985,098           | 1,715,256         | 1,138,499         | 524,061              | 781,109              |
| E           | 1,000,924,086       | 22,613,981        | 2,990,443         | 20,580,238           | 4,191,677            |
| Avg. degree | 504.22              | 26.37             | 5.25              | 78.54                | 10.73                |
| #Labels     | 7                   | 5                 | 47                | 7                    | 7                    |
| #train      | 70,000              | 75,958            | 31,703            | 20,684               | 10,398               |

参数设置：

对于所有方法，随机梯度下降的小批量大小设置为1；以起始值  $\rho_0 = 0.025$  和  $\rho_t = \rho_0 (1 - t / T)$  设定学习速度， $T$  是小批量或边缘样品的总数；为了公平比较，语言网络嵌入的维度被设置为200；而其他网络中，默认设置为128；其他的默认参数设置包括：LINE的负采样  $K = 5$ ，样本总数  $T = 100$  亿(LINE),  $T = 200$  亿(GF)，窗 Walk，每顶点行走  $\gamma = 40$ ；所有的嵌入向量

定量结果：

1.语言网络：

两个应用程序用于评估学习嵌入的有效性：词类比和文档分类。

Table 2: Results of word analogy on WIKIPEDIA data.

| Algorithm     | Semantic (%) | Syntactic (%) | Overall (%) | Running time |
|---------------|--------------|---------------|-------------|--------------|
| GF            | 61.38        | 44.08         | 51.93       | 2.96h        |
| DeepWalk      | 50.79        | 37.70         | 43.65       | 16.64h       |
| SkipGram      | 69.14        | 57.94         | 63.02       | 2.82h        |
| LINE-SGD(1st) | 9.72         | 7.48          | 8.50        | 3.83h        |
| LINE-SGD(2nd) | 20.42        | 9.56          | 14.49       | 3.94h        |
| LINE(1st)     | 58.08        | 49.42         | 53.35       | 2.44h        |
| LINE(2nd)     | 73.79        | 59.72         | 66.10       | 2.55h        |

词类比：给定一个单词对  $(a, b)$  和一个单词  $c$ ，该任务旨在找到一个单词  $d$ ，使得  $c$  和  $d$  之间的关系类似于  $a$  和  $b$  之间的关系，或表示为： $a : b \rightarrow c : ?$ ；表2使用在维基百科语料库或维基百科词汇网络上学习的词汇的嵌入来报告单词类比的结果，对于图分解，每对单词之间的权重被定义为同时出现次数的对数，这导致比共同出现的原始值更好的性能。对于DeepWalk，尝试不同的截止阈值将语言网络转换为二进制网络，并且当所有边缘保留在网络中时，实现最佳性能。同时与最先进的词嵌入模式SkipGram进行比较。直接从原始维基百科页面学习嵌入词，也隐含地是矩阵分解法。

Table 3: Results of Wikipedia page classification on WIKIPEDIA data set.

| Metric   | Algorithm     | 10%     | 20%     | 30%     | 40%     | 50%     | 60%     | 70%     | 80%     | 90%     |
|----------|---------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Micro-F1 | GF            | 79.63   | 80.51   | 80.94   | 81.18   | 81.38   | 81.54   | 81.63   | 81.71   | 81.78   |
|          | DeepWalk      | 78.89   | 79.92   | 80.41   | 80.69   | 80.92   | 81.08   | 81.21   | 81.35   | 81.42   |
|          | SkipGram      | 79.84   | 80.82   | 81.28   | 81.57   | 81.71   | 81.87   | 81.98   | 82.05   | 82.09   |
|          | LINE-SGD(1st) | 76.03   | 77.05   | 77.57   | 77.85   | 78.08   | 78.25   | 78.39   | 78.44   | 78.49   |
|          | LINE-SGD(2nd) | 74.68   | 76.53   | 77.54   | 78.18   | 78.63   | 78.96   | 79.19   | 79.40   | 79.57   |
|          | LINE(1st)     | 79.67   | 80.55   | 80.94   | 81.24   | 81.40   | 81.52   | 81.61   | 81.69   | 81.67   |
|          | LINE(2nd)     | 79.93   | 80.90   | 81.31   | 81.63   | 81.80   | 81.91   | 82.00   | 82.11   | 82.17   |
|          | LINE(1st+2nd) | 81.04** | 82.08** | 82.58** | 82.93** | 83.16** | 83.37** | 83.52** | 83.63** | 83.74** |
| Macro-F1 | GF            | 79.49   | 80.39   | 80.82   | 81.08   | 81.26   | 81.40   | 81.52   | 81.61   | 81.68   |
|          | DeepWalk      | 78.78   | 79.78   | 80.30   | 80.56   | 80.82   | 80.97   | 81.11   | 81.24   | 81.32   |
|          | SkipGram      | 79.74   | 80.71   | 81.15   | 81.46   | 81.63   | 81.78   | 81.88   | 81.98   | 82.01   |
|          | LINE-SGD(1st) | 75.85   | 76.90   | 77.40   | 77.71   | 77.94   | 78.12   | 78.24   | 78.29   | 78.36   |
|          | LINE-SGD(2nd) | 74.70   | 76.45   | 77.43   | 78.09   | 78.53   | 78.83   | 79.08   | 79.29   | 79.46   |
|          | LINE(1st)     | 79.54   | 80.44   | 80.82   | 81.13   | 81.29   | 81.43   | 81.51   | 81.60   | 81.59   |
|          | LINE(2nd)     | 79.82   | 80.81   | 81.22   | 81.52   | 81.71   | 81.82   | 81.92   | 82.00   | 82.07   |
|          | LINE(1st+2nd) | 80.94** | 81.99** | 82.49** | 82.83** | 83.07** | 83.29** | 83.42** | 83.55** | 83.66** |

Significantly outperforms GF at the: \*\* 0.01 and \* 0.05 level, paired t-test.

社交网络：

与语言网络相比，社交网络更加稀缺；将每个节点分配到一个或多个社区的多标签分类任务来评估顶点嵌入；随机抽取不同百分比的顶点进行训练，其余用于评估。结果在10次不同运行中进行平均。

Table 5: Results of multi-label classification on the FLICKR network.

| Metric   | Algorithm        | 10%     | 20%     | 30%     | 40%     | 50%     | 60%     | 70%     | 80%     | 90%     |
|----------|------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Micro-F1 | GF               | 53.23   | 53.68   | 53.98   | 54.14   | 54.32   | 54.38   | 54.43   | 54.50   | 54.48   |
|          | DeepWalk         | 60.38   | 60.77   | 60.90   | 61.05   | 61.13   | 61.18   | 61.19   | 61.29   | 61.22   |
|          | DeepWalk(256dim) | 60.41   | 61.09   | 61.35   | 61.52   | 61.69   | 61.76   | 61.80   | 61.91   | 61.83   |
|          | LINE(1st)        | 63.27   | 63.69   | 63.82   | 63.92   | 63.96   | 64.03   | 64.06   | 64.17   | 64.10   |
|          | LINE(2nd)        | 62.83   | 63.24   | 63.34   | 63.44   | 63.55   | 63.55   | 63.59   | 63.66   | 63.69   |
|          | LINE(1st+2nd)    | 63.20** | 63.97** | 64.25** | 64.39** | 64.53** | 64.55** | 64.61** | 64.75** | 64.74** |
|          | LINE(1st+2nd)    | 63.20** | 63.97** | 64.25** | 64.39** | 64.53** | 64.55** | 64.61** | 64.75** | 64.74** |
| Macro-F1 | GF               | 48.66   | 48.73   | 48.84   | 48.91   | 49.03   | 49.03   | 49.07   | 49.08   | 49.02   |
|          | DeepWalk         | 58.60   | 58.93   | 59.04   | 59.18   | 59.26   | 59.29   | 59.28   | 59.39   | 59.30   |
|          | DeepWalk(256dim) | 59.00   | 59.59   | 59.80   | 59.94   | 60.09   | 60.17   | 60.18   | 60.27   | 60.18   |
|          | LINE(1st)        | 62.14   | 62.53   | 62.64   | 62.74   | 62.78   | 62.82   | 62.86   | 62.96   | 62.89   |
|          | LINE(2nd)        | 61.46   | 61.82   | 61.92   | 62.02   | 62.13   | 62.12   | 62.17   | 62.23   | 62.25   |
|          | LINE(1st+2nd)    | 62.23** | 62.95** | 63.20** | 63.35** | 63.48** | 63.48** | 63.55** | 63.69** | 63.68** |
|          | LINE(1st+2nd)    | 62.23** | 62.95** | 63.20** | 63.35** | 63.48** | 63.48** | 63.55** | 63.69** | 63.68** |

Significantly outperforms DeepWalk at the: \*\* 0.01 and \* 0.05 level, paired t-test.

引用网络：

通过GF和LINE两种方法对引用网络进行评估。还通过多标签分类任务评估顶点嵌入。我们选择7个流行会议，包括AAAI，CIKM，ICML，KDD，NIPS，SIGIR和WWW作为分类类别。

DBLP(AuthorCitation) Network：





Table 7: Results of multi-label classification on DBLP(AUTHORCITATION) network.

| Metric   | Algorithm     | 10%               | 20%                | 30%                | 40%                | 50%                | 60%                | 70%                | 80%                | 90%                |
|----------|---------------|-------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Micro-F1 | DeepWalk      | 63.98             | 64.51              | 64.75              | 64.81              | 64.92              | 64.99              | 64.99              | 65.00              | 64.90              |
|          | LINE-SGD(2nd) | 56.64             | 58.95              | 59.89              | 60.20              | 60.44              | 60.61              | 60.58              | 60.73              | 60.59              |
|          | LINE(2nd)     | 62.49<br>(64.69*) | 63.30<br>(65.47**) | 63.63<br>(65.85**) | 63.77<br>(66.04**) | 63.84<br>(66.19**) | 63.94<br>(66.25**) | 63.96<br>(66.30**) | 64.00<br>(66.12**) | 63.77<br>(66.05**) |
| Macro-F1 | DeepWalk      | 63.02             | 63.60              | 63.84              | 63.90              | 63.98              | 64.06              | 64.09              | 64.11              | 64.05              |
|          | LINE-SGD(2nd) | 55.24             | 57.63              | 58.56              | 58.82              | 59.11              | 59.27              | 59.28              | 59.46              | 59.37              |
|          | LINE(2nd)     | 61.43<br>(63.49*) | 62.38<br>(64.42**) | 62.73<br>(64.84**) | 62.87<br>(65.05**) | 62.93<br>(65.19**) | 63.05<br>(65.26**) | 63.07<br>(65.29**) | 63.13<br>(65.14**) | 62.95<br>(65.14**) |

Significantly outperforms DeepWalk at the: \*\* 0.01 and \* 0.05 level, paired t-test.

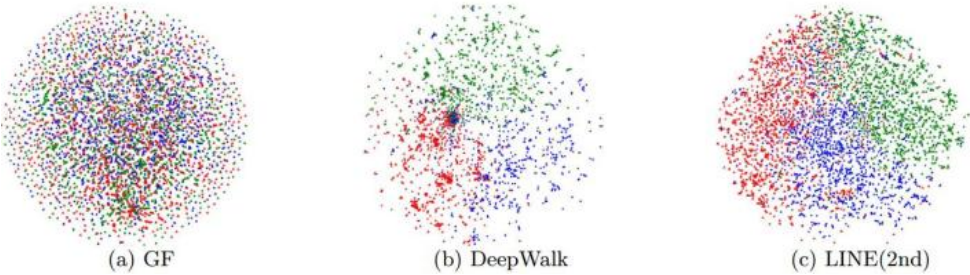
DBLP(PaperCitation) Network :

Table 8: Results of multi-label classification on DBLP(PAPERCITATION) network.

| Metric   | Algorithm | 10%                | 20%                | 30%                | 40%                | 50%                | 60%                | 70%                | 80%                | 90%                |
|----------|-----------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Micro-F1 | DeepWalk  | 52.83              | 53.80              | 54.24              | 54.75              | 55.07              | 55.13              | 55.48              | 55.42              | 55.90              |
|          | LINE(2nd) | 58.42<br>(60.10**) | 59.58<br>(61.06**) | 60.29<br>(61.46**) | 60.78<br>(61.73**) | 60.94<br>(61.85**) | 61.20<br>(62.10**) | 61.39<br>(62.21**) | 61.39<br>(62.25**) | 61.79<br>(62.80**) |
| Macro-F1 | DeepWalk  | 43.74              | 44.85              | 45.34              | 45.85              | 46.20              | 46.25              | 46.51              | 46.36              | 46.73              |
|          | LINE(2nd) | 48.74<br>(50.22**) | 50.10<br>(51.41**) | 50.84<br>(51.92**) | 51.31<br>(52.20**) | 51.61<br>(52.40**) | 51.77<br>(52.59**) | 51.94<br>(52.78**) | 51.89<br>(52.70**) | 52.16<br>(53.02**) |

Significantly outperforms DeepWalk at the: \*\* 0.01 and \* 0.05 level, paired t-test.

训练结果：



四、综述

LINE模型具有精心设计的客观功能，保留了一阶和二阶接近度，相互互补。并提出了一种有效和有效的边缘抽样方法进行模型推理；解决了加权边缘随机梯度下降的限制，而不影响效率。此外，除一阶和二阶之外更高的相似度也是LINE模型算法在未来能够更加拓宽的方面。异构网络的嵌入，也是研究的方向之一。

发布于 2017-06-11

深度学习 ( Deep Learning )

文章被以下专栏收录



RUC智能情报站

本专栏由中国人民大学信息学院的赵鑫副教授 (微博 @赵鑫RUC) 所带领的研究小组维...

关注专栏

推荐阅读

《Binarized Neural Networks: Neural Networ...

转载请注明出处：西土城的搬砖日常  
论文链接：Binarized Neural Networks: Training Networks with Wei

25

20 条评论

分享

收藏

...



《Generative Adv Text to Image Syr

转载请注明出处：西土城  
原文链接：《Gener

20 条评论


切换为时间排序

写下你的评论...

 陈——1 年前


膜拜大佬！

赞

 HTLiu1 年前


其实在实现中，O1和O2都先边采样，再负采样的。

赞

 Phoenix Cat11 个月前


这篇文章其实写得很成问题，首先一阶相似那个公式就意义不明，最后为了突出“一阶”相似强行把公式中的 $u_k$ 改成 $u_k$ ，简直匪夷所思。二阶相似性直接抄NLP但至少没问题，就是本质上除了采样概率就跟skip-gram一毛一样了，想想都知道不可能语言数据集上outperform word2vec那么多.....实际测评也是如此，实验室同学反复追问作者后承认做了论文中没有提到的大幅度预处理.....

3

 Zheng Wang 回复 Phoenix Cat11 个月前


其实LINE的一阶和二阶相似性都是等价于矩阵分解的（证明可见[1707.05926] Equivalence between LINE and Matrix Factorization），在我们的分析中发现LINE其实只考虑了"existing link relationships"，当然这篇论文在15年还算是蛮有亮点的。

赞 查看对话

 Phoenix Cat 回复 Zheng Wang11 个月前


是等价于矩阵分解啊就一个对称一个非对称.....而且我记得这个前人说过呀.....好像刘知远还是谁组里的一篇，之前我们组的一个学长也写过一篇不过没中.....

赞 查看对话

 Phoenix Cat 回复 Zheng Wang11 个月前


我说写得不好是他文中的公式跟做得差别太大，等价于矩阵分解这种事情应该再原文中就说明白.....

赞 查看对话

 Zheng Wang 回复 Phoenix Cat11 个月前


恩，估计是因为一方面其实质上只考虑了"existing links"，另一方面因为sgd这种方法本身就不稳定

赞 查看对话

 Phoenix Cat 回复 Zheng Wang11 个月前

sgd不稳定？你是说那个采样的方法吗？我觉得那部分分析挺不错的，至少挺务实的

赞 查看对话

 Zheng Wang 回复 Phoenix Cat10 个月前

SGD方法最大的问题就是其每一步的更新都不能保证全局的loss值下降（因为它是局部更新，而不是全局更新），第二个问题就是步长的选择的问题了。相对于GD当步长足够小的时候每一步更新都能保证loss的减少，SGD真的是不太稳定。



LINE的SGD分析的确是挺务实的，赞一个。

👍 1    💬 查看对话

 issac Syndrome 回复 HTLiu 6 个月前

貌似除了多个了context\_embedding，其他的实现都是一样的？

👍 赞    💬 查看对话

 HTLiu 回复 issac Syndrome 6 个月前

对，用的同一个函数，一阶二阶更新一样。。

👍 赞    💬 查看对话

 歌者默然 回复 HTLiu 2 个月前

在计算O2的时候，为什么要引入节点的context\_embedding呢？根据二阶损失的定义，可以直接根据一阶相似度来计算二阶损失啊？

👍 赞    💬 查看对话

 HTLiu 回复 歌者默然 2 个月前

不能呀，二阶里面公式是一个节点的context\_emb 与另一个节点的向量内积计算的。只是计算公式一样

👍 1    💬 查看对话

 歌者默然 回复 HTLiu 2 个月前

谢谢你的回复。我其实有点想不明白这里为什么要引入节点的context\_emb这个东西，应该怎么去理解呢？根据作者前面的definition 3部分二阶损失的定义，可以先计算出每个节点关于其他节点的一阶相似度，得到一个n维向量，然后可以根据两个节点的这个n维向量计算相似度得到二阶相似度。

👍 赞    💬 查看对话

 歌者默然 回复 HTLiu 2 个月前


另外关于论文中的一阶损失的定义我觉得有些不太严谨，希望可以跟你探讨一下。在定义一阶相似度的时候，使用sigmoid函数来处理向量的内积得到边的概率，作者把所有边的这种概率认为是一种分布，然后与经验分布计算kl散度。我的问题是，经验分布显然满足归一化条件，但是sigmoid函数的输出是0-1之间，这里的p1的分布好像不能保证归一化吧？是不是也应该像p2那样进行归一化呢？

👍 赞    💬 查看对话

 HTLiu 回复 歌者默然 2 个月前

context\_emb 这个向量的引入，可以参考word2vec的原理，应该是同一个意思。大概就是一个节点有两个角色，一个是作为邻居节点，一个是中心节点

👍 赞    💬 查看对话

 歌者默然 回复 HTLiu 2 个月前

恕我鲁钝。word2vec里的context指的就是目标词语周围skip-window里的词语吧，这里的context向量居然跟目标节点向量——对应，实在是有点费解。而且感觉跟二阶损失完全沾不上边啊。。。

👍 赞    💬 查看对话

 HTLiu 回复 歌者默然 2 个月前

不是，word2vec的原理中，每个词有两个Embedding，一个是中心的Embedding，另一个就是Context Embedding，也就是第二层神经网络的权重矩阵。一阶信息是节点的邻居，这里的二阶其实是邻居的邻居，即两个节点的共同节点很多，就可能比较相似，这个也是它的动机。p2的定义就是一个节点的邻居条件概率。

👍 1    💬 查看对话

 HTLiu 回复 歌者默然

👍 25    💬 20 条评论    ➦ 分享    ★ 收藏    ...

对，关于一阶的 $p_1$ 的定义，说成分布的确不太合适。只能说是为了输出到概率范围，方便与后续经验概率计算loss



👍 赞    💬 查看对话

 歌者默然 回复 HTLiu

2 个月前

感谢你的耐心解答！

👍 赞    💬 查看对话