

最新： 添加了PTranse(EMNLP 2015) 的源码!

评测结果

以下是我们自己实现的方法在数据集FB15k和WN18上的测试结果:

FB15k

Model	MeanRank(Raw)	MeanRank(Filter)	Hit@10(Raw)	Hit@10(Filter)
TransE(paper)	243	125	34.9	47.1
TransH(paper)	212	87	45.7	64.4
TransR(n=50)	198	77	48.2	68.7
TransE(Our, n=50)	210	82	41.9	61.3
TransE(Our, n=100)	205	63	47.9	70.2
PTranse (ADD, 2-step)	200	54	51.8	83.4
PTranse (MUL, 2-step)	216	67	47.4	77.7
PTranse (RNN, 2-step)	242	92	50.6	82.2
PTranse (ADD, 3-step)	207	58	51.4	84.6

WN18

Model	MeanRank(Raw)	MeanRank(Filter)	Hit@10(Raw)	Hit@10(Filter)
TransE(paper)	263	251	75.4	89.2
TransH(paper)	318	303	75.4	86.7
TransR	238	225	79.8	92.0
TransE(Our)	251	239	78.9	89.8

数据

我们按照源码里输入的格式，在data.zip文件中提供了用于链接预测的数据集FB15k和WN18。原始数据可在此处下载：

FB15k, WN18在这里发布”Translating Embeddings for Modeling Multi-relational Data (2013).” [\[Download\] \(https://everest.hds.utc.fr/doku.php?id=en:transe\)](https://everest.hds.utc.fr/doku.php?id=en:transe)

FB13, WN11在这里发布”Reasoning With Neural Tensor Networks for Knowledge Base Completion”. [\[Download\] \(http://cs.stanford.edu/~danqi/data/nips13-dataset.tar.bz2\)](http://cs.stanford.edu/~danqi/data/nips13-dataset.tar.bz2)

New York Times Corpus: 此数据用于文本关系抽取，发布在”Modeling relations and their mentions without labeled text”。此数据应先在LDC (<https://catalog ldc.upenn.edu/LDC2008T19> (<https://catalog ldc.upenn.edu/LDC2008T19>))处获得。

FB40k [\[Download\] \(http://pan.baidu.com/s/1c0xrtVa\)](http://pan.baidu.com/s/1c0xrtVa)

数据集需要放在 data/ 文件下，并按照如下的格式要求，一共有6个文件：

- train.txt: training file, format (e1, e2, rel).
- valid.txt: validation file, same format as train.txt
- test.txt: test file, same format as train.txt.
- entity2id.txt: all entities and corresponding ids, one per line.
- relation2id.txt: all relations and corresponding ids, one per line.
- e1_e2.txt: the top-500 entity pairs which are calculated by TransE. [\[Download\] \(https://pan.baidu.com/s/1c2iLtmg\)](https://pan.baidu.com/s/1c2iLtmg)

源码

源码在文件夹 TransE/, TransR/, CTransR/下.

编译

只需要在目录 ./ 下输入make并执行

训练

为了训练，需要执行以下步骤：

TransE: 执行在目录 TransE/ 下面的 Train_TransE 程序

TransH: 执行在目录 TransH/ 下面的 Train_TransH 程序

TransR:

- 训练TransE的unif method用于初始化
- 执行在目录 TransR/ 下面的 Train_TransR 程序

CTransR:

- 训练TransR的unif method用于初始化
- 在目录 cluster/ 下运行bash文件 run.sh, 参数为relation number, 从而对训练数据中的三元组聚类。例如： bash run.sh 10
- 执行在目录 CTransR/ 下面的 Train_cTransR 程序

在运行Train_TransE, Train_TransR, Train_CTransR时，你也可以改变以下这些参数：

-size : the embedding size k, d

-rate : learning rate

-method: 0 - unif, 1 - bern

测试

For testing, you need to follow the steps below:

为了测试，你需要执行以下步骤：

TransR: 执行在目录 TransR/ 下的 Test_TransR 程序，参数为 ‘method’

CTransR: 执行在目录 CTransR/ 下的 Test_CTransR 程序，参数为 ‘method’

这将在 test.txt 上评测并且报告 mean rank 和 Hits@10

引用

如果你使用了此源码，请在参考文献中标明引用以下论文：

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, Xuan Zhu. Learning Entity and Relation Embeddings for Knowledge Graph Completion. The 29th AAAI Conference on Artificial Intelligence (AAAI'15).[pdf]

(http://nlp.csai.tsinghua.edu.cn/~lzy/publications/aaai2015_transr.pdf)