

方法：

(1) 利用 <https://github.com/phanein/deepwalk> 和 `reddit.coo` 中的边信息生成带边节点的 embedding, embedding 是 64 维的。<https://github.com/phanein/deepwalk> 不能利用边的权重信息, 所以 `reddit.coo` 的边权重信息在生成 embedding 时没有发挥作用。

(2) 把每个节点表示成 666 维向量, 前 602 维对应 `reddit.feature` 中的 602 维向量, 后 64 维是步骤 (1) 中得到的 embedding。在 (1) 中, 没有边的节点是没有 embedding 的, 对于这些没有边的节点, 后 64 维全用 0 进行填充。

(3) 利用 `sklearn.neural_network` 的 `MLPClassifier` 对节点进行分类。`reddit.label` 文件有后 19297 个节点的类别信息 (41 维的 one-hot 向量), 将类别信息作为分类器的输出。分类器的输入为已知类别的节点的 666 维向量表示。利用已知类别的节点的向量表示和类别信息对分类器进行训练, 得到训练好的分类器。将前 4000 个未知类别的节点的 666 维向量作为训练好的分类器的输入, 用分类器对 4000 个未知类别的节点进行分类, 得到 41 维 one-hot 向量。

*分类器的参数:

```
(hidden_layer_sizes=(100,100),activation='logistic', solver='adam',learning_rate_init=0.0001,
max_iter=1000)
```

*代码在 `1701214038pred.py` 中。