

r06922115

鄭皓謙

HW1 report

Task 1

hw1我嘗試了node2vec以及GCN train on degree count兩種embedding，最後選擇使用node2vec 當作graph embedding。為seen-test產生了一些fake link，配合uniform negative sampling訓練classifier。後面的link prediction classifier則是nn base的方法。

Classifier

將兩個node的graph embedding 接起來當作輸入，經過多層的linear with selu activation function最後用binary classification loss做訓練。

Embedding

這是這次作業的重點，最後是利用training跟seen-test訓練node2vec當作點的特征，丟到classifier。

Negative sampling

只對training set做uniform distribution，sample一些不存在的link當作negative sample。

嘗試與心得

negative sampling應該相當影響這次的作業，而我沒有發現這部分的重要，只使用了uniform sampling，而是嘗試了不同node2vec的參數，但由於沒有validation set的緣

故，只根據uniform sampling，沒有適合的方式評估embedding的好壞，後來試了幾組node2vec的參數後，就開始研究GCN。

最後GCN的部分以失敗告終，感覺是實作的問題，先嘗試了out-degree當作regression目標，訓練了幾次GCN都overfit得相當嚴重。GCN的部分，首先利用了one-hot encoding作為feature以及neighbor count作為目標，訓練簡單的L2 regression，結果非常的容易overfit，嘗試了不同層的GCN，將中間的不同層數的hidden vector做L1 L2 norm regularization，經過幾種徒勞的嘗試後，選擇放棄GCN。

原本嘗試產生test node的fake link，計算testing set每個node的out-degree，然後再uniform sample training set，作為50% missing的fake link，但是沒有比較好就放棄了。後來覺得應該是產生fake link的方式不夠好，後來在task2使用fake link，結果就好一些。