

# 社群網路與推薦系統 HW2

## Report

賴廷瑋 F44054045

### 目錄

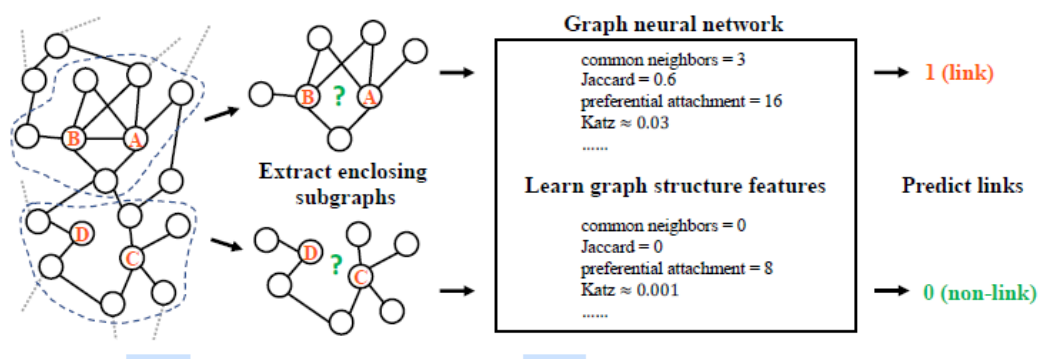
Introduction .....	2
Methodology .....	2
Implementation Detail .....	3
Node labeling .....	4
Experimental Analysis .....	7
參數說明 .....	7
Insights .....	7
Conclusion .....	10
Novelty .....	10
Findings Summarization .....	10
Future work .....	10
Citations .....	10

## Introduction

針對本次作業的 Link Prediction 任務，使用 Muhan Zhang 與 Yixin Chen 兩人於 2018 年所提出之 Link Prediction Based on Graph Neural Networks 中的 SEAL 監督式深度模型。實作過程大多皆參照作者提出的做法與架構實踐，並同時利用 Pytorch Geometric 實踐，上傳競賽成績約為 0.8

## Methodology

Learning from Subgraphs, Embeddings and Attributes for Link prediction (SEAL), 為該篇論文作者提出對 Link Prediction 任務設計出的深度模型架構。該模型利用由兩節點形成的 subgraph 中的節點位置訊息，以及在 subgraph 中參與節點的 Attribute 或是 Embedding 作為 feature，預測這兩個節點之間是否存在 Link 的關係。



上圖由論文中擷取。

圖中可知，從兩點所展開的 subgraph 中的各點所形成的 node position feature 以及 node attribute 或 node embedding 即成為該 subgraph 的 feature matrix，而透過後續的 Graph Neural Network 將此 subgraph 分類為 0 或 1，即有無連結產生，特別要注意的是，該圖方框中顯示的 jaccard, katz...等特徵僅為示意，並非實際，僅以表示由 subgraph 表示的 graph feature 可能可以抓到如 kaccard, katz...等訊息，導致更好的分類表現。

## Implementation Detail

SEAL 方法主要核心在於 subgraph feature 的建構，論文作者稱此步驟為 node labeling，即針對 subgraph 中每一個節點產生 structural label，表達節點在 subgraph 中的相對位置訊息。

## Node labeling

### Criteria

1. Target nodes  $x$  和  $y$ (欲針對進行分類的兩節點)，固定被 label 成 "1"。
2. 除了 target nodes 但在 subgraph 中的節點  $i, j$  有相同的 Label，如果  $d(i, x) = d(j, x)$  and  $d(i, y) = d(j, y)$ ，這麼做的目的是，讓與中心兩點距離相同的節點有著相同的 label，而反映節點之間的相對關位置訊息和在 subgraph 中結構的重要性。

### Method

作者針對 node labeling 提出了 Double-Radius Node Labeling (DRNL)方法，具體方式如下：

1. 將要預測的 target nodes  $x, y$  label 為 1
2. 任何節點若與中心兩點距離為 1，即  $(d(i, x); d(i, y)) = (1; 1)$ ，則將此節點 label 為 2，若距離為  $(1, 2)$  或  $(2, 1)$ ，則標記為 3，距離為  $(1, 3)$  或  $(3, 1)$ ，標記為 4， $(2, 2)$  或  $(2, 2)$ ，則標記為 5， $(1, 4)$  或  $(4, 1)$ ，標記為 6， $(2, 3)$  或  $(3, 2)$ ，標記為 7，以此類推。
3. 在計算上，DRNL 的好處即有 Hashing function 可以幫忙標記，減

少計算時間，hashing function 爲：

$$f_l(i) = 1 + \min(d_x, d_y) + (d/2)[(d/2) + (d\%2) - 1]$$

其中  $d_x := d(i, x)$ ,  $d_y := d(i, y)$ ,  $d := d_x + d_y$

## Node Attribute

除了 graph feature 外，SEAL 亦可將 subgraph 中各節點的 attribute 納入，爲模型提供更多的訊息。作法是將 subgraph 的 graph feature 和其中的 node attribute 做 concatenation。

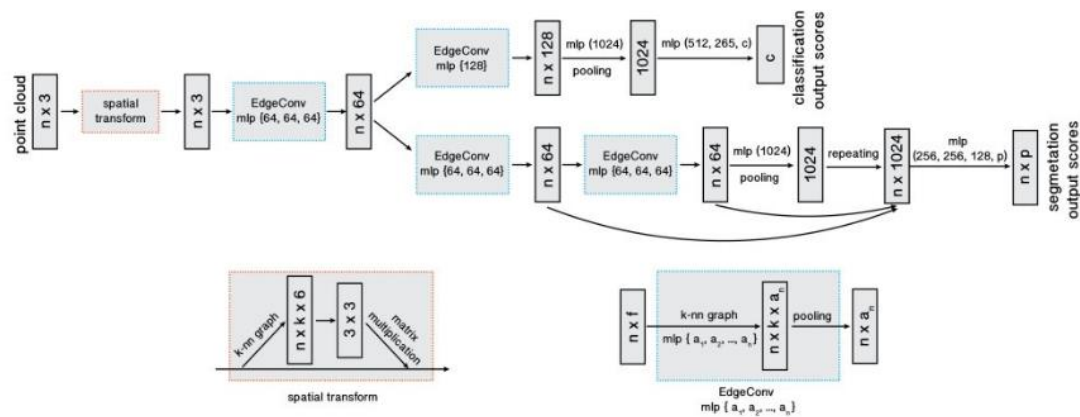
## MODEL

得到每組 target nodes 所產生的 graph feature 後，本篇論文作者利用 GNN 模型中 DGCNN 架構進行深度模型訓練與預測。

### DGCNN 介紹

1. 解決的問題：DGCNN 針對點雲的結構做特徵學習，獲取足夠的局部訊息，用於分類和分割等 TASK。
2. 方法：提出新的神經網路 EdgeConv，其捕捉了局部領域的訊息，作爲特徵，並可以透過堆疊 EdgeConv 提取到全局訊息。

3. 架構如下：



上圖中，top branch 用於 classification 的任務，也是本次 Link prediction 中使用的架構。針對 Classification 的模型，input 為  $n$  個 data，在 EdgeConv layer 層中計算每個點的 edge feature(size= $k$ )，並將其 aggregate。經過多層 EdgeConv 後，至最後一層 EdgeConv layer 將所有從 EdgeConv 的 output 做 aggregation，形成一維的 classification score。

而針對 EdgeConv block 的部分，input 維度為  $n \times f$ ，過程中計算這  $n$  個資料各自 edge feature，透過利用 multi-layer perceptron(mlp)，而此 mlp 有著 layer neuron 數： $\{a_1, a_2, \dots, a_n\}$ ，最後經過 neighboring edge feature pooling 生成 tensor with shape  $n \times a_n$ 。

## Experimental Analysis

### Experimental Parameters

參數	值
Num_hops	2
Batch_size	32
Hidden_channels	64
Num_layers	3
lr	1e-4
epoch	30

### 參數說明

1. Num\_hops: target node x,y 由 edge index 隨機跳躍步數(形成 subgraph)
2. Num\_layers: DGCNN 中 EdgeConv 層數

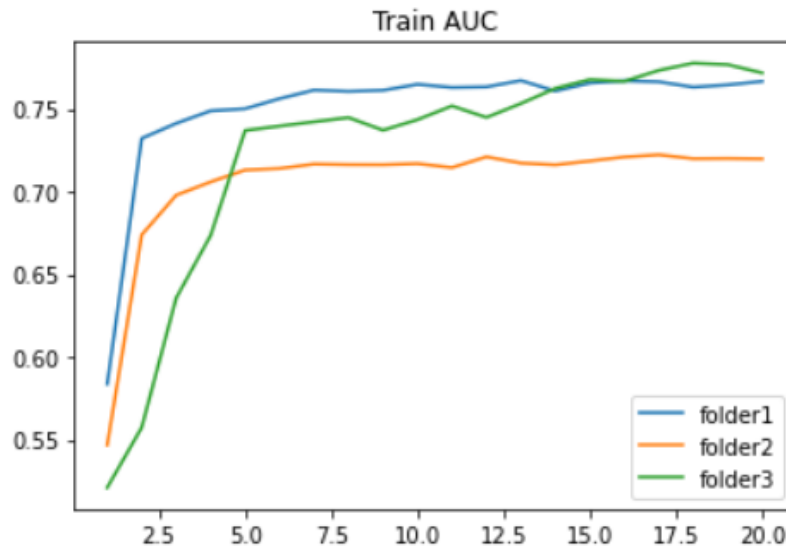
### Insights

1. 僅利用 graph feature(node labeling)，無使用 node attribute

實驗過程中，針對三份 dataset 的模型訓練皆訓練 20 epochs 後發

現，若無考慮 node attribute 進行模型訓練，則在進行 7 次 epoch 後

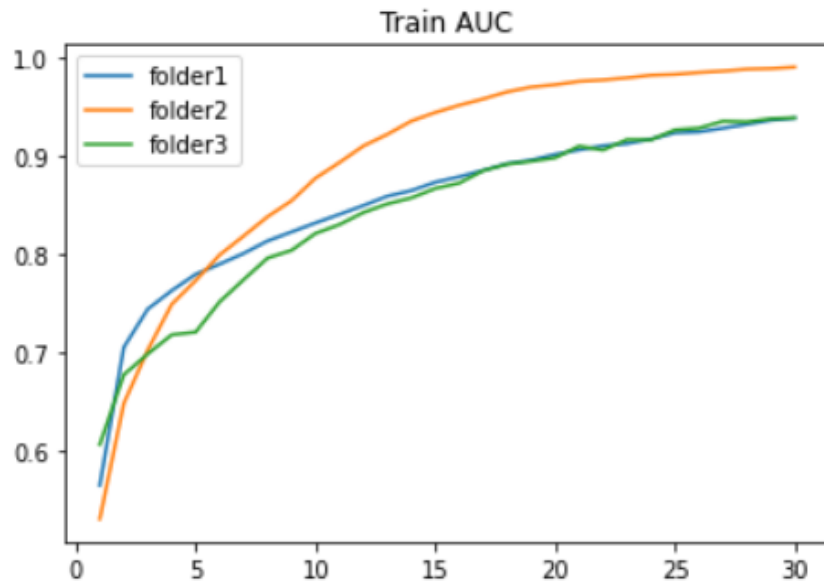
三個資料集之 AUC 則收斂至 0.7 至 0.8 之間左右。



## 2. 加入 node attribute 後

加入 node attribute 的特徵後，可以發現三個資料集訓練的 AUC 皆可收斂至 0.9 以上，但有趣的是，經過上傳至競賽後發現，**TEST** 的 AUC 以及 AP 結果在與有無加入 node attribute 並無顯著差異，儘管加入 node attribute 後約莫平均 AUC, AP 有 0.05 的提升。表示加入 node attribute 後的模型有 overfit 的現象(降低 hidden\_size 或 batch\_size 皆對結果沒有顯著提升)。





而這不滿意的預測結果我歸納有兩種可能：

1. Node attribute 高維(1000 多維)導致模型容易 overfit
2. 本次模型並未加入 node embedding 的特徵，未完全發揮 SEAL 的潛力。

## Conclusion

### Novelty

SEAL 結合了節點的圖形訊息，並無用到全部的 graph 進行預測，大大降低了演算時間，並且引入了 EdgeConv 的作法，針對點雲做分類，且 SEAL 亦能結合 node embedding 或 node attribute 增加預測的準確度。

### Findings Summarization

1. 節點 attribute 維度過高可能導致此模型預測穩定度下降
2. SEAL 主要核心為 graph feature 的生成，而 node embedding 和 node attribute 僅為輔助（雖本次作業未親測加入 node embedding 對於預測結果是否有明顯提升）

### Future work

若能針對 node attribute 有高維現象而導致將低預測穩定性有額外方法，則 SEAL 可發揮它真實的預測能力。

### Citations

Link Prediction Based on Graph Neural Networks, Muhan Zhang, 2018