

隨機特徵圖神經網路之穩健表格異常偵測

Robust Anomaly Detection with Random-Feature

Graph Neural Networks

指導教授：李政德

專題成員：唐文蔚

開發工具：Python、Pytorch

測試環境：Colab(Ubuntu)

一、簡介：

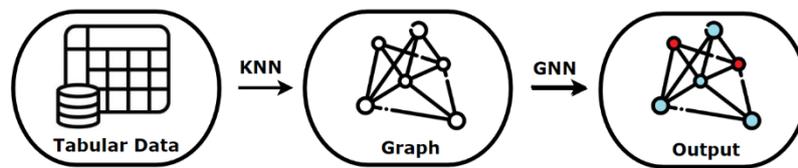
本專題旨在改進 LUNAR 模型，提升其在表格資料異常偵測中的穩健性。

LUNAR 模型使用圖機器學習方法來偵測異常資料，然而，若是訓練資料中混入異常資料就會導致準確率下降。本專題將特徵分成多個子集，分別訓練不同的模型，綜合各模型的結果來進行異常偵測，並額外引入 GraphMAE 和 MLP Layer 做特徵重建，以提升模型的穩健度和準確度。

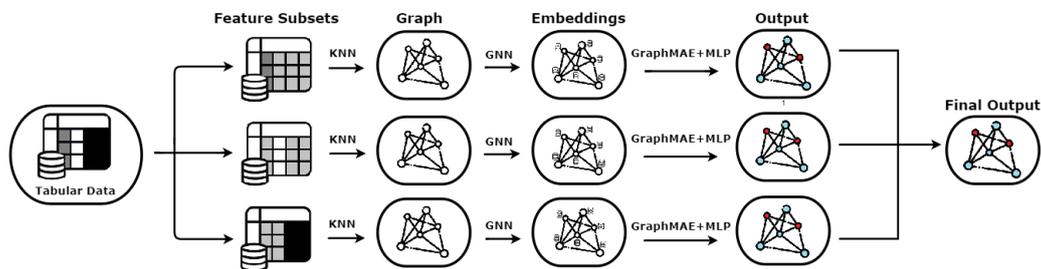
特徵分割：將表格數據中的特徵分成多個子集，針對每個特徵子集，分別訓練不同的 GNN 模型，並將各子集的結果綜合進行異常資料的判斷。

在嘗試多種分群方法後發現在隨機分群、且有 Overlap 的情況下表現最好。

GraphMAE 和 MLP：將 GNN 的輸出結果作為 Embedding 進行特徵重建，最後經過 MLP Layer 產出異常值結果。



圖一：LUNAR 模型架構



圖二：改進後的模型架構

通過將特徵分成多個子集並分別訓練多個 GNN 模型，本專題提出的改進方法提升了 LUNAR 模型在表格異常偵測中的穩健度和準確率，顯示了其在實際應用中的潛力和價值。

二、測試結果：

第一個表格為用多個 Subset 的模型和原始的模型比較，使用 Subset 後雖然沒有明顯提升，但還是有效果。

第二個表格是針對混入異常資料的表格資料，可以看到有用 Subset 的模型可以有較高的準確度。

第三個表格是針對減少 Training Data 的比較結果。

Target 3 – Origin Summary

Dataset	LUNAR	K = 3	K = 6	K = 9
HRSS	92.17	91.06	91.76	92.06
MI-F	84.37	83.89	83.93	83.13
MI-V	96.73	97.98	99.30	99.66
OPTDIGITS	99.76	99.99	99.94	99.71
PENDIGITS	99.81	100.00	100.00	97.76
SATELLITE	85.35	95.77	97.47	96.44
SHUTTLE	99.97	99.99	99.81	99.98
THYROID	85.44	88.97	90.56	88.75

Target 3 – Target1 Summary

Target1	P = 10			P = 30			P = 50					
	K = 3	K = 6	K = 9	K = 3	K = 6	K = 9	K = 3	K = 6	K = 9			
HRSS	87.13	92.00	93.69	93.94	78.33	83.78	86.85	88.82	70.10	81.26	83.23	83.45
MI-F	50.34	50.69	51.79	50.01	46.54	44.95	46.53	44.89	47.19	43.78	41.90	41.27
MI-V	77.15	77.66	85.70	85.38	64.10	67.18	70.72	71.83	59.85	64.64	66.68	67.43
OPTDIGITS	94.57	99.63	99.54	99.03	68.44	66.37	80.97	86.43	50.49	56.78	67.35	68.79
PENDIGITS	99.56	99.95	99.94	99.86	53.95	76.11	90.73	92.34	25.20	46.19	78.16	73.54
SATELLITE	51.05	68.73	76.33	73.95	45.64	63.48	67.14	64.34	49.94	57.39	60.39	61.83
SHUTTLE	39.34	96.93	98.49	99.53	90.22	86.95	87.14	93.44	86.11	86.13	80.18	89.51
THYROID	71.36	87.98	85.19	91.02	70.09	95.25	86.12	87.08	66.38	79.39	85.73	89.15

Target 3 – Target2 Summary

Target1	P = 50			P = 70			P = 90					
	K = 3	K = 6	K = 9	K = 3	K = 6	K = 9	K = 3	K = 6	K = 9			
HRSS	82.49	80.40	79.75	79.16	86.66	85.75	83.53	85.90	90.39	90.09	89.93	83.83
MI-F	84.11	83.07	79.38	80.59	82.22	79.07	81.49	80.70	84.16	81.89	81.87	83.84
MI-V	95.33	98.86	99.25	98.37	95.61	99.07	98.64	99.53	96.41	98.09	99.33	99.42
OPTDIGITS	99.39	99.97	99.96	99.74	99.44	99.79	99.37	99.95	99.44	99.71	100.00	99.97
PENDIGITS	99.84	99.86	99.90	100.00	99.84	99.99	99.40	99.89	99.83	99.96	100.00	99.99
SATELLITE	84.14	93.98	95.19	96.59	84.61	97.26	94.54	97.09	86.01	96.90	95.13	95.39
SHUTTLE	99.86	99.99	99.83	99.98	99.97	99.97	99.98	99.95	99.94	99.97	99.98	99.99
THYROID	87.93	91.26	92.43	91.87	85.23	87.95	83.06	86.03	83.68	88.89	85.51	92.97

圖三：準確率比較圖