基於圖神經網路之隱私保護表格資料生成

Privacy-persevering Tabular Data Synthesis with Graph Neural Networks

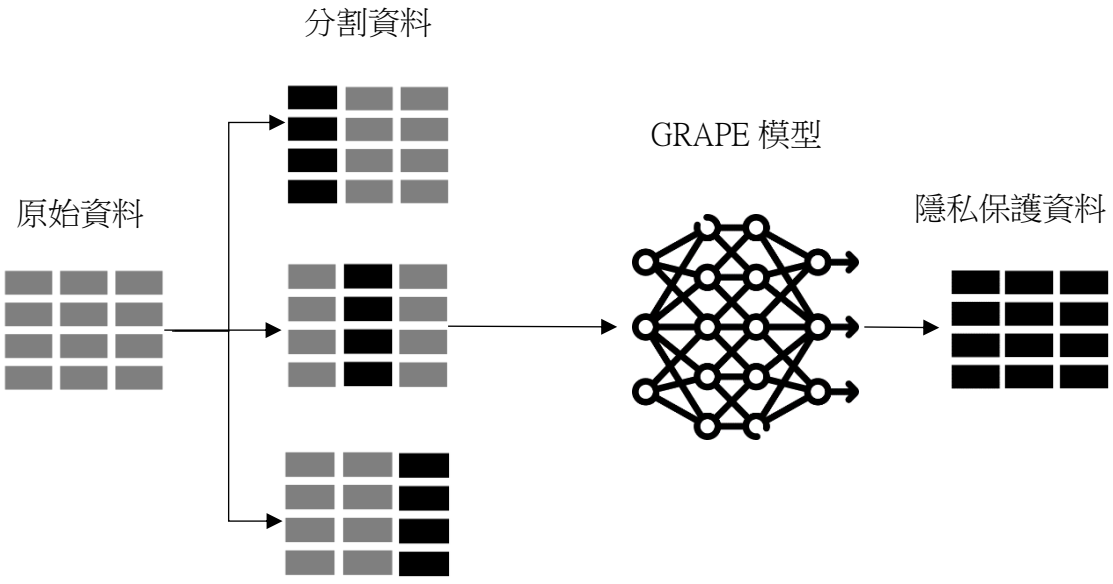
指導教授：李政德
專題成員：呂宏家
開發工具：Python、Jupyter notebook
測試環境：Python 3.11.5

一、簡介：

藉由圖神經網路 GRAPE 的表格資料合成，一般來講，銀行或是醫療機構他們想要把表格的資料釋出給大家使用，但是表格的資料涉及使用者的個資隱私，銀行或醫院不想公開真實的資料，因此我們就必須設計方法來協助合成表格的資料，也就是假的資料，但資料必須符合真實資料的一些特性，來讓銀行跟醫院釋出讓機器學習的模型使用。

將原始資料透過輪流遮罩不同的部分，並輸入 GRAPE 模型產生特徵補值，最後再把生成資料整合，以達到隱私保護的效果。

以下為系統架構圖：



二、測試結果：

在圖四中，我們將原始資料輸入並產生結果後，針對原始資料和新產生的資料，隨機將一個特徵做遮罩，並使用 Random Forest 來測試新的資料能否反推回原始資料的特徵值。

	0	1	2	3	4	5	6	7	8
0	540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28.0	79.99
1	540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28.0	61.89
2	332.5	142.5	0.0	228.0	0.0	932.0	594.0	270.0	40.27
3	332.5	142.5	0.0	228.0	0.0	932.0	594.0	365.0	41.05
4	198.6	132.4	0.0	192.0	0.0	978.4	825.5	360.0	44.30
...	...	...	...	...	...	...	...	...	...
1025	276.4	116.0	90.3	179.6	8.9	870.1	768.3	28.0	44.28
1026	322.2	0.0	115.6	196.0	10.4	817.9	813.4	28.0	31.18
1027	148.5	139.4	108.6	192.7	6.1	892.4	780.0	28.0	23.70
1028	159.1	186.7	0.0	175.6	11.3	989.6	788.9	28.0	32.77
1029	260.9	100.5	78.3	200.6	8.6	864.5	761.5	28.0	32.40

圖二：原始資料

	0	1	2	3	4	5	6	7
0	525	94	102	243	163	438	188	120
1	520	89	108	246	148	438	197	145
2	323	288	105	479	87	270	112	196
3	344	281	112	486	92	272	85	153
4	193	246	97	349	100	303	348	130
...	...	...	...	...	...	...	...	...
1025	263	232	312	331	210	181	301	137
1026	309	102	374	353	230	144	344	124
1027	153	267	132	340	173	213	349	125
1028	171	329	101	278	160	353	304	133
1029	275	226	283	375	212	169	299	132

圖三：經隱私保護後生成的資料

```
MSE of masked data from original data:  3.0887637690809067

MSE of masked data from generated data:  1177.0220239077717
```

圖四：與原始資料相比