

電子研究紀錄簿

系統時間(台北):2021/03/24 17:33:20 登入者: 余承歡

管理與統計

撰寫入口

機密
CONFIDENTIAL

紀錄檢視

首頁 > 管理與統計 > 查閱管理 > 紀錄列表 > 紀錄檢視

上一篇

尚未完成相似度比對

2021/3/23 上午 11:52:21 頁碼: 22 撰寫者: 王若宇(已查閱)

計畫代號: 無

主題: GAN's Data plot 100k entry

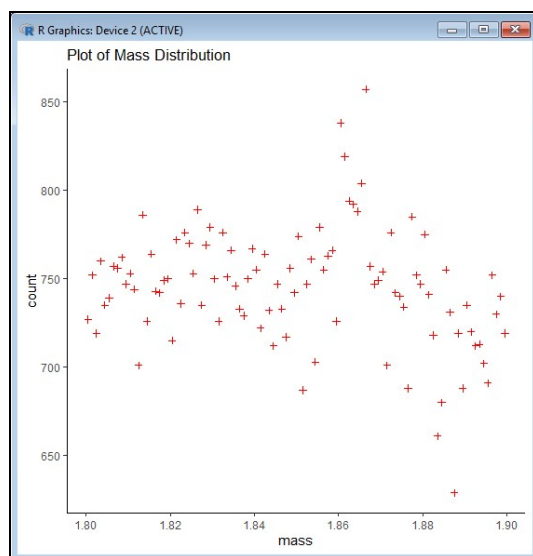
關鍵字/標籤: GAN;ML;python

GAN's Data plot 100k entry

接續上一篇的紀錄，這次是針對100k entry的資料做GAN的訓練與資料的生成，除了原來的檔案，tabgan依舊如同上次的模型係數設定一樣，生成兩種，ctgan模型則是生成一種。

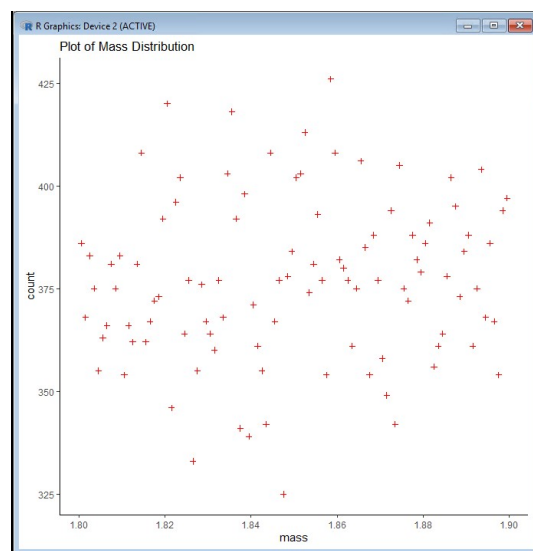
首先是原來的資料，約295萬筆，100001個entry(0-100000)

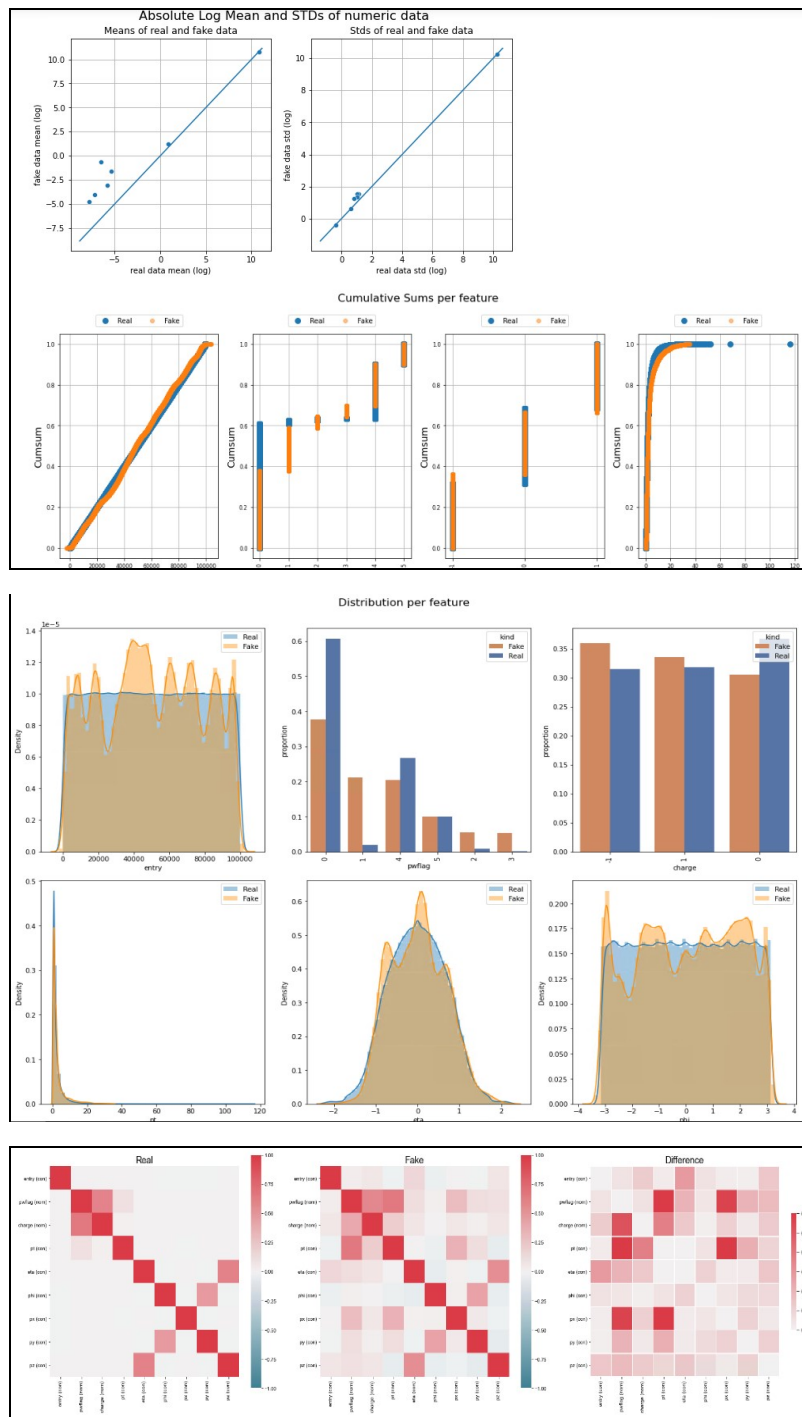
LEP1Data1992_recons_aftercut-MERGED.root.20200630.top100kEntries



總共4430867筆資料，99998個entry(0-100000，可能中間有某幾個entry並沒有生成到)

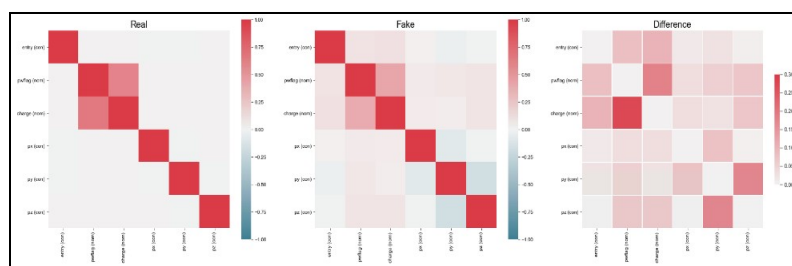
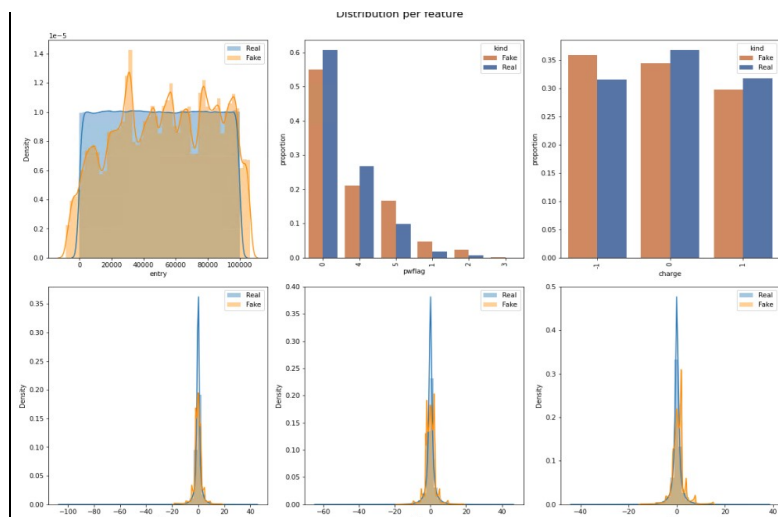
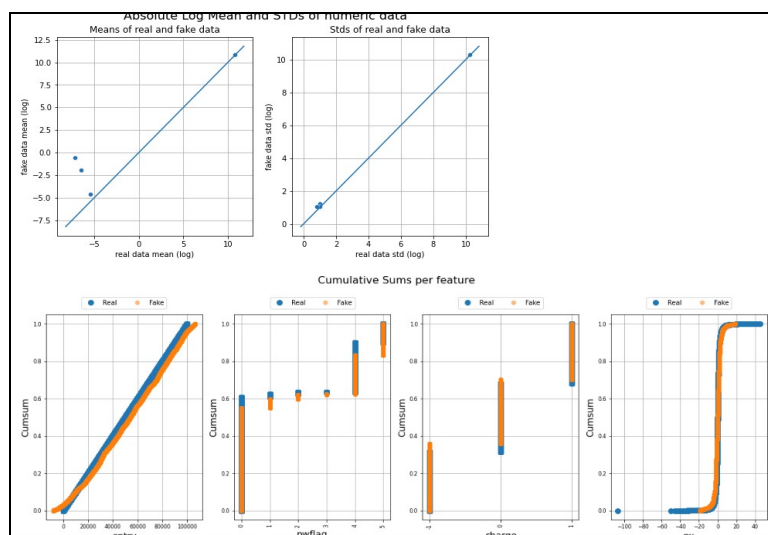
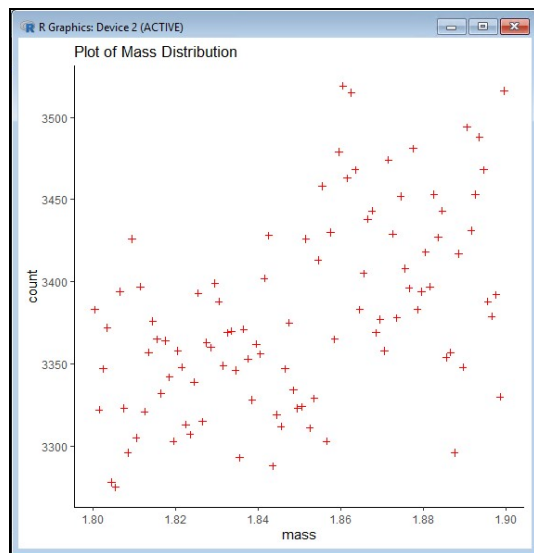
GAN_ctgan_100kentry_sorted_cut_add3





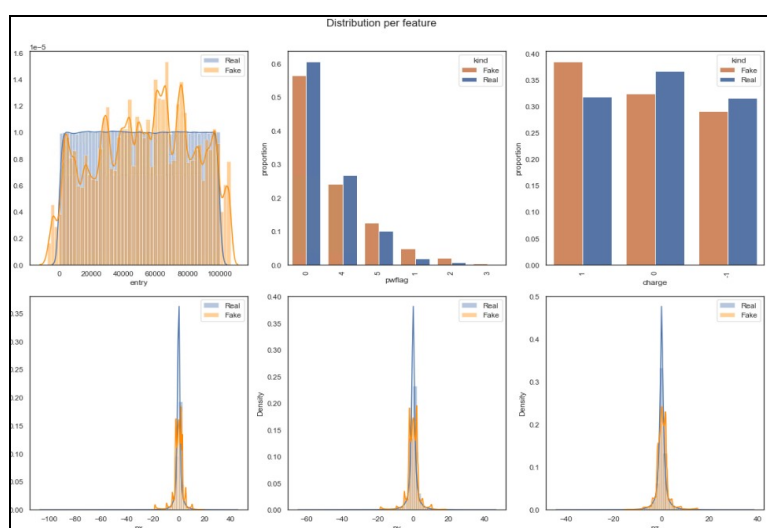
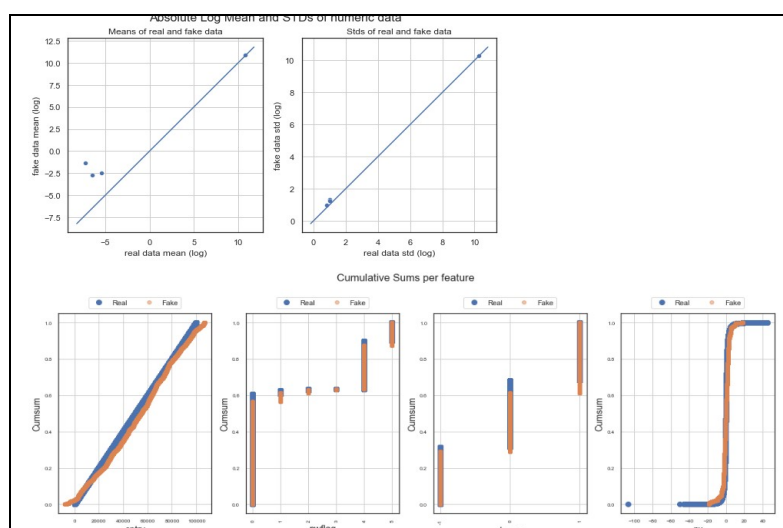
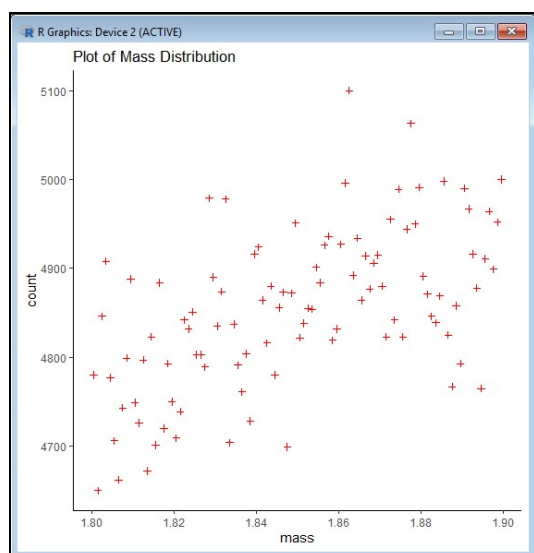
總共7839949筆資料・100001個entry

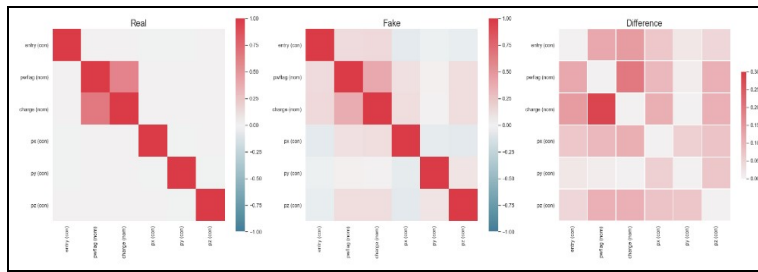
GAN_tabgan_100kentry_sorted_cut_add3



總共8217297筆資料 · 100001個entry

GAN_tabgan_detailed_100kentry_sorted_cutt_add3





結論：

可以觀察到，ctgan生成的資料，分布非常的鬆散，但是密集的地方看起來有一點點趨勢樣子，但是還是太多離群值，而且pwflag這個cate觀察real / fake比較，可以看到其實類別比例不太好。

Tabgan的模型看起來比較像是有pattern的感覺，資料看起來不會太分散，但是趨勢還是跟原版有差，只有在count 1.87左右有抓到最大值，GAN生成的1.88的數值並非最低值，因為tabgan的傾斜趨勢，讓這裡跑掉了。

在table evaluator的分析中，tabgan的生成資料觀察起來，勉強可以接受，但是在R的plot圖裡，還是有明顯的差距。

全部展開

全部收合

參考文件(0)

查閱結果(1)

分享 (0)

挑選計畫編號

回到紀錄列表

回到查閱管理

版權所有©2020 資訊工業策進會 | 限本會同仁公務使用禁止對外公開

財團法人資訊工業策進會機密資料 禁止複製、轉載、外流 III CONFIDENTIAL DOCUMENT DO NOT COPY OR DISTRIBUTE

業管部門 企推處 聯絡窗口：龔素嬌 02-6631-8614 資訊處 資訊服務櫃檯：電話：02-66318185 / [Google](#) / [Facebook](#) 系統使用說明書 | 撰寫入口