


KDD CUP 2019期末報告

e123789g



第二十三組：電機碩一 M10707601 彭柏
資工碩一 M10715046 徐子璿

Outline

1

Data Preprocessing

2

Features and Models

3

Results

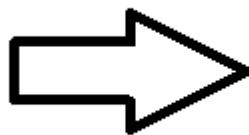
4

Difficulties

Data Preprocessing

- Req_time take **day** and **hour** as feature(以小時作為間隔)
- Transform [O,D] to [O1,O2]and [D1,D2]

o	d
116.58,40.08	116.20,39.91
116.32,39.89	116.27,39.86



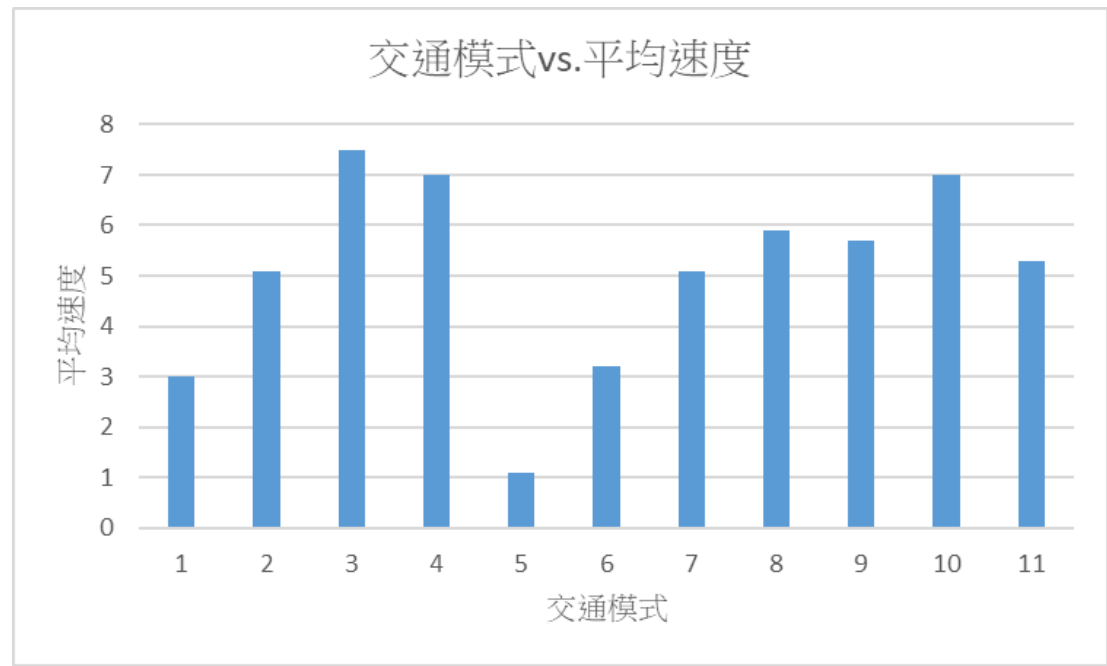
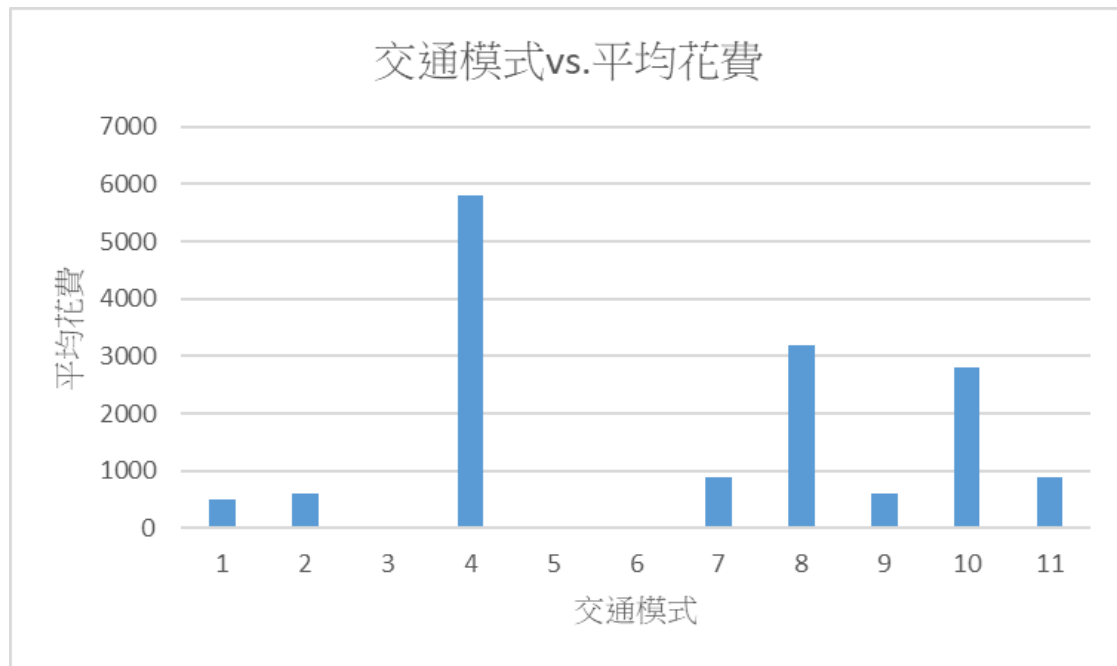
	o2	o3	o4
116.58	40.08	116.2	39.91
116.32	32.89	116.27	39.86

- 轉換經緯度成距離
- 有些沒有pid,將所有one-hot vector表示為0

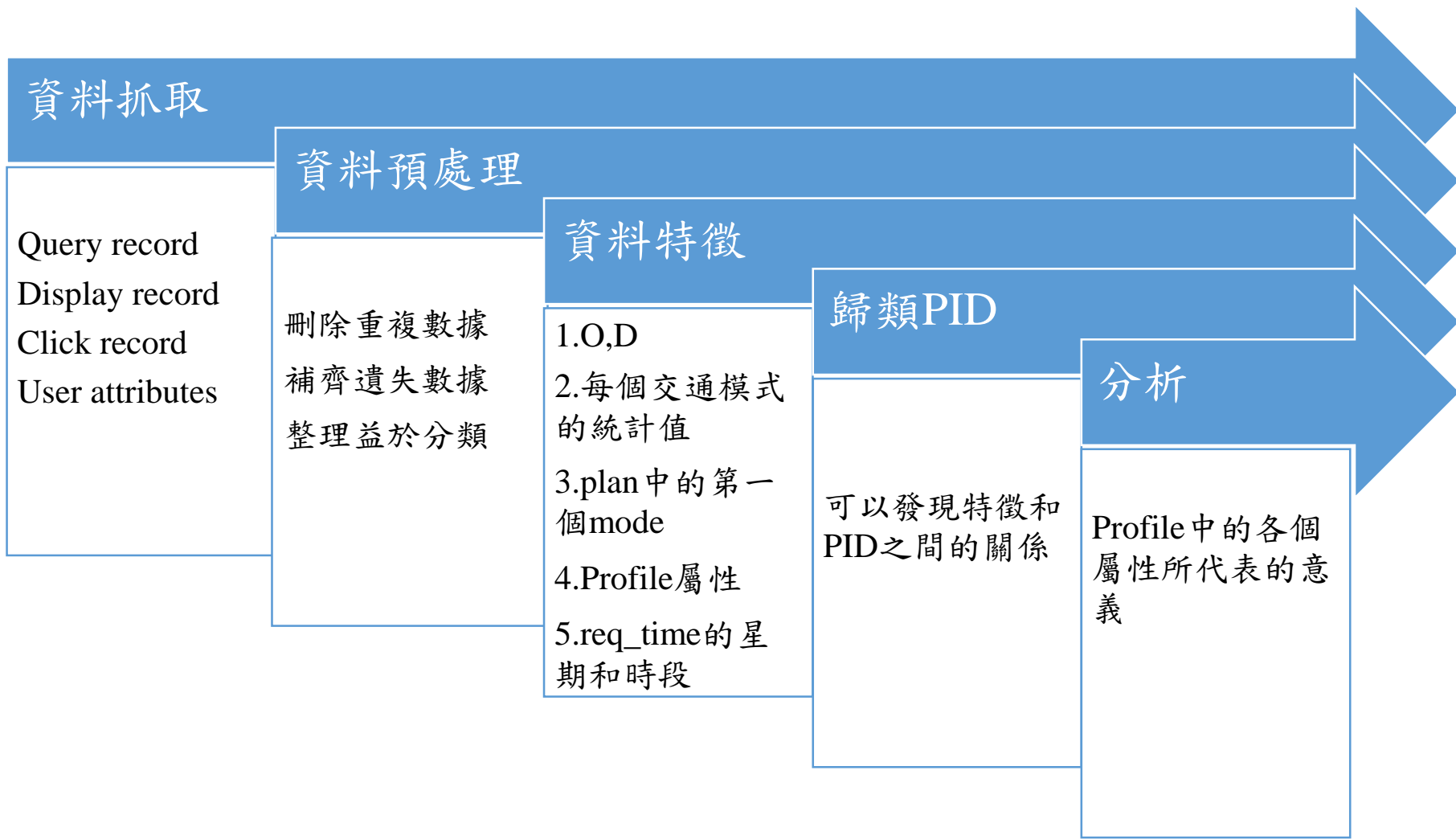
Features

- 一些SID並沒有PID，所以這些SID預測結果就給百度地圖推薦的第一個
- 各種交通方式的所需時間、價格和距離的統計值
- 推測交通模式
(mode 5可能是步行、mode 6可能是騎腳踏車、mode 3可能是自駕)
- 天氣變化不大，故沒使用天氣資料

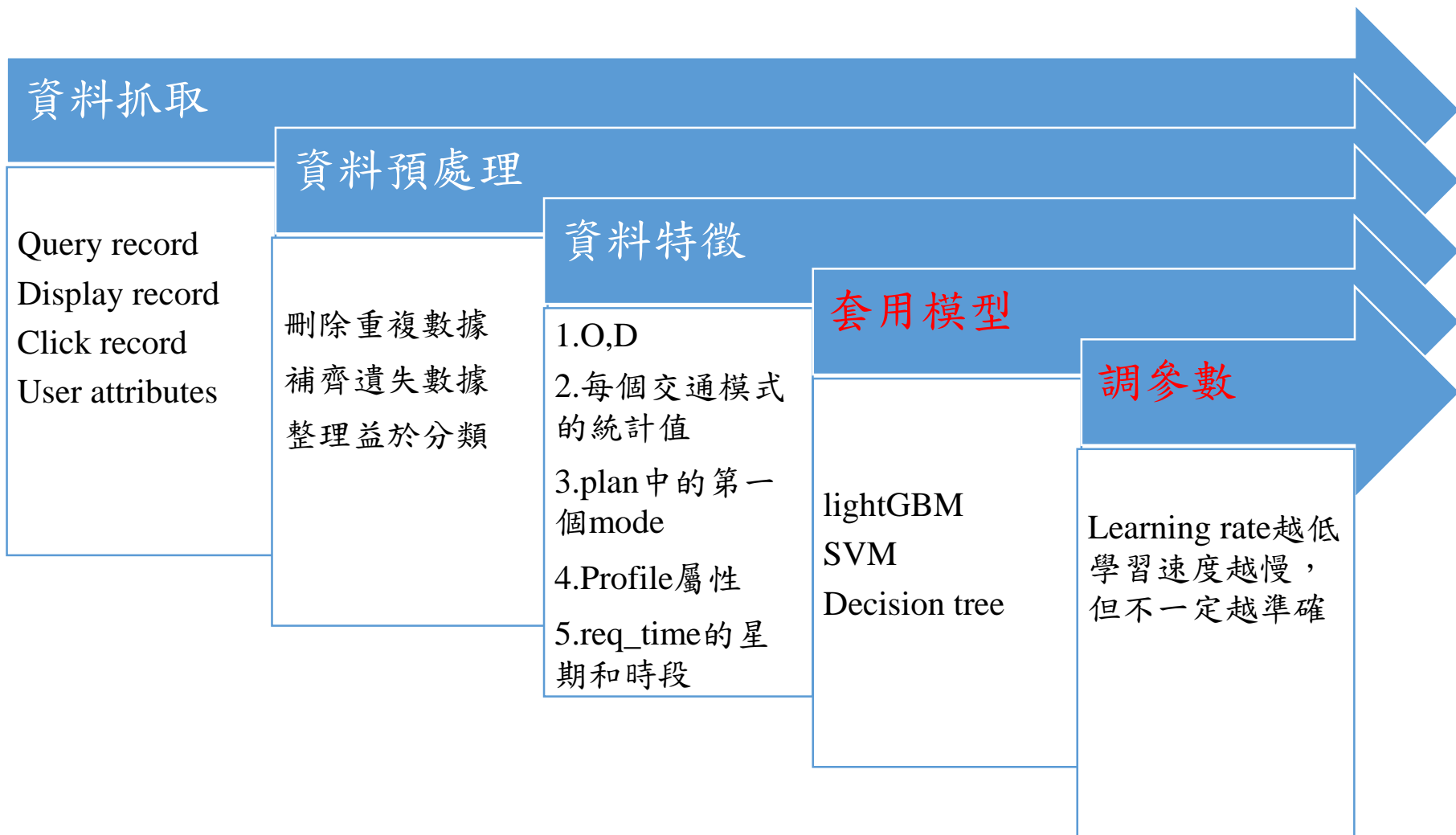
Features



Features



Features





Models

- LightGBM
 - 重要超參數設定
 - `Learning_rate = 0.05`
 - `max_depth = 8`
 - `num_leaves = 61`





Models

- SVM

`kernel='rbf'`

`class_weight='balanced'`

- Decision tree

`criterion='gini'`

`splitter='best'`



Results

LightGBM	learn rate=0.05	learn rate=0.1	learn rate=0.03
num_leaves=61	0.687065	0.686397	0.6871
num_leaves=70	0.687179		
num_leaves=90	0.687456	0.687077	0.687421
num_leaves=100	0.687413	0.686955	0.687129
num_leaves=150	0.687238		
num_leaves=200	0.68738		
SVM	decision tree		
0.685932233	0.687706		

Results

Rank	Up/Down	Participant	Score
541	14 ▼	艾萨miu	0.68864115
542	14 ▼	leifive	0.68863228
543	14 ▼	一零零8611	0.68863037
544	14 ▼	ddwxy2	0.68862542
545	14 ▼	ljs_wf	0.68853592
546	14 ▼	Ccy徽	0.68847401
547	14 ▼	e123789g	0.68842036
548	13 ▼	73926038314	0.68837237
549	13 ▼	yamadatakashi1910	0.68833500
550	13 ▼	SilenceGTX	0.68832717

困難與改進

- 其他model 的測試
- 參數的調整
- Pid資料的分析困難
- 線下結果與線上結果不如預期

分工

- 彭柏(M10707601):
 - Data preprocessing
 - Feature extraction
 - Presentation
- 徐子璿(M10715046):
 - Model training
 - Debug
 - Presentation

Thanks for listening!