

Predicting Prices of Taiwanese Commercial Real Estate

Sharlene Chen, Shih-Ya Chang,
Vivian Lee, Jenny Wei

Mentor: Jennifer Lin





Introduction



Issue and Hypothesis



當前問題

1. 較少針對“商用不動產”準確或成熟的預測模型。
2. 當前不動產研究與價格預估，多著重在不動產“自身狀況”(如：屋齡、樓層、是否有電梯、與是否有隔間等)。惟其他外在因素，如周邊機能、綠覆率等，對於房價應也有相當程度的影響。
3. 外部條件多以觀察市場趨勢納入評估，缺乏客觀量化分析，易受質疑準確性。

專題假設

不動產自身條件與外在機能對於其價格都有影響

Expected Results



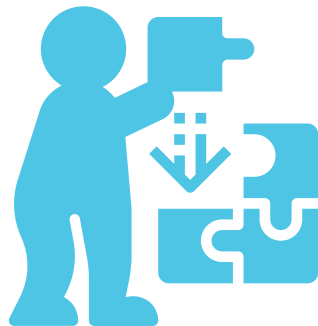
預期結果

1. 周邊環境與機能，都市空間規劃都會影響商用不動產價格。
2. 透過多種建物本身及外在條件特徵(**Features**)，建立深度學習模型(**Deep Learning Model**)，來預測商用不動產價格。



ISSUE

1. 現況較少針對「商用不動產價格」的預測模型，多偏重住宅市場分析模型。
2. 且既有模型偏重交易標自身條件進行量化分析，缺乏納入外部條件共同模擬。
3. 外部條件多以觀察市場趨勢納入評估，缺乏客觀量化分析，易受質疑準確性。



透過預測模型，協助市場預測商用房地產價格

- 以量化、大數據分析方式，綜觀了解影響房價的因素。
- 透過模型預測可更精確、客觀了解商用不動產行情，減少資訊不對等及減少估價誤差。



Dataset & Data Preprocessing

Dataset

分為數值型資料(Numerical data)

與 圖像資料(Image data)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	鄉鎮市區	交易標的:土地位置	土地移轉	都市土地	非都市土地	非都市土地	交易年月	交易筆棟	移轉層次	總樓層數	建物型態	主要用途	主要建材	建築完成	建物移轉	建物現況		
2	The village	transaction land secto	land shift	the use	zo the non-m	non-metre	transaction	transaction	shifting	le total floor	building st	main use	main buil	constructi	building sl	Building p	bu	
3	文山區	房地(土地)	臺北市文	26.97	住				1090406	土地1建	四層	十五層	住宅大樓	住家用	鋼筋混凝	1020103	229.91	3
4	文山區	房地(土地)	臺北市文	10.32	住				1090317	土地1建	三層	六層	華廈(10層	住家用	鋼筋混凝	850503	37.31	2
5	文山區	房地(土地)	臺北市文	39.64	都市: 其他:交通用地-公共				1090418	土地2建	六層	十九層	住宅大樓	住家用	鋼筋混凝	1000214	131.15	3
6	中正區	房地(土地)	臺北市中	4.26	商				1090413	土地2建	九層	十四層	住宅大樓	住家用	見其他登	970513	38.17	1
7	文山區	房地(土地)	臺北市文	44.57	住				1090328	土地3建	四層	六層	華廈(10層	住家用	鋼筋混凝	791127	123.66	3
8	文山區	房地(土地)	臺北市文	9.19	都市: 其他:市場用地(公共				1090428	土地1建	三層	七層	套房(1房	見使用彰	鋼筋混凝	770212	62.86	1
9	文山區	房地(土地)	臺北市文	30.65	住				1090417	土地2建	五層	五層	公寓(5樓	住家用	鋼筋混凝	711016	124.25	3
10	文山區	房地(土地)	臺北市文	9.84	住				1090417	土地1建	五層	五層	公寓(5樓	住家用	鋼筋混凝	690201	30.16	5
11	文山區	房地(土地)	臺北市文	15.86	住				1090419	土地2建	十二層	十二層	住宅大樓	住家用	鋼筋混凝	981202	106.36	2
12	大同區	房地(土地)	臺北市大	12.47	商				1090411	土地1建	四層	十三層	住宅大樓	住家用	鋼筋混凝	1020826	146.47	2
13	大同區	房地(土地)	臺北市大	29.25	住				1090130	土地1建	陽台, 二	四層	公寓(5樓	住家用	鋼筋混凝	600603	90.69	2
14	大同區	房地(土地)	臺北市大	27.33	住				1081214	土地1建	三層	三層	公寓(5樓	含以下無	加強磚造	570328	72.96	4
15	萬華區	房地(土地)	臺北市萬	27.5	住				1090422	土地1建	二層	四層	公寓(5樓	住家用	鋼筋混凝	620528	66.64	3
16	萬華區	房地(土地)	臺北市萬	12.12	住				1090422	土地1建	三層	四層	公寓(5樓	住家用	鋼筋混凝	640428	38.04	3
17	大同區	房地(土地)	臺北市大	20.72	都市: 商				1090326	土地2建	一層, 騎	十三層	店面(店面	見其他登	鋼筋混凝	960215	144.06	1
18	中正區	房地(土地)	臺北市中	23.53	住				1090415	土地2建	九層	十三層	住宅大樓	住家用	鋼筋混凝	1060302	153.32	2



Dataset

分為數值型資料(Numerical data) 與 圖像資料(Image data)

數值型資料 (一)

- 台北市與新北市，2017 第三季至2021第二季不動產交易資料。(近三年)
- 裡面包含“鄉鎮市區”、“地址”、“建物材料”、“樓層數”、“是否有電梯”、“建物格局”、“總價”、“建物移轉尺寸”等不同資訊。

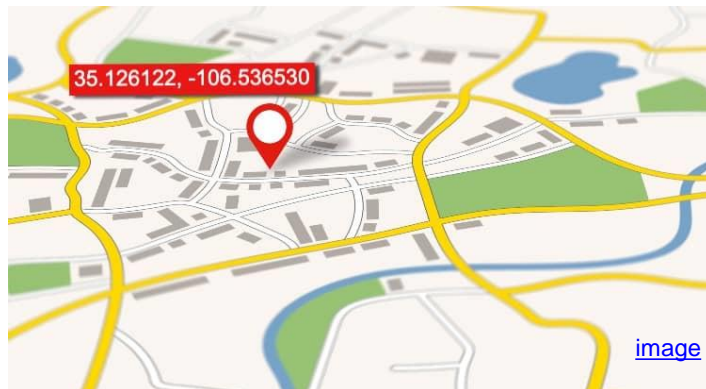
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	鄉鎮市區	交易標的	土地位置建	土地移轉總	都市土地使	非都市土地	非都市土地	交易年月日	交易筆數	移轉層次	總樓層數	建物型態	主要用途	主要建材	建築完成年	建物移轉總	建物現況格	建物現況格	建物現況格	建物現況格	有無管理組	總
2	The villages	transaction si	land sector	ps land shifting	the use zonin	the non-metr	non-metropo	transaction ye	transaction ps	shifting level	total floor nu	building state	main use	main building	construction t	building shift	building pres	building pres	building pres	building pres	Whether then tot	
3	板橋區	房地(土地+)	新北市板橋	31.25	住			1070614	土地1建物1	三層	四層	公寓(S樓含)	住家用	加強磚造		79.04	3	2	1	有	無	2
4	土城區	土地	柑林段101D-	998.91		特定專用區	農牧用地	1070509	土地1建物0	車位0		其他				0	0	0	0	有	無	2
5	土城區	土地	安和段87地1	815.29	其他			1070730	土地1建物0	車位0		其他				0	0	0	0	有	無	2
6	土城區	土地	安和段90地1	14.2	其他			1070730	土地3建物0	車位0		其他				0	0	0	0	有	無	2
7	土城區	土地	安和段100-1	14.2	其他			1070730	土地3建物0	車位0		其他				0	0	0	0	有	無	2
8	板橋區	土地	介壽段913地	16	其他			1070725	土地1建物0	車位0		其他				0	0	0	0	有	無	2
9	板橋區	房地(土地+)	新北市板橋	24.61	住			1070722	土地1建物1	二層	十五層	住宅大樓(11	住家用	鋼筋混凝土	990115	198.74	3	2	2	有	有	2
10	板橋區	房地(土地+)	新北市板橋	14.12	住			1070701	土地1建物1	五層	十五層	住宅大樓(11	見其他登記	鋼筋混凝土	1070209	105.26	2	2	1	有	無	1
11	板橋區	房地(土地+)	新北市板橋	24.68	住			1070720	土地1建物1	一層	五層	公寓(S樓含)	住家用	鋼筋混凝土造		98.49	3	1	1	有	無	1
12	板橋區	房地(土地+)	新北市板橋	18.6	住			1070702	土地1建物1	四層、陽台	五層	公寓(S樓含)	住家用	鋼筋混凝土造		72.19	3	2	1	有	無	1
13	土城區	房地(土地+)	新北市土城	27.07	住			1070703	土地1建物1	四層、陽台	五層	公寓(S樓含)	住家用	鋼筋混凝土造		86.4	3	2	1	有	無	1
14	土城區	房地(土地+)	新北市土城	32.34	住			1070614	土地1建物1	四層、陽台	四層	公寓(S樓含)	住家用	鋼筋混凝土造		87.04	5	3	2	有	無	1
15	土城區	房地(土地+)	新北市土城	9.36	住			1070705	土地1建物1	四層、陽台	十九層	套房(1房)	住家用	鋼筋混凝土	1020503	59.19	1	1	1	有	有	1
16	板橋區	房地(土地+)	新北市板橋	11.08	住			1070712	土地1建物1	五層	十三層	住宅大樓(11	住家用	鋼筋混凝土	871120	76.83	2	1	1	有	有	1
17	土城區	房地(土地+)	新北市土城	34.18	住			1070717	土地1建物1	二層	四層	公寓(S樓含)	住家用	鋼筋混凝土	650724	97.22	3	2	4	有	無	1
18	板橋區	房地(土地+)	新北市板橋	19.25	住			1070720	土地1建物1	四層	四層	公寓(S樓含)	住家用	鋼筋混凝土		50.04	5	0	5	有	無	1
19	土城區	房地(土地+)	新北市土城	24.06	住			1070701	土地1建物1	三層	五層	公寓(S樓含)	住家用	鋼筋混凝土	701221	81.74	3	2	1	有	無	1
20	板橋區	房地(土地+)	新北市板橋	29.63	住			1070517	土地1建物1	十六層	二十三層	住宅大樓(11	見其他登記	鋼筋混凝土	1070104	206.1	3	2	2	有	有	1
21	板橋區	房地(土地+)	新北市板橋	25.02	住			1070523	土地1建物1	五層	二十一層	住宅大樓(11	見其他登記	鋼筋混凝土	1070104	149.3	3	2	2	有	有	1
22	板橋區	房地(土地+)	新北市板橋	13.17	住			1070523	土地1建物1	五層	二十一層	住宅大樓(11	見其他登記	鋼筋混凝土	1070104	80.6	2	2	1	有	有	1
23	板橋區	房地(土地+)	新北市板橋	26.34	住			1070520	土地1建物1	十三層	二十一層	住宅大樓(11	見其他登記	鋼筋混凝土	1070104	184.04	3	2	2	有	有	1
24	板橋區	房地(土地+)	新北市板橋	24.36	住			1070523	土地1建物1	一層	二十三層	住宅大樓(11	見其他登記	鋼筋混凝土	1070114	147.62	3	2	2	有	有	1
25	板橋區	房地(土地+)	新北市板橋	25.02	住			1070523	土地1建物1	一層	二十三層	住宅大樓(11	見其他登記	鋼筋混凝土	1070114	179.26	3	2	2	有	有	1
26	板橋區	房地(土地+)	新北市板橋	24.36	住			1070519	土地1建物1	十二層	二十四層	住宅大樓(11	見其他登記	鋼筋混凝土	1070104	173.5	3	2	2	有	有	1
27	板橋區	房地(土地+)	新北市板橋	13.83	住			1070523	土地1建物1	四層	二十四層	住宅大樓(11	見其他登記	鋼筋混凝土	1070104	110.48	2	2	1	有	有	1

Dataset

分為數值型資料(Numerical data) 與 圖像資料(Image data)

數值型資料 (二)

- 透過資料庫中的文字，運用Heremap API找到相對應的經緯度，使用Google Neaby Search 以及周圍機能特徵，如：是否有學校、餐廳、醫療院所、交通轉運站及其相對應數量等



Dataset

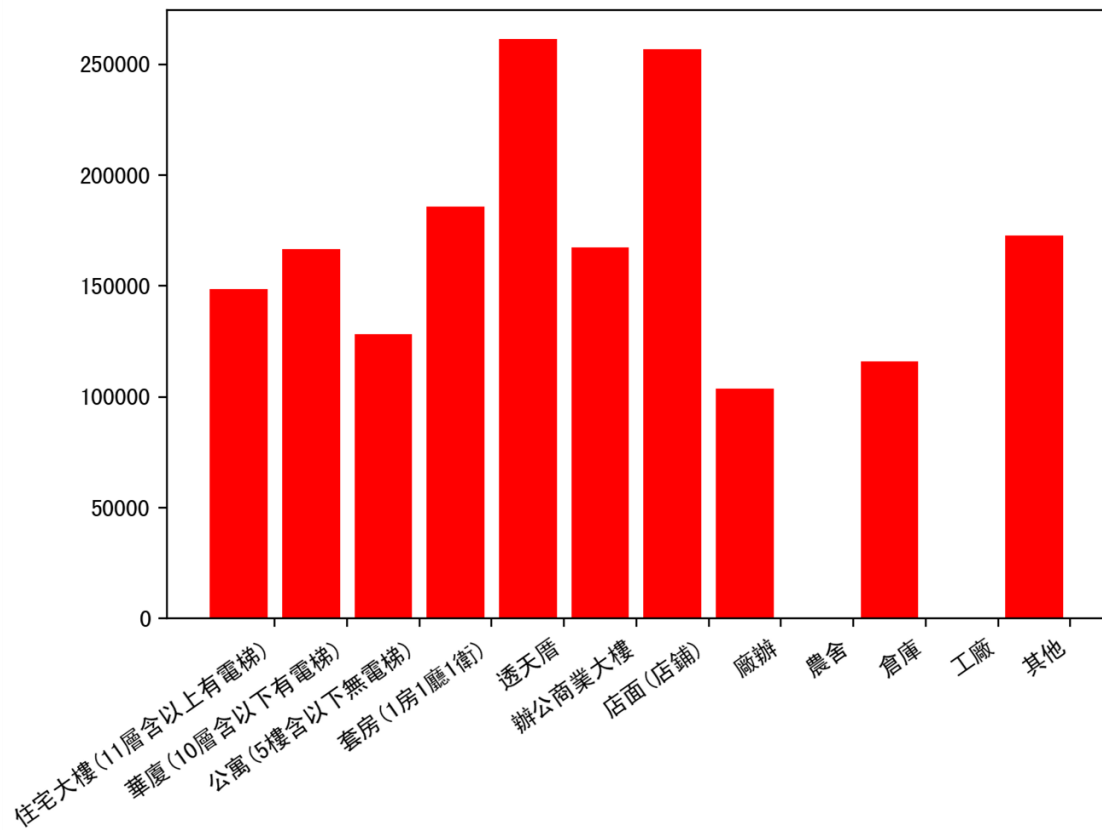
分為數值型資料(Numerical data) 與 圖像資料(Image data)

資料處理

- 將文字轉為數值、剔除有缺失欄位的data資料，並進行data preprocessing:
 1. Convert string to numerical value(int/float/double..)
 2. Deal with missing data
 3. One-Hot Encoding
 4. Standardization
- 根據特徵(Feature)相關性，篩選項目

lat	lng	zoning	total_floors	floors_area	room	hall	bathroom	compartmented	management_committee	unit_price	unit_berth_price	total_berth_price	main_building_area	auxiliary_building_area	balcony_area
25.02398	121.42394	14.26	12	191.73	0	0	2	1	1	65978	32.18	0	116.73	0.0	0.0
25.06209	121.43308	22.21	4	77.14	3	2	1	1	1	54446	0.0	0	72.35	0.0	4.79
25.07846	121.38825	22.9	4	76.25	2	1	1	1	0	43279	0.0	0	76.25	0.0	0.0
25.18262	121.44844	7.74	27	102.52	2	1	1	1	1	69920	15.42	650000	51.8	2.8	4.89
25.18657	121.43498	10.89	26	152.48	3	2	2	1	1	63959	30.84	1300000	70.14	5.23	7.71
25.187	121.43629	8.71	27	127.94	2	1	1	1	1	68589	30.84	920000	56.75	3.47	6.13
25.08044	121.49077	6.22	15	61.87	2	1	1	1	1	126071	0.0	0	40.29	0.0	3.01
25.01322	121.51863	21.78	29	257.27	0	0	0	0	0	204913	31.32	2300000	130.59	9.5	8.89
24.94649	121.3815	14.56	19	170.04	3	2	2	1	1	81963	34.89	0	83.76	0.0	12.94
24.93415	121.37126	19.29	12	198.05	0	0	0	0	0	77556	33.13	0	110.28	2.72	3.84

- 不同建物型態的平均單價



Dataset

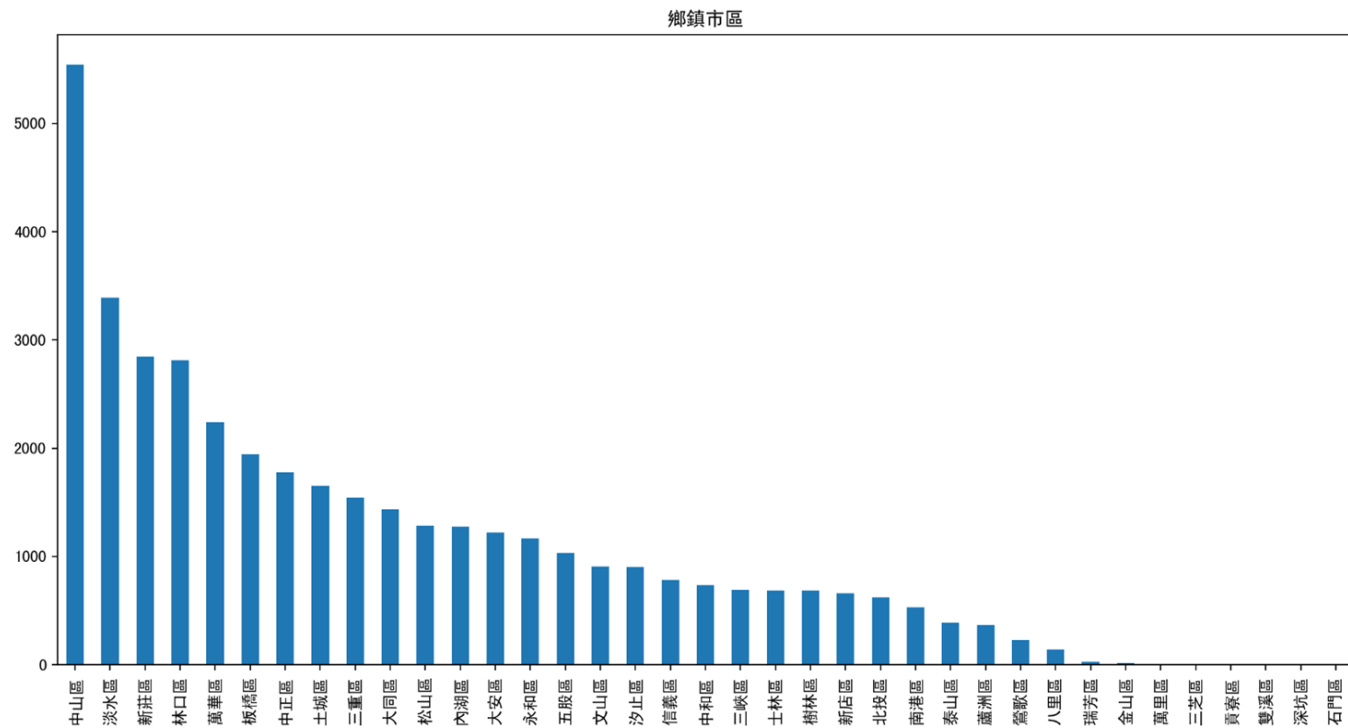
分為數值型資料(Numerical data) 與 圖像資料(Image data)

資料處理

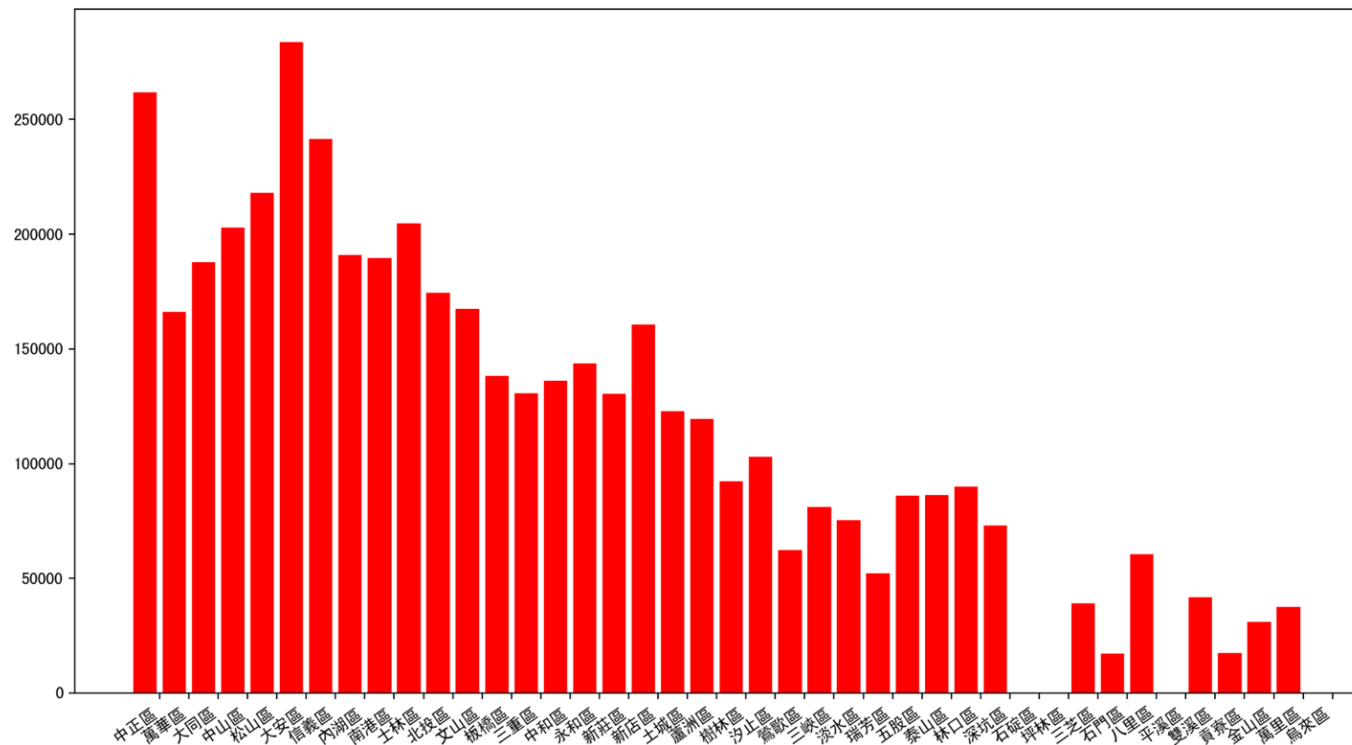
- 將文字轉為數值、剔除有缺失欄位的data資料，並進行data preprocessing:
 1. Convert string to numerical value(int/float/double..)
 2. Deal with missing data
 3. One-Hot Encoding
 4. Standardization
- 根據特徵(Feature)相關性，篩選項目

lat	lng	zoning	total_floors	floors_area	room	hall	bathroom	compartmented	management_committee	unit_price	unit_berth_price	total_berth_price	main_building_area	auxiliary_building_area	balcony_area
25.02398	121.42394	14.26	12	191.73	0	0	2	1	1	65978	32.18	0	116.73	0.0	0.0
25.06209	121.43308	22.21	4	77.14	3	2	1	1	0	54446	0.0	0	72.35	0.0	4.79
25.07846	121.38825	22.9	4	76.25	2	1	1	1	0	43279	0.0	0	76.25	0.0	0.0
25.18262	121.44844	7.74	27	102.52	2	1	1	1	1	69920	15.42	650000	51.8	2.8	4.89
25.18657	121.43498	10.89	26	152.48	3	2	2	1	1	63959	30.84	1300000	70.14	5.23	7.71
25.187	121.43629	8.71	27	127.94	2	1	1	1	1	68589	30.84	920000	56.75	3.47	6.13
25.08044	121.49077	6.22	15	61.87	2	1	1	1	1	126071	0.0	0	40.29	0.0	3.01
25.01322	121.51863	21.78	29	257.27	0	0	0	0	1	204913	31.32	2300000	130.59	9.5	8.89
24.94649	121.3815	14.56	19	170.04	3	2	2	1	1	81863	34.89	0	83.76	0.0	12.94
24.93415	121.37126	19.29	12	198.05	0	0	0	0	1	77556	33.13	0	110.28	2.72	3.84

● 各區交易筆數



● 各區平均單價



Dataset

分為數值型資料(Numerical data) 與 圖像資料(Image data)

圖像資料

透過Heremap API找到的交易標的之經緯度，運用mapbox找到對應之地圖

The image shows a web interface for configuring a Mapbox map. On the left, there are input fields for map settings:

- style**: Mapbox Streets
- ☐ Use my own public style
- width**: 256 in pixels
- height**: 256 in pixels
- center**: longitude 121.5436, latitude 25.0332
- zoom**: 15
- Optional parameters**: bearing 0, pitch 0
- overlays**: + add overlay
- ☒ @2x render map at 2x scale

On the right, the **Request URL** is displayed:

```
https://api.mapbox.com/styles/v1/mapbox/streets-v11/static/121.5436,25.0332,15,0/256x256?access_token=pk.eyJ1Ijoic2hhcmx1bmUwMTEzIiw1YSI6ImNrcTNhYzQwbTBoa2oydm8wb2oxdWZlZkxkfQ.n8hExcAc4aVh5wk33Idfg
```

Below the URL is a map view of a street grid in Taipei, Taiwan, with various landmarks and street names visible. The map includes a compass and zoom controls.

Dataset

分為數值型資料(Numerical data) 與 圖像資料(Image data)

共18項特徵、1項用來檢測預測準確率的標籤(房價)

建物
條件
(16)

Town(行政區)
transaction_type(交易標的)
total_floors(總樓層)
building_state(建物型態)
building_materials(建材)
Room(房間數)
hall(廳數)
bathroom(衛數)

total_berth_price(車位總價)
main_building_area(主建物面積)
auxiliary_building_area(附屬建物面積)
balcony_area(陽台面積)
Elevator(有無電梯)
building_age(屋齡)
transaction_first_floor(是否移轉一樓)
shifting_level(移轉樓層類型)

外部
條件
(1)

point_of_interest
(包含以下 features) :

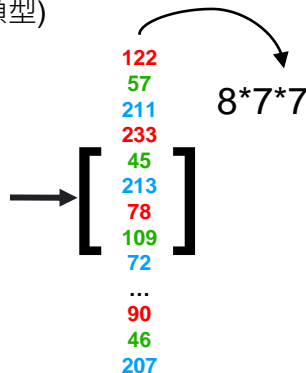
travel agency
establishment
clothing store
home goods store
store
local government office

natural feature
health
tourist attraction
transit station
food
university

圖像
張量
(1)



Input



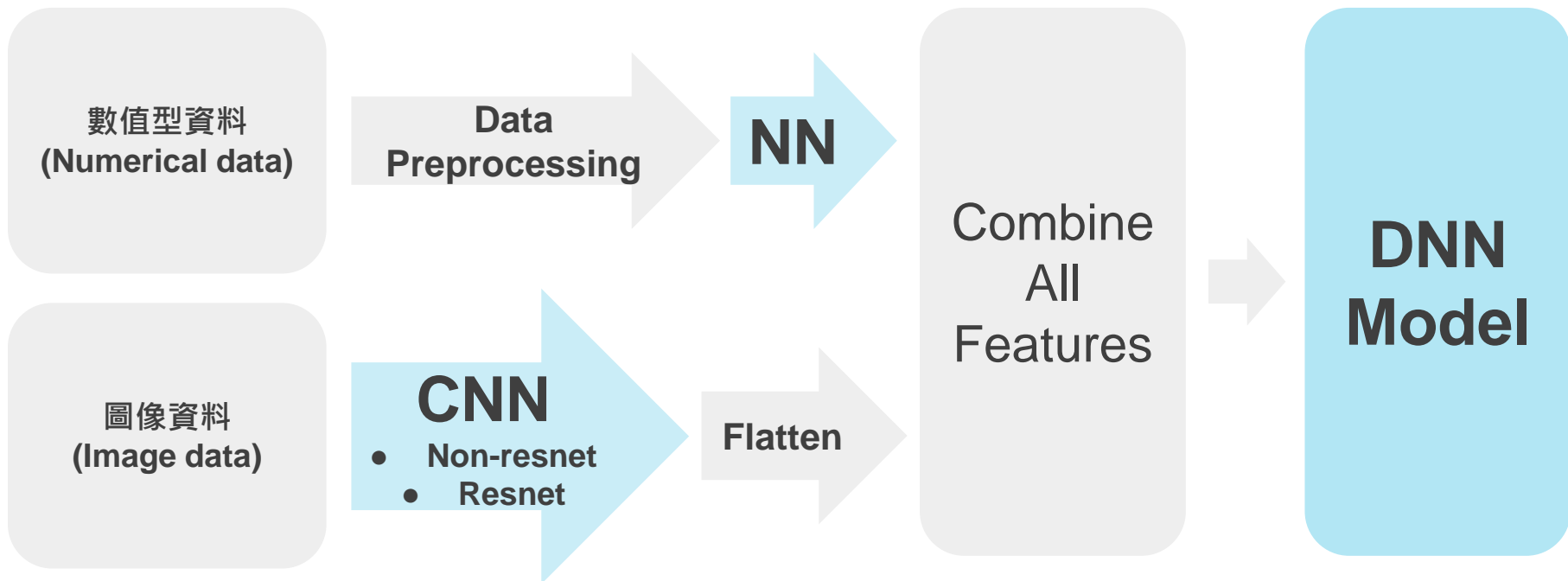
標籤

unit_price
(元/平方公尺)



Dataset

Algorithm and Training

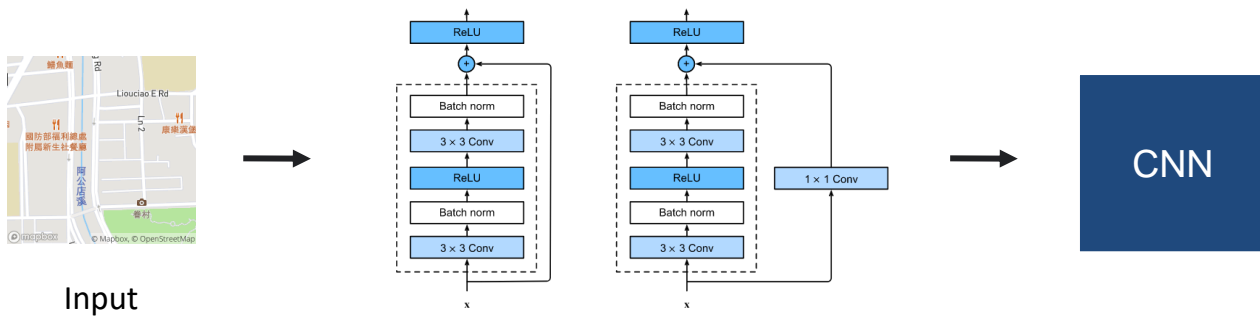


Dataset

分為數值型資料(Numerical data) 與 圖像資料(Image data)

圖像資料處理 - model架構

1. 將每個圖像經由先經由Resnet 50 pretrained 前6層
2. 在進入自己寫的CNN reduce dimension
3. 最後flatten成tensor(張量)，並將tensor結合到其他data資料中

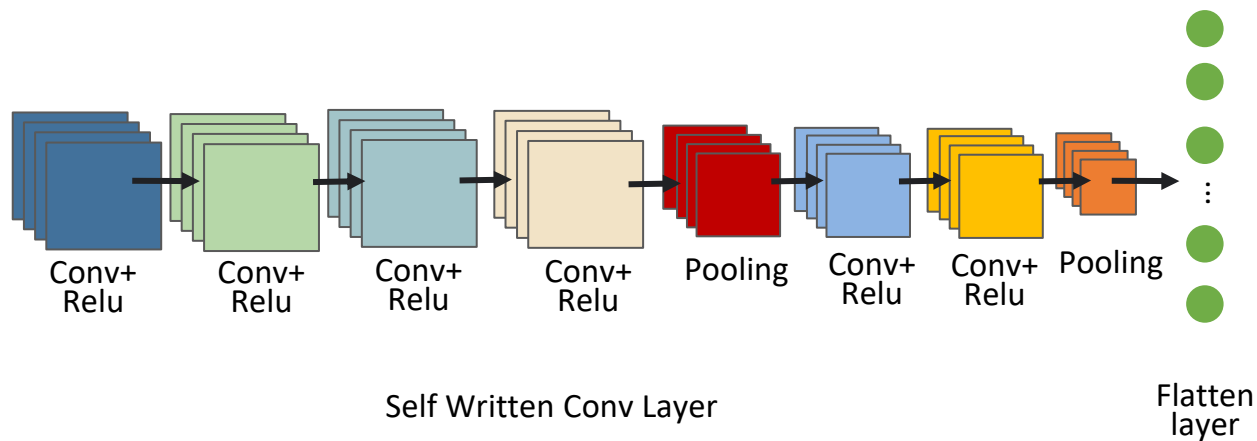


Resnet50 First 6 Layers

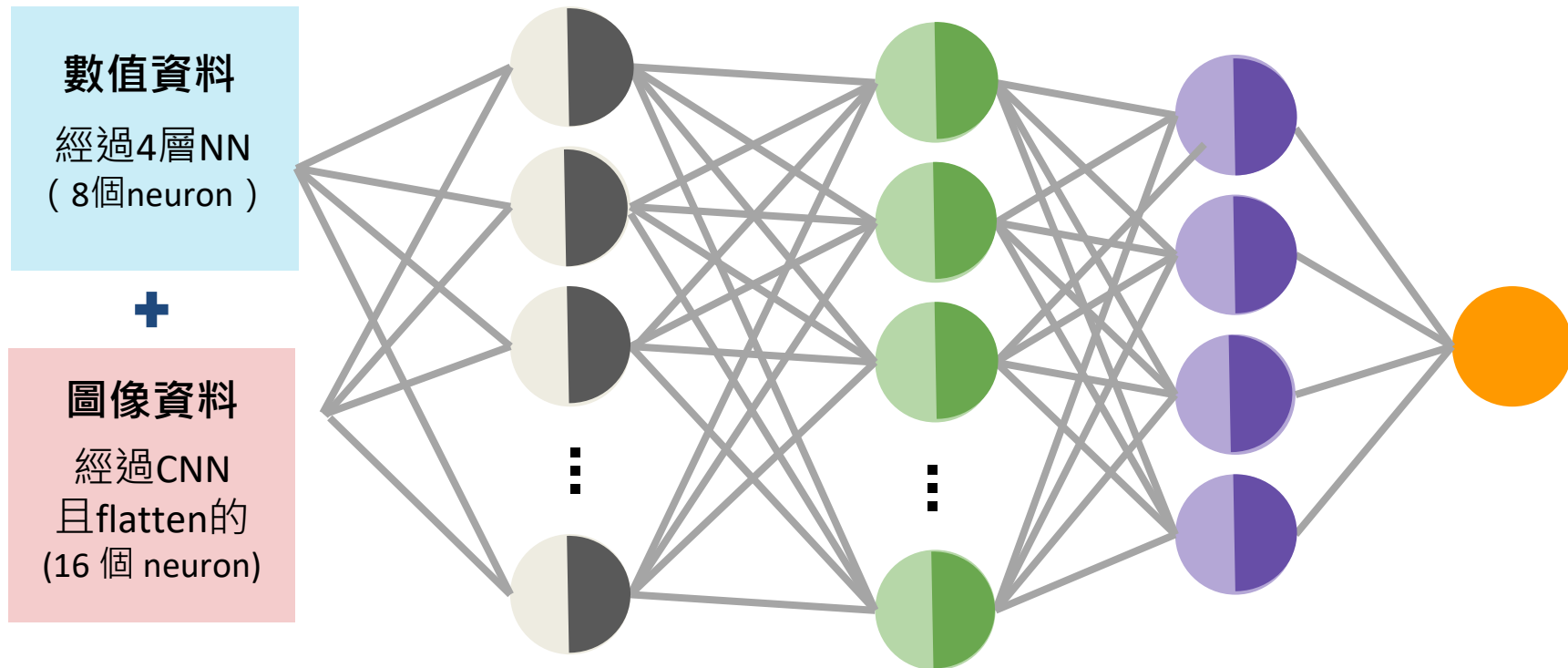
Dataset

分為數值型資料(Numerical data) 與 圖像資料(Image data)

圖像資料處理 - CNN架構



Model-DNN





Testing and Result



Testing and Recommendation System

模型訓練結果

數值型資料(Numerical data)

- L1 Loss(誤差絕對值) : 13,000元/平方公尺 (~ 39,000元/坪)
10% (每筆交易平均值約為150,000元/平方公尺)
- 約為近一年大安捷運站半徑500公尺內出現的價差範圍

Concatenated Data : 圖像資料(Image data)+數值型資料(Numerical data)

- L1 Loss(誤差絕對值) : 26,000 元/平方公尺 (~ 78,000元/坪)
- 圖像資料有助於模型的運算

Testing and Recommendation System

模型訓練結果 (圖像資料有助於模型的運算)

- Numerical Data

```
Iteration 300, loss = 0.1921
Checking accuracy on validation set
Epoch 3, Eval loss: 0.48551075048529346, l1 loss: 36617.55859375

Iteration 100, loss = 0.0314
Checking accuracy on validation set
Epoch 26, Eval loss: 0.2300190421714948, l1 loss: 16378.8134765625

Iteration 200, loss = 0.0636
Checking accuracy on validation set
Epoch 26, Eval loss: 0.23153304476613942, l1 loss: 16517.373046875

Iteration 300, loss = 0.0957
Checking accuracy on validation set
Epoch 26, Eval loss: 0.21603125135678086, l1 loss: 15540.1396484375

Iteration 400, loss = 0.1251
Checking accuracy on validation set
Epoch 26, Eval loss: 0.18043599643803745, l1 loss: 12850.1201171875
```

- Image + Numerical Data

```
Iteration 100, loss = 0.0473
Checking accuracy on validation set
Epoch 3, Eval loss: 0.3694531695346612, l1 loss: 28385.349609375

Iteration 200, loss = 0.0944
Checking accuracy on validation set
Epoch 3, Eval loss: 0.37538337104582375, l1 loss: 28387.1328125

Iteration 300, loss = 0.1453
Checking accuracy on validation set
Epoch 3, Eval loss: 0.3758517073069005, l1 loss: 28781.69921875

Iteration 400, loss = 0.1931
Checking accuracy on validation set
Epoch 3, Eval loss: 0.39483148066294677, l1 loss: 30359.599609375

Iteration 500, loss = 0.2377
Checking accuracy on validation set
Epoch 3, Eval loss: 0.35445074504510515, l1 loss: 26676.701171875

Iteration 600, loss = 0.2852
```




Numerical Data

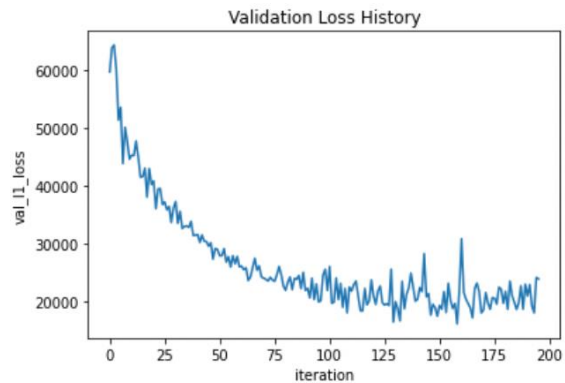
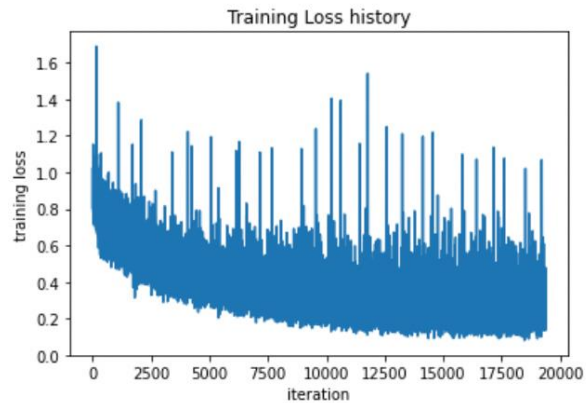
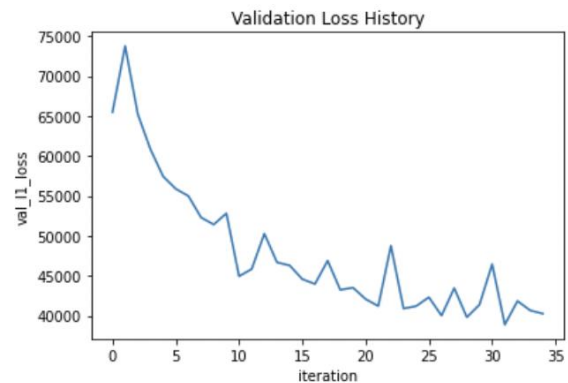
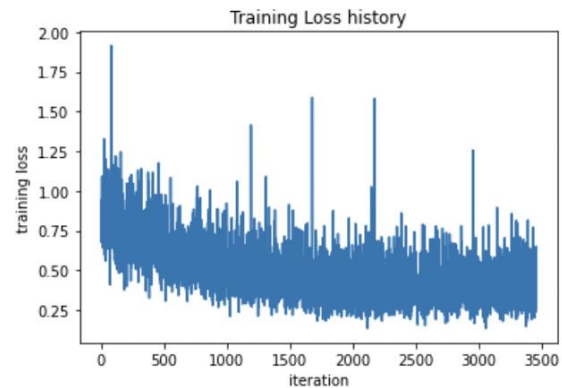


Image + Numerical



Dicussion and Future Work



Discussion and Conclusion

01

How to decrease loss?

數值型資料(Numerical data)Numeric data

- 觀察特徵(Feature)相關性，篩選要進入訓練模型的項目。
- 數據型資料較圖像單純，升高維度容易降低每項特徵的資訊量，故調降初始訓練維度(78-->128修正為78-->60)，再逐層降低收斂。
- 考量要與圖像資料共同訓練，故收斂至8項Neuron(Output Layer)
(原本是收斂至1項Neuron，但會太過簡化數值型資料而無法與複雜的圖像資料結合一起訓練)

圖像資料(Image data):

- 使用訓練好的模型(Pretrained)：Resnet。
原本使用自己的訓練模型會花太多時間
考量本次使用圖像資料類型，故只使用Resnet前六層，以訓練輪廓特徵為主。
- 再使用CNN降維(512-->16)，與數值型資料接軌。



Discussion and Conclusion

02

How we tune hyperparameters?

- Optimizer: Adam
- Decrease learning rate ($=1e-5$)
- Add weight decay ($=0.99$)
- Add epochs ($=28$)



Future Work

研究範圍之房價差異大

台北市、新北市範圍內原始交易資料房價差距大，故未來可縮小研究範圍，提高資料間的相關性。

訓練不同類型的圖像

可使用不同類型、主題的地圖，如衛星影像，提高圖像複雜度，用更多的資訊練去訓練模型。



A background image showing a business handshake over a desk with financial charts and a laptop. The handshake is the central focus, with one hand wearing a watch. Below the hands are several documents featuring bar and pie charts. To the right, a laptop keyboard is partially visible. The scene is brightly lit, suggesting an office environment.

WISH YOU A GOOD DEAL!