**中文摘要**

近年來有許多統計方法被應用於房價的預測，諸如線性回歸、隨機森林等機器學習技術，在眾多技術中又以深度學習的表現最為突出，有許多研究都使用類神經網路進行房價的預測，並且在準確率上具有良好的表現。然而，大多數研究的訓練資料都只關注於交易標的本身的特徵值，使用的特徵包含房屋條件如屋齡、衛浴數、房間數等變數作為模型的輸入，忽略了交易標的於地理空間中的位置以及鄰近空間的環境特徵，然而真實世界並不是完全均質的空間，某些現象於空間中存在差異現象，尤其房地產交易是屬於在空間中變化劇烈的情境，進而降低模型於非均質研究區內的泛化能力。有鑑於此，本研究試圖結合卷積類神經網路與多層感知器，提出一個能夠將標的物周圍特徵值作為模型輸入的類神經網路架構，期望能透過提供模型相關的空間特徵以提高模型面對非均質區域的泛化能力。並且同時建立兩個對照組，前者由地理加權迴歸構成，後者則由與本研究所提出之模型相同結構的多層感知器組成，並將三組模型進行基準比較，以此驗證卷積類神經網路為模型提供空間相關的特徵有助於模型準確率。本研究將選擇桃園市桃園區作為研究區，研究區間為2015年1月至2018年12月，使用內政部所提供之實價登錄資料作為房屋特徵資料來源，並且選用哨兵二號衛星影像、桃園市國土利用調查、公車與火車站點距離作為環境特徵資料來源。

研究結果顯示，以深度學習為基礎的多層感知器模型，其調整後R平方高於地理加權回歸約0.103，與多層感知器相比，其調整後R平方提升了約0.102，意味著額外建立環境特徵提取管道的確能提升模型解釋力，並且由於類神經網路的高容錯性，使這模型更適合被應用於雜異性高的情境中。

英文摘要

In recent years, many statistical methods have been applied to the prediction of housing prices, including machine learning techniques such as linear regression, random forest, etc. Among those methodologies, deep learning has achieved outstanding performance. With excellent prediction accuracy, many researchers had applied the artificial neural network to the prediction of housing prices. However, most of them use tabular data only focusing on the feature of the house transaction record itself, ignoring the fact that the house price could be influenced by the surrounding environment, which eventually brings the problem of spatial heterogeneity to the table. Therefore, this study aims to combine the convolution neural network with multilayer perception, proposing a mixed input neural network structure that also extracts the distribution of surrounding variables. Hoping to improve the generalization of the model by downsizing the influence of spatial heterogeneity. This study chooses Taoyuan City, Taoyuan district as the main research area, research period ranging from 2015/01 to 2018/12. We use Actual Price Registration data released by the Ministry of the Interior, R.O.C as the real estate feature data source. In regard to the environment feature, we use satellite images of Sentinel-2, land use inventory data provided by the MoI, R.O.C., and distances to public transportation.

The research findings show that the multi-layer perceptron model has an adjusted R-squared higher than the geographically weighted regression by approximately 0.103. Compared to the multi-layer perceptron, the adjusted R-squared of the former has improved by about 0.102. This means that establishing an additional environmental feature extraction pipeline can enhance the model’s explanatory power. Moreover, due to the high fault tolerance of the neural network, this model is more suitable for application in highly heterogeneous contexts.

**目 錄**

[第一章 緒論 7](#_Toc132799664)

[1-1 研究背景與動機 7](#_Toc132799665)

[1-2 研究目的 9](#_Toc132799666)

[1-3 研究限制 9](#_Toc132799667)

[1-4 研究範圍 10](#_Toc132799668)

[第二章 文獻回顧 10](#_Toc132799669)

[2-1 傳統估價方法 10](#_Toc132799671)

[2-1-1 比較法 10](#_Toc132799672)

[2-1-2 成本法 11](#_Toc132799673)

[2-1-3 收益法 12](#_Toc132799674)

[2-1-4 特徵價格法 13](#_Toc132799675)

[2-2 地理加權迴歸 14](#_Toc132799676)

[2-2-1 地理加權迴歸模型 14](#_Toc132799677)

[2-2-2 地理加權迴歸於不動產估價之應用 15](#_Toc132799678)

[2-3 深度學習技術 15](#_Toc132799679)

[2-3-1 類神經網路理論 16](#_Toc132799680)

[2-3-2 類神經網路應用於不動產估價 18](#_Toc132799681)

[第三章 研究方法 19](#_Toc132799682)

[3-1 研究流程 19](#_Toc132799687)

[3-2 研究資料 20](#_Toc132799688)

[3-2-1 實價登錄2.0資料前處理流程 22](#_Toc132799689)

[3-2-2 環境特徵資料前處理 25](#_Toc132799690)

[3-3 演算法 25](#_Toc132799691)

[3-3-1 最小平方法與地理加權迴歸 26](#_Toc132799692)

[3-3-2 類神經網路架構設計 26](#_Toc132799693)

[3-3-3 類神經網路訓練與基準測試 29](#_Toc132799694)

[第四章 結果與討論 32](#_Toc132799695)

[4-1 皮爾森相關係數矩陣 32](#_Toc132799697)

[4-2 線性迴歸模型結果 32](#_Toc132799698)

[4-2-1 普通最小平方法結果 32](#_Toc132799699)

[4-2-2 地理加權迴歸結果 33](#_Toc132799700)

[4-3 類神經網路訓練結果 33](#_Toc132799701)

[4-3-1 多層感知器結果 34](#_Toc132799702)

[4-3-2 混合輸入模型結果 35](#_Toc132799703)

[4-3-3 基準測試 36](#_Toc132799704)

[第五章 結論與建議 38](#_Toc132799705)

[5-1 結論 38](#_Toc132799707)

[參考文獻 40](#_Toc132799708)

**表目錄**

[表 1 實價登錄欄位資訊 21](#_Toc132799709)

[表 2 模型超參數 31](#_Toc132799710)

[表 3 訓練演算法 32](#_Toc132799711)

[表 4 普通最小平方法結果 34](#_Toc132799712)

[表 5 地理加權迴歸結果 34](#_Toc132799713)

**圖目錄**

[圖 1 研究範圍圖 10](#_Toc132799714)

[圖 2 人工神經元模型 16](#_Toc132799715)

[圖 3 類神經網路架構示意 17](#_Toc132799716)

[圖 4 卷積運算示意圖 18](#_Toc132799717)

[圖 5 研究流程圖 20](#_Toc132799718)

[圖 6 資料前處理程序一 22](#_Toc132799719)

[圖 7 資料前處理程序二 23](#_Toc132799720)

[圖 8 資料前處理程序三 23](#_Toc132799721)

[圖 9 資料前處理程序四 24](#_Toc132799722)

[圖 10 模型示意圖 27](#_Toc132799723)

[圖 11 卷積類神經網路架構圖 28](#_Toc132799724)

[圖 12 多層感知器架構圖 29](#_Toc132799725)

[圖 13 權重分享層架構圖 29](#_Toc132799726)

[圖 14 皮爾森相關係數矩陣 32](#_Toc132799727)

[圖 15 多層感知器訓練損失圖 34](#_Toc132799728)

[圖 16 多層感知器訓練調整後R平方圖 35](#_Toc132799729)

[圖 17 混合輸入模型訓練損失圖 36](#_Toc132799730)

[圖 18 混合輸入模型訓練調整後R平方圖 36](#_Toc132799731)

[圖 19 基準比較結果圖 37](#_Toc132799732)

**公式目錄**

[公式 1 15](#_Toc132799733)

[公式 2 15](#_Toc132799734)

[公式 3 27](#_Toc132799735)

# 第一章 緒論

## 研究背景與動機

臺灣房價於過去數十年來呈現不斷增長的趨勢，因傳統觀念購屋與投機炒房等行為使房市火熱，隨之而來的高房價問題使購屋需求高的年輕族群毫無能力與機會購買，而新冠疫情的衝擊更使失業與薪資停滯的情況惡化。臺灣於2011年之前沒有公開房屋交易紀錄等相關資訊，使消費者面臨交易資訊不透明的市場，大多數消費者較少有實際的購屋經驗，常因先備知識不足與資訊不對稱而蒙受不必要之損失，為使不動產交易資訊透明化，立法院於2011年12月份三讀通過「地政士法部分條文修正案」、「平均地權條例部分條文修正案」、「不動產經紀業管理條例部分條文修正案」等法案，以此作為不動產買賣資訊登記之依據，使房地產價格透明化。

除了一般購屋者外，銀行等金融機構對於不動產估值的需求也是不容忽視，金融業者審核不動產信貸時，傾向委外由民間之專業不動產估價公司進行標的估值，然而過去估價師所使用的估價方法有著些許瑕疵，如比較法、成本法與收益法等，此類方法的缺點在於估價過程中個人主觀意識過重與所涉及的專業知識龐雜，導致估價結果無法完全反映實際市場價格，雖然也有較為客觀的特徵價格法，但標的物與實際價格為非線性關係，一般的線性迴歸無法有效地建立出可用於反應市場價格的模型，所估值出的價格無法完全使人信服。

近年來機器學習與深度學習等技術蓬勃發展，由於此類模型能較好的建立出非線性映射模式，因此被大量用於房地產價格預測的相關研究中，然而，大多數研究所使用到的訓練資料僅關注於交易標的本身的特徵值，模型的輸入特徵包含交易標的的屋齡、格局、建材與資產類型等屬性資料，忽略了交易標的於地理空間中的座標與鄰近空間的特徵，房地產又屬於區位影響占比重的資產類型，時常因為所位於的區域而對價格產生巨大影響，換言之，若模型忽略空間等相關資訊，會導致模型面臨空間非穩態問題，進而降低模型於非均質研究區內的泛化能力。此外，聯合國世界衛生組織（World Health Organization, WHO）於1961年的住房公共健康專家會議報告中就指出人類居住環境需要基本的生理基礎，其中包含了許多環境指標，如空氣品質、環境噪音、公園綠地等因子，這些都是會影響人類居住上生心理健康的因素，因此消費者於購房時，除了標的物本身的特徵外，標的物周圍的環境特徵也成為人們越發在意的條件，如何將環境特徵作為模型預測的參考來源，已成為不可或缺的一環。

有鑑於此，本研究試圖結合卷積類神經網路與多層感知器，由前者作為標的物周遭空間特徵的輸入管道，為模型提供相關的空間特徵，並由後者作為標的物本身的特徵值輸入管道，為模型提供基本的特徵來源，期望透過此種方法提高模型面對非均質區域的泛化能力，從而提升標的物估值準確率。

## 研究目的

綜觀過去將深度學習技術應用於不動產價格預測之研究無法考量到標的物周圍的環境變數，本研究之目的在於提出一個可將環境特徵納入模型輸入，為類神經網路添增輸入特徵之方法，並且與常見之模型方法進行基準測試。本研究欲達成之目的有以下三項：

1. 藉由內政部所提供之實價登錄2.0資料進行資料前處理，包含資料解析與邏輯檢和，並從中擷取出交易紀錄之屬性與空間資料，以此作為多層感知器的資料來源。
2. 擷取研究區內之環境與社會經濟特徵並藉由資料前處理轉換為網格型態，以此作為卷積類神經網路的資料來源，預計使用常態化差值植生指標（Normalized Difference Vegetation Index, NDVI）、常態化差值水指標（Normalized Difference Water Index, NDWI）、農業用地面積占比、建地面積占比、交通用地占比、桃園市公車站點距離、桃園市火車站點距離，用於提供交易標的周圍之環境特徵給模型。
3. 結合卷積類神經網路與多層感知器進行本研究之模型設計，並額外建立兩組模型，分別為以地理加權回歸為主的線性回歸，以及與多層感知器結構相符的類神經網路，以此做為基準測試與殘差分析的對照組。

## 研究限制

以下為本研究於資料取得過程與研究過程時的限制。

1. 本研究所使用之房屋資料均來自內政部實價登錄2.0資料集，並且只使用交易資產類別為住宅大樓（11層含以上有電梯）、公寓（5樓含以下無電梯）、華廈（10層含以下有電梯）、套房（1房1廳1衛）。
2. 因部分所使用之網格資料須由衛星影像計算取得，然而因取象時間與雲覆問題，導致無法衛星影像取向時間無法與實價交易紀錄時間完美重疊，因此有機率導致模型於基準比較時的偏差。
3. 本研究預計建立三組模型進行基準比較，其中一組為地理加權迴歸，其餘兩組基於類神經網路，本研究只對預測結果的準確率與泛化能力進行比較，不針對模型內參數進行解釋。

## 研究範圍

研究範圍設定為桃園市桃園區內，桃園區為桃園市內成交紀錄分布較為平均，且資料數量較豐富的區域，並且區域內房價於空間上的變化劇烈，適合用於探討空間異質性問題，研究區請參考圖 1。

|  |
| --- |
|  |
| 圖 1 研究範圍圖 |

第二章 文獻回顧



## 傳統估價方法

「現代估價師於勘估標的物時所使用的方法包含比較法、收益法、成本法與特徵價格法等」（Yeh, I. C, 2018），我國法規「不動產估價技術規則」的第三章中詳細記載前三種方法的估價方式，並且於其他章節明確定義不動產估價作業程序及相關規則，以下將針對經常使用之估價方法進行介紹，並評估各項方法之優缺點。

### 比較法

我國法規「不動產估價技術規則」第三章第一節中便記載著比較法的定義，於第18條中描述到「比較法是以比較標之價格為基礎，經過比較、分析與調整等，用以推算勘估標之價格的方法。」，並且於21條中描述比較法之估價程序：

* 1. 蒐集並查證比較標的相關資料。
  2. 選擇與勘估標的條件相同或相似之比較標的。
  3. 對比較標的價格進行情況調整及價格日期調整。
  4. 比較、分析勘估標的及比較標的間之區域因素及個別因素之差異，並求取其調整率或調整額。
  5. 計算勘估標的之試算價格。
  6. 決定勘估標的之比較價格。

Pagourtzi et al.（2003）認為比較法是最被廣泛使用的估價方法，但是其效果極度依賴於交易資料之可用性、精準度及完整度。Yeh, I. C.（2018）認為比較法中的比較案例價格在應用於標的物前必須要由權重修正，然而這些權重的估算缺乏合理且客觀的方法，完全由估價者的主觀意識來決定。

「不動產估價技術規則」第26條也僅提到「經比較調整後求得之勘估標的試算價格，應就價格偏高或偏低者重新檢討，經檢討確認適當合理者，始得作為決定比較價格之基礎。檢討後試算價格之間差距仍達百分之二十以上者，應排除該試算價格之適用。」，代表法規上的權重規範也僅針對調後的比較標的是否於合理範圍內，沒有對權重的估算方法有明確規範。由此可知，由於缺乏明確權重規範，比較法於價格調整階段會涉及大量個人主觀意識，缺乏較為客觀的估算。

### 成本法

我國法規「不動產估價技術規則」第三章第二節中便記載著成本法的定義，於第28條中描述到「指求取勘估標的於價格日期的重建成本或重置成本，扣減其累積折舊額或其他應扣除部分，用以推算勘估標的價格之方法。」，並於49條中描述比較法估價之程序：

1. 蒐集資料。
2. 現況勘察。
3. 調查、整理、比較及分析各項成本及相關費用等資料。
4. 選擇適當方法推算營造或施工費。
5. 推算其他各項費用及利潤。
6. 計算總成本。
7. 計算建物累積折舊額。
8. 計算成本價格。

Pagourtzi et al.（2003）認為成本法適合對特殊標的物進行估值，經常被用來針對無法在市場上尋找到相似案例的標的物，然而估值出的結果易與市場價格偏離，Dotzour, M.（1990）認為成本法的缺點在於其計算折舊時所使用的估算法是根據估價者的主觀意識，並且會因標的物的年齡增加而產生更大的估算錯誤，因此不利於用於老舊標的物的估值。Mason, J. J.（1993）則認為成本法需要考量到的專業知識過於深厚，並且許多特殊標的物無法被重建，因此透過此法進行估價的難度較大。

### 收益法

「收益法的運用原理是從投資的角度將不動產未來預期會產生之收入，以適當的收益資本化率（capitalization rate, cap rate）與折現率（yield rate）折算回現在，用以推估不動產價值」（張桂琪, 2019），我國法規「不動產估價技術規則」第三章第二節中便記載著收益法的定義，於第28條中描述到「收益法得採直接資本化法、折現現金流量分析法等方法。」，直接資本化法與折現現金流量分析法分別於第29條「指勘估標的未來平均一年期間之客觀淨收益，應用價格日期當時適當之收益資本化率推算勘估標的價格之方法」與第31條「指勘估標的未來折現現金流量分析期間之各期淨收益及期末價值，以適當折現率折現後加總推算勘估標的價格之方法。」，並且於第34條中描述收益法估價之程序：

1. 蒐集總收入、總費用及收益資本化率或折現率等資料。
2. 推算有效總收入。
3. 推算總費用。
4. 計算淨收益。
5. 決定收益資本化率或折現率。
6. 計算收益價格。

採收益法評估不動產價格時，因為收益價格等於客觀淨收益除以收益資本化率，若收益資本化率之決定不夠客觀、嚴謹，將會大大影響所評估的不動產收益價格，甚至影響不動產估價報告使用者的決策以及對評估結果的信賴感（陳詠婷, 2018）。

### 特徵價格法

特徵價格法（Hedonic Price Method, HPM）又被稱為特徵價格理論或特徵迴歸，是一種使用商品特徵作為推估商品市場價格的方法（Herath, S, 2010），特徵價格理論認為，人們消費財貨、服務所擁有的屬性以從中獲得效用，因而效用是財貨、服務的屬性所構成的函數（蔡明哲, 2015）。

以房地產為例，每個標的物由各式各樣的特徵所組成，並且會因為其所在的地理位置而反映至價格上，標的物特徵的存在，代表著房地產不是同質性高的商品，不同特徵的組合以及買家對於這些特徵的個人評價又有所差異，以上種種因素都使估價過程困難重重，因此特徵價格法作為解決此問題的常見方法，能夠將房屋價格分解至各項因子，並預測其價格（Sirmans, S, 2005）。

Selim, H.（2009）就使用特徵價格法預測土耳其都市與鄉村不動產價格，分別達到調整R平方0.546與0.556，並且整體R平方達到0.643，Król, A.（2015）也使用特徵價格法對波蘭Wrocław地區進行不動產價格的預測，使用當地最大的不動產資料庫OtoDom作為資料來源，並達成調整R平方0.7543的結果，Xiao, Y et al. （2015）也使用特徵價格法進行中國北京市的不動產價格預測，並於結果達成調整R平方0.8384，由此可知，特徵價格法作為以迴歸為基礎的不動產估價方法，有別於前幾節所提到的比較法等方法，是屬於高度量化的統計方法，在房價的估算上大量排除了人為主觀想法等因素的介入。然而，此類以線性迴歸模型作為預測模型的方法，忽略了交易標的於空間的座標，以及鄰近的空間特徵，使模型沒有考量到空間異質性（Spatial heterogeneity），進而使模型面臨空間非穩態問題（Spatial nonstationarity），有鑑於此，許多研究開始將目光投向以解決空間非穩態為前提的模型。

## 地理加權迴歸

過去許多研究都使用HPM對不動產價格進行預測，然而此種方法只關注房屋特徵與價格之間的映射，無法將標的物周圍的環境變數納入模型考量，將評估標的區域當成完全的同質區看待，然而不動產為異質性高的商品，其所伴隨的環境變數自然不會是完全同質。

長久以來空間推估模型一直是地理資訊科學的關注重點，其中迴歸方法常被用來探討自變數與依變數之間的線性關係，或用以推論預測依變數，地理加權迴歸（Geographically Weighted Regression, GWR）是一種地域型迴歸分析方法，將變數的空間相依性特徵納入迴歸參數推估的考量，分別針對預測目標周圍的統計單元建構空間的權重矩陣，將非穩態的空間變化關係整合至迴歸模型中，有別於以往全域型的最小二平方法，地理加權迴歸分析是一種局部估計方法，可提升整體模型推估的準確性（Brunsdon et al., 1996 ; Fotheringham et al., 2002）。

### 地理加權迴歸模型

GWR模型如公式 1所示：

|  |  |  |
| --- | --- | --- |
|  |  | 公式 1 |

*i*代表一個空間座標的索引，則代表其相應的空間座標，代表該筆資料的依變數，與分別代第*k*個自變數權重以及自變數，與則分別代表常數項以及誤差項。除此之外，GWR的特點在於每個用於*i*點推估的資料點自變數權重會根據空間距離而變化，而這個變化是由空間權重矩陣（Spatial weighted matrix, ）所決定，自變數權重推估如公式 2所示：

|  |  |  |
| --- | --- | --- |
|  |  | 公式 2 |

代表根據空間權重矩陣推估出來的自變數權重，代表轉置後的自變數矩陣，則代表該推估點相較其他資料點的空間權重矩陣。

### 地理加權迴歸於不動產估價之應用

過去有許多研究將GWR應用於各種情境上，並且大部分研究都將GWR與全域型的迴歸分析進行基準測試，以此驗證將GWR應用於異質性高的區域時，能夠考量到該地的空間異質性，從而降低空間非穩態問題所帶來的模型偏誤。

Massimo, D. E et al.（2018）就使用GWR對義大利南部城市Reggio Calabria進行不動產價格推估，並且使用AICc模式進行空間帶寬優化，最終達成調整後R平方0.697的結果。黃于祐（2008）也使用GWR對臺北市12個行政轄區於民國八十八年一月至九十三年十二月的不動產交易資料進行建模與推估，結果顯示相較於全域型迴歸分析，GWR的模型解釋能力明顯提升2.1%，標準誤下降了6.11。Zhang, S et al.（2019）也使用GWR對江蘇省南京市的不動產價格進行推估，使用CV模式對空間帶寬進行優化，並且同時建立最小二平方法（Ordinary Lease Squares, OLS）模型，結果顯示GWR達到調整後R平方0.619，相較於OLS達到的0.508，其模型準確率更勝一籌。

由上述研究可看出，全域型迴歸模型雖然能建立自變數與應變數之間的線性映射，但是此種作法的缺點就是無法考量空間異質性於模型映射之間的影響，於是許多研究將GWR引入，並且將其與全域型迴歸進行基準測試，結果顯示此種作法的確能夠部分地提升模型解釋力。然而，現實生活中的映射關係往往是非線性的，GWR雖然能夠考量到空間異質性問題，但終歸是屬於線性模型，如預測不動產價格等應用情境都是極其複雜的，普通的線性迴歸無法很好的建立特徵與價格之間的映射。

## 深度學習技術

地理加權迴歸與特徵價格法等模型都是基於線性迴歸為核心，並且均被應用於不動產估價中，然而不動產交易中的標的物特徵與市場價格並不是單純的線性關係，即使排除非正規交易等因素，標的物特徵與市場價格仍是較為複雜的非線性映射，因此使用基於迴歸分析為主的方法無法很好的建立一個高解釋力模型。有鑑於此，許多研究開始將機器學習與深度學習等技術應用於不動產價格預測上，其中又以深度學習中的類神經網路的表現最為突出，許多研究證實其預測準確率優於許多線性方法。

### 類神經網路理論

類神經網路（Artificial Neural Network, ANN）是一種強大的深度學習方法，它由多個神經元（Neuron）所組成，並且可以很好的運用於非線性映射模式的建立。其架構是模仿人體大腦中的神經網路，由輸入層、隱藏層和輸出層組成。每一層都由數個神經元所組成，這些神經元接收上一層傳遞過來的輸入，不同層之間的神經元會互相連結，每個連結都代表一個權重，輸入會經過權重運算與激發函數後形成輸出。這些權重於訓練開始前隨機生成，並且在反向傳播（Back propagation, BP）的過程中進行權重的修正，以進行模型的學習。圖 2為人工神經元的數學模型。

|  |
| --- |
|  |
| 圖 2 人工神經元模型 |

每個神經元都會接收來自上一層神經元的輸出，並且經過權重計算後進行加總，此時得到*net*輸出，接著將*net*輸入進後方的激發函數（Activation function）得到*out*輸出。激發函數在類神經網路中扮演著非常重要的角色，它可以使神經元的輸出非線性化，並且增強神經網路的表達能力。常用的激發函數包括sigmoid函數、tanh函數、ReLU函數以及softmax函數等。其中，sigmoid函數通常用於二分類問題，softmax函數常用於多分類問題。圖 3為類神經網路的架構示意圖。

|  |
| --- |
|  |
| 圖 3 類神經網路架構示意 |

類神經網路由輸入層、隱藏層與輸出層組成，輸入層可看作預測目標的特徵值，以本研究為例就是不動產標的物的特徵值，隱藏層則是負責類神經網路大部分的運算，每個隱藏層可由數個神經元組成，並且能依照需求更換不同的激發函數，最後則是輸出層，輸出層的神經元數量會根據類神經網路處理的任務不同而有所變化，若類神經網路是被設計用來處理多分類任務（Multi-classes Classification），那輸出層的神經元數量就會是一個以上，然而本研究欲使用類神經網路進行不動產價格預測，如此一來就是迴歸任務（Regression），輸出結果只會有一個連續數值，因此本研究所建立之類神經網路之輸出層只會由一個神經元組成。

上述介紹到由神經元組成的隱藏層又稱為全鏈結層（Fully connected layer）或Dense層，然而類神經網路中有許多種變體，卷積類神經網路（Convolutional Neural Network, CNN）便是由卷積層（Convolution layer）所組成的類神經網路，被大量應用於影像辨識和語音分析領域的研究，CNN使用數個卷積核（Kernel）對輸入資料進行特徵提取，每個卷積核可視作一個神經元，其運算方式與神經元相同，差別在於卷積層執行的是卷積運算，搭配超參數Stride，使卷積核於資料上進行平移，圖 4為卷積運算的示意圖。

|  |
| --- |
|  |
| 圖 4 卷積運算示意圖 |

從圖 4可看到卷積核會從左至右、從上至下的移動，每次移動所計算的結果都會儲存成下一個二維張量，我們稱之為特徵地圖（Feature map），而這個特徵地圖也會做為下一個卷積層的輸入。

### 類神經網路應用於不動產估價

綜觀國內外有關於類神經網路的研究都有相對高的準確率，類神經網路被廣泛應用於許多應用情境，特別是不動產價格的預測上。Jiang, Z.（2019）就使用爬蟲技術蒐集上海地區的不動產交易資料，使用類神經網路作為模型，於2100次參數更新時結束訓練，並達到94.59%的準確率。Mukhlishin, M. F. et al.（2017）使用了模糊邏輯、類神經網路與K-近鄰演算法分別進行印尼不動產價格的預測，於測試資料達到平均77%的準確率。Xiao, L.（2019）也使用類神經網路進行不動產價格預測，使用上海統計年鑑做為資料來源，並且搭配主成分分析（Principal Component Analysis, PCA）對資料進行降維，以減少類神經網路收斂所需的時間，最終於測試資料集上達到97.04%的準確率。

此外，國內也有許多研究利用內政部提供的實價登錄資料作為類神經網路的訓練資料，黃惠芬（2017）就使用實價登錄資料，從中擷取高雄市2013至2016年高交易量地區的房屋特徵值，並使用類神經網路進行價格預測，最終於自住型集合住宅類別的R平方達到0.85以上，張哲文（2017）也使用實價登錄資料搭配類神經網路對彰化縣透天厝進行價格預測，使用到土地面積、交易日期、建物面積、建物年份、人口數、人口密度與交易座標作為模型輸入特徵，最終達到R平方0.8。

許多將類神經網路應用於不動產價格預測的研究都取得良好的結果，然而上述許多研究中模型所使用到的輸入特徵，僅僅是描述了交易標的本身的特徵，有部分研究雖然額外提供了交易標的經緯度，然而這些變數所提供的資訊都只圍繞在標的物上，並沒有將標的物周圍的環境變數納入模型考量。聯合國世界衛生組織於1961年的住房公共健康專家會議報告中就指出人類居住環境需要基本的生理基礎，其中包含了許多環境指標，如空氣品質、環境噪音、公園綠地等因子，這些都是會影響人類居住上生心理健康的因素，因此消費者於購房時，除了標的物本身的特徵外，標的物周圍的環境特徵也成為人們越發在意的條件。因此本研究將基於原有的多層感知器架構，額外開發卷積類神經網路架構，作為交易標的周圍環境特徵的特徵提取途徑，試圖將環境特徵納入不動產價格預測中。

第三章 研究方法



## 研究流程

本研究首先會針對2015年1月至2018年12月的實價登錄2.0資料設計資料前處理流程，以進行後續的模型建立。為了進行基準測試，本研究將建立三個不同的模型，分別為地理加權迴歸、多層感知器和混合輸入類神經網路。在進行基準測試之前，本研究會進一步分析三個模型的預測能力，並評估其準確度。若模型的預測結果未達預期，則本研究將進行模型超參數的調整並重新訓練，以提高模型的準確度。

本研究的研究流程如圖 5所示，從資料前處理到模型建立，並進行準確性測試和超參數優化，最終獲得三組模型。

|  |
| --- |
|  |
| 圖 5 研究流程圖 |

## 研究資料

立法院於2020年12月份三讀通過「平均地權條例部分條文修正案」、「地政士法部分條文修正案」、「不動產經紀業管理條例部分條文修正案」等地政三法，至此實價登錄2.0正式登場，此次修法將原有的實價登錄資料進行更完整的揭露，包含標的物的門牌與地號資訊、預售屋即時申報與合約納管等資訊，並且於房屋所有權轉移登記完成內30日得主動向主管機關申報，進行交易相關資訊登錄。所需登錄的資訊如表 1：

表 1 實價登錄欄位資訊

|  |  |  |
| --- | --- | --- |
| **連續資料** | **類別資料** | **描述性資料** |
| 總樓層數 | 鄉鎮市區 | 交易標的 |
| 移轉層次 | 都市土地使用分區 | 土地位置/建物門牌 |
| 交易筆棟數 | 非都市土地使用分區 | 備註 |
| 建物移轉總面積(平方公尺) | 非都市土地使用編定 | 編號 |
| 建物現況格局-房 | 有無管理組織 |  |
| 建物現況格局-廳 | 建物型態 |  |
| 建物現況格局-衛 | 主要用途 |  |
| 車位移轉總面積(平方公尺) | 主要建材 |  |
| 主建物面積 | 建物現況格局-隔間 |  |
| 附屬建物面積 | 車位類別 |  |
| 陽台面積 | 電梯 |  |
| 交易標的橫坐標 | 有無備註欄(Y/N) |  |
| 交易標的縱坐標 |  |  |
| 土地移轉總面積(平方公尺) |  |  |
| 車位總價(元) |  |  |
| 總價(元) |  |  |
| 單價(元/平方公尺) |  |  |
| 建築完成年月 |  |  |
| 交易年月日 |  |  |

表 1為本研究對實價登錄資料各欄位進行變數分類，主要分為連續變數、類別變數與描述性資料，由於此份資料來源複雜，有部分欄位的內容未進行正規化，無法直接用做模型輸入。此外，由於實價登錄資料具有一定複雜性，資料的邏輯檢核也是必須特別關心的重點部分，以此來排除許多不合理的交易紀錄。有鑑於此，本研究為此資料設計了一套資料前處理的流程，詳細內容將於下一節進行描述。

### 實價登錄2.0資料前處理流程

本研究所設計之資料前處理流程共有五個標準程序，每一個程序都是為了特定目標而設計，並且每個程序環環相扣缺一不可，接下來本研究會針對每個程序進行解說，並且附上流程圖。

首先是進行單一縣市多年期資料的schema檢查，若發現某年度之資料欄位有異動，則會由人工重新評估並決定是否要重新設計研究schema或是捨棄該欄位，若確認多年期的欄位都沒有異動，則會進行多年期資料的merge，最終產出一份該縣市多年期的實價登錄資料，流程請參考圖 6：

|  |
| --- |
|  |
| 圖 6 資料前處理程序一 |

接著進入程序二，程序二主要被設計用來解析交易內容的相關資料，由於實價登錄資料許多欄位充斥著未標準化的資料，而這些未標準化的資料無法做進一步的加值應用，於是程序二是不可或缺的一環，其解析的欄位包含交易年月日、建築完成年月、總樓層數、移轉層次、電梯、有無管理組織、建物現況格局-隔間、交易筆棟數、都市土地使用分區、非都市土地使用分區、非都市土地使用編定、建物型態、停車位型態等欄位，流程請參考圖 7：

|  |
| --- |
|  |
| 圖 7 資料前處理程序二 |

接著進入程序三，此程序被設計用來檢查解析結果的狀態，解析結果狀態主要分三種情形，第一種為因為資料內容錯誤而導致程式出現執行期錯誤，第二種為因為內容缺失導致無法判斷結果，第三種則為解析成功，此程序若解析結果不成功，則會將該筆資料排除，並無法進入接下來的流程，程序三流程可參考圖 8：

|  |
| --- |
|  |
| 圖 8 資料前處理程序三 |

接著進入程序四，此程序被設計用來進行邏輯檢核，其中較特別的項目為備註檢核，此操作是為了排除非正規交易，如親友與員工或其他特殊關係間之交易、向政府機關承購之案件、包含公共設施保留地之交易等紀錄，都會於備註欄位進行註記，其餘則是針對移轉層次與建物型態進行樓層高度的邏輯檢核，若上述流程結束後發現該筆資料有邏輯異常，則會將該筆資料排除，並無法進入接下來的流程，程序四的流程可參考圖 9：

|  |
| --- |
|  |
| 圖 9 資料前處理程序四 |

最後進入程序五，此關卡是被設計用來計算移轉面積單價，由於實價登錄資料的單價是使用建物總移轉面積加上停車位移轉面積做為分母，去和總價進行單價的計算，本研究預測的價格是以建物本身為主，因此需排除掉停車位所佔的移轉面積，因此價格與移轉面積之關係式為(建物移轉總面積-車位移轉面積)\*單價+車位價格=總價。

### 環境特徵資料前處理

除了交易標的本身的特徵外，本研究的另一個主要目標為將交易標的周圍的環境特徵納入模型輸入，提供模型預測的額外資訊，此類環境特徵資料可以分為很多面向，但本研究目前暫且朝自然環境與社會經濟資料著手，接下來會分別講解兩種類型的資料前處理方法。

首先是自然環境資料，本研究使用以衛星影像進行計算後取得的常態化差值植生指標、常態化差值水指標，這份資料分別能反映出居住環境的公園綠地、水體與建物樣態。

本研究使用哨兵2號（Sentinel-2）衛星的資料作為衛星影像的資料來源，哨兵2號為歐洲太空總署哥白尼計畫的一個地球觀測任務，常用於森林監測、水質監測、土地覆蓋變化偵測、天然災害管理等遙測分析，該任務由兩顆相同的哨兵衛星組成，分別為哨兵2A與哨兵2B，所拍攝多光譜的影像總共有13個波段，空間解析度為60公尺，搭配其高時間解析度，適合用來進行長時間的土地覆蓋觀測，本研究預計取得哨兵2號於2015/01至2018/12間每個月的衛星影像資料，並且取紅光與近紅外光波段進行NDVI的計算，主要取像範圍以桃園區為主，大約有55至60平方公里。

接著是社會經濟資料，本研究使用內政部主計處社會經濟資料服務平台所提供之桃園市國土利用調查統計資料，使用農業、建地、運輸用地比例作為輸入，這份資料能反映出居住環境周圍的土地利用樣態，土地利用與房價為高度相關的映射。但由於此份資料為村里多邊形資料，需要將其轉換為網格型態才能輸入至模型，因此本研究會將此份國土利用調查資料轉換為與哨兵2號空間解析度相同的10公尺網格資料。除此之外，本研究也擷取桃園市內的公車站與火車站站點，並計算每個網格與公車站、火車站的直線距離，以此作為公共運輸設施之可及性參考。

## 演算法

本章節會針對本研究所使用之演算法與流程進行詳細解說，本研究預計對研究區建置三個模型，分別為以線性迴歸為主的地理加權迴歸、多層感知器、混合輸入類神經網路。三組模型都將於相同的資料集上進行訓練，並且最終以調整後R平方作為基準比較之依據。

### 最小平方法與地理加權迴歸

本研究首先將2015年1月至2018年12月之實價登錄資料進行資料前處理，接著使用普通最小平方法（Ordinary Least Squares, OLS）對研究區內的實價登錄資料進行建模，所使用到的自變數包含移轉層次、房間數量、廳數量、衛浴數量、建物移轉面積與總樓層高等6個自變數，並且依變數為扣除車位後的房屋單價（價格/平方公尺），本研究所使用之OLS模型可參考公式 3。

|  |  |  |
| --- | --- | --- |
|  |  | 公式 3 |

為依變數，則為第n筆資料的第i個自變數，為第i個自變數的參數，為常數項，為殘差項。對研究區內進行OLS建模後，本研究首先會對模型的診斷結果進行分析，主要判斷Koenker（BP）Statistic指標是否具有顯著性，若其顯著則代表模型於研究區內存在空間非穩態問題，並且根據此結果進一步判斷Joint Wald Statistic是否顯著，若顯著則代表整體模型具有顯著性，根據上述步驟，若模型於研究區內存在空間非穩態問題，則本研究接續進行地理加權迴歸建模。

本研究所使用之地理加權迴歸模型可參考公式 1，預計使用距離空間帶寬，並且選用黃金比例優化法進行空間帶寬的優化，優化指標則使用AICc作為空間帶寬的評分優化指標，並且於空間帶寬優化完成後，產製出各資料點之空間權重矩陣，最後用於區域迴歸式的估算，計算模型於研究區內的調整後R平方，用以作為後續基準比較的依據。

### 類神經網路架構設計

本研究設計之類神經網路主要由三部分組成，分別為以卷積層為主的卷積類神經網路，以及由全鏈結層為主的多層感知器，以及連接這兩個模型的權重分享層，卷積類神經網路於進入權重分享層之前會經過扁平層，將卷積層輸出的特徵地圖（Feature Map），轉換為特徵向量，用以輸入至權重分享層，模型示意結構請參考圖 10。

|  |
| --- |
|  |
| 圖 10 模型示意圖 |

接下來本研究會針對類神經網路各部分進行詳細解說，首先是由卷積層所組成的卷積類神經網路，主要負責對交易標的周圍的環境變數進行特徵提取，此架構由四層卷積層組成，每層的卷積核大小都是3\*3，並且將stride設定為1，使輸出的特徵地圖大小維持一致，並且每層的輸出都會經過Relu激發函數，此外，為了避免模型於訓練過程中遇到過度擬合（Over-fitting）的狀況，本研究設計於每層卷積層都搭載Dropout層，此操作可使類神經網路於每次計算損失時，以特定機率忽略隱藏層中的部分神經元，並且於反向傳播時也不更新這些被忽略神經元的參數，以此抑制類神經網路過度擬合於訓練資料集，導致其於驗證集與測試集上無法達到預期的成效，本研究所設計之卷積類神經網路架構可參考圖 11。

|  |
| --- |
|  |
| 圖 11 卷積類神經網路架構圖 |

接著是由全鏈結層所組成的多層感知器，此類神經網路架構主要負責提取交易標的本身的特徵，此架構之輸入層由6個神經元組成，代表本研究輸入的6個交易標的特徵，隱藏層主要由四層全鏈結層所組成，第一層有256個神經元，第二層有128個神經元，第三層有64個神經元，第四層有32個神經元，每層的輸出都會經過Relu激發函數，並且為了避免過擬合狀況發生，本研究也於每層全鏈結層搭載Dropout層，以降低過擬合發生的機率，本研究所設計之多層感知器架構可參考圖 12。

|  |
| --- |
|  |
| 圖 12 多層感知器架構圖 |

最後是匯聚兩個模型的權重分享層，此層的輸入為上述兩個模型的輸出，目的為整合兩種不同資料來源的特徵，將這些特徵進行提取，並且最後作為模型預測的依據，此層由三層全鏈結層組成，第一層有32個神經元，第二層有16個神經元，第三層有1個神經元，前兩層搭載Relu激發函數與Dropout層，僅輸出層沒有搭載任何轉換，模型結構圖請參考圖 13。

|  |
| --- |
|  |
| 圖 13 權重分享層架構圖 |

除了卷積類神經網路與多層感知器的混合體外，本研究為了驗證將卷積結構納入模型的效果，額外建立一個只由多層感知器與參數共享層的類神經網路，此模型只會使用到交易標的的特徵作為輸入，最後再將兩個模型進行基準比較，以驗證加入卷積層以及環境特徵是否有助於房價的預測。

### 類神經網路訓練與基準測試

此段會詳細講述本研究於模型訓練時所劃分的資料集、使用到的所有超參數（Hyperparameters）以及訓練流程。

首先進行資料集劃分，本研究預計按照6：2：2的比例劃分出訓練集、驗證集與測試集，每份資料集內所含不同建物類別的資料都會平均分散，使不同類型的建物資料均勻分布在這三個資料集間，訓練集會用做模型的參數更新，驗證集則是用於訓練過程時，確保模型沒有發生過擬合的狀況，測試集則是用於訓練結束後進行模型成效計算。

有關於超參數的部分，可參考表 2，此份超參數設定為實驗可控參數，會於實驗過程中進行調整，本研究暫時使用以往在訓練類神經網路時常用的超參數，待後續實驗結果出爐後再進行優化。

表 2 模型超參數

|  |  |
| --- | --- |
| Learning rate | 0.01 |
| Batch size | 32 |
| Epoch | 10000 |
| Optimizer | Adam(betas=(0.9, 0.999)) |

本研究會於模型訓練時，將訓練集用作參數更新，驗證集則用來檢測訓練過程時模型是否發生過擬合，本研究暫定為10000次訓練迭代，並且使用L1 loss（Mean Absolute Error Loss）作為損失函數。

除了原有的超參數外，本研究也針對卷積層提出額外的超參數，主要為輸入特徵的維度與實際地理維度，由於多維張量的維度與模型結構為固定關係，因此於更改輸入特徵的地理維度時，同時也需要對模型結構進行修改，Van den Berg, M. et al. （2015）就針對40文獻進行居住環境與環境影響範圍的探討，因此本研究於網格資料上，預計以每筆資料點為中心，向外延伸400公尺裁切出一個大小為400公尺\*400公尺的矩形範圍，對應到本研究所設定的10公尺網格，輸入張量的維度約為(7, 40, 40)，其中第一個參數為輸入特徵的數量。

本研究於上述模型都訓練完畢後會進行基準測試，測試指標為調整後R平方，若包含卷積層的模型準確率不如預期，則本研究會回頭進行超參數調整，以及模型結構重新設計。訓練流程演算法可參考表 3。

|  |
| --- |
| 表 3 訓練演算法 |
| |  | | --- | | **Algorithm 1** Model training algorithm | | **Input:** *Xa,* attribute feature dataset. *Xr*, raster feature dataset.  **Input:** *θm,* multi-layer perception. *θmi*, mixed input model.  **Input:** *m*, batch size of a minibatch. *r*, cropped size of raster feature.   |  |  | | --- | --- | | 1: | Initialize*θm*, *θmi* | | 2: | **for** epoch **do** | | 3: | **for** number of iteration **do** | | 4: | Sample minibatch of data from ***Xa***with size ***m***. | | 5: | Sample minibatch of data with coordinates of ***Xa*** from ***Xr***with size ***m***. | | 6: | Perform gradient descent to ***θm*** using minibatch of ***Xa***. | | 7: | Perform gradient descent to ***θmi*** using minibatch of ***Xa, Xr***. | | 8: | **end for** | | 9: | **if** accuracy of *θmi* > *θm* **do break** | | 10: | **end for** | | |

第四章 結果與討論



皮爾森相關係數矩陣

本研究首先以整個實價登錄資料集為基礎，計算各變數之間的皮爾森相關係數，並且以此做為模型自變數選擇的參考，產製出的皮爾森相關係數矩陣可參考圖 14，根據此結果，本研究選擇總樓層數、建物移轉面積、土地移轉面積、建坪面積、移轉樓層、陽台面積等六個房屋特徵作為自變數。

|  |
| --- |
|  |
| 圖 14 皮爾森相關係數矩陣 |

線性迴歸模型結果

本小節列出了基於以線性迴歸模型為基礎的結果，本研究首先進行普通最小平方法建模後，根據Koenker(BP) Statistic檢定來判斷是否接續進行地理加權迴歸建模，並且對兩者的解釋力進行比較。

### 普通最小平方法結果

本研究首先對整個研究區進行OLS建模，建模結果可參考表 3，獲得調整後R平方0.291，Koenker(BP) Statistic檢定顯著，意味著模型有空間非穩態的問題存在，此外，Joint Wald Statistic檢定也為顯著，代表整體模型是顯著的。由於OLS於研究區內存在空間非穩態問題，導致模型整體解釋力不佳，因此本研究接續使用地理加權迴歸進行建模。

|  |
| --- |
| 表 4 普通最小平方法結果 |
| |  |  |  |  | | --- | --- | --- | --- | | Multiple R-Squared[d] | 0.291 | Adjusted R-Squared[d] | 0.291 | | Joint F-Statistic[e] | 2324.959 | Degrees of freedom | 0.000000\* | | Joint Wald Statistic[e] | 10546.402 | Degrees of freedom | 0.000000\* | | Koenker(BP) Statistic[f] | 657.728 | Degrees of freedom | 0.000000\* | | Jarque-Bera Statistic[g] | 38839.665 | Degrees of freedom | 0.000000\* | |

### 地理加權迴歸結果

本研究接著對研究區進行GWR建模，建模結果可參考表 4，獲得調整後R平方0.637，AICc優化至730299.212，並且使用黃金搜索法對空間帶寬進行最佳化。藉由使用地理加權迴歸建模，可以大幅提升模型整體解釋力。除此之外，使用地理加權迴歸建模可以更好地降低空間異質性的影響，並考慮到研究區不同地區之間可能存在的差異。此外，透過黃金搜索法對空間帶寬進行最佳化，可以更準確地選擇最適合模型的參數，提高模型的預測準確度。

|  |
| --- |
| 表 5 地理加權迴歸結果 |
| |  |  | | --- | --- | | R-Squared[d] | 0.641 | | Adjusted R-Squared[d] | 0.637 | | AICc | 730299.212 | | Sigma-Squared | 134270922.531 | | Sigma-Squared MLE | 132847613.226 | | Effective Degrees of Freedom | 33511.957 | |

類神經網路訓練結果

本小節列出了以類神經網路模型為基礎的結果。在本研究中，本研究設計了兩個不同的類神經網路模型，並以房屋單價預估作為案例，以探究不同方法的成效。首先，我們使用全鏈結層為主的多層感知器模型，該模型僅使用房屋特徵作為輸入，並以建物每平方公尺單價作為輸出。另外，我們還開發了一種基於多層感知器的混合輸入類神經網路模型，其中包含卷積層作為特徵提取管道。經過多次超參數調整和重新訓練，最終對這兩個模型進行基準比較。

值得注意的是，由於類神經網路的複雜性，我們需要經常調整模型的參數以獲得最佳的結果。在本研究中，我們對這兩個模型的超參數進行了多次最佳化，以確保它們在預測房屋單價的準確性和可靠性。

### 多層感知器結果

首先是多層感知器的訓練結果，根據模型損失收斂的趨勢，在第1350次 epoch 時我們終止了訓練。如圖 15所示，訓練損失逐漸下降，代表多層感知器在訓練過程中逐漸最佳化。此外，本研究使用調整後的 R 平方來評估模型的性能，可參考圖 16。

根據實驗結果結果，在訓練集上，多層感知器模型的最大調整後 R 平方達到了0.846，在驗證集上達到了0.744。這代表我們的模型在預測房屋單價方面具有較高的準確性和可靠性，相較於傳統的線性回歸模型。值得一提的是，我們在訓練過程中使用了正規化和損失函數優化等技術，以避免過度擬合和提高模型的泛化能力。

|  |
| --- |
|  |
| 圖 15 多層感知器訓練損失圖 |
|  |
| 圖 16 多層感知器訓練調整後R平方圖 |

### 混合輸入模型結果

接著為混合輸入模型的訓練結果。和多層感知器模型一樣，本研究在第1350次 epoch 時結束了訓練。根據模型損失收斂的趨勢，訓練損失逐漸下降，如圖 17所示。此外，我們同樣使用調整後的 R 平方來評估模型的性能，如圖 18所示。

根據實驗結果，在訓練集上，混合輸入模型的最大調整後 R 平方達到了0.941，在驗證集上達到了0.846。這個結果比多層感知器更好，代表混合輸入模型在預測房屋單價方面具有更高的精確度和可靠性。我們使用卷積層作為環境特徵提取管道，這種方法可以有效地為模型提供交易標的周邊的環境資訊

綜上所述，我們的研究使用了兩個基於類神經網路模型的方法來預測房屋單價。混合輸入模型表現優於多層感知器模型，這代表環境特徵提取管道的設計對模型的性能具有顯著提升。

|  |
| --- |
|  |
| 圖 17 混合輸入模型訓練損失圖 |

|  |
| --- |
|  |
| 圖 18 混合輸入模型訓練調整後R平方圖 |

### 基準測試

結束上述各模型的訓練後，本研究將各模型的準確率表現進行總體比較，並且歸納為。在線性迴歸的架構底下，相較於普通最小平方，地理加權迴歸模型解釋力更勝一籌，然而在深度學習的理論基礎下，本研究所設計與訓練出的模型，在模型解釋力上明顯優於傳統線性模型，以多層感知器為例，在驗證集時便達到了調整後R平方0.744，明顯高於地理加權迴歸約0.103。

此外，以深度學習為基礎的模型，其中又以本研究所設計之混合輸入模型的解釋力最高，於驗證集達到調整後R平方0.846，相較於沒有卷積類神經網路的多層感知器，高出了約0.102。這意味著混合輸入模型可以更準確地預測房屋單價。這可能是因為混合輸入模型結合了多種資料類型的特徵，包括房屋特徵與交易標的周圍的空間特徵，使得模型可以更全面地建立高解釋力的映射。另外，卷積類神經網路在處理網格類型的資料方面具有其特殊的特徵提取方法，因此在特徵提取的過程中加入卷積類神經網路也能夠增強模型的表現。總體而言，本研究所提出的混合輸入模型在解釋能力和準確性方面均優於其他控制組模型。

|  |
| --- |
|  |
| 圖 19 基準比較結果圖 |

第五章 結論與建議



## 結論

以往透過傳統估價方法進行估價需要具備大量先備知識，過程中需涉及許多人為主觀判斷，導致結果會有偏誤的發生，因此近幾年有許多量化方法被應用於不動產估價上，由最一開始以線性迴歸為主的特徵價格法，到近幾年逐漸成熟的深度學習方法，前者雖然能反映輸入與輸出之間的線性關聯，但於模型解釋力上稍嫌不足，後者則能透過反向傳播，對資料集進行深度擬合，於準確率上能達到更好的成果。

然而許多以深度學習為主的研究，只關注於交易資料本身，忽略了交易標的周圍的環境特徵，房地產又為受區位影響佔比重的商品，因此本研究基於以往的多層感知器模型，額外建立以卷積層為主的環境特徵提取管道，期望透過提供模型額外的資訊，以提升整體解釋力，並且此方法可以被應用在其他領域中。

研究結論能歸納為以下幾點：

1. 額外建立環境特徵提取管道的確能提升模型解釋力，並且由於類神經網路的高容錯性，使這模型更適合被應用於雜異性高的情境中。若僅使用房屋特徵作為模型輸入，則忽略了環境因子對房價的影響，這可能導致模型的預測能力下降。然而，透過環境特徵提取管道，可以將複雜的環境因子轉換為相對簡單的數值特徵，有效地位模型提供相關資訊用以進行判斷。因此，本研究所建立的混合輸入模型在實際應用中具有較高的穩定性和可靠性，並且可以更好地將區位因素納入依變數的預測中。
2. 本研究於環境特徵資料中使用到NDVI、NDWI、農業用地面積占比、建地面積占比、交通用地占比、桃園市公車站點距離、桃園市火車站點距離等環境因子，可以很好的提升模型的準確率。這些環境特徵資料可以幫助本研建構出效能更好的模型。例如，NDVI和NDWI可以提供關於植被和水體的資訊，而農業用地面積占比、建地面積占比、交通用地占比可以反映出該地區的土地利用與發展的狀況。此外，公車站點和火車站點距離也可以提供該地區交通便利性的資訊。這些環境因子的結合，可以更全面地反映房價受到哪些因素的影響，利於模型進行進一步的預測。

參考文獻

Bidanset, P. E., & Lombard, J. R. (2014). Evaluating spatial model accuracy in mass real estate appraisal: A comparison of geographically weighted regression and the spatial lag model. *Cityscape*, *16*(3), 169-182.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996)."Geographically weighted regression: a method for exploring spatial nonstationarity". Geographical analysis, 28(4), 281-298.

Dotzour, M. (1990). An empirical analysis of the reliability and precision of the cost approach in residential appraisal. *Journal of Real Estate Research*, *5*(1), 67-74.

Fotheringham, Stewart A., Chris Brunsdon, and Martin Charlton. Geographically Weighted Regression: the analysis of spatially varying relationships. John Wiley & Sons, 2002.

Herath, S., & Maier, G. (2010). The hedonic price method in real estate and housing market research: a review of the literature.

Jiang, Z., & Shen, G. (2019, November). Prediction of house price based on the back propagation neural network in the keras deep learning framework. In *2019 6th International Conference on Systems and Informatics (ICSAI)* (pp. 1408-1412). IEEE.

Król, A. (2015). Application of hedonic methods in modelling real estate prices in Poland. In *Data Science, Learning by Latent Structures, and Knowledge Discovery* (pp. 501-511). Springer, Berlin, Heidelberg.

Mason, J. J. (1993). Under the microscope: the cost approach. *The Appraisal Journal*, *61*(1), 116.

Massimo, D. E., Giudice, V. D., Paola, P. D., Forte, F., Musolino, M., & Malerba, A. (2018, May). Geographically weighted regression for the post carbon city and real estate market analysis: a case study. In *International Symposium on New Metropolitan Perspectives* (pp. 142-149). Springer, Cham.

Mukhlishin, M. F., Saputra, R., & Wibowo, A. (2017, November). Predicting house sale price using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor. In *2017 1st International Conference on Informatics and Computational Sciences (ICICoS)* (pp. 171-176). IEEE.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: a review of valuation methods. *Journal of Property Investment & Finance*, *21*(4), 383-401.

Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert systems with Applications*, *36*(2), 2843-2852.

Sirmans, S., Macpherson, D., & Zietz, E. (2005). The composition of hedonic pricing models. *Journal of real estate literature*, *13*(1), 1-44.

Van den Berg, M., Wendel-Vos, W., van Poppel, M., Kemper, H., van Mechelen, W., & Maas, J. (2015). Health benefits of green spaces in the living environment: A systematic review of epidemiological studies. *Urban forestry & urban greening*, *14*(4), 806-816.

Xiao, L., & Yan, T. (2019, November). Prediction of house price based on rbf neural network algorithms of principal component analysis. In *2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)* (pp. 315-319). IEEE.

Xiao, Y., Chen, X., Li, Q., Yu, X., Chen, J., & Guo, J. (2017). Exploring determinants of housing prices in Beijing: An enhanced hedonic regression with open access POI data. *ISPRS International Journal of Geo-Information*, *6*(11), 358.

Yeh, I. C., & Hsu, T. K. (2018). Building real estate valuation models with comparative approach through case-based reasoning. *Applied Soft Computing*, *65*, 260-271.

Zhang, S., Wang, L., & Lu, F. (2019). Exploring housing rent by mixed geographically weighted regression: A Case study in Nanjing. *ISPRS International Journal of Geo-Information*, *8*(10), 431.

陳詠婷（2018）。不動產收益資本化率研究：以中山區商辦大樓為例。輔仁大學金融與國際企業學系金融碩士論文。取自<https://hdl.handle.net/11296/64h3e8>。

黃于祐（2008）。台北市房價影響因素之空間分析-地理加權迴歸方法之應用。國立臺北大學都市計畫研究所碩士論文。取自<https://hdl.handle.net/11296/2562c8>。

黃惠芬（2017）。以類神經網路方法建構房價估價模型-以高雄市實價登錄資料為例。國立高雄應用科技大學金融系金融資訊研究所碩士論文。取自<https://hdl.handle.net/11296/2k9yv5>。

張哲文（2019）。實價登錄資料庫結合類神經網路推估房地產市價。國立中興大學土木工程學系碩士論文。取自<https://hdl.handle.net/11296/xwhe53>。

張桂琪（2019）。不動產收益資本化率研究-市場萃取法之分析。國立政治大學地政學系碩士論文。取自<https://hdl.handle.net/11296/54jz53>。

蔡明哲（2015）。台灣都會區氣候條件與空氣品質之價值評估-特徵價格法之應用。國立台北大學自然資源與環境管理研究所碩士論文。取自<https://hdl.handle.net/11296/r6fn54>。