



國立台灣科技大學
營建工程系
博士學位論文

校舍耐震資料庫之資料探勘

Data Mining on Aseismic School Building Database

研 究 生：高偉格

學 號：D9505501

指導教授：陳鴻銘博士

中 華 民 國 一 零 二 年 七 月 七 日

校舍耐震資料庫之資料探勘

學 生：高偉格

指 導 教 授：陳鴻銘博士

中文摘要

台灣國家實驗研究院地震工程研究中心（NCREE）在九二一大地震後，與教育部合作評估全台灣的各級學校校舍之耐震能力，在此計畫執行的過程中產生了大量的評估與調查資料，因此 NCREE 便建立了一個校舍耐震能力資料庫來收集各種相關的資料，收集了包括校舍的各種設計參數、材料強度、校舍現況及年齡、技師的評估與補強建議方案、實際補強的金額與補強方法等。收集的校舍資料數量龐大，除了當初設計的目的之外，應該還潛藏難以由人直接判斷取得的知識（knowledge）、模式（pattern）。資料探勘（Data Mining）就是用來分析這種數量龐大的資料，從中找出潛藏的知識的相關技術的統稱，本研究之目的即為利用資料探勘技術來發掘潛藏於此校舍耐震資料庫中的知識，本研究從資料探勘的四種主要分析方法：回歸、分類、分群、關聯出發，分別探討各種方法在此資料庫中何可能的分析方向，有哪些可能的潛藏知識，並進行分析，最後得到了三個有用的預測模型，分別為校舍耐震能力預測模型、校舍破壞模式預測模型以及校舍補強經費預估模型。

Data Mining on Aseismic School Building Database

Student: Kao, Wei-Ko

Advisor: Dr. Hung-Ming Cheng

ABSTRACT

In general, the aseismic ability of buildings is analyzed using nonlinear models. To obtain buildings' aseismic abilities, numerical models are constructed based on the structural configuration and material properties of buildings, and their stress responses and behaviors are simulated. This method is complex, time-consuming, and should only be conducted by professionals. In the past, soft computing techniques have been applied in the construction field to predict the particular stress responses and behaviors; however, only a few studies have been made to predict specific properties of entire buildings. In this study, a Weighted Genetic Programming system is developed to construct the relation models between the aseismic capacity of school buildings, and their basic design parameters. This is based on information from the database of school buildings, as well as information regarding the aseismic capacity of school buildings analyzed using complete nonlinear methods. This system can be further applied to predict the aseismic capacity of the school buildings.

誌謝



目 錄

論文摘要	I
Abstract	II
誌謝	III
目錄	IV
圖目錄	VII
表目錄	VIII
1 緒論	1
1.1 動機與目的	1
1.2 研究方法	3
1.3 論文架構	5
2 相關研究	6
3 校舍耐震資料庫	7
3.1 典型校舍	7
3.2 收集範圍	8
3.2.1 初步評估	9
3.2.2 詳細評估	9
3.2.3 補強設計	11
3.2.4 補強工程	11
4 資料探勘	13

4.1	資料前處理方法	16
4.2	資料探勘方法	16
4.2.1	迴歸方法	17
4.2.2	分群方法	19
4.3	探勘結果驗證方法與指標	21
4.3.1	驗證方法	21
4.3.2	結果指標	21
5	校舍資訊與耐震能力之關係模型	23
5.1	Is 值與校舍設計之關係模型	24
5.1.1	資料前處理	24
5.1.2	資料探勘	28
5.1.3	驗證	30
5.1.4	結果	31
5.2	CDR 值與校舍設計之關係模型	35
5.2.1	資料前處理	35
5.2.2	資料探勘	41
5.2.3	結果	47
5.3	D _{is} R 值與校舍設計之關係模型	51
5.3.1	資料前處理	51
5.3.2	資料探勘	51
5.3.3	結果	51

6	校舍資訊與破壞構件之關係模型	52
6.1	資料前處理	52
6.2	資料探勘	56
6.3	結果	56
7	校舍資訊與補強經費之關係模型	58
7.1	資料前處理	59
7.2	資料探勘	65
7.3	結果	66
8	結論與討論	67
8.1	結論	67
8.2	未來展望	67
	參考文獻	68
	授權書	70



圖 目 錄

圖 3.1	典型校舍	7
圖 4.1	CRISP-DM 流程圖 (c)Kenneth Jensen, CC BY-SA 3.0	14
圖 4.2	類神經網路結構圖 [1]	18



表 目 錄

表 5-1	Cross-validation result of the prediction model	33
表 5-2	Sequencing analysis of prediction of aseismic ability	35
表 6-1	Result of the Crack Component Model	56
表 7-1	補強工程參考單價之計價方式	59
表 7-2	Result of the Retrofit Cost Model	65



第 1 章 緒論

1.1 動機與目的

中小學校舍是使用人數密度極高的建築，且在國家防災的規劃上，許多的學校都是災害發生時，用來收容與暫時安置災民的重要場所，因此其建築物之可靠度相當重要，應比一般建築物有更高的要求，然而一九九九年九月二十一日發生的南投集集大地震，造成台灣將近半數校舍受損，讓學校校舍耐震能力不足的隱憂浮現，根據統計，中小學遭到損壞者共計 656 所，約佔全國學校總數的五分之一，其中南投地區更是有多數的校舍半毀或全毀，所幸地震發生時間為半夜，校舍並未在使用時間，並沒有因為校舍的受損而造成學生的傷亡。

現在中小學學校校舍的耐震性不足其問題最主要在於有很大比例的校舍屋齡已經很大，建造時使用的是建造當時的建築規範，其對於耐震能力的要求以現今的建築法規來看已經顯得不足，二來老舊校舍也會因為長期的使用以及經過數次天災、地震的影響，而使其可靠度下降，因此教育部國教司在九二一大地震後，便開始了校舍耐震能力補強計畫，並與國家實驗研究院國家地震工程研究中心（National Center for Research on Earthquake Engineering, NCREE）合作執行，目的在找出所有耐震能力有疑慮的學校校舍，請專業人士來評估現有校舍建築的耐震能力，根據評估結果來判斷是否有安全疑慮需要補強，抑或是需要拆除重建。

現有結構非線性分析模擬技術之發展已經能掌握結構系統於地震力作用下之整體反應與其細部構件之行為，使的非線性分析成為評估結構耐震力最詳細可靠的方法，美國 FEMA-273 [2] 之規範已建議使用非線性推垮分析來評估結構物於不同程度地震力作用下之破壞形式與可靠度，然而非線性推垮分析非常的耗時且非常昂貴，因其分析之正確性仰賴完整且詳細的數值模型以及專業人士的操作與結果判讀，模型的建立則同樣也是高度仰賴專業人力的耗時工作，要對全國所有校舍都進行專業的詳細評估實在是有所困難，且耗時長久，在即時地震威脅不會消失的情況下，要對全國所有的學校校舍都進行非線性推垮分析，實在是緩不濟急。

因此校舍耐震能力補強計畫是用三階段的篩選機制來快速的篩選出耐震能力有疑慮的校舍，進行詳細的耐震能力評估並進入後續的處理，這個篩選流程的第一階段為全國中小學校校舍之普查與建檔，且同時調查校舍的基本資訊，第二階段則是初步評估，是委託各個地方的專業人士，如土木、結構技師、建築師等，到各學校透過 NCREE 所設計之初步評估表格，快速的推算出校舍的耐震能力參考值，藉以判斷該校舍是否有安全疑慮，如有安全疑慮則需要進入下一個階段的評估流程。第三個階段是詳細評估，這個階段使用的評估方法就是非線性推垮分析，由專業人士去現地調查，建立出完整詳細的數值模型，之後使用非線性推垮分析方法來模擬校舍受到地震力之行為，評估校舍的耐震能力，最後才根據評估結果決定校舍是否需要補強，如果需要補強，則進入到之後的補強流程。

在這個計畫執行的過程中，不同階段的評估都會產生大量的校舍相關資料，例如初步評估會有校舍的基本結構參數：長度、深度、樓層數、梁柱之尺寸及數量、樓地板面積以及校舍現狀等等資料，詳細評估則會有更詳細的如材料強度、優先破壞的構件、破壞地表加速度等等，後續的補強流程還會產生如補強工法、不同工法的補強量、補強經費等等資料，數量龐大，因此國家地震工程研究院便建置了一個校舍耐震資料庫，收集此計畫執行間產生的各種校舍資料，目前資料庫收集有全台灣兩萬多棟校舍的設計、評估與工程相關資料，其主要用途雖為輔助校舍耐震能力補強計畫，然而此一大量的資料，應當還可以從中挖掘出難以由人工觀察判讀的隱含知識。

資料探勘 (Data mining) 此一研究領域的發展是為了因應資料庫系統以及資料倉儲系統的發展、資料量的急遽成長以及越來越複雜的資料性質，因而越來越難從收集的資料中獲取有用的知識的情形。資料探勘的方法包括統計、線上分析處理 (OLAP、on-line analytical processing)、情報檢索 (information retrieval)、機器學習 (machine learning)、模式識別 (pattern recognition) 等，由前段敘述可以得知，校舍耐震資料庫內的資料量非常多，不只校舍數量龐大，收集的資料屬性也非常多，其中隱含的知識難以直接由人眼觀察取得，如果可以使用資料探勘技術，從其各種分析方法的特性出發，配合各種實務上的需求，應當可以從此資料庫中找出部分隱含的校舍建物知識，不過資料探勘分析需要每筆資料都能夠用相

同的形式，並且用固定數量且有限的資料屬性來描述資料實體的特性。由於不同建築物的結構差異可能很大，沒有一個標準的形式可以只使用有限的資料屬性就描述所有的建築物，因此以往的研究均難以對大量的建築物資料進行資料探勘分析，不過校舍建築中，有很大的比例有相似的結構形式，這些校舍都為一字形，隔間為一間一間連著，外面有走廊，樓梯間、廁所通常在末端，樓層數不超過五層樓，有些校舍雖然非一字形，較為複雜可能是 L 字形或是 U 字形，但是也可拆分為數個一字形形式的校舍，而由於有這些常見的形式，可以把校舍建築的資料屬性特徵化，用少量的資料屬性就可以正確的描述這種常見形式的校舍建築，資料屬性數量不會隨著建築物規模的擴大而增加，這樣的資料形式讓大量校舍建築物的資料探勘分析成為可能，而 NCREE 所建立的校舍耐震資料庫中收集的校舍資料即為使用這種形式來描述校舍設計參數之資料，因此本研究之研究目的即為基於校舍的建築形式以及已經收集大量校舍資料的校舍耐震資料庫，利用各種資料探勘方法來分析並尋找此意資料庫中，難以人工觀察判讀的隱含知識。



1.2 研究方法

本研究之研究方法可以分為三個階段，第一個階段為分析規劃階段，此一階段的主要目標為假設各種可能的隱含知識，並且定義出不同隱含知識的探勘方式，Fayyad [3] 依照不同資料探勘技術之特性，分出迴歸、分類、分群、關聯四大類，本研究的第一階段即根據此四種知識形式，以及校舍耐震補強計畫的執行流程與需求，假設並定出各種可能透過資料探勘技術取得的校舍耐震資料庫隱含的知識以及探勘方法，其中，迴歸形式的可能取得知識包括了校舍耐震能力預測、校舍破壞模式預測、校舍補強經費預測等，分類形式的可能知識包括了校舍是否需要補強的預測，分群形式的知識則是校舍的類型歸類條件，關聯式法則形式的可能知識則是校舍設計參數與其現狀的關連性。

第二階段則是根據假設的各種隱含知識和資料探勘規劃，實際進行資料探勘的分析和測試，最後的第三階段則是探勘結果的驗證和隱含知識的整理，基於此一流程，本研究最後得到了三個有一定可靠度的校舍耐震資料庫的隱含知識，分

別為：

- 校舍資訊與校舍設計關係模型
- 校舍資訊與破壞構件之關係模型
- 校舍資訊與補強經費之關係模型

校舍耐震能力預測模型為本研究最主要的資料探勘目標，因為校舍耐震能力補強計畫當中，最重要的資訊就是校舍的耐震能力，傳統上，如果要取得可靠的校舍耐震能力，需要由專業的技師來評估，其過程需要先到現場調查，根據調查的資續建立完整的結構數值模型，並使用非線性推垮分析，其過程耗時且所費不貲，因此現在校舍耐震能力補強計畫是以分階段篩選的機制，先讓所有校舍進行一個較為簡單的初步評估，再根據初步評估的結果來決定哪些校舍的耐震能力可能比較不足夠，需要詳細的非線性分析，才真的對這些校舍進行詳細的非線性分析與耐震能力評估，然而這種方法有個缺點是其初步的評估方法無法完全反映出校舍的耐震能力，可能有校舍已經因為年代久遠造成耐震能力低落，然而卻無法在初步評估的結果中真實的反映出來，因此，如果有一個方法可以快速的得到更為可靠的評估數據，甚至可以當作詳細評估的參考，將可以大大的加速校舍耐震能力補強計畫的進行。

除了數值化的校舍耐震能力，本研究還建立一個模型，可以對校舍受到地震力時，優先破壞的構件進行預測，這個資訊可以幫助對校舍進行耐震能力評估的專業技師對目標的校舍弱點先有一些初步了解，不但可以協助詳細評估的進行，對於校舍補強設計的方式也有一定程度的幫助。

最後，由於校舍補強所需的經費龐大，因此校舍耐震能力補強計畫不可能在短期內就把所有耐震能力不足的校舍都完成補強，實務上會需要估算各個校舍補強所需的經費，排定預算，然後才知道不同預算年度能夠完成多少的校舍補強作業，因此校舍的補強經費在校舍補強計畫的決策中，是一個非常重要的數字，傳統的經費預估方法是由過往的經驗、數據和所欲補強校舍的規模作為依據，經由一些推估和統計所計算出來的，如果能夠建立一個預測模型，經由校舍的基本資

料就可以得到準確的補強經費預測值，那便可以大大的加速校舍耐震能力補強計畫決策者的決策速度，也可以讓計畫執行人員能更快的了解補強作業的規模。

1.3 論文架構

本論文共分為八章，各章內容分別介紹如下：

1. 緒論：說明本研究之動機及目的。
2. 相關研究：回顧與本研究相關之文獻。
3. 校舍耐震資料庫：介紹校舍耐震資料庫之架構與其所收集資之資料。
4. 資料探勘：介紹資料探勘技術以及校舍耐震資料庫之資料探勘規劃，另外還介紹本研究使用到的各種演算法。
5. 耐震能力與校舍設計關係模型：詳述本研究第一個使用資料探勘方法找到的隱含知識，耐震能力與校舍設計參數間的關係模型。
6. 校舍設計、現況與破壞構件之關係模型：詳述使用資料探勘技術找出校舍現況、設計參屬等屬性與其遇到地震力時，可能先受力破壞構件之關係模型之方法與過程
7. 校舍設計、現況與補強經費之關係模型：詳述使用資料探勘技術找出校舍現況、設計參數等屬性與其可能需要之補強經費間關係模型之方法與過程
8. 結論：結果探討與未來展望。

第 2 章 相關研究



第 3 章 校舍耐震資料庫

學校是人才培育的場所，也是緊急災難時，居民避難的主要地方，但台灣地區學校建築在每次大地震來時之損壞卻非常嚴重，尤其是老舊校舍，因興建年代久遠，其設計所依據之規範較為老舊，耐震能力可能遠低於現今結構耐震安全上之要求，故教育部已委託國家地震工程研究中心進行全國學校校舍之耐震能力評估與補強研究，此計畫建立了校舍建築耐震能力評估補強機制及施行的流程與細節，並且已經對全國學校校舍耐震能力作了全面性的普查，篩選出耐震能力有疑慮之校舍，並儘速透過補強或拆除新建的手段來提昇校舍的耐震能力。而此計畫進行期間所產生之資料，均收集到此一校舍耐震資料庫中。

3.1 典型校舍



我國之校舍建築有極大比例在校舍的結構形式、幾何尺寸等都有相似的結構，其平面配置多如圖 3.1，這類校舍為一字型的長形建築，教室一間一間排列，教室外有走廊，走廊多有柱，且樓層數多不超過五樓，此一類型之校舍統稱為典型校舍，典型校舍在受地震力時，通常破壞樓層都在一樓，而且是沿著長向破壞。而因為其結構形式單純，只需要少量的屬性便可以完整的描述建築物的結構，不用完整詳細的記錄所有的樑柱等構件之個別尺寸強度與位置。

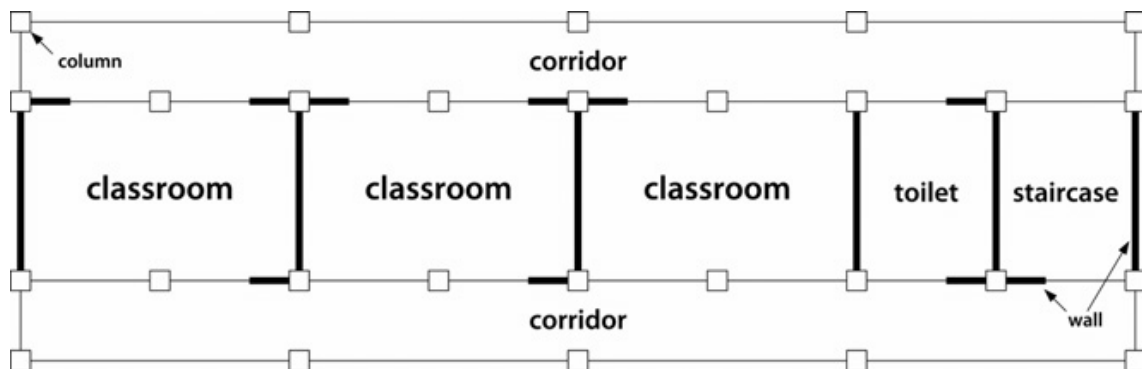


圖 3.1: 典型校舍

3.2 收集範圍

此一校舍耐震資料庫中收集有台灣各級學校，不同詳細程度的耐震能力相關的調查資料，此一資料庫乃是由國家地震工程研究中心在執行教育部委託之「全國各級學校舍耐震能力補強計畫」時所建。而除了耐震能力相關的結構評估資料，本資料庫還收集了評估階段之後的，包括補強設計、補強工程發包與結案的相關資料。

結構耐震能力評估的方法可分為兩類，第一類為初步評估（preliminary evaluation），其通常的方法是基於結構物之設計及現況填寫評估表，所填寫資料再依評估公式計算出結構物耐震能力之評分等級或指數，此類方法之評估速度快，但是結果可靠度較低，主要的目的是對大量結構物之耐震能力作排序與篩選，這類型評估方法所使用之評估表與評估公式通常是使用已經收集的其他結構物資訊，用數值統計的方法迴歸，或是根據基本的結構耐震能力供需比以及專業人士相關的經驗設計得到的，其所適用之結構物類型也有所限制，評估表並不能套用到各種類型的結構物上；另一類評估方法為詳細評估（detailed evaluation），此類方法為對結構物進行詳細的結構耐震分析，通常是使用結構分析程式以電腦數值模擬結構物遇到地震時的非現性行為，並根據模擬結果為依據，準確詳細的檢驗評估出結構物的耐震能力，這類方法的可靠度較高，但是花費相當昂貴，所需要的時間也比使用評估表要來的久很多。此二類耐震評估的方式通常相輔相成，結合成為標準的結構耐震能力評估程序，其即是先對所有需評估之結構物先以初步評估作耐震能力之評分與排序，以篩選出其中耐震能力有疑慮者，再對其以詳細評估的方式做詳細的檢驗。

由於中小學校舍數量龐大，若直接大量投入人力物力，可能造成大量之資源浪費，也無法快速的鎖定耐震能力不足之校舍建築，故針對有效達成此一校舍耐震評估標準需求以及基於上述標準耐震評估程序之精神，國家地震工程研究中心提出之解決程序為；經由學校總務人員之簡易調查及工程專業人員之初步評估，有效的將校舍結構之耐震能力排序，以縮小問題之規模，對於耐震堪慮之校舍，依嚴重程度，由工程專業人員，進行結構耐震之詳細評估，倘尚符合補強之經濟

效益，即進一步作耐震補強之設計，若不符合補強之經濟效益，則將之列為拆除重建。在此過程中會產生的校舍資料最重要的為：初步評估、詳細評估、補強設計、補強工程等四組資料。

3.2.1 初步評估

ask 趙哥

初步評估相較於詳細評估的模擬運算而言，是簡化許多的耐震能力評估方式，評估的方式是使用地震中心根據過往資料與相關規範所設計而得，其設計的基礎與詳細評估的 CDR 值相同，皆為校舍的耐震能力與耐震需求的比值，主要的對象為典型校舍，假設校舍之破壞位置為長向的一樓，耐震需求為則為建築物需要能承受其所在地所可能發生的 475 年週期最大地震之地震力，由推估的校舍載重以及地震之地表加速度推估而得。而耐震能力則是根據校舍之柱、牆等尺寸，並根據實驗數據推估所得到的等效強度，兩者相除後再乘上調查人員根據校舍現況所填入之調整參數 Q ，所得到的 I_s 即為使用初步評估表格所評估之校舍耐震能力。

3.2.2 詳細評估

美國針對鋼筋混凝土（RC）建築物之耐震能力規範 ATC-40 [4] 中，建議用來評估建築物耐震能力之方法稱為容量震譜法（capacity spectrum method），此一評估方法可以分為兩個部分，第一部分是進行側推分析（pushover analysis）取得容量震譜（capacity spectrum），流程如 Figure 1 所示，第二部分是根據建築物所在地的各種相關資訊和規範取得需求震譜（demand spectrum），接著使用容量震譜法（capacity spectrum method）以取得建築物性能點，根據性能點（performance point）座標帶入公式計算可以得到建築物的破壞地表加速度（collapse ground acceleration）以及耐震能力指標（aseismic ability index），如 Figure 2 所示。

求取容量震譜的側推分析（pushover analysis）需要先建立建築物完整的非線

性分析模型 (nonlinear structure model)，此模型要由進行評估的專業人士根據建築物的幾何尺寸資訊來建立基本的建築物框架。接著根據建築物的材質計算建築物自身重量以及需要承載的人員和配設物件之重量，並將其附加到柱、牆等承載元件上。接著最重要的是設定建築物各構成元件如樑、柱、牆受力時的非線性行為，這些力與變形的關係多為實驗得到或是利用其他分析模擬方法得到的。至此，側推分析 (pushover analysis) 要用的數值模型才算建立完成。而側推分析 (pushover analysis) 其程序是依照規範，使用建議之側力，逐步地給結構模型施加增量外力，將結構物向某一方向推動，每次增量即進行一次結構分析計算每一構件之應力 (stress) 與變位 (strain)，之後與上一次的分析結果累加即可得到每一個構件於此受力階段之反應，並判斷構件是否破壞，例如開裂 (crack)、降伏 (yield)，甚至達到極限強度 (ultimate force) 等，之後各構件依其破壞程度更新其行為，例如改變其勁度，或者將已發生破壞的構件從結構模型中抽離，如此重複分析直到結構不穩定而崩垮為止，側推分析完成後可以得到建築物之容量震譜 (capacity spectrum)。

由於典型校舍通常的破壞模式為沿著校舍的長向破壞，也就是走廊的方向，因此側推分析所施加的側力也就是沿著長向施加，因此長向也被稱為 X 向，另外一個方向，也就是校舍較短的那一邊，則是 Y 向，而施加力量的方向則是正向，和力量相反的方向則是反向。

需求震譜 (demand spectrum) 是根據建築物環境現況，參考規範製成之地震需求頻譜。參考的環境狀況如土層種類，堅硬的土層可以讓建築物有較好的抗震能力。另外靠近斷層的建築物在地震發生時往往會因為斷層於地震時產生的反應而受到嚴重的損害，因此建築物與斷層的距離也納入參考的資料之一，稱為近斷層效應 (special effects of near-source earthquakes)，其他還需要的資料有地層資料、地震震區等，綜合這些資料並參考規範即可得到該建築物所需符合之需求震譜。

容量震譜法 (capacity spectrum method) 最後是將需求震譜 (demand spectrum) 以及容量震譜 (capacity spectrum) 轉換格式後相疊，求取建築物之性能點 (performance point)，性能點之物理意義為該建築物在特定水準地震下所能承受之最大變位 (strain) 及剪力 (shear force)，但是由於結構物受地震力作用進入非線

性行為時，結構物的阻尼效應會產生消能的作用，因此還要視情況對需求曲線進行折減，此為一個迭帶運算的過程，可能需要不斷折減需求曲線，並進行性能點（performance point）正確性的檢核，直到檢核通過，才能得到建築物真正之性能點（performance point）如 Figure 2(d)，由建築物性能點（performance point）之座標帶入公式可以求得建築物的破壞地表加速度（collapse ground acceleration） AC ，詳細評估所使用的耐震能力值為 CDR ，其定義如下：

$$CDR = \frac{AC}{AD} \quad (3-1)$$

AD 為根據建築物所在地的並依據規範所得到的，建築物所需要能承受的 475 年週期最大地震所產生之地表加速度， CDR 大於 1 代表此一棟建築物能夠承受該地區 475 年一遇的最大地震。



3.2.3 補強設計

ask 趙哥

如果校舍經過詳細評估過後，負責的專業人士認為此棟校舍確實有安全疑慮，但尚可以補強，不需要拆除重建時，則該棟校舍需要進入補強設計，補強設計階段中，負責的專業人士需要依據詳細評估的結果，設計出兩種不同的補強方案，並且對不同的補強方案進行詳細的耐震能力評估，因此這個階段的資料包括兩組補強設計的資料，以及兩組補強後的耐震能力詳細評估資料。補強設計收集的資料主要為使用的補強工法以及補強量，主要收集的補強工法包括了：擴柱、剪力牆等，

3.2.4 補強工程

ask 趙哥

補強工程部分的資料則包括了實際補強工程的相關資料，其中包括了採用的

補強設計，工程發包的相關資料，例如工程經費、發包時間、驗收報告等等。



第 4 章 資料探勘

過往在各種不同商業、工業、學術等領域都隨著產業逐步的數位化，有各種資料庫甚至資料倉儲系統的建置與資料收集，例如零售業的交易紀錄，各種學術實驗的結果記錄等等，而隨著時間推移，這些資料庫和倉儲系統收集的資料數量都成長的非常龐大，資料間的關係和複雜度也越來越複雜，這些資料除了當初建置的目的和資料的統計分析外，研究人員還開始思考更進一步的資料應用，希望能把資料中以往隱藏不可見的知識找出來，因此資料探勘（Data Mining）此一研究領域就因應而生。資料探勘的方法包括統計、線上分析處理（OLAP、on-line analytical processing）、情報檢索（information retrieval）、機器學習（machine learning）、模式識別（pattern recognition）等。

資料探勘的流程也有多種規範，其中最廣為被使用的是 CRISP-DM [5]（Cross Industry Standard Process for Data Mining），是由 SPSS 以及 NCR 兩大廠商在 1990 年開始發展的，它的流程架構如圖 4.1 所示，將一個完整的資料探勘流程分為六個步驟，分別為：

1. **定義領域問題**（Business Understanding），CRISP-DM 所定義的資料探勘最初的步驟是瞭解並定義此一資料探勘所希望解決的問題。
2. **定義分析資料**（Data Understanding），瞭解問題之後，就要定義要解答此一問題所需要的資料為何，並且著手收集。
3. **資料前處理**（Data Preparation），這個步驟包括了資料的由於現實世界的資料都會有很多雜訊在其中，因此在真正的訓練並建立模型之前，需要先把雜訊去除，包括了不合理的資料、多餘欄位，甚至是透過一些方法來凸顯資料中本來較不明顯的特定性質。
4. **建立模型**（Modeling），選擇適合該問題的資料探勘方法對前處理過的資料進行分析與模型建立。

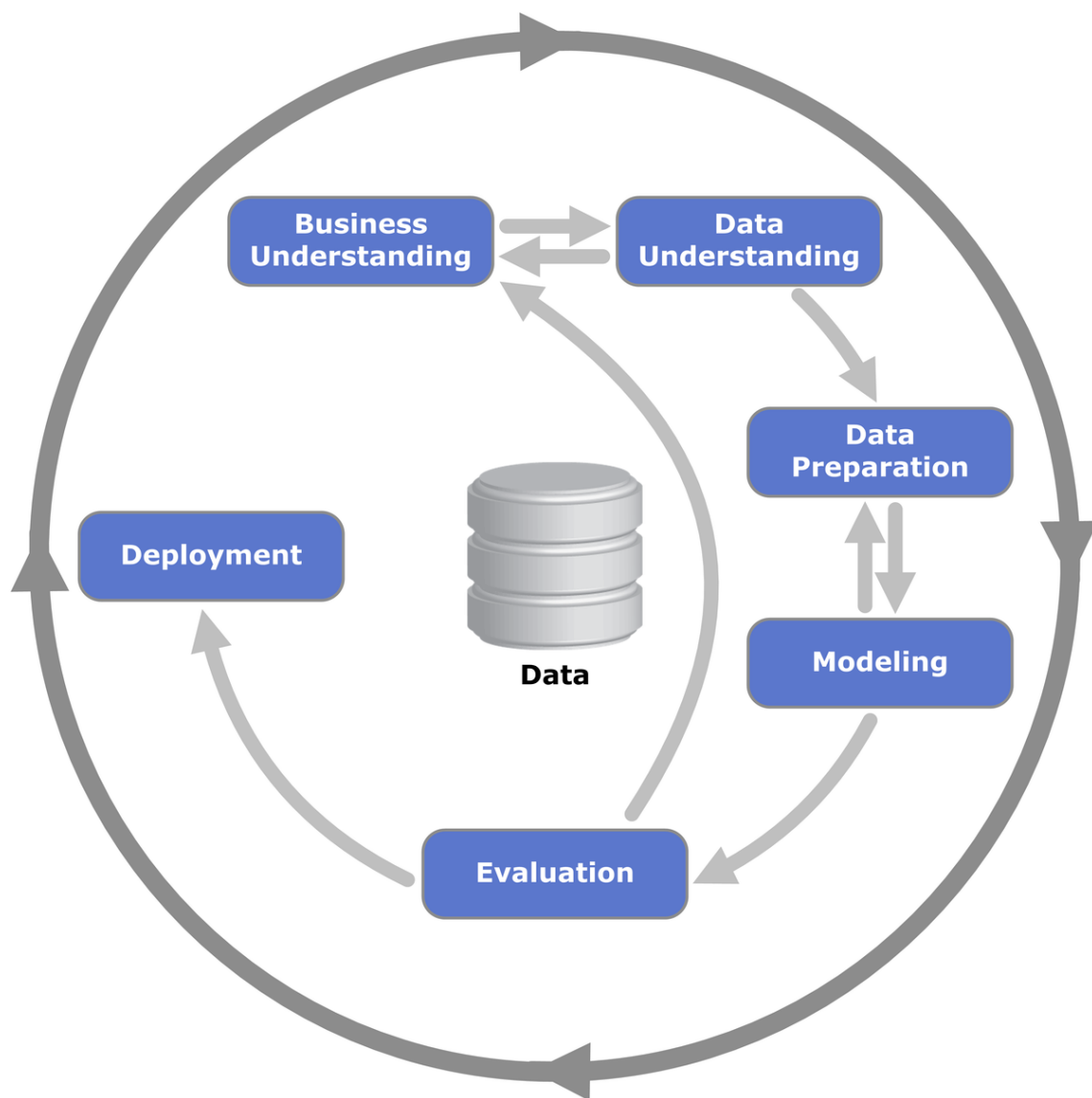


圖 4.1: CRISP-DM 流程圖 (c)Kenneth Jensen, CC BY-SA 3.0

5. **評估模型 (Evaluation)**，評估前一步驟中所建立的模型品質是否符合需，多是將資料分為訓練集與測試集兩組來驗證模型的可靠度。
6. **應用模型 (Deployment)**，將模型產出的知識實際納入應用，抑或是將探勘結果整理成完整的報告。

本研究之流程較 CRISP-DM 之流程稍有不同，是先基於一個現有的特定領域的校舍耐震資料庫作分析，找出校舍耐震資料庫中各種潛藏知識的可能性，定義出希望能夠解決的問題，接著分析不同的問題，從校舍耐震資料庫中挑選出相關的資料屬性，然後接著進行資料前處理、建立模型、評估模型幾個步驟。

資料探勘技術方法繁多，Fayyad [3] 根據其處理的問題形式，將資料探勘的方法分為分類、分群、迴歸尋找關聯等四種主要的問題類型，分類方法處理的問題是在用來判斷資料的類別，而且這些類別是已知的類別，例如將所有的校舍資料分類成有安全疑慮和沒有安全疑慮的就是屬於分類問題。分群問題和分類問題有點相似，一樣是將資料分成數個群組，最主要的差異是分群問題的各個群組特性一開始並不清楚，分群方法是將資料根據其屬性數值為依據，把相似的放在同一個群組，不同群組的特性是要在分出群組後進行分析才會得到。迴歸問題就是要用迴歸方法來從資料的屬性中，找出特定屬性與其他屬性間的關係模型，這些屬性間的關係可能是非線性的，而且沒有解析解的關係模型，因此常見的方法是用統計迴歸的方式，用現有的資料來迴歸得到，又或著是用像類神經網路之類的機器學習方式，拿現有的資料下去學習已得到關係模型，以校舍耐震資料庫來說，校舍耐震能力指標的預測就是一種迴歸問題，因為校舍耐震能力指標與其校舍的設計參數間的關係就是一個非線性關係，要得到兩者之間的非線性模型就需要用到迴歸問題的處理方法，迴歸問題也是最常見的資料探勘問題種類。最後一種是尋找屬性間的關聯，這種問題的主要目標在尋找不同筆資料屬性間所存在的關係，舉例來說，使用校舍耐震資料庫的資料來作關聯分析，可能可以去尋找像是：五層樓的校舍的校舍長度深度有什麼趨勢，或是民國八十到九十年之間的校舍的校舍走廊設計是否偏好有走廊柱等。本研究在確定主要的探勘目標後，使用的資料探勘方式為迴歸為主，分類分群為輔助。

以下分別介紹本研究所使用到的各種分析方法，除建立模型的各種訓練和演算法外，還包括資料前處理和驗證所使用的分析方法和驗證指標。

4.1 資料前處理方法

資料前處理方法常見的目的有：找出重要性較高的屬性、凸顯資料特性、剔除特異資料點等；本研究主要的資料過濾方法為資料的合理性分析，因此使用的前處理方法較少，只有主成分分析法。

主成分分析

PCA, a very common data preparation method, can identify very important attributes among various attributes. The goal is to convert the original variables through vector transition into mutually independent variables of a linear combination. The ideal situation is that principal components obtained from linear combination retain most of the information of original variables.

4.2 資料探勘方法

資料探勘的方法分為分類、分群、迴歸尋找關聯 [3]，其中迴歸是最常使用到的一種，本研究的主要目標皆可以歸納為迴歸類的問題，因此使用到的迴歸演算法最多，接著才是作為輔助用的分類和分群演算法。

資料探勘所建立的模型可以簡單的定義為：

$$y = f(x) \quad (4-1)$$

$f(.)$ 就是透過資料探勘分析學習而得到的模型函數， x 代表輸入參數，輸入參數是為各種已知的屬性資料，例如可以簡單透過量測得到的校舍尺寸資訊，

y 則是代表輸出參數，也就是希望透過這個模型所得到的，比較難取得的屬性資料，例如需要經過詳細評估才能得到的校舍 CDR 值。

4.2.1 迴歸方法

Generalized Linear Model

廣義線性模型是由 Nelder and Wedderburn [6] 所提出，比起迴歸分析（simple regression）更為彈性，此模型是假設資料點的分佈有一分佈模式，且輸入參數 x 與輸出參數 y 之間的關係是由一連結函數（Link Function）建立，如 log function、power function 等，其定義之 x, y 間之關係模型如下：

$$g(E(y)) = x\beta + O, y \sim F \quad (4-2)$$

$g(\cdot)$ 是為所選的鏈結函數， O 是偏移（offset）變數， F 則是 y 的分佈模型，其是用牛頓法（Newton-Raphson Method）不斷的調整 β 使的 $x\beta + O$ 逼近 $g(E(y))$ ，最後最接近的方程式即為 x, y 兩者的關係式。比起迴歸分析，此方法還需要了解 y 值分佈狀況，選擇出最適合的分佈函數，並假設 x, y 間的鏈結函數形式，雖然越多的參數選擇代表了更多的模型不確定性，但廣義線性模型卻能夠提供比迴歸分析更廣的應用範圍，也可能得到更接近真實的關係模型。

Support Vector Machine

SVM 最早是 Boser [7] 等人，在 1992 年的 COLT（Computational Learning Theory）所提出，SVM 是一個基於統計學習理論的分類方法，用來處理二元分割的問題，其原理是將原本無法線性分割的問題轉換到一個不同維度的空間（kernel）後，假設該空間存在一超平面（hyperplane），可以正確的將資料分開，並將尋找此一超平面的問題轉換為一最佳化問題，求解後即可得到二元分割邊界的方程式。而後 Harris Drucker, et. al., [8] 將此二元分割問題轉換為迴歸分析問題，

故 SVM 也可以處理迴歸問題。

類神經網路

類神經網路 (Artificial Neural Networks)，其是希望能模擬建構出人腦內的神經網路，以處理各種複雜的問題，人類大腦是由大約千兆個神經元 (Neuron) 所構成，而每個神經元又會和其他約一萬個神經元連結，構成一個龐大且複雜的神經網路，這樣複雜的一個神經網路讓人類可以學習並了解各種事物與知識。McCulloch and Pitts [9] 所提出的模型為後續類神經網路發展的雛形，一個標準的類神經網路可以分為輸入層 (input layer)、隱藏層 (hidden layer)、輸出層 (output layer)，其結構如圖 4.2，輸入層負責接受各種求解問題需要的量化數據和資料，經由隱藏層的不斷自我更新學習的模型處理過後，在輸出層就可以得到想要的解答，類神經網路可以處理的問題種類多樣，其模型的品質多數也都不錯，缺點是學習時間長，且得到的模型為一個黑盒子，難以解釋其物理或是數學模型上的意義。

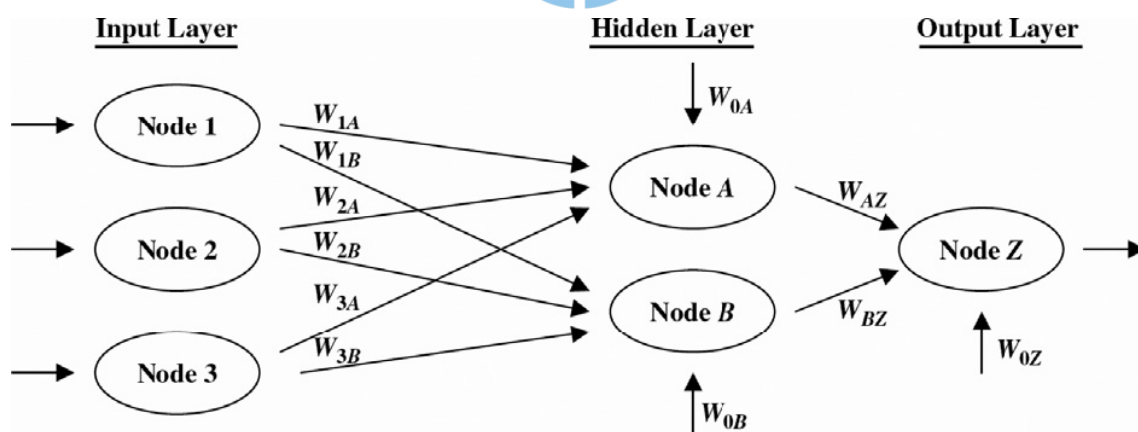


圖 4.2: 類神經網路結構圖 [1]

Genetic Programming

Weighted Genetic Programming

4.2.2 分群方法

兩步分群

modify required

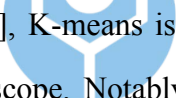
由於各種學術、商業研究所處理的問題資料量越來越大，分群演算法所需要的運算時間也急遽的成長，為了讓分群演算在處理大量的資料時也能有良好的效能表現，而發展出兩步分群（Two-Step Clustering），其是由 Zhang, Ramakrishnan 與 Livny [10] 所提出，此一方法分為兩個主要步驟：第一步是先把資料依據其與相鄰資料的相適度來排序並分成數個小群集，接著第二步再使用階層式分群方法，將這些群集慢慢組合或拆分直到達成停止條件。兩步分群的演算費時和資料量呈線性關係，因此資料量成長時也不會讓運算時間成長到無法應用的程度。

Based on of the massive volume of basic data for school buildings in the database, this study chooses two-step clustering method. The basic concept was first proposed by Zhang, Ramakrishnan and Livny [10] for handling large amounts of data. This method has two major steps. The first step sequences data and pre-clusters sequences into small subclusters based on the similarity of adjacent data, thereby reducing the amount of data. The second step divides several small subclusters into the desired number of clusters using a hierarchical clustering method. The hierarchical clustering method then combines close subclusters slowly until the stop condition is met. The computing speed of this method is influenced slightly by the volume of data.

K-means

K-means 是由 MacQueen [11] 所提出，也是最常被使用的分群演算法之一，屬於機器學習方法，其主要步驟為：

1. 使用者決定要將資料分為幾個群集，群集數為 K 。
2. 將資料隨機分到 K 個群組，並且計算出此時分群狀態下的每個群組的中心點。
3. 講每個資料點重新分群到，選擇群組中心離它最近的那個群組。
4. 用新的分群狀態計算出新的群組中心點。
5. 重複步驟 3 和 4 直到每個資料的群組都穩定不再有變化。



As proposed by MacQueen [11], K-means is one of the most common clustering methods and has a wide application scope. Notably, it is a machine learning method; its principal steps are as follows.

1. A user indicates that data should be grouped into K clusters.
2. Data are divided randomly and equally into K groups and the center of each cluster are calculated.
3. Each bit of data should find the proximal center of a cluster and update its cluster label that it belongs to.
4. Recalculate the new cluster center.
5. Repeat steps 3 to 4 until the cluster centers of all data do not change.

4.3 探勘結果驗證方法與指標

4.3.1 驗證方法

10 fold cross validation

4.3.2 結果指標

決定係數 Coefficient of determination

決定係數（Coefficient of determination）又稱為 R^2 ，其公式為：

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = \frac{SS_{reg}}{SS_{tot}} = \frac{\sum (\hat{y}_i - \tilde{y})^2}{\sum (y_i - \tilde{y})^2} \quad (4-3)$$

其中 y_i 是關係模型輸出屬性之值， \hat{y}_i 則是該輸出屬性之實際值， \tilde{y} 則是所有資料的輸出屬性實際值之平均，透過此一指標可以了解輸出屬性中有多少比例的資訊是由輸入屬性的變量所產生的，也代表著關係模型的正確性，而其值恰巧為關係係數 R （correlation coefficient）的平方，透過關係係數可以了解實際的輸出屬性數值與透過關係模型得到的推估值之間的線性關係，線性關係越高表示兩者之間越接近，也代表著關係模型所建立關係之正確性。

平均絕對百分比誤差 MAPE

MAPE (Mean Absolute Percentage Error) 的公式如下：

$$MAPE = \frac{\sum |y_i - \hat{y}_i|}{N} \times 100\% \quad (4-4)$$

其中 N 是資料的總數， y_i 是使用探勘得到的關係模型所求得的輸出屬性預測值， \hat{y}_i 則是該屬性的實際值，此一指標代表了模型產出結果的平均誤差，可以呈現模型的準確度，數值越低代表模型品質越好。

均方根誤差 RMSE

均方根誤差（Root Mean Squared Error，RMSE）定義如下：

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{N}} \quad (4-5)$$

其中 N 是資料的總數， y 是透過資料所建立的關係模型所求得的輸出屬性預測值， \hat{y} 則是輸出屬性的實際值，此一指標代表了模型產出結果的平均誤差，數值越低越好，和 MAPE 相比，其差異在 MAPE 指表現了模型輸出數值的誤差平均，而 RMSE 還包含了誤差量的離散度資訊在內，MAPE 表現相同的模型，其單筆資料誤差值分布越離散，RMSE 的表現會越差，其可接受範圍則要根據輸出屬性的數量級和問題複雜度而定。



命中率 Hit Rate

命中率（Hit Rate）是用來判斷關係模型的正確率的，其定義為：

$$\text{Hit Rate} = \frac{\sum I\{(1 - \alpha)y_i \leq \hat{y}_i \leq (1 + \alpha)y_i\}}{N} \quad (4-6)$$

其中 $I\{L\} \in \{0, 1\}$ ，如果 L 為真，則 $I\{L\}$ 為 1，反之則為 0， y 是透過資料所建立的關係模型所求得的輸出屬性預測值， \hat{y} 則是輸出屬性的實際值，而 α 為命中率的容許誤差，且 $0 \leq \alpha \leq 1$ ，如果 $\alpha = 0.1$ 則表示誤差 10% 內都算是有預測模型有預測命中。Hit Rate 越高代表關係模型的結果越好。

第 5 章 校舍資訊與耐震能力之關係模型

在校舍耐震資料庫中，不同階段的調查資料有不同的校舍耐震能力指標，這些指標都是以數值形式來量化校舍建築物的耐震能力，有了此一數值後，可以快速找出有安全疑慮的校舍，根據耐震能力排序，甚至推算會需要多少預算來補強多少棟校舍等等，都可以使用此數值作為依據來達成，可以說是非常重要的校舍特性，然而要取得此一數值非常耗時耗力，如果有其他快速便宜的方法可以取得此一數值，那便可以大幅度的減少校舍耐震能力補強作業的所需要的時間與經費，因此本研究的第一個資料探勘目標，便是找出預測耐震能力索引的預測模型，得到此一預測模型後，便可以根據校舍的設計參數與現況作為輸入參數，快速的得到該校舍的耐震能力索引參考值。

初步評估階段資料的耐震能力指標是 I_s 值，此一數值為專業人士於現場良策、調查後，使用國家地震中心的專家根據過往的實驗數據統計分析後，所設計出的一個評估表計算而得到的，為一初步的評估值，較詳細評估所計算的耐震能力指標可靠度來的低，但仍是校舍補強計畫初期的篩選工作中非常重要的數值。詳細評估和補強設計階段的耐震能力指標則是 CDR 值， CDR 值是由專業人士根據校舍的設計與實際狀況建製結構模型，並進行非線性的推垮分析所得到的，此數值與建築物之設計參數為高度非線性的關係，也是最接近校舍建築物實際耐震能力的量化指標。

而除了數值量化的耐震能力指標外，本研究還將「校舍是否需要補強」(D_{isR}) 這個二元指標作為耐震能力來作為校舍耐震資料庫中的第三個耐震能力指標， I_s 、 CDR 、 D_{isR} 這三個耐震能力指標的預測模型，即為本研究所取得的第一個校舍耐震資料庫隱含知識，以下便針對不同指標的預測模型的資料探勘流程與結果詳細介紹。

5.1 I_s 值與校舍設計之關係模型

校舍耐震資料庫中的初步評估資料表的耐震能力索引是為 I_s 值，初步評估是校舍耐震能力補強作業中，很重要的篩選過程，因為建築物的詳細評估所需要的金額極高，評估需要的時間也很長久，因此需要一個速度、價錢和可靠度都可以接受的評估方法，用比較短的時間找出耐震能力有疑慮的校舍，盡早處理。而地震中心所設計的初步評估表即為一個速度、價錢和可靠度都可以接受的評估方法，只要根據建築物的設計參數，像是樓層數、長度、深度、柱子尺寸以及校舍現況等，就可以快速的得到一個極具參考價值的耐震能力指標，而其計算的原理也與詳細評估的耐震能力指標索引 CDR 值一樣，為耐震容量與耐震需求的比。找出 I_s 值與校舍設計參數之間的關係模型，可以進一步分析初步評估表中各種數據的重要性，並將結果回饋到此一初步評估表，將初步評估表調整的更為精簡，負責評估的人員也可以更快速的完成此意評估表格。

The proposed prediction model, based on data-mining technology, has six steps – business understanding, data understanding, data preparation, modeling, evaluation, and deployment – as recommended in the cross industry standard process for data mining (CRISP-DM) proposed by Chapman et al. (2000) (Fig. 9). The objective of this work is to construct a prediction model for the aseismic ability indices of school buildings and collect complete data. Hence, the major tasks are data preparation, modeling, and evaluation. Fig. 10 shows the procedural plan based on needs and the methods chosen.

5.1.1 資料前處理

首先，從國震中心的校舍耐震資料庫中，將各棟校舍的 I_s 值與其校舍的相關資料挑出並整合在一起，第一步是過濾掉明顯不合理的錯誤資料，這些資料通常是人為的錯誤造成的，而過濾的方法則是使用國陣中心建議的過濾條件：

- 校舍總長或總深度超過兩百公尺
- 有任意牆厚度超過五十公分

- 有任意柱的長或寬超過一百公分
- 柱間垮距超過八公尺或小於兩公尺

大約有八百筆資料符合這組基本的過濾條件，而在初步的過濾之後，前處理的第二個步驟則是減少資料屬性的維度，主成分分析（Principal component analysis）是很常使用的資料前處理方法，這個分析方法可以找出所有資料屬性中，重要度較高的屬性，它是把原始的多個資料屬性，透過向量轉換的方式，線性組合出主要的資料屬性。

Roughly 800 bits of valid data are obtained. After filtering using these conditions. Principal component analysis (PCA) is then used to reduce the dimensionality of data attributes. Notably, PCA, a very common data preparation method, can identify very important attributes among various attributes. The goal is to convert the original variables through vector transition into mutually independent variables of a linear combination. The ideal situation is that principal components obtained from linear combination retain most of the information of original variables.

校舍的分群則是基於校舍的設計模式，目標是要找出校舍建築物的幾種特定設計模式，也就是要分析校舍的幾何設計參數、柱量、牆量等等，經過不斷的測試及調整，最後找出了五個主要成分屬性：

- 走廊柱資訊
- 教室柱資訊
- 強設計資訊
- 牆量資訊
- 柱量資訊

Clustering analysis of school buildings is based on the design patterns of school buildings; the goal is to find several design patterns. Hence, the attributes for analysis are the

geometric information of school buildings, such as dimensions and quantity of walls and columns, and width and height of school buildings. Other attributes, such as year of construction and locale, may not be incorporated into PCA analysis. After continuous testing and adjustment, five major attributes are obtained and the importance of constitutional fields is considered as the basis for naming. The five attributes are as follows: corridor column information; classroom column design information; wall design information; data for number of walls; and, data for number of classroom columns.

在前面的處理完成後，還使用分群分析將校舍分群，以幫助後續的 I_s 值關係模型的建立，分群的目標是將校舍依照不同的設計模式特性區分開來，將相近設計的校舍放在同一群集，分群的依據是主成分分析法所找出的五種主要校舍設計相關屬性，圖十一是使用 SPSS Clementine 進行此一分群分析時的節點設計圖，分別使用的 K-means 和 Two-Step 兩種分群演算法，其參數設定如下：

After data preparation, clustering analysis is utilized to find hidden design patterns of school buildings. This study uses the K-means and two-step clustering methods. Fig. 11 shows the node deployment for clustering using the SPSS Clementine (2007) software. The parameter choices for the two methods are as follows.

K-means

此一分群方法最重要的設定參數是初始的群集數 k ，雖然有許多研究都在研究如何找出最佳的 k 值，但是目前仍沒有一個方法可以宣稱它找到 k 值是最佳的，唯有對該領域的專業了解以及詳細的測試才能得到最佳的 k 值，在本分析案例中，K-means 分群的停止條件是設定為 20 次迭代，如果 k 太大，那會造成分群結果無法在 20 次迭代內收斂，如果 k 小於 6，則分群的結果就可以在 20 次迭代內完整的收斂，每筆資料都會故底定在所屬的群集內，不在變動，最後挑選的 k 值是 5，而根據此意參數的分群結果可以找出三個主要的校舍群集，分別的比例是 28%、56% 和 16%。

The most important parameter with this clustering method is the initial group, k .

Although many researchers have developed methods for choosing the initial K value, no method can confirm that it has found the best K value. The best K can be found only based on researcher understanding and problem testing. In this study, K-means clustering is set to stop after 20 iterations. If the K value is too large and cannot be completely converged after 20 operations, some data points will continue to change the clusters to which they belong. When the K value is reduced to <6 , Clustering models can be completely converged after 20 iterations, become stable, and no longer change cluster label of all data bit. The K value chosen is 5, and three major clusters are obtained. The remaining two clusters have few data and are considered outliers. The distribution of three major clusters are 28%, 56%, and 16%.

Two-Step

Two-Step 分群有兩個優點，一是複雜度不高，運算時間與資料數量間之關係為線性關係，第二個優點就是不需要由人工決定分群的群數，演算法即可自己根據資料狀況決定，操作人員只需給予上下限，在本分析中，上下限的設定為最少兩個群集，最多八個群集，而最後的分析結果是所有的資料都被分到兩個群集中，分別佔了 54% 和 45%。

This clustering method has two features. The first is enhanced scalability. The algorithm has low complexity. Computing time does not grow nonlinearly as data volume increases. The other feature is that it can determine the number of clusters, unlike K-means clustering, which requires manual designation of parameters. However, this work can designate the upper and lower limits for the number of clusters. This work sets the limit to 2–8 clusters based on experience with K-means clustering. Consequently, all data are divided into two clusters, accounting for 54% and 45% of all data.

此資料探勘分析還用了十群交叉驗證來驗證結果的可靠度，因此資料前處理的最後一個步驟就是將整理好的資料隨機分為十組。

This work uses 10-fold cross validation to validate the prediction model. After prepa-

ration, data are grouped by first dividing data randomly and equally into 10 clusters. One cluster is then chosen as a dataset for validation and the remaining nine clusters are combined into one training dataset.

5.1.2 資料探勘

完成資料前處理，將校舍的群集分好之後，才開始建立 I_s 值與校舍設計參數的關係模型，本分析使用了廣義線性模型、線性回歸和類神經網路三種分析方法，每種方法都有三種分析資料群，分別為先經過 K-means 分群的資料、先經過 Two-step 分群的資料及沒有先經過分群的資料。圖 12 為使用 SPSS Clementine 分析時的節點設計圖。以下分別對三種分析方法的參數設定作說明。

The prediction model is built only after clustering is completed. This work uses a GLM, simple regression, and ANNs to build the prediction model based on three groups of data – not clustered in advance, clustered by K-means, and clustered by the two-step method in advance. In total, nine prediction models are generated. Fig. 12 shows the node configuration within SPSS Clementine. Below are the parameters chosen for the three methods.

Generalized linear model

廣義線性模型是假設在輸入參數和預測目標之間有一個可以用連結函數表達的關係，這個連結函數可能是指數函數、對數函數、Logistic 函數等，經由一些測試資料的測試，我們選擇使用對數函數作為連結函數，並且根據實際的資料分布選擇了常態分布作為輸入參數的分布函數，而關係模型的分布也選擇常態分布，因其表現較其他分布形式較好。

The GLM assumes a relationship between input variables and a predictor; this relationship can be built by a link function such as identity function, log function, logit function, or power function. After available link functions testing on some data, the pre-

diction model performs best when using log function as the link function; hence, this work chose the log function. The distribution function of the predictor is based on the actual distribution of data. This work chose the normal distribution, which is close to the real data distribution. The prediction model constructed using the normal distribution performed better than those with other distribution types.

Simple regression

線性回歸是選擇使用最小平方根法來建立校舍建築物的設計參數與其耐震能力 I_s 之間的關係，這也是最常使用的回歸方法之一。

Simple regression in this work uses the least square method by adopting the building design parameters as independent variables X and the aseismic ability of buildings as dependant variables Y . The linear equation between regressed design parameters and aseismic indices serve as the model for predicting aseismic ability Y based on building design parameters X .

Artificial neural networks

類神經網路需要決定的參數包括隱藏層的數量、每層的神經元數量、學習率、停止條件等，除了直接設定神經網路的參數，還有一些方法可以使用，例如 dynamic、multiple 或是 prune method 可以用來調整並找出最佳的神經網路大小和結構，dynamic method 是從一個小型的神經網路開始（兩個隱藏層、每層兩個神經元），慢慢成長，並且比較成長前後的神經網路效能與結果，Multiple method 則是同時產生各種不同的神經網路，並且一起訓練到達停止條件，然後在從中挑選出表現最好的一個，而 Prune Method 則是從一個大的類神經網路開始，慢慢的把重要度低的神經元節點拿掉。本分析最後挑選的是 Exhaustive Prune Method，是 Prune Method 的一種修改形式，對於節點的篩選要求較高，是所有方法中最花時間的，但是通常也可以找到最好的結果。其他的類神經網路設定參數為：初始的神經網路為兩層隱藏層，其中一層有 30 個神經元、一層有 20 個神經元，停止條

件為 250 個訓練循環，在這個設定下，Exhaustive Prune Method 是表現最好的方法。

Generally, ANNs must decide on such parameters as number of hidden layers, number of neurons in each layer, learning rate, and stop condition. Aside from directly setting these parameters, methods such as dynamic, multiple, and Prune methods are available for adjusting and finding the optimal size and structure of the neural network. The dynamic method starts with a small neural network (two hidden layers with two neurons for each layer), expands network size gradually, and decides on the further expansion based on model performance before and after expansion. The multiple methods constructs multiple neural networks simultaneously, trains all neural networks to reach the “stop condition,” and then selects the group with the best performance. In contrast with the dynamic method, which slowly builds a large neural network from a small one, the Prune method first builds a large network and then removes neurons with low importance based on training. This work chooses the exhaustive Prune method, a special application of the Prune method. The initial neural network has two hidden layers, one with 30 neurons and the other with 20 neurons. The stop condition is set to 250 training cycles. Under this limit, the prediction model built by the exhaustive Prune method performs best.

5.1.3 驗證

驗證有兩個主要的目的，一是確保資料探勘找到的關係模型的可靠度，而不會找到只適用於該組訓練資料集的關係模型，第二個目的是可以用來作為比較不同分析方法的指標數據，本分析使用的驗證方式是十群交叉驗證，這個方法將所有的資料等分成十份，每次挑選九組出來作為訓練資料集，留下一組作為驗證資料集，如此可以得到十組模型以及其可靠度的指標，求此十組指標平均值即可得到代表此關係模型的可靠度代表值，而本分析所選擇的指標有三個，線性關係、絕對平均誤差（Mean Absolute Prediction Error, MAPE）以及 hit rate。

Validation work has two purposes. The first is to ensure model reliability instead

of to generate only good performance during data training. The second is to serve as a benchmark for comparing the performance of different prediction models. This work uses 10-fold cross validation to assess and compare the performance of prediction models. This method divides a fixed amount of data into 10 groups, conducts 10 rounds of model building and validation, chooses a different group of data for testing, trains the model with remaining nine groups of data, and uses test group data to validate model accuracy. After validation for 10 times, the accuracy of the 10 models is obtained and their average is taken as the accuracy of this algorithm. This study uses linear correlation, mean absolute prediction error (MAPE), and hit rate as the indices for comparing the prediction model performance.

5.1.4 結果

此資料探勘分析會產生九個校舍耐震能力與校舍建築設計資訊間的關係模型，包括直接使用 GLM、線性回歸和類神經網路的三組，混合使用 K-means 或是 Two-Step 分群的六組，要比較其優劣我們使用了 R^2 、MAPE 和 hit rate 三個指標，並配合十群交叉驗證，其中線性關係 R^2 之公式為：

This work constructed nine prediction models; three are directly generated by the GLM, simple regression, and ANNs. The mixed model of K-means and two-step clustering generated three prediction models. Hence, nine models were obtained and 10-fold cross validation is used to compare the performance of the three reference indices – R^2 , MAPE, and hit rate. Notably, R^2 , the linear correlation is

$$R = \frac{\sum (\hat{y}_i - \bar{\hat{y}})^2}{\sum (y_i - \bar{y})^2} \quad (5-1)$$

其中 y_i 是校舍的實際耐震能力指標 I_s 值， \hat{y}_i 則是透過此關係模型的到的推估耐震能力指標值， $\bar{\hat{y}}$ 則是所有資料的 I_s 值之平均，透過此一公式即可得到實際的 I_s 值與透過關係模型得到的推估值之間的線性關係，線性關係越高表示兩者之

間越接近，也代表著關係模型的正確性。MAPE 的公式為：

where y_i is the aseismic CDR of school buildings obtained using nonlinear analysis of the database, \hat{y}_i is the CDR obtained from the prediction model, and \bar{y} is the average aseismic CDR of school 651 buildings obtained using nonlinear analysis. The correlation between aseismic CDR of school buildings obtained via the prediction model and nonlinear analysis can be determined based on linear correlation. A high CDR indicates a strong correlation and many opportunities to make correct predictions. The MAPE is derived as

$$MAPE = \frac{\sum \frac{y_i - \hat{y}_i}{y_i}}{N} \quad (5-2)$$

其中 N 是資料總數，MAPE 的是用來表示關係模型誤差之數值，由於關係模型不可能完全沒有誤差，即使有非常高的線性關係也是會有誤差，因此會使用 MAPE 作為判斷其誤差程度的參考。hit rate 的定義是：

where N is the number of samples. The MAPE is used to judge the degree of error of prediction models as the prediction result always has errors, although the prediction mode has an adequately high R^2 . The hit rate is derived as hit rate

$$hit\ rate = \frac{\sum I\{(1 - \alpha)y_i \leq \hat{y}_i \leq (1 + \alpha)y_i\}}{N} \quad (5-3)$$

其中 $0 \leq \alpha \leq 1$ ，且 $I\{L\} = 1$ 。hit rate 是用來判斷關係模型的正確率的。在本分析中，我們設定了兩個 α 值作為 hit rate 指標用，分別是 0.1 和 0.2，表 5-1 列出了這九個關係模型使用這三個指標配合十群交叉驗證所得到的數值，表現最好的關係模型是先使用 K-means 分群再使用 GLM 所建立的，第二好的則是先使用 Two-step 分群再使用 GLM 所建立的，圖 13 是先使用 K-means 分群再使用 GLM 所建立模型的實際 I_s 值與使用模型得到的 \hat{I}_s 值的比較圖，資料點的回歸取現的斜率非常接近 1，可以看得出來兩者之間的相關度非常高。如果單看模型的線性關係表現，類神經網路的表現比 GLM 和線性回歸都要來的好，這也可以

驗證校舍的耐震能力與其設計參數之機善一個非線性的關係，而雖然 GLM 整體的排名較 ANN 來的好，但是 hit rate 卻是 ANN 表現的比較好，但是看到 MAPE 又會發現 ANN 的 MAPE 較大，因此我們建議在 I_s 值的關係模型的挑選，可以依據應用的需求來決定，如果需要較高準確率的時候，建議使用 ANN，如果是需要降低整體的誤差，則建議使用 GLM。如果先使用分群方法將校舍資料根據設計參數分出不同群集後，再對不同群集分別探勘其耐震能力與設計參數的關係模型，結果會比沒有先分群要來的好一些，探討其原因，是因為典型校舍已經是校舍建築物的一個子集合，而此子集合的特性已經非常接近，因此再進行分群也不會有顯著的改善。

表 5-1: Cross-validation result of the prediction model

	K-means ANNs	Two-step ANNs	ANNs	K-means Regression	Two-step Regression	Regression	K-means GLM	Two-Step GLM	GLM
R	80.21%	81.78%	80.76%	72.16%	72.16%	72.15%	87.41%	87.11%	87.05%
MAPE	28.69%	26.64%	26.51%	46.06%	46.05%	46.05%	24.68%	24.70%	24.71%
hit_rate(0.2)	53.12%	54.29%	54.31%	40.16%	40.17%	40.21%	48.82%	48.64%	48.52%
hit_rate(0.1)	27.71%	28.72%	28.75%	21.13%	20.98%	21.09%	25.05%	25.16%	25.16%
Rank	6	4	3	9	7	8	1	2	4

When $0 < a < 1$ and $IL = 1$, hit rate is utilized to determine the percentage of data predicted correctly by the prediction model, that is, prediction model accuracy. In this work, the hit rate is ranked by setting a equal to 0.1 and 0.2, which are utilized as two assessment indices that average and rank the performance of accuracy. Table 1 lists the assessment indices of the nine prediction models. The prediction model that performs best is that built using the GLM with K-means clustering. The second best prediction model is that built via the GLM with two-step clustering. Fig. 13 compares the actual CDR and CDR obtained using the K-means and GLM prediction models. The slope of the regression curve equation approaches 1, indicating a strong correlation between school building design data and aseismic ability of building. However, the scattered distribution of actual data points corresponds to a high R^2 and high MAPE. After a thorough comparison of the nonlinear analysis by ANNs, the GLM, and linear analysis by simple regression, the ANNs perform better than the GLM and linear analysis by simple regression all aspects, confirming that the design parameters of school buildings have a nonlinear relationship with aseismic ability, which conforms to the fact that the aseismic CDR in this work is obtained using

nonlinear analysis. Although the GLM ranks high in comprehensive assessment, its hysteresis is worse than that of ANNs; ANNs also have a higher MAPE. Hence, it is possible to determine which prediction method is suitable based on actual needs when predicting the aseismic ability of school buildings. We recommend using ANNs for accurate prediction of the aseismic abilities of school buildings; however, the drawback in using ANNs is that the prediction model generated is a black box. We recommend using the GLM to minimize total error. When building prediction models by clustering first and then comparing the performance of the three assessment methods, the prediction model built with clustered data performs slightly better than those built directly, indicating that traditional school buildings are already a subcluster of various architectural patterns. One feature of subclusters is their weak correlation with the aseismic ability of school buildings. Hence, information added to the cluster will not markedly improve prediction model quality.

由於校舍補強預算和時間有限，因此其執行的優先順序就會依照評估得到的耐震能力作為參考排序，因此本分析還用此實際的應用作為另一個評量關係模型優劣的指標，此指標將所有的校舍依照其 I_s 值排序後，照順序等分成 10 群，另外在用關係模型得到的 \hat{I}_s 排序，一樣照順序等分為 10 群，接著比較每筆校舍所分配到的群集，如果實際所屬的群集編號和關係模型得到的群集編號一樣，則誤差 (error) 為 0，如果差了一號，則誤差為 1，差了兩號則誤差為 2，表 5-2 即為九組關係模型的排序誤差結果。可以發現表現最好的一組仍然為先使用 K-means 分群再用 GLM 探勘所得到的關係模型，其誤差小於 1 的資料比例為 70.8%，誤差小於 2 的資料比例則有 88.9%。

The budgets and priorities for reinforcing school buildings are based on the aseismic abilities of school buildings. This work analyzed the sequencing result of aseismic ability of school buildings by sequencing school buildings based on CDR values, dividing them into 10 equal zones, and comparing the zone number of actual and predicted values. Table 2 shows the zoning result. When Error = 0, the predicted and actual values have the same zone number; when Error = 1, predicted and actual values are in adjacent zones; when Error = 2, predicted and actual values are separated by one zone. The prediction

model built by the GLM with K-means clustering performs best. The zoning error of this prediction model <1 is 70.8%, and the zoning error <2 is 88.9%, indicating that the prediction model already has sufficient accuracy when sequencing is used.

表 5-2: Sequencing analysis of prediction of aseismic ability

Error	K-means ANNs	Two-step ANNs	ANNs	K-means Regression	Two-step Regression	Regression	K-means GLM	Two-Step GLM	GLM
0	32.8%	35.0%	33.9%	33.4%	34.1%	33.3%	35.6%	34.9%	35.0%
1	36.0%	35.2%	36.5%	35.4%	34.2%	35.5%	35.2%	35.9%	35.7%
2	17.5%	15.2%	16.2%	16.6%	17.1%	16.4%	18.1%	17.3%	17.5%
>2	13.7%	14.6%	13.5%	14.6%	14.7%	14.8%	11.1%	11.9%	11.8%
Rank	5	5	4	7	9	8	1	2	2

5.2 CDR 值與校舍設計之關係模型

校舍耐震資料庫中，詳細評估表的耐震能力索引 CDR 值是用來評估校舍是否需要補強、甚至是拆除的最重要依據，此數值的取得非常耗時耗力，且與校舍結構材料、設計與現況等參數之間為高度非線性的關係，如果能夠取的此一關係模型，對於校舍耐震能力補強計畫的進行，可以有很大的幫助。

5.2.1 資料前處理

人工輸入的資料很容易發生單位錯誤或是格式不正確的資料，而資料的品質對於資料探勘的結果優劣影響很大，因此需要在資料前處理時把這些問題都處理過，資料前處理的最主要目標是讓處理過的資料能夠很確實的反映出分析問題所要尋找的資料特性，使用的方法包括資料過濾、屬性過濾、新屬性的生成等，首先，本分析中第一個前處理工作就是資料過濾，要把有問題的資料，不合理的資料都過濾掉，過濾的條件則是依照國震中心所建議的典型校舍特性的過濾條件：

Manually inputting data may result in incorrect units or formats because controlling data quality in the real world is difficult. Hence, it is necessary to pre-process data before building the relational model. Quality, as an important part of soft computing and data analysis, has considerable influence on subsequent analytic results or even on the reliabil-

ity of the generated model. Apart from the actual data, the researcher also refers to expert advice from NCREE for data pre-processing. The main target of data pre-processing is to ensure the accuracy and adjustment of the data in a format that clearly reflects the target of analysis. Pre-processing includes data screening, property screening, and new property synthesis. Data screening is divided into two stages. The first stage is the rationality screening of the data. Pre-existing mistakes are unavoidable because the data in the School Building Database were entered manually; the obligation is to identify such mistakes. Most of the school buildings were I-type shaped buildings; a very common design. . With regard to the properties of these school buildings, the NCREE (2005) suggests the following screening conditions:

- 校舍深度介於六到二十公尺之間
- 校舍垮距介於二到八公尺之間
- 每間教室至少有一個垮、兩個柱
-
- 校舍的破壞地表加速度是長向較小
- Total depth of the school building should exceed 20 meters or is less than 6 meters
- The span exceeds 8 meters or is less than 2 meters
- The number of spans for a single classroom is less than 1
- The number of columns in the classroom is low
- The collapse ground acceleration of the major direction is greater than that of the minor direction

由於並不是所有的校舍都有進行過詳細評估，因此第二個步驟就是要挑選出同時有基本設計資料以及詳細評估的 *CDR* 值的資料，接著，由於校舍資料庫中

的原始資料，每棟校舍都有上百個資料屬性，因此接著要對屬性進行篩選與合併，與國震中心之專家討論過後，挑選過濾掉大部分的屬性，但是仍然有約三十多個屬性，因此為了能夠繼續減少資料屬性，本分析接著還繼續對這些資料的屬性進行分類，根據其屬性分布挑選出一組典型校舍子集來分析，這個挑選出來的子集其特性為：

In the second stage, choosing school buildings with both basic design parameters and minimum destruction ground acceleration is necessary because not all school buildings have detailed information. According to the raw data in the seismic assessment database for school buildings, each data set contains hundreds of properties. Based on our judgment with expert which are non-structural and low importance, and synthesize some properties with similarity. There are still more than 30 properties left after this reduction process. This study further classifies school building records into subsets based on similarities in property values, and chooses one subset with major population as the data set for further studying. After the classification of school buildings, we try to do further reduction and finally determine a set of key properties which is optimal to represent the seismic characteristics of individual school buildings. The choice is based on data distribution, and a subset that correctly represents I-shaped school buildings. The features of this subset adopted in this study are listed below:

- 沒有走廊柱
- 只有一種教室柱
- 校舍沒有 RC 牆
- 沒有四面圍束磚牆
- 沒有三面圍束磚牆
- No corridor columns
- Only use one type of classroom column

- School buildings have no RC walls
- School buildings have no brick walls with four-side confinement
- School buildings have no brick walls with three-side confinement

最後前處理完成的資料總共有 107 筆。並且依照專家建議挑選出 12 個屬性：

- | | |
|---------|-----------|
| • 樓層數 | • 教室柱斷面積 |
| • 鋼筋強度 | • 是否有走廊 |
| • 混凝土強度 | • 二樓樓地板面積 |
| • 校舍長度 | • 總樓地板面積 |
| • 校舍深度 | |
| • 教室柱數量 | |



並編為 P_1 到 P_{12} ，其中 S_{DS} 是校舍的設計地表加速度， S_{D1} 則是一秒週期的設計譜加速度，都是分析校舍耐震能力時非常重要的參數，另外也根據專家的建議，使用現有的資料組合出兩個新的屬性，分別是校舍的教室數量 (P_{13}) 和校舍建築物的垮距 (P_{14})，以下介紹各輸入屬性：

樓層數 建築物的樓層數直接反映了該建築物結構所需乘載的重量的規模，樓層數越高表示其結構所需要能乘載的重量越多，同時，不同樓層的建築物結構和高度均有顯著差異，其在承受相同水平側力時所產生之變形量也會有所不同，因此將其納入輸入參數。

鋼筋強度 鋼筋混凝土校舍結構物之鋼筋要乘載其所受到的拉力，例如校舍受力變形產生的拉力，樑乘載樓版重量時的拉力等，因此其強度直接影響了結構元件所能承受的最大拉力，間接的也影響了建築物整體的結構強度，因此將其納入輸入參數。

混凝土強度 鋼筋混凝土校舍結構物之混凝土要乘載其所受到的壓力，例如校舍受力變形或是乘載重量所產生的壓力等，因此其強度直接影響了結構元件所能承受的最大壓力，間接的也影響了建築物整體的結構強度，因此將其納入輸入參數。

校舍長度、校舍深度 校舍的長度和深度資訊可以表現出校舍的規模及結構形式，並且能夠從中推測出結構物所需乘載的重量，因此將其納入輸入參數中

是否為雙邊走廊 本研究已經將分析的資料集限制在沒有走廊柱的校舍範圍內，而是否有雙邊走廊則影響到校舍短向的配重方式，如果是雙邊走廊表示其配重較平均，如果是單邊走廊則其走廊側之教室柱所需要乘載的重量會較大。

二樓樓地板面積 由於典型校舍形式固定，每層樓之樓地板面積差異不大，因此可以使用二樓樓地板面積作為代表每層樓之樓地板面積以及其所佔地坪，此一屬性乘上樓層數即為總樓地板面積，直接反映了結構物所需乘載的重量規模，但其獨立存在時仍有其物理意義，因此仍然保留此一屬性。

總樓地板面積 總樓地板面積直接反映了結構物所需乘載的重量規模，是非常重要的參數，因此納入輸入屬性。

教室柱數量、教室柱斷面積 教室柱數量以及其斷面積可以代表校舍結構物中，乘載重量的能力，在評估結構物的耐震能力時也是非常重要的資訊，因此將此二屬性納入本分析的輸入屬性中。

設計地表加速度 設計地表加速度表示該地區的建築物結構所需要能承受的地震力，是判斷結構物耐震能力是否足夠時的重要依據，因此將其納入輸入屬性中。

一秒週期的設計譜加速度

教室數量、垮距 校舍的教室數量以及垮距兩個屬性是在調查時並沒有調查的資料，不過可以由其他的屬性資料中推算得到，而在跟國震中心的專家討論過後，認為可以將此兩個屬性也加入此分析的輸入屬性，因其都可以代表校舍

結構物的結構資訊，教室數量代表著典型校舍的隔間情形，也包含了隔間牆的分布，而垮距則可以代表校舍結構物中柱子的平均乘載量。

After finalizing the data for the first and second stages, 107 datasets conform to the above condition. Twelve properties are then chosen for the screened data with reference to expert advice, displayed as P1 to P12 in Table 1. In addition to the screening based on existing properties, this study synthesizes two new properties, P13 and P14. They represent the number of classrooms and number of spans for a single classroom, respectively, based on expert advice and existing data. In P11, SDS stands for design spectral response accelerations at short-periods, and in P12, SD1 stands for design spectral response accelerations at 1 sec. These two parameters represent the magnitude of the seismic force at the building's location, and are very important parameters for analyzing the aseismic ability of buildings by non-linear analysis.

前處理的最後一個步驟是資料的正規化，因為不同屬性的物理意義不一樣，其數值的數量級差異也很大，如果有一個重要參數其數量級很小，那就會增加探勘的難度，因此為了減少數量級差異造成的影響力落差，要先把參數經過正規化處理，正規化的方法很多，可以把屬性的數值調整至 0 到 1 的區間，也可以用標準差和平均值來作調整，其最重要的目標是要保留原始數值的分布，本分析則是對 P_2 、 P_3 、 P_7 、 P_9 和 P_{10} 除以 1000，讓這些屬性的數值和其他屬性的數值比較接近。

The last step of data preprocessing is normalization. The purpose of normalization is to balance the impacts of the parameters in different scales. If an input parameter has small values of mean and standard deviation, but is of high importance and if the result is also sensitive to this parameter, then it is necessary to use data normalization to prevent its influence from being overshadowed by other larger scale parameters. Normalization methods include converting the data into the range of 0 to 1, using the maximum and minimum values, and converting data to the standard deviation of its mean. The normalization principle adopted in this paper is to retain the original values as far as possible, so only a

few parameters with large values, such as P2, P3, P7, P9 and P10, are divided by 1000 to make their scale comparable to other parameters.

完成以上所有的前處理工作後，可以得到 107 筆校舍資料，每筆資料有 14 個屬性，本分析就是針對這個資料集作分析。

After the three pre-processing steps, 107 datasets and 14 properties were obtained. The subsequent analysis was based on this dataset.

5.2.2 資料探勘

Genetic Programming

雖然校舍耐震能力的資料中，最能代表校舍耐震能力的數值是 *CDR* 值，但是由於其組成成分中的 *AD* 值的數值是根據建築物所在地而決定的，和校舍設計、建築方式有關係的參數則是 *AC*。因此本分析使用資料探勘所尋找的知識為校舍設計參數與校舍破壞地表加速度 *AC* 間的關係模型，首先是使用 GP 方法來進行資料探勘，使用 GP 分析的結果可以作為了解模型複雜度的參考。運算樹的層數影響了所建構模型的複雜度，越多層可以建構出越複雜的模型，然而卻可能會造成無法收斂、運算時間太長或是太過複雜的模型等問題，本分析先使用不同的運算樹層數來進行測試和調整，最後根據運算所需時間和模型的結果表現，選擇了 4 和 5 層兩個運算樹層數來進行後續的資料探勘，在本分析中使用了多組不同的參數設定進行測試與探勘分析，最後的結果是使用 200 組基因、5000 次的演化，並且交配率是 0.8、突變率是 0.1，而交配函數使用的是 Scattered 函數，突變函數使用的是 Adaptive Feasible 函數，兩者都是非常通用的函數，可以處理各種不同面向的問題。表二是此分析中 GP 演算法的資料探勘結果的 RMSE（均方根誤差）值。

An AC model was built for this study to represent the relationship between the basic design parameters of school buildings, and minimum destruction ground acceleration. GP was the first model to be used, and based on a preset number of tiers for different operation

trees; it can result in relational equations with different degrees of complexity. In this case, several operation trees with different number of tiers were tested, and it was found that the most suitable number is either four or five. Having a low number of tiers leads to reduced complexity of the relation model and hence, poor performance. Conversely, large numbers result in many difficulties, such as convergence problems, time-consuming progressive computation, and a very complicated relationship model. The optimum setting is 200 populations of 5000 progressive iterations, a crossover rate of 0.8, and a mutation rate of 0.1. The crossover function used in this paper is the scattered function. The mutation function is the adaptive feasible function. Both these functions can be applied to solve many different problems. The scattered function diversifies the child layer after crossover. The adaptive feasible function is suitable for the constrained minimization problem. This setting was chosen after the analysis was conducted 30 times. Table 2 shows the root mean square (RMSE) of the model generated.

RMSE 是本分析案例中最主要用來判斷模型品質的索引，定義如下：

RMSE is the index used in the current study to judge the quality of models, and is defined as the equation below:

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{N}} \quad (5-4)$$

其中 N 是資料的總數， y 是使用探勘得到的關係方程式所求得的建築物耐震能力 AC 值， \hat{y} 則是詳細評估的非線性分析所得到的建築物 AC 值，建築物的破壞地表加速度分布在 0.04 至 0.5 之間，NCREE 專家則建議 RMSE 值需要低於 0.04，此模型才夠有能力用來鑑別校舍的耐震能力優劣與否，表二中的控制組則是使用 SPSS Clementine 的類神經網路所建立的模型，類神經網路的設定為使用 Exhaustive Prune method 來發展神經網路，初始神經網路為兩層隱藏層，分別有 30 個和 20 個神經元，訓練 250 個週期，結果的 RMSE 是 0.041，非常接近目標的 0.04，本分析使用了 WGP 來建立校舍的 AC 值與其設計參數間的關係模型，結果的模型其品質和使用類神經網路所建立的模型非常接近，然而類神經網路所

建立的模型其複雜度太高，且其機制為一黑盒子，而 WGP 方法產生的模型就只是一個代表方程式的運算樹，可以非常快速的將結果拿到各種不同環境或平台中應用，

where n is the number of datasets, y is the estimated value obtained from the equation, and \hat{y} is the actual unit-less deviation index value (the smaller the better). The ground acceleration of the minimum destruction obtained from the nonlinear analysis was distributed between 0.04 and 0.5, and therefore, the relationship model has a sufficient recognition rate. Experts from NCREE recommend that the RMSE must be below 0.04. The control group is the relationship model obtained from the artificial neural network. In applying the relationship model constructed by SPSS Clementine and choosing the Exhaustive Prune method to adjust the number of tiers and nodes, the initial neural network has two hidden tiers with 30 and 20 neurons, respectively. The neurons have been trained for 250 iterations, and those with a low degree of importance are removed during the training period based on the situation. The resulting RMSE is 0.041, which is close to the target of 0.04. In this paper we use WGP to create an aseismic ability prediction model for real school buildings. The quality of our model is similar to models built using Artificial Neural Networks. However, Artificial Neural Network based models are complicated; their mechanism is in a black box. The WGP model, on the other hand, is just an equation of the building's design parameters and its aseismic ability. Thus, it can easily be ported to other platforms and programming languages for use in many applications. The optimum model obtained from the GP pattern is represented by the relationship equation below. Function nodes in a tree topology, displayed in Figure 8, uses several symbols and text to represent the F of that node.

“+” 代表 $f = x_1 + x_2$;

“-” 代表 $f = x_1 - x_2$;

“×” 代表 $f = x_1 \times x_2$;

“÷” 代表 $f = x_1 / x_2$;

“pow” 代表次方函數 $f = x_1^{x_2}$

結果使用 GP 方法所建立的關係模型其表現並不好，RMSE 只能達到 0.056，而根據圖九可以發現，其實使用 GP 方法並沒有正確的建立 AC 和校舍設計參數間的關係模型，因為他只使用了三個輸入參數，而且其方程式也非常的線性，可以推斷出此一個關係模型的複雜度很高，而 GP 方法並不能正確的找出這個關係。

The performance of this model is not ideal because the RMSE can only reach 0.056. Based on Figure 9, the relationship model generated did not correctly build the relationship between the design parameters of school buildings, and ground acceleration of the minimum destruction. As only three input parameters were used, it resulted in the minimum equation as the linear equation. This can be attributed to the fact that this seismic ability model for school buildings has high complexity, and the application of GP pattern alone cannot obtain the relationship between them. The WGP pattern was then used to build the relationship equation between the design parameters of school buildings, and ground acceleration of the minimum destruction because it could be used for relationships that are more complex than with the GP pattern.

$$AC = \frac{P_3}{P_{11}} + P_3 P_{12} - P_3 P_{12}^3 \quad (5-5)$$

Weighted Genetic Programming

WGP 是本分析第二個使用的方法，和 GP 一樣選擇了五層的運算樹，其他的基因演算的參數也都和 GP 分析中使用的一樣，200 組基因演化 5000 代，交配率是 0.8，突變率是 0.1，而唯一不一樣的參數則是權重 w 的範圍，設定為 +10 到 -10 間，在此設定下分析 30 次，從中挑出表現最好的一組關係模型，其 RMSE 直在表二中，圖 10 是此關係模型的運算樹結構，其符號定義與 GP 方法運算樹使用的一樣，唯一不同的是黑色實心點，代表的是 $f = w_1 x_1$ ，此關係模型的方程式如下：

WGP was used as the second model, which also chose five tiers of the operation tree. The optimum setting of GA is the same with GP: 200 populations of 5000 progressive

iterations, a crossover rate of 0.8, and a mutation rate of 0.1. The best group was chosen after the analysis was conducted 30 times. Contrary to GP, the weight was set from +10 to -10 within the weighted (w) range. Table 2 shows the RMSE of the model generated, and the four-tier optimum equation is shown as Equation (9). The tree topology generated is displayed in Figure 10, and uses the same symbol as the GP tree topology to represent the same operator. The other symbol that was used is a black solid dot, which represents $f=w \times l$.

$$AC = (165P_{10}^{8.86P_6}P_4^{4.86} + \frac{22.5P_8 + 39.5P_{10}}{P_{10}})^{-98.6P_4P_8^{-1.3}-133P_{10}-0.05P_7^{1.38}P_{10}} \quad (5-6)$$

表二列出了 WGP 方法所建立的校舍設計參數與其最所破壞地表加速度 AC 間關係模型的 RMSE 值，其值為 0.039，比使用類神經網路所建立的關係模型表現還要好，圖 12 是此模型的推估值與實際值的比較圖，可以看得出來比使用 GP 所建立的模型要來的好。

Table 2 shows the RMSE of the optimum relationship equations generated from the two patterns. The RMSE reached 0.039, which is better than the performance of the model built by the artificial neural networks in the contrast group. Figure 12 displays the comparison between the estimated, and the actual values of the model. By comparing the results, the model constructed by the WGP pattern is superior to the model constructed by the GP pattern.

本研究還分析了此關係模型所使用到的輸入參數，表三列出了此關係模型所有使用到的輸入參數， S_{DS} 和 S_{D1} 並沒有被使用到，因其兩者都是和耐震需求相關的參數，而本分析的主要目標是校舍的地震承受能力，也就是耐震能力相關的部份，而此結果也正可以符合預期，而其餘所有的參數都有被使用到，也可以代表前處理所選擇到的參數都有其重要性。

The parameters entered are analyzed based on the equation obtained from WGP, and

Table 3 shows the input parameters obtained from the optimum relationship equations. SDS and SD1 were not used, as both are relevant to the demand of the CDR. For this study, the target is to estimate the Capacity (Aseismic Ability Index), which is irrelevant to Demand. Hence, this result conforms to the expectations. As all of the remaining input parameters were used, this indicates that the parameters chosen at the data processing stage are important.

Capacity Index Formulation Tuning

CDR 是詳細評估所得到的，校舍建築物的耐震能力指標，是校舍結構所能承受的地震力和規範所規定，校舍所在位置所需要能承受的地震力的比值，CDR 大於 1 表示其耐震能力足夠，如果 CDR 等於 1 則表示耐震能力與需求相等，然而 I_s 值的數量級和 CDR 不一樣， I_s 是百分系統，100 表示耐震能力與需求相等，不過 I_s 內其實還有一個調整參數 I 為 1.25，而此時如果 CDR 是 1 則 I_s 是 80，這個關係可以用來將 I_s 值轉為 C_E 值，其物理意義和 AC 相同。

Similar with the CDR obtained from the detailed estimation, the aseismic capacity index of school buildings is the demand ratio of the aseismic capacity, excluding the ability unit and measure. CDR is directly compared to the ground acceleration, and IS is the estimated force ratio. If CDR is greater than 1, then the building has sufficient aseismic capacity. If CDR is 1, the building has an aseismic capacity equal to the demand. However, it should be noted that IS is a hundred-mark system, and 100 indicates that the capacity is equal to the demand. IS also needs to consider the usage coefficient I , of buildings. When IS equals 1.25, CDR is 1 and IS is 80. This relationship can be described by a formula that converts IS into C_E , which has the same meaning as the ground acceleration of the minimum destruction.

$$C_E = f(I_s) = \frac{I_s \times Demand \times I}{80} \quad (5-7)$$

I 是確保 C_E 值能趨於保守用的係數，在 NCREE 的設計中，校舍的 I 定為常數，為 1.25， C_E 即為最小破壞地表加速度，不過其可靠度較詳細評估所得到的數值低，因此本研究下一個分析目標即為修正初步評估的公式，為 C_E 加上一個調整因子 T ，可以讓其結果更接近詳細評估所得到的最小破壞地表加速度。

I is the usage coefficient that intends to keep the seismic design at the conservative side, and prevents miscalculations caused by the insufficient estimation of the earthquake destruction. For this study, I is set to a constant of 1.25 for school buildings. Demand is set to different recommended values based on the positions of the school buildings. It represents the minimum ground acceleration that the buildings can withstand in the area. The School Building Database contains the demand data that was determined by engineers, based on the actual situations. This study adds T to C_E as a revised formula, such that it is closer to the minimum destruction ground acceleration AC , of school buildings obtained from nonlinear analysis.



$$AC = C_E + T \quad (5-8)$$

基於之前的結果，本分析只使用 WGP 來分析，基因演算的參數也和之前一樣，4 或 5 層的運算樹，200 初始基因，5000 個演化週期，交配率為 0.8，突變率為 0.1，訓練出 30 組模型後，從中挑選出最好的模型，結果 T 的模型為一個 4 層的運算樹，如圖 11。

Based on the analysis above, this study only uses the WGP pattern to build the model, and this can be applied to complicated situations. Parameters chosen are: the operation tree has four to five tiers, 200 populations of 5000 progressive iterations, crossover rate of 0.8, mutation rate of 0.1, and choosing the best group after the analysis was conducted 30 times. The optimum T generated from the four-tier operation tree is shown in the equation

below, and the tree topology is displayed in Figure 11.

$$AC = C_E + 0.54 \frac{P_3}{P_2} - 24.86 \frac{P_6^{0.36P_3}}{P_5^{2.6P_9}} - (37.84 \frac{P_5}{0.67P_2 + 5.52P_7}) [(-0.67P_{12})^{1.23P_7/P_{13}}] \quad (5-9)$$

5.2.3 結果

表二顯示了 NCREE 初步評估 I_s 值轉換成詳細評估 AC 的等價數值 C_E 與校舍實際進行詳細評估所得到的最小破壞地表加速度 AC 相比的 RMSE，為值 0.067，距離目標的 0.04 還有一段距離，圖 13 是 C_E 與 AC 的比較圖，可以發現其趨勢相近，加上修正因子 T 之後，RMSE 降低為 0.045，圖 13 顯示了 $C_E + T$ 和 AC 的比較，可以發現其趨勢更為接近，因此可以確定修正因子確實可以將初步評估的結果修整的更好。

Table 2 shows the RMSE after the revision. The RMSE of the IS formula used by NCREE in the preliminary appraisal is 0.067. Although there is still a gap between the target of 0.04 and this value, the emphasis is on the degree of relationship between them, and the main target is to screen out the school buildings with degrees of higher risk. Figure 13 shows the comparison between the CE converted from IS of NCREE, and the destructive ground acceleration obtained through linear analysis. Even though the data have the correct directional tendency, the model obtained from the GP method is better because of higher deviation, and a linear relationship. Subsequent to adding T to the revised formula, RMSE is reduced to 0.045. Figure 13 shows the comparison, and it can be seen that the directional tendency is quite close, thus reducing the deviation. A good revision effect is observed, making the screening result of the preliminary appraisal more accurate.

表三呈現了此分析中所有使用到的輸入參數，四層運算樹和五層運算數分別使用到的參數分開表示，其中：樓層數 (P_1)、總樓地板面積 (P_{10}) 和單間教室的垮數 (P_{13}) 並沒有被使用到，其中前兩個參數對於校舍結構的耐震能力來說非常重要，在 I_s 的推估過程中已經有使用到，而本分析的修正因子並沒有使用到可以

顯示 NCREE 的初步評估已經很完整正確的表現出這兩個參數所造成的影響，而單間教室垮數的資訊則可能在初步評估的過程中已經將該資訊表現在評估公式內了，因此修正因子 T 的公式內並不包含它，不過此參數在前一個分析中依然有使用到，因此仍然有其重要性。

Table 3 shows the input parameters used by the equation obtained from the WGP method, and by the optimum relationship equations with four to five tiers. The number of floors (P1), total floorage (P10), and number of spans in a single classroom (P13) were not used because the number of floors and total floorage, which are highly important, were already used in the IS formula. Hence, IS has been correctly included in the two properties above, and the number of spans in a single classroom can be inferred as hidden in the formula. Hence, T in the revised formula will not use this input parameter. However, the input parameter is still used to directly construct the relationship equation in the previous case, and has a certain degree of importance.

In order to estimate the aseismic ability of a building, we need to obtain detailed information about its geometric dimensions, the properties of materials used, etc. and we may need to perform sampling and other experiments to obtain certain information e.g. the properties of materials used. Based on this information, a structural model of the building has to be constructed and then non-linear pushover analysis is used to perform estimation. This process is time-consuming, and technicians are required. It usually takes about one month to estimate the aseismic ability of a building.

在校舍資訊與耐震能力之關係模型之分析中，本研究提出了基於 WGP 方法所建立的關係模型，只要透過校舍結構的基本參數就可以推估校舍的耐震能力，其前處理的工作可以分為兩個部分，第一個部分是資料的過濾與挑選，其而過濾與挑選的條件是基於專家的建議，第二個部分是找出不同資料屬性的重要性，第一步是先將和校舍結構無關以及明確不重要的資料屬性剔除，接著藉由將校舍分成不同子集，並逐步減少資料屬性，最後選出了 14 個資料屬性，然後基於這個子集進行資料探勘分析，相較於原來上百個資料屬性，本分析達到了一個非常高效的資料壓縮，減少了非常多的資料雜訊。

In this study, we adopted the WGP method to construct a relation model, which uses the basic design parameters of school buildings to estimate their aseismic ability. The pre-processing efforts carried out for the proposed prediction model can be divided into two parts. The first part is a data filter for selecting out data sets which are reasonable as typical school buildings. The selection conditions of this data filter is proposed based on judgment with expertise in structural engineering. The second part is the determination of the key properties used in the proposed model. We first eliminate some properties which are non-structural and low importance, and synthesize some properties with similarity. We further classifies school building records into subsets based on similarities in property values, and chooses one subset with major population as the data set for further studying. Based on this subset, we try to do further reduction and finally determine 14 properties which are optimal to represent the seismic characteristics of individual school buildings. In comparison with hundreds of properties in the original data, a very high reduction ratio is reached.



WGP 最大的特色就是其產生的模型結果是使用數學方程式的形式來呈現，因此可以很簡單的就將產出的模型拿到不同的系統或平台中應用，除 WGP 外，本分析也使用了 GP 來進行分析，經過比較可以發現 WGP 方法所建立的模型的表現要比 GP 方法的好很多，可以顯示其對於複雜問題的處理能力是更為優秀的，而其 RMSE 也可以低於 0.04，比類神經網路的表現還優秀，透過預測值與實際值的比較圖，也可以發現其趨勢都是正相關，驗證了使用挑選出的 14 個參數和 WGP 方法所建立的關係模型是可以在校舍補強工作上使用的。

Same as GP, a key characteristic of WGP is the resulting model is in the form of mathematic equation, which is very easy and convenient to be applied in engineering practice, and thus is more practical than other soft computing methods. This study also applies GP on the same set of school building data for predicting the building's collapse ground acceleration. Compared to the prediction results by using GP, the accuracy of WGP model is much better. This case shows WGP can handle problems with more complexities than GP. Regarding the application of WGP method, this study performs 5000 iterations of

revolutions for test models with operation tree from two to five levels respectively, and determines the optimized values of model parameters, such as the crossover rate and mutation rate, for this application case through repeated tests. The RMSE of the resulting model achieves less than 0.04. This accuracy in prediction is comparable with the model using Artificial Neural Networks (ANNs). In addition, the graph compared the actual values with predict values shows the distribution trend of the data points are consistent with the expected direction. The result verification indicates the proposed WGP-based model successfully establishes the relation between the 14 input properties and the output property, building's collapse ground acceleration. In practice, the proposed model can be applied directly and efficiently for preliminary assessment of school buildings for seismic capacities.

In addition to directly inferring the aseismic capacity of school buildings from the design parameters, the current study revised the aseismic capacity index formula of school buildings designed by NCREC using GPS. The formula was to provide greater accuracy with a smaller deviation. This model can help decision makers with issues that are related to the aseismic capacity of school buildings, as well as estimating the disaster loss, during a disaster in a timely manner. These applications require that the estimation of the aseismic capacity of many school buildings be generated rapidly, and this would be impossible if traditional non-linear structural analysis were applied.

5.3 D_{isR} 值與校舍設計之關係模型

「校舍是否需要補強」此一指標其數值之基礎即為詳細評估的耐震能力指標 CDR 值， CDR 值超過 1 表示其耐震能力尚符合安全規範，反之，則是有安全疑慮，需要進一步補強或是拆除。而 D_{isR} 則是標記各校設耐震能力是否足夠的二元指標，也是教育部的校舍耐震能力補強計畫中，前期篩選工作的最主要目標。找到這個指標與校舍設計、現況等參數的關係模型，與 I_s 值和 CDR 值關係模型一樣，此一關係模型對於校舍耐震能力補強計畫中，初期的篩選工作可以有很大

的助益，也可以輔助決策者編定預算、快速的根據狀況決定不同計畫年度補強的規模等。

5.3.1 資料前處理

5.3.2 資料探勘

5.3.3 結果



第 6 章 校舍資訊與破壞構件之關係模型

校舍耐震評估最準確的階段為進行詳細評估作業，但進行校舍詳細評估須歷經非常多之流程如圖 1.2，花費很多時間，如能有一輔助工具能於執行校舍詳細評估前期即能粗估校舍之可能破壞構件與其破壞模式，將有助於土木技師與專業人員進行作業時能先有初步結果之掌握，與校方或主管機關進行洽談時也能更加迅速。依前段之敘述，本研究之子主題(預測破壞構件及其模式)主要是希望能於校舍詳細評估初期能有一些關於詳細評估結果之初步掌握，因此研究將採取耐震評估作業中花最少經費的階段(校舍普查作業)與詳細評估之評估結果，並以鋼筋混凝土建築、單邊走廊且廊外無柱、無地下室、詳細評估顯示為已送出之資料做為此模型資料選擇之條件篩選，藉由已存在校舍資料庫內之大量資料，以找出期間關聯性達成研究之預測目的。



6.1 資料前處理

本預測模型其主要功能為輔助土木技師與專業人員於執行校舍詳細評估前期即能粗估校舍之可能破壞構件與其破壞模式。校舍耐震評估最準確的階段為進行詳細評估作業，但進行校舍詳細評估須歷經非常多之流程如圖 1.2，花費很多時間，藉由此模型之預測輔助，將有助於土木技師與專業人員進行作業時能先有初步結果之掌握，與校方或主管機關進行洽談時也能更加迅速。

依本章 3-2 節敘述，本研究需使用校舍資料庫內校舍普查與詳細評估這兩個資料表，經連結兩表後發現，資料庫內並無記錄校舍跨距之資料，因此本研究依據此規則「校舍長度/(一樓教室柱/2-1)」計算創造出一新欄位，另外校舍之長度與深度亦有填反之現象，因此本研究透過人工方式將此現象做修正。

本模型採取耐震評估作業中花最少經費的階段(校舍普查作業)與詳細評估之評估結果，並以鋼筋混凝土建築、單邊走廊且廊外無柱、無地下室、詳細評估顯示為已送出之資料做為此模型資料選擇之條件篩選，資料經條件篩選完後，開始

進行探勘模型之屬性欄位選擇，本模型挑選之屬性集如圖 4.3，選擇之欄位分為輸入變數與輸出變數，本模型之輸出變數為破壞構件與其破壞模式，輸入變數則是我們用來預測輸出變數之所用變數，表 4.8 為本預測模型屬性集內之屬性介紹與選擇說明。

- | | |
|----------------|--------------------|
| • 校舍長度 | • 用途係數 |
| • 校舍深度 | • X 向 1F 三面圍束磚牆斷面積 |
| • 樓層數 | • 有無三面圍束磚牆 |
| • 地上層樓地板總面積 | • 有無四面圍束磚牆 |
| • 一樓教室柱根數 | • 有無磚牆 |
| • 一樓走廊外柱斷面積和 | • 同時有三面和四面圍束磚牆 |
| • 475 年設計地表加速度 | • 有無 RC 牆 |
| • 校舍垮距 | |



校舍長度 中小學校舍中之典型校舍其建築形式較為固定且普遍存有耐震能力較弱之現象，因此本研究推測在這些典型校舍中，某些特定尺寸(校舍長度)之校舍應與其極限強度下會破壞之構件與其破壞模式存有一定之相關性。因此本研究將此屬性納入屬性集做為本探勘模型之輸入屬性。

校舍深度 中小學校舍中之典型校舍其建築形式較為固定且普遍存有耐震能力較弱之現象，因此本研究推測在這些典型校舍中，某些特定尺寸(校舍深度)之校舍應與其極限強度下會破壞之構件與其破壞模式存有一定之相關性。因此本研究將此屬性納入屬性集做為本探勘模型之輸入屬性。

樓層數 加入此屬性，其構想為假設這些校舍中應該有集中幾層樓之建築與其極限強度下會破壞之構件與其破壞模式存有一定之相關性。因此本研究將此屬性納入屬性集做為本探勘模型之輸入屬性。希望藉由其中隱藏之關聯性達到目標之預測。

地上層樓地板總面積 校舍之樓地板面積亦屬於建築形式之一環，因此可視為該校舍之特有形式，本研究假設在典型校舍中存有特定樓地板面積大小有其普遍對應之耐震能力或某些尺寸之樓地板面積可能存有接受補強較高的機率。

一樓教室柱根數 教室柱之根數間接可之校舍樓地板面積之大小或規模，因此本研究將之視為影響工程經費之因素之一，經將此屬性納入探勘模型中測試後，也確實發現此屬性對預測之正確率有一定之提升。

一樓走廊外柱斷面積和 台灣中小學校舍其建築形式較為固定，教室柱之多寡或其尺寸亦可視為建築形式中之一環，因此本研究推測在這些校舍中，教室柱應該存有某些特定總斷面積大小之校舍其耐震能力較差，因此本研究將此屬性納入屬性集做為本探勘模型之輸入屬性。

475 年設計地表加速度 由校舍所在位於查出其工址短週期設計水平譜加速度之 0.4 倍 ($0.4S_{DS}$)

校舍垮距 資料庫內並無記錄校舍跨距之資料，此屬性為本研究依據「校舍長度/(一樓教室柱/2-1)」計算創造出之一新欄位。

用途係數 該屬性為紀錄該校舍是否作為緊急避難使用，以決定其值為 1.25 或 1.5，為反映其因重要性所需之安全係數。如該校舍當初規劃為緊急避難使用，那其當初設計強度就會比較高，本研究設想不同之用途係數其可能受到之破壞模式可能有某種關係，因此將之加入本屬性集內。

X 向一樓三面圍束磚牆總斷面積 磚牆對校舍能提供部分耐震能力，本研究試過單獨放入三面圍束、四面圍束與同時放入三面與四面圍束磚牆，經分別試驗不同之屬性集，建立多個不同的模型，其結果以單獨使用三面圍束磚牆總斷面積做為輸入屬性，其預測之結果最佳。

有無三面圍束磚牆 此屬性為校舍資料庫不存在之欄位，因經測試許多屬性集與不同探勘方法發現磚牆對預測結果有一定影響力，但並不是所有校舍均有磚牆，因此本研究規劃創造出一新欄位，紀錄有無三面圍束磚牆，其內之值為

依照原三面圍束磚牆欄位內有值且非 0 或 NULL，其產生之新欄位，其值設為 1，相反則設為 0。

有無四面圍束磚牆 此屬性為校舍資料庫不存在之欄位，因經測試許多屬性集與不同探勘方法發現磚牆對預測結果有一定影響力，但並不是所有校舍均有磚牆，尤其是四面圍束磚牆，因此本研究規劃創造出一紀錄有無四面圍束磚牆之新欄位，期望能對此模型之預測目標提升一定之正確率，該屬性之值為依照原四面圍束磚牆欄位內有值且非 0 或 NULL，其產生之新欄位，其值設為 1，相反則設為 0。

有無磚牆 此屬性為校舍資料庫不存在之欄位，因經測試許多屬性集與不同探勘方法發現磚牆對預測結果有一定影響力，但並不是所有校舍均有磚牆，且磚牆有分三面圍束與四面圍束，經測試單獨存有三面圍束與四面圍束其預測結果均不同，因此本研究規劃創造出一新欄位，紀錄校舍有無磚牆，其內之值為依照原三面圍束與四面圍束磚牆欄位，只要任一欄位內有值且非 0 或 NULL，其產生之新欄位，其值設為 1，相反則設為 0。

同時有無三面和四面圍束磚牆 此屬性為校舍資料庫不存在之欄位，因經測試許多屬性集與不同探勘方法發現磚牆對預測結果有一定影響力，經測試校舍是否同時存有三面圍束與四面圍束磚牆對其預測結果，存有一定影響力。因此本研究規劃創造出一新欄位，紀錄校舍有無同時建有三面和四面圍束磚牆，其內之值為依照原三面圍束磚牆與四面圍束磚牆欄位內，同時有值且非 0 或 NULL，其產生之新欄位，其值設為 1，相反則設為 0。

有無 RC 牆 此屬性為校舍資料庫不存在之欄位，因經測試許多屬性集與不同探勘方法發現 RC 牆對預測結果有一定影響力，但並不是所有校舍均有 RC 牆，因此本研究規劃創造出一新欄位，紀錄有 RC 牆，其內之值為依照原 RC 牆欄位內有值且非 0 或 NULL，其產生之新欄位，其值設為 1，相反則設為 0。

6.2 資料探勘

本研究採用支援向量機進行資料探勘之建模，選擇此探勘方法，主因為本研究之預測目標，破壞構件及其破壞模式其資料類型為「名義」，名義之資料類型其最適合之探勘方法為支援向量機，因此本研究採用此探勘方法，使用前處理完之 897 筆資料並挑選屬性集內之欄位，將整體資料分為訓練區 70% 與測試區 30%，做為探勘模型之訓練方法。最後再挑一組結果較好之模型保留，並以另外挑選之獨立 50 筆校舍資料進行結果測試，測試結果敘述於第五章。本模型之訓練結果如表 6-1。

表 6-1: Result of the Crack Component Model

破壞元件	資料集	正確率
X 正向樑破壞	訓練集	76.97%
	測試集	82.14%
X 正向柱破壞	訓練集	86.52%
	測試集	86.90%
X 正向窗台柱破壞	訓練集	79.21%
	測試集	77.38%
X 正向 RC 牆破壞	訓練集	99.32%
	測試集	98.25%
X 正向磚牆破壞	訓練集	80.34%
	測試集	80.95%

6.3 結果

本探勘模型使用支援向量機並分別針對耐震詳細評估資料表之性能點狀態下最嚴重破壞樓層之主要破壞桿件及其破壞模式進行預測，其各項之預測正確率如下：

X 正向樑破壞 97.06%

X 正向柱破壞 100%

X 正向窗台柱破壞 94.12%

X 正向 RC 牆破壞 100%

X 正向磚牆破壞 82.6%

最後並根據每棟校舍於性能點狀態下，各項破壞構件其破壞模式進行正確率之計算，根據耐震詳細評估資料表內有五項破壞構件之欄位，因此本研究假設該校舍於每項破壞構件之欄位如預測正確，則加正確率百分之二十，共五項欄位，如各項接預測正確，則該校舍針對其性能點狀態下最嚴重破壞樓層之主要破壞桿件及其破壞模式之預測正確率為百分之百。其 34 棟總體校舍針對 5 種破壞構件及其破壞模式之平均預測正確率為 95.3%。本預測目標其欄位屬性為類別型態，因此無法以其他評估指標進行比較，一般評估指標只能用來評估數值型資料。34 棟校舍於破壞構件及其破壞模式之預測資料與結果如，X 正向樑破壞於表 5.6 中欄位名稱為 D_pXdestoryB，X 正向柱破壞於表 5.6 中欄位名稱為 D_pXdestoryC，X 正向窗台柱破壞於表 5.6 中欄位名稱為 D_pXwindowsC，X 正向 RC 牆破壞於表 5.6 中欄位名稱為 D_pXRCwall，X 正向磚牆破壞於表 5.6 中欄位名稱為 D_pXBwall，其餘之欄位說明如表 5.5。



第 7 章 校舍資訊與補強經費之關係模型

校舍建築物的補強經費對於全國校舍的補強工作來說是非常重要的。一個數值，和校舍的耐震能力並列因其直接表示了主管機關要讓該棟校舍符合耐震安全規範所需要的成本，教育部為了提升全國高中職及中小學校舍耐震能力，歷年來已經訂定及執行了各種評估與補強計劃。教育部將計畫委託國震中心執行，國震中心與教育部研擬商討後，再訂出執行計畫交由各縣市政府辦理。計劃的訂定需要資料的輔助，計劃的執行也需要定期的計劃報告以掌握計劃的執行進度與效力。全國校舍總數，經目前耐震資料庫中校舍普查作業已記錄之校舍數量統計共有 24930 棟，如果考慮尚未執行校舍普查作業之校舍，其實際校舍總數將更多。面對如此大量之校舍，進行總體或縣市為單位之經費估算，需耗費大量時間與人力進行估算與統計。且依照目前「高中職及國中小校舍結構耐震能力詳細評估作業規範」規定，主管機關須等到詳細評估之承攬廠商提送期末報告書，才能得知該校舍較精確之可能補強費用。由此可知，校舍補強經費的取得也是一個非常耗時且困難的流程，要提早在校舍進入詳細評估前就推估出校舍的補強經費，以往只能透過簡單的經驗公式推估，因此本研究還試著找出校舍的基礎設計參數與其補強經費間的關係模型，透過這個模型，應該可以從校舍的初步評估資料得到一個可靠度交高的補強經費推估值。

主管機關在提撥補強校舍的耐震能力計畫的經費時，是每個年度依據需求估算，編列該年度的固定經費，根據教育部公布九十九年九月十六日修正之國中小校舍「補強工程」之經費支用範圍及參考單價計價方式，「補強工程」費用，包含了「直接補強工程費」、「合理之間接修復費」、「工程管理費」、「補強設計及監造費」等費用支出。依公告之估價方式將校舍分為「已完成補強設計」與「尚未完成補強設計」，歸類為「已完成補強設計」之校舍，其補強工程經費依補強設計審查通過或補強設計報告之金額，教育部原則將據以如數核定補助經費。但其補強工程單價超出 4000 元/ m^2 者，則應經補強設計期末審查或特別審查通過並附審查意見表，始能獲教育部如數核定補助經費；否則將依單價 4000 元/ m^2 核定補助經費。歸類為「尚未完成補強設計」之校舍（凡未附補強設計審查意見表、未附補

強設計報告重點摘要部分、或佐證資料不齊全者，皆屬尚未完成補強設計)：其「補強工程」之參考單價計價方式依其總樓地板面積所屬級距範圍之不同，分別計算如表 7-1 所示，其中「計價面積」係由「實際面積」以無條件進位為 2 的倍數之方式轉換而成；因此計價面積為 2 的倍數且為整數。例如：實際面積 601 m^2 ，應轉換為計價面積 602 m^2 。

表 7-1: 補強工程參考單價之計價方式

建築物樓地板總面積	補強工程費用計算方式
不足 600 m^2	實際面積 $\times 4000\text{ 元}/\text{m}^2$
600 m^2 以上，不足 3600 m^2	計價面積 $\times (-0.5 \times \text{計價面積} + 4300)\text{ 元}/\text{m}^2$
3600 m^2 以上	實際面積 $\times 2500\text{ 元}/\text{m}^2$

此一補強工程參考計價方式的主要依據是校舍的樓地板面積，但是實際影響補強經費的因素應該不只是校舍的樓地板面積，建築物的牆柱設計、施工品質、建造年代等等都是可能的影響因素，因此用此計價方式所得到的補強費用預估值是較為不可靠的數字，而此一預算編列的方式，編列的年度預算難以在事後追加，使用可靠度不足的數字作為參考，會讓編列的預算與實際需求失準，而且每年固定的預算已經會讓每個年度能夠補強的校舍數量有所限制，對於不知何時會再發生的地震威脅，絕對是希望能夠儘快將最危險的校舍找出並做出對應處置，校舍耐震能力補強的經費編列和分配的影響就非常大大，不過實際上在編列預算時，校舍都還沒進行詳細評估、也還沒有建議的補強方案，因此只能透過參考資料和初步評估的資料來估算，如果可以在編列預算時，就知道每間校舍的耐震能力、補強可能需要的經費，那麼主管機關在編列和分配預算的時候就可以非常有效率的來編列。

7.1 資料前處理

本分析其構想為使用校舍評估階段初期之已調查資料來針對評估階段後期結果做預估。校舍耐震之補強工程經費須等到工程完成且承攬廠商提送竣工報告才能得知最後實際施工所花經費，因研究目標為達成前期預測之目的，所以需挑選

補強施工以前之補強階段資料。觀校舍耐震評估與補強之流程如圖 1.1，可知進行補強施工之前尚須進行校舍普查、初步評估、詳細評估與補強設計，經了解各階段評估作業之內容與用途，校舍普查執行之人員為各大專院校之土木系學生協助執行，且其所記錄之資料也較其他各階段為少，補強設計為補強施工之上一階段，根據「高中職及國中小校舍結構耐震能力補強設計作業規範」規定，補強設計成果報告書中須編製工程預算書，因此即可於補強設計階段得知該補強工程之所需經費。為達研究之預測目的，本研究將採用初步評估與詳細評估（不含耐震補強方案）結果做為本探勘模型之選擇資料，並以典型校舍、鋼筋混凝土建築、單邊走廊且廊外無柱、無地下室、補強工程經費介於 100 至 1000 萬之間、初步評估與詳細評估均顯示為已送出之資料做為此模型資料選擇之條件篩選。

本模型採用初步評估與詳細評估（不含耐震補強方案）結果做為本探勘模型之選擇資料，並以典型校舍、鋼筋混凝土建築、單邊走廊且廊外無柱、無地下室、補強工程經費介於 100 至 1000 萬之間、初步評估與詳細評估均顯示為已送出之資料做為此模型資料選擇之條件篩選。模型經資料選擇與前處理之後建立資料集，下一步即是進行模型屬性之選擇，建立屬性集，本模型挑選之屬性集如下：

- 一樓樓地面積
- 柱等效強度
- 軟弱層顯著性
- 裂縫鏽蝕滲水等程度
- 短柱嚴重性
- 調整因子
- 二樓樓地板面積深
- 二樓樓地板面積長
- 樓層數
- 二樓以上樓地板面積
- 一樓教室柱根數
- 一樓走廊外柱斷面積和
- 非結構牆
- X 正向樑破壞
- X 正向柱破壞
- X 正向窗台柱破壞
- X 正向磚牆破壞
- X 正向性能點之屋頂最大位移
- X 正向性能點之等效基本週期
- X 正向性能點之基底剪力

- 校舍耐震容量需求比

以下為本預測模型屬性集內之屬性介紹與選擇說明：

一樓樓地面積 依據教育部現行之補強工程計價公式如表 7-1，其經驗公式以樓地板面積為主要參數，而其可靠度雖然較差，但仍然是主管機關一個重要的參考數值，本分析所建立之模型為捕強工程經費與校舍設計參數間的關係，而校舍之樓地板面積也是一個非常重要的設計參數，因此本分析將此屬性拿來作為建立此關係模型之輸入屬性。

柱等效強度 柱之等效強度 T_{AC} 其計算公式為：

$$T_{AC} = (4 + 1.8NF) \times ClaAc + (2.4 + 1.08NF) \times CorAc + 2.6 \times InsAc \quad (7-1)$$

其中 NF 為樓層數， $ClaAc$ 為一樓教室柱總斷面積， $CorAc$ 為一樓走廊外柱總斷面積， $InsAc$ 為一樓隔間柱總斷面積，其公式為根據三種類結構柱之單位斷面積極限剪力牆度計算公式 [12] 之加總，本分析將此屬性加入探勘模型之最初原因為該公式內有教室柱、走廊外柱與隔間柱於一樓的總斷面積，並且因補強施工其所補強之面積愈大其所花錢就愈多。本分析經建立多組模型與各種屬性集之測試均發現該屬性對模型之預測結果有一定影響。

軟弱層顯著性 [13] 若結構物之一樓因為使用性等考量，而使得二樓以上 RC 牆或磚牆於一樓中斷，致使一樓之極限層剪力強度與勁度降低，將造成地震力作用時變形集中，以致於韌性用盡，建築物就發生軟弱層破壞。故本表格依據牆體中斷的程度折減其對應之耐震能力，若 2/3 以上牆體中斷，則耐震能力折減為 0.8 倍；若 1/3 至 2/3 之牆體中斷，則耐震能力折減為 0.9 倍；若 1/3 以下之牆體中斷，則不折減其耐震能力。依此敘述得知軟弱層顯著性會影響結構物之耐震能力，結構物之耐震能力愈低，極可能所需補強之工程花費就愈高，因此本分析將此屬性納入本屬性集。

裂縫鏽蝕滲水等程度 [13] 鋼筋混凝土構材若具有裂縫，代表混凝土品質不良或強度不足；保護層不足等因素使得鋼筋鏽蝕膨脹，鋼筋鏽蝕將會降低構材之強度，鋼筋鏽蝕膨脹亦會導致混凝土剝落，並加速鋼筋鏽蝕的程度，這些因素都會影響結構物的耐震安全，故以結構物整體之裂縫鏽蝕滲水等程度作為調整項目。若稍有裂縫鏽蝕滲水等情形，則耐震能力折減為 0.95；若裂縫鏽蝕滲水等情形較為嚴重，則耐震能力折減為 0.9；若無，則不折減其耐震能力。根據上述，建物若有裂縫鏽蝕滲水等情形會影響耐震能力，耐震能力愈差就有可能提升所需花費之工程經費，因此本分析將之納入輸入屬性。

短柱嚴重性 [13] 一般老舊校舍之柱箍筋間距多為 20cm 至 30cm 左右，其剪力強度不高，且老舊校舍於設計時假設為純梁柱系統，並沒有考慮教室窗台及樓梯廁所等牆壁開氣窗所造成之短柱效應，然而這種短柱效應將會使得剪力容量不足之柱於地震時發生非預期之剪力破壞，導致結構韌性不足，若該校舍有過多之柱受到短柱效應之影響，將易造成校舍瞬間倒塌。故若校舍因窗台或氣窗造成短柱現象之柱根數達到全部柱根數之 50% 以上，則耐震能力折減為 0.9 倍，若不足 50% 則不予折減其耐震能力。值得注意的是，短柱嚴重性具有方向性，故評估時只需考慮評估方向之短柱比率是否超過一半即可，另一方向開窗等因素造成之短柱效應不需考慮。根據上述得知短柱現象為影響校舍耐震能力之原因之一，因此將之納入輸入屬性。

調整因子 [13] 典型校舍初步評估表之六項調整因子調查項目，其功能為對初步評估耐震能力產生增減，包括：平面及立面對稱性、軟弱層顯著性、裂縫鏽蝕滲水等程度、變形程度、平面耐震性以及短柱嚴重性，定義 q_1 至 q_6 分別代表六項調整因子，其中軟弱層顯著性、裂縫鏽蝕滲水等程度和短柱嚴重性三項調整因子因對耐震能力之影響較為直接，所以也分別被納入本分析之輸入屬性，而除此六項調整因子外，還有定義一整體調整因子 Q 為上述六項調整因子之乘積，整體調整因子代表所有調整因子對於耐震能力的折減或增加的幅度，其物理意義為校舍的現況對於校舍耐震能力影響的程度，對於校舍耐震能力來說也是很重要的一个參數，且經測試後證實該屬性對預測之正確性有提升，因此本分析也將其納入輸入屬性。

二樓樓地板面積深 依據教育部現行之補強工程計價公式如表 7-1，其公式以樓地板面積為主要參數，又面積之算法為長乘以寬（深），因此亦可將它視為影響校舍補強工程經費的因素之一。本分析所預測之目標為實際施工之工程總經費，與計價公式所估經費有直接關聯性，因此本分析將此屬性拿來作為本預測模型之輸入屬性。

二樓樓地板面積長 依據教育部現行之補強工程計價公式如表 7-1，其公式以樓地板面積為主要參數，又面積之算法為長乘以寬（深），因此亦可將它視為影響校舍補強工程經費的因素之一。本分析所預測之目標為實際施工之工程總經費，與計價公式所估經費有直接關聯性，因此本分析將此屬性拿來作為本預測模型之輸入屬性。

樓層數 樓層數愈高其總樓地板面積愈多，其總樓地板面積愈多，依補強工程計價公式如表 7-1。其工程經費預算就可能愈多，因此樓層數亦可視為影響工程經費之影響原因之一。

二樓以上樓地板面積 依據教育部現行之補強工程計價公式如表 7-1，其公式以樓地板面積為主要參數，本分析所預測之目標為實際施工之工程總經費，補強工程計價公式為經費預算之計算方式，估算出來之經費預算往往與實際施工所花經費不太一樣，預算可能高於實際施工也可能低於實際施工，但因為目標都是估算工程經費，因此本分析將此屬性拿來作為本預測模型之輸入屬性。

一樓教室柱根數 由校舍教室柱之數量可以間接得知校舍樓地板面積之大小或規模，且柱子的數量可以間接代表校舍乘載能力的大小，因此本分析將之視為影響工程經費之因素之一，經將此屬性納入探勘模型中測試後，也確實發現此屬性對預測之正確率有一定之提升。

一樓走廊外柱斷面積和 走廊外柱總斷面積愈大其意義可能有二：其一，斷面積愈大可能代表樓地板面積愈大，而根據補強經費估算的經驗公式，樓地板面積愈大其施工所需經費即可能愈高。其二，如果校舍樓地板面積大小一樣，但其廊外柱斷面積和比較大，代表其走廊外柱可能斷面積比較大，如果要進行

擴柱補強工法，其工程所花經費就會比較高。綜合以上原因，本分析即嘗試將此屬性納入探勘模型中，並且經測試後效果不錯因此保留此屬性。

非結構牆 此屬性為記錄該校舍，有無非結構牆，並於分析時所建立模型是否有將該非結構牆模擬成其他等值桿件進入結構模型內。因有無非結構牆可能會影響校舍之耐震能力，因此將之加入輸入變數。

X 正向樑破壞、X 正向柱破壞、X 正向窗台柱破壞、X 正向磚牆破壞 此四屬性為詳細評估資料表中之記錄欄位，記錄詳評後校舍可能優先破壞的構件以及其破壞的形式，此分析中根據實際詳細評估資料中的破壞趨勢選擇了包括 X 正向樑、X 正向柱破壞、X 正向窗台柱以及 X 正向磚牆四種構件的破壞模式，區分為無破壞、剪力破壞、撓剪破壞、撓曲破壞四種破壞形式，不同之破壞形式其需補強的方式會有所不統，其所需經費也不一樣，因此本分析將之列為本模型之輸入屬性。

X 正向性能點之屋頂最大位移 [14] 進行詳細評估時，若採用推垮分析，則該欄位則須依 X 及 Y 方向分別輸入屋頂最大位移值。本分析僅挑選採用推垮分析之校舍進行建模，因屋頂之最大位移為影響校舍耐震能力的因素之一，因此也將該屬性列入。

X 正向性能點之等效基本週期 [15] 基本週期為結構物等效單自由度系統的動力參數（性能目標地表加速度），此單自由度系統在性能目標地表加速度的設計地震作用下，其動力反應將是已設定的性能需求，因此基本週期亦是影響耐震能力評估結果的因素之一。

X 正向性能點之基底剪力 基底剪力為側推分析中計算容量曲線之重要參數之一，校舍之耐震能力可能影響補強之工程經費，因此本分析嘗試將此屬性放入本模型中進行測試，其結果對模型預測之正確率有提升，因此將它保留至屬性集之中。

校舍耐震容量需求比 [16] 校舍建築物進行耐震初步評估後，得到一耐震指標分數 I_s ，此耐震指標可作為是否需進入下一階段評估工作的參數，同樣地，經

由專業人員完成詳細評估後，得到建築物長短向最大可抵抗之地表加速度，除以該校舍 475 年回歸期設計地表加速度，稱之耐震容量需求比（Capacity Demand Ratio, CDR），此數值若小於 1 時，則判定該棟建築須進行補強；反之，該棟建築耐震能力暫無疑慮，但仍須定期檢視。根據上述得知 *CDR* 可視為該校舍之耐震能力與正常安全值之距離，如果數值小於 1 且離 1 愈遠，代表需替該校舍提升的耐震能力愈多，所需花的經費就愈多，因此亦可視為影響補強工經費之因子。

7.2 資料探勘

本分析採用類神經網路與 CHAID 進行資料探勘之建模，選擇此二種探勘方法，主因為本分析之預測目標，總工程經費其資料類型為「連續」，連續之資料類型其最適合之探勘方法為類神經網路，另外屬於決策樹方法分支之 CHAID 法經初步測試應用於本分析之預測目標，其效果不錯，因此本分析以此二種探勘方法，使用前處理完之 197 筆資料，將整體資料採亂數選取方式分為 70% 為訓練區與 30% 為測試區，先行根據既有知識挑選資料集內之欄位並配合使用試誤法直到找出最佳做為探勘模型之屬性集，做為模型訓練方法。最後再挑一組結果較好之模型與屬性集保留。本模型之訓練結果如表 7-2，因本模型之整體資料較少，其結果之線性相關程度目前呈現比較低，但模型最後進行另外 50 筆校舍資料測試時會將 197 筆資料全數作為模型之訓練集，其線性相關之結果將大為提升。

表 7-2: Result of the Retrofit Cost Model

	ANNs		CHAID	
	Training	Testing	Training	Testing
R	0.689	0.705	0.708	0.736

7.3 結果

本分析用來判端模型優劣的指標為 R^2 和 MAPE，MAPE (Mean Absolute Percentage Error) 的公式如下：

$$MAPE = \frac{\sum |y_i - \hat{y}_i|}{N} \times 100\% \quad (7-2)$$

其中 N 是資料的總數， y_i 是使用探勘得到的關係模型所求得的預測值， \hat{y}_i 則是該屬性的實際值，以此分析為例， y_i 就是該棟校舍所需要的補強經費，MAPE 用來表示模型預測值的誤差平均，數值越小表示模型品質越好，本分析最後選擇的模型是使用類神經網路來建立，並且不分訓練測試集，其 MAPE 為 31.8%， R^2 (相關係數) 為 0.9004，可以發現 R^2 的表現很好，表示此模型有確實的呈現校舍參數與其補強費用間的關係趨勢，然而 MAPE 表現較差，顯示仍然有些影響補強經費的因素尚未被找出。



第 8 章 結論與討論

8.1 結論

Conclusion

8.2 未來展望

Feature



参 考 文 献

- [1] D. T. Larose, *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2005.
- [2] B. S. S. C. (US) and A. T. Council, *NEHRP guidelines for the seismic rehabilitation of buildings*, vol. 1. Federal Emergency Management Agency, 1997.
- [3] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” *AI magazine*, vol. 17, no. 3, p. 37, 1996.
- [4] Applied Technology Council (ATC), *Seismic evaluation and retrofit of concrete buildings*, vol. 1. Report No. SSC 96-01: ATC-40, 1996.
- [5] C. Shearer, “The crisp-dm model: the new blueprint for data mining,” *Journal of data warehousing*, vol. 5, no. 4, pp. 13–22, 2000.
- [6] J. A. Nelder and R. W. M. Wedderburn, “Generalized Linear Models,” *Journal of the Royal Statistical Society. Series A (General)*, vol. 135, no. 3, pp. 370–384, 1972.
- [7] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT’92)* (D. Haussler, ed.), (Pittsburgh, PA, USA), pp. 144–152, ACM Press, jul 1992.
- [8] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, “Support vector regression machines,” *Advances in neural information processing systems*, vol. 9, pp. 155–161, 1997.
- [9] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

- [10] T. Zhang, R. Ramakrishnan, and M. Livny, “Birch: an efficient data clustering method for very large databases,” in *ACM SIGMOD Record*, vol. 25, pp. 103–114, ACM, 1996.
- [11] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, 1967.
- [12] 蘇耕立, “灣中小學校舍結構耐震能力初步評估方法之探討,” Master’s thesis, 國立台灣大學土木系, jul 2008.
- [13] “國民中小學典型校舍耐震能力初步評估法,” Tech. Rep. NCREE-03-049, 國家地震工程研究中心, dec 2003.
- [14] “全國中小學校舍結構耐震評估與補強資料庫建置 (一),” Tech. Rep. NCREE-08-035, 國家地震工程研究中心, dec 2008.
- [15] “鋼筋混凝土建築物耐震能力詳細評估分析方法 (推垮分析),” Tech. Rep. NCREE-09-015, 國家地震工程研究中心, jun 2009.
- [16] “全國中小學校舍結構耐震評估與補強資料庫建置 (二),” Tech. Rep. NCREE-09-026, 國家地震工程研究中心, dec 2009.

