



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：陳鴻明博士

中華民國一零二年七月七日

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有什么可能的分析方向，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。



ABSTRACT

Distributed querying and monitoring systems have been widely studied in recent years. These systems aim to maintain data sources, such as data set or log files, and allow users to query over those data sources. When the data sources are highly related and users only care some statistic results, like the sum or the average, it is consumed to transmit all data sources via the network. To minimize the network consumption, in-network aggregation technique is proposed. However, this technique is subject to some known attacks, such as the injection attack and the pollution attack. Prior works only considered the settings that data sources are trusted while the network is not. We study the way to relax the limitation and guarantee the aggregate queries robust to malicious or faulty data sources (also called polluted data sources).



誌

謝

首先誠摯的感謝指導教授陳明明博士，老師悉心的教導使我得以一窺 WSN 的深奧，不時的討論並指點我正確的方向，使我在這些年中獲益匪淺。老師對學問的嚴謹更是我輩學習的典範。本論文的完成另外亦得感謝老師們大力協助。因為有你們的體諒及幫忙，使得本論文能夠更完整而嚴謹。兩年裡的日子，實驗室裡共同的生活點滴，學術上的討論、言不及義的閒扯、讓人又愛又怕的宵夜、趕作業的革命情感、因為睡太晚而遮遮掩掩閃進實驗室……，感謝眾位學長姐、同學、學弟妹的共同砥礪，你/妳們的陪伴讓兩年的研究生活變得絢麗多彩。最後絕對不能忘記最了解、最支持我的家人——我的父親、母親及姊姊，在我喪失動力之時，隨時都能給予我心靈上無窮盡的關心與鼓勵，讓我有勇氣堅持到最後，完成研究的旅途。還有很多曾經幫助過我的朋友，因為有大家的幫助，我才能有今天的成果。想要感謝的人真的太多太多，就只有感謝上天了！



目 錄

論文摘要	I
Abstract	II
誌謝	III
目錄	IV
圖目錄	VI
表目錄	VII
1 緒論	1
1.1 動機與目的	1
1.2 研究方法	2
1.3 論文架構	3
2 相關研究	5
3 校舍耐震資料庫	6
3.1 典型校舍	6
3.2 收集範圍	7
3.2.1 初步評估	8
3.2.2 詳細評估	8
3.2.3 補強設計	9
3.2.4 補強工程	10
4 資料探勘	11
4.1 Generalized Linear Model	11
4.2 Support Vector Machine	12
4.3 Artificial Neural Networks	12

4.4	Genetic Programming	13
4.5	Weighted Genetic Programming	13
4.6	K-means	13
4.7	Two-Step Classification	13
5	耐震能力與校舍設計關係模型	14
5.1	Is 值與校舍設計之關係模型	14
5.1.1	資料前處理	15
5.1.2	資料探勘	15
5.1.3	結果	15
5.2	CDR 值與校舍設計之關係模型	15
5.2.1	資料前處理	15
5.2.2	資料探勘	15
5.2.3	結果	15
5.3	D _{isR} 值與校舍設計之關係模型	15
5.3.1	資料前處理	16
5.3.2	資料探勘	16
5.3.3	結果	16
6	破壞構件預測	17
7	補強經費預測	18
8	結論與後續工作	19
8.1	Future Work	19
	參考文獻	20
	授權書	21

圖 目 錄

圖 2.1	The diagram of “prototypical PHI query”	5
圖 3.1	典型校舍	6



表 目 錄

表 1.1	The relation of aggregation overhead between different techniques . . .	4
-------	---	---



第 1 章 緒論

1.1 動機與目的

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，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，然而學校校舍的安全問題也因此浮現，學校的校舍建築物，除了在平時會有大量的學生在內使用外，也有很多校舍是兼作為緊急時期的避難或安置所，因此其安全性之需求應高於其他一般建物，然而有很多的校舍建築屋齡已經很大，以現今的建築法規來看，其耐震能力也顯得不足，因此教育部在九二一大地震後，便成立了校舍耐震能力補強計畫並與國家地震工程研究院合作執行，目標在找出所有安全性有疑慮的學校校舍，補強其安全性，甚至是拆除重建，而在計畫執行的過程中，產生了大量的校舍相關的資料，這些資料主要為校舍的設計參數和評估資料，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，此一資料庫收集了全台灣兩萬多棟校舍的設計與評估資料，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，其中應當還有各種隱含的知識，難以由人工觀察判讀以取得。

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，根據其取得的知識形式，可以分為四種：迴歸、分類、分群、關聯式法則。由前段敘述可以得知，校舍耐震資料庫內的資料量非常多，其中隱含的知識難以直接由人工觀察取得，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。

在各式建築物結構中，校舍為結構形式上較具有規則性之一類，大多數的校舍建築之設計形式相近，為一字型、外有走廊、隔間形式類似，因此可以將建築物原本複雜多樣化的設計、尺寸等資訊轉為數個代表屬性，此一特性讓校舍建築物整體特性預測模型的建置變成可能，可以使用各種軟式運算 (heuristic computing) 技術來建置預測模型。台灣的國家地震工程研究中心 (NCREE, National

Center for Research on Earthquake Engineering) 已建立有一校舍耐震資料庫，此一資料庫收集有全台灣約兩萬棟校舍的各種資料，除了學校和校舍的基本資料如學校位置，校舍用途、使用人數等，還收集了各校舍的結構、設計資訊，如梁柱尺寸 (beam column design)、數量、設計形式 (design pattern)、樓層數、教室分布等，這些資料屬性皆為針對校舍特定建築形式之特性所定。

本研究之目的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，針對校舍耐震資料庫挖掘隱藏於其中的隱含知識。

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，資料探勘技術依照其特性，可以分為迴歸、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隱含於校舍耐震資料庫中的知識，並且以預測模型的形式呈現，分別為：

- 校舍耐震能力預測模型
- 校舍優先破壞構件預測模型
- 校舍補強經費預估模型

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，根據調查的資績建立完整的結構數值模型，並使用非線性的分析軟體分析，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，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，本研究還建立一個預測模型，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

網路發展興盛至今，小至個人，大至政府單位與各機關組織，都相當仰賴網路的使用，但許多人仍然對資安危機意識較低，針對資訊安全產品的投資也相對較少，加上對於資訊安全軟體工具缺乏有系統的整理，以致於未能有效運用。為此，本手冊蒐集整理相關開放源碼（Open Source）的資訊安全軟體工具，並透過專業人員實際操作演練，加以彙整並集結成冊，希冀透過本手冊的幫助，不僅能給予初學者對於資安工具軟體初步認識，也讓資訊從業人員在資訊安全工具上能有更多的選擇與應用。

資安開放源碼軟體的發展，往往會公開其發展技術及運用的原理，配合程式碼的開放，使得開放源碼軟體具有相當大的彈性，並根據個人使用情況所需，進行軟體的編修與整合，以求適應各種作業環境所需。使用開放源碼軟體所需負擔的金錢成本，遠低於商業付費軟體，可降低企業組織對資訊科技產品的部分支出，不需要過度仰賴軟體製造商的技術支援與更新，也能減少相對應的軟體開發時程。由於目前多數的資安開放源碼軟體的開發多為國外組織，因此較缺乏中文化介面，且部分軟體工具的使用，需要具備相當程度的專業知識，並非人人皆可輕易上手。本手冊擬透過中文化的工具介紹，減緩國內使用者入門的負擔。

由於現今網路環境日益複雜，遭受網路攻擊的事件層出不窮，網路安全越來越受到各界重視。網路掃描是網路安全的根本，也是攻擊者對目標主機進行攻擊的首要步驟，因此，了解網路掃描的攻擊與防禦，將有助於網路管理者提升網域的安全管理。此外，網路流量代表所有網路訊息的傳送，能提供管理者即時了解網路狀況，藉此檢視網路情況正常與否。本手冊將針對以上兩類的開放源碼軟體，逐一介紹其功能、安裝、操作與軟體評比，令讀者對相關的資訊軟體能有所了解，並進一步應用於資訊安全的監測與控管。以下即對網路掃描及流量監控兩大類軟體，進行整理與原理說明。

如表 1.1 所示。

表 1.1: The relation of aggregation overhead between different techniques

	Space usage of root aggregator	Communication overhead	Query requirement
Traditional warehouse	n	$O(n)$	$O(n)$
AM-FM sketch technique	$\log a$	$O(\log n)$	$O(a \log n)$
“prototypical PHI query”	$\log a$	$O(\log n)$	$O(\log n)$

第 2 章 相關研究

作業系統指紋辨識的方法，可分為主動式作業系統指紋辨識（Active OS Fingerprinting）與被動式作業系統指紋辨識（Passive OS Fingerprinting）。主動式作業系統指紋辨識，主動對目標主機送出自製的探測封包，並根據回傳的反應做判斷依據，軟體工具 Nmap 與 Xprobe2 即屬於此類。Nmap 主要控制 TCP 的參數值，做為探測用封包；Xprobe2 則是著重於送出 ICMP 封包，利用邏輯樹斷定作業系統的類型。被動式作業系統指紋辨識是監聽網路上目標主機的封包往來做為判斷的依據，P0f 即屬於被動式，相對於主動式作業系統指紋辨識較不易被人察覺。不論是主動式或被動式的作業系統指紋辨識，皆利用 TCP/IP 堆疊進行辨識，包括封包存活時間（time to live, TTL）、Window Size、最大分割大小（Maximum Segment Size）、不分段標記（Don't Fragment flag）、Window Scale Option 等，因為不同的作業系統的 fingerprint 有所不同，所以可做為判定作業系統的依據。如圖 2.1 所示。

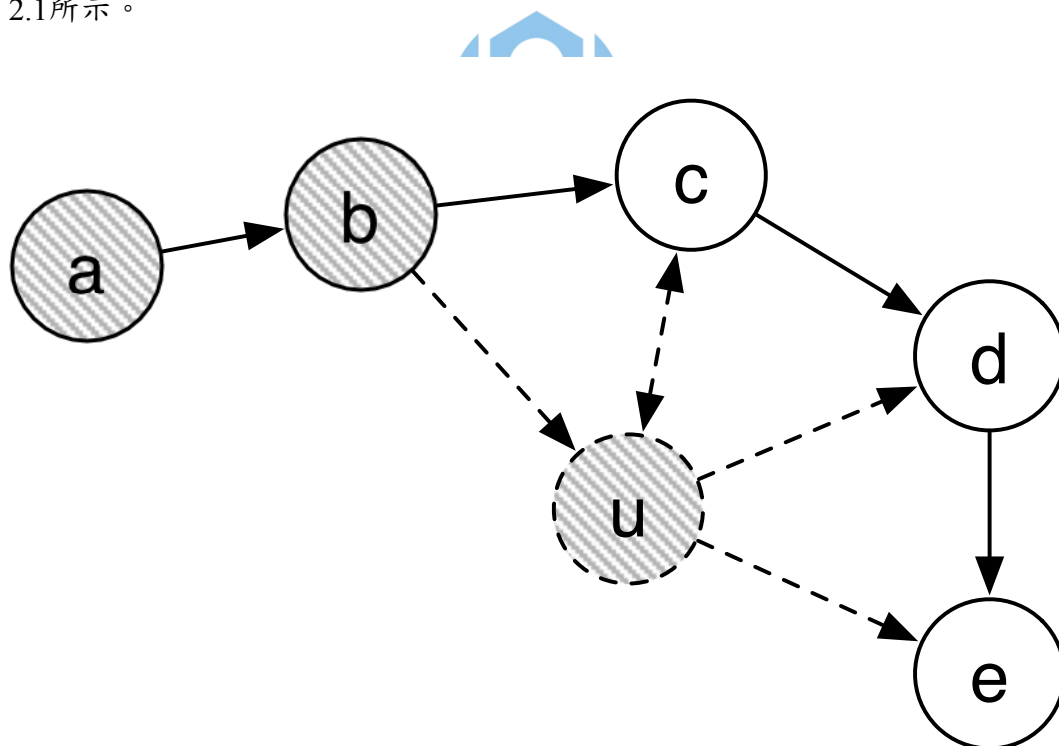


圖 2.1: The diagram of “prototypical PHI query”

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。

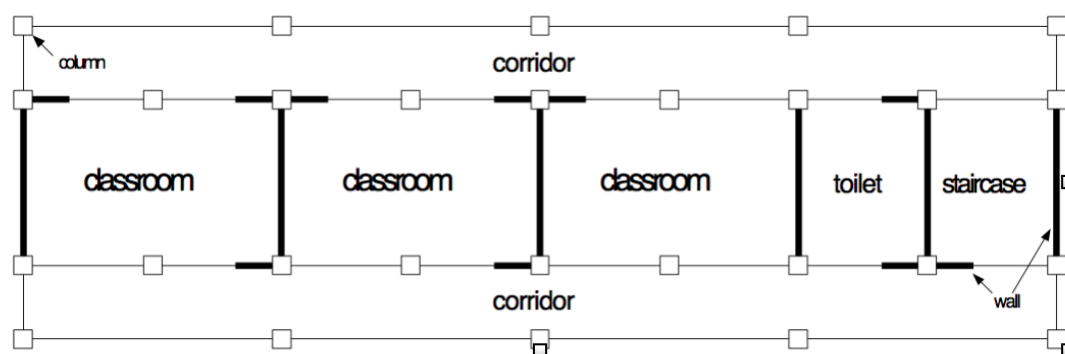


圖 3.1: 典型校舍

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（NCREE）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 for text

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 [1] 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)。

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

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

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

資料探勘技術方法繁多，Fayyad [2] 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。

本研究後期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，以下分別介紹各種使用的的分析方法：

4.1 Generalized Linear Model

廣義線性模型是由 Nelder and Wedderburn [3] 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 X 與 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 XY 關係模型如下：

$$g(E(y)) = x\beta + O, y F \quad (4-1)$$

$g(\cdot)$ 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 XY 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 XY 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。

4.2 Support Vector Machine

SVM 最早是 BOSER [4] 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., [9] 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。

4.3 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 [5] 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)，輸入層 (input layer) 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 (hidden layer) 的不斷自我更新學習的模型處理過後，在輸出層 (output layer) 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。

4.4 Genetic Programming

4.5 Weighted Genetic Programming

4.6 K-means

As proposed by MacQueen [6], 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.



4.7 Two-Step Classification

Based on of the massive volume of basic data for school build- ings in the database, this study chooses two-step clustering method. The basic concept was first proposed by Zhang, Ramakrishnan and Livny [7] 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.

第 5 章 耐震能力與校舍設計關係模型

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」($D_{is}R$) 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 、 $D_{is}R$ 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。

5.1.1 資料前處理

5.1.2 資料探勘

5.1.3 結果

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

5.2.2 資料探勘

5.2.3 結果

5.3 $D_{is}R$ 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 $D_{is}R$ 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 破壞構件預測



第 7 章 補強經費預測



第 8 章 結論與後續工作

本論文蒐集了各類資訊安全工具軟體，目的是為了讓更多使用者了解資訊安全的重要性，以及如何更有效的運用網路資源。透過一系列的資訊安全基礎知識及專有名詞解釋，搭配軟體的安裝及實作步驟，讓初級使用者能更容易跨越資訊安全議題的門檻，對駭客攻防與妨礙相關知識有更深的了解。

對進階使用者而言，本手冊也針對開放源碼工具做介紹，大部分工具都有釋出其原始碼，並歡迎有能力的使用者開發出更完善的程式。另外，使用者也可以結合不同功能性的軟體，自行開發出一套符合其需求的軟體，例如利用作業系統辨識工具搭配弱點掃描工具，能夠更快的找出目標主機的系統漏洞，以發揮 1+1 大於 2 的功效。

8.1 Future Work

使用安全工具軟體開創一個美好和諧的社會。



參 考 文 獻

- [1] Applied Technology Council (ATC), *Seismic evaluation and retrofit of concrete buildings*, vol. 1. Report No. SSC 96-01: ATC-40, 1996.
- [2] 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.
- [3] 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.
- [4] 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, July 1992.
- [5] 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.
- [6] 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.
- [7] 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.

