



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。



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	V
表目錄	VI
1 緒論	1
2 Preliminaries	4
3 研究方法	5
4 實驗設計	6
5 實驗結果與分析	7
6 結論與後續工作	8
6.1 Future Work	8
參考文獻	9
授權書	10



圖 目 錄

圖 2.1	The diagram of “prototypical PHI query”	4
-------	---	---



表 目 錄

表 1.1	The relation of aggregation overhead between different techniques . . .	3
-------	---	---



第 1 章 緒論

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。

根據資料探勘的四種知識形式，以及校舍耐震補強計畫的需求，反推回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，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，基於這些可能取得的知識假設，經過不斷的分析並測試，根據取得結果的可靠度來作篩選，最後建立四種可以幫助校舍耐震評估作業進行的預測模型，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，根據調查的資續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。

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 [1]。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 [2]，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路的安全管理。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 [3-6]。

Insecure.org 網站曾於 2003 年及 2006 年間調查各使用者喜愛的工具軟體，其中 2006 年收到 3,243 位受訪者的回覆，受訪對象涵括各界對資安工具有持續研究與發展的學者及廠商，包括 Insecure.org 自身、研究網際網路議題的機構、發展開放源碼軟體的組織，與其他著名的資安網站（如 Open Source Security、Honeypots 和 IDS Focus 等），並根據調查結果選出前 100 大網路安全工具（Top 100 Network Security Tools）。本手冊從中篩選了數套較廣泛應用且屬於開放源碼的工具軟體進行蒐集整理，以提供使用者參考學習使用。以下針對各工具軟體予以介紹相關的資訊安全基礎知識（包括專業術語解釋、專有名詞解析）、軟體安裝與使用方式，以及防駭相關知識。使用者將能透過本手冊掌握並熟悉更多相關熱門工具，且對資訊安全攻防技術有更進一步的認識。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”	$\log a$	$O(\log n)$	$O(\log n)$

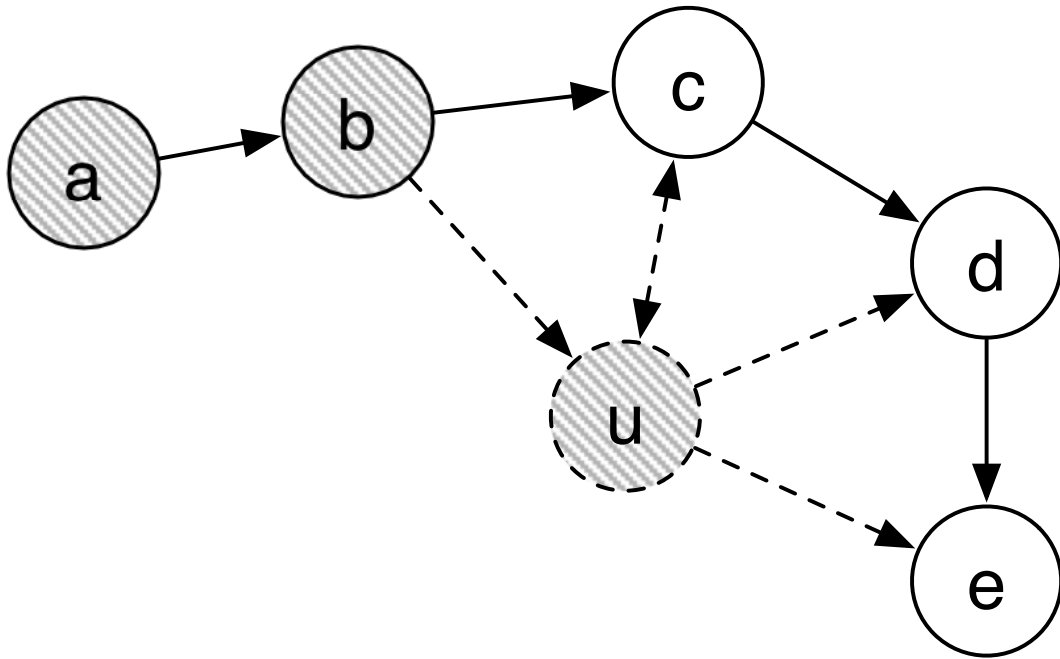


圖 2.1: The diagram of “prototypical PHI query”

第 2 章 Preliminaries

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。

第 3 章 研究方法



第 4 章 實驗設計



第 5 章 實驗結果與分析



第 6 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。

6.1 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] M. N. Garofalakis, J. M. Hellerstein, and P. Maniatis, “Proof sketches: Verifiable in-network aggregation,” in *Proceedings of IEEE 23rd International Conference on Data Engineering (ICDE)*, (Istanbul, Turkey), pp. 996–1005, Apr 2007.
- [2] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, “Tag: A tiny aggregation service for ad-hoc sensor networks,” in *Proceedings of the 5th Symposium on Operating Systems Design and Implementation (OSDI)*, vol. 36, (Boston, Massachusetts, USA), pp. 131–146, ACM, Dec 2002.
- [3] Y. Kotidis, V. Vassalos, A. Deligiannakis, V. Stoumpos, and A. Delis, “Robust management of outliers in sensor network aggregate queries,” in *Proceedings of the 6th ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE)*, (Beijing, China), pp. 17–24, ACM, Jun 2007.
- [4] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos, “Online outlier detection in sensor data using non-parametric models,” in *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB)*, (Seoul, Korea), pp. 187–198, VLDB Endowment, Sep 2006.
- [5] B. Sheng, Q. Li, W. Mao, and W. Jin, “Outlier detection in sensor networks,” in *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, (Montreal, Quebec, Canada), pp. 219–228, ACM, Sep 2007.
- [6] D. Wagner, “Resilient aggregation in sensor networks,” in *Proceedings of the 2nd ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN)*, (Washington DC, USA), pp. 78–87, ACM, Oct 2004.

