

內部稽核運用資料探勘 FP-growth演算法之研究 —以國軍人員申領休假補助費 缺失態樣為例

傅祖銘

關鍵字：內部稽核、薪餉次系統、資料探勘、
FP-growth演算法、WEKA軟體

摘要

本次研究主要是在內部稽核時運用資料探勘中的 FP-growth演算法找出國軍人員申領休假補助費（以下簡稱休補費）缺失態樣中的頻繁項，並利用WEKA軟體系統加以確認各缺失態樣中之頻繁項，進而增加稽核的效率，未來期透過國軍主財資訊雲端服務網陸續建置及整合，透過自動化的稽核程序及不斷強化財務收支及會計帳務資訊集中管控，配合電腦運用以輔助稽核人員查察，以期在持續性的稽核作業下，能有效預警執行單位，避免違失案件產生。透過稽核人員與系統維運人員的研討及精進作業流程，提高財務單位薪餉發放整體效益，並兼顧官兵同仁權益，有效發揮財務管理精神及內部審核功能，在執行上亦逐年檢討改善，使得內部控制機制更為精進完備，進而健全財務業務。

壹、緒論

在全球化資訊及企業管理不斷演進下，我國公部門面臨了各項外在及內在環境之諸多挑戰，例如人力資源管理及配置，財務資源之管理運用等等，有鑑於此，紛紛致力於管理及制度層面之改革與創造，無不建立並積極尋求有效之內部控制及內部稽核等機制，以增進服務之效能，為我國公部門之重要施政要點。惟在面對資訊科技快速發展，在國防事務上以傳統人工方式進行的內部控制及內部稽核程序及作業，恐已無足夠能量因應其工作，較不易及時反應內部作業之漏洞及即時預警財務業務之風險，為提升內部稽核的價值及效率，應加強持續性稽核的應用與發展。

以自動化作業減輕稽核人員人工作業負擔

是持續性稽核的概念之一，該項稽核步驟能夠減少一般事後查核的項目數量，並加強或創造一般事前查核的效益，提供及時回饋及即時預警的查核效能，以達到持續性系統監控的預防成果，能提升稽核工作的價值。剖析現行主財作業演進，在現今主財人員流動下，仍能有效降低財務作業違失，並加強官兵服務之效果，都顯示著內部控制及內部稽核的重要性。

貳、文獻探討

內部控制及與其相關之議題一直是國內外研究政府施政效能的主要內容之一，具備更多解釋能力的「持續性稽核與監控」概念也開始受到重視。不過，風險管控是有限度的，如何在沒有規律的風險上做好資訊分析及相關因應作為，亦是本文研究的重點之一。

資訊管理平台的精進對於內部稽核作業的強化是具有實效的，現行國軍主財資訊雲端服務網的建置漸趨完整，其網頁鏈結之特性較傳統稽核作業具有較佳的線上溝通能力，同時安全檢查的工作亦可同步及時進行，加上具有自動化及電腦化處理的型態，較以往人工彙整資料所產生的時間等稽核作業成本將大幅降低，相對增加各種面向稽核的需求，減少缺誤產生。

影響內部控制的因子相當廣泛，且彼此間可能亦存在著複雜的關係，故本文藉由透過文獻探討分析。區分為六小節，第一節透過研讀內部稽核及相關之關聯性，並探討國軍主財資訊雲端服務網「薪餉次系統」內部稽核制度，第二節對國軍主財資訊雲端服務網之薪餉次系統簡介，第三節透過資料探勘依使用者的需求，從資料中萃取出有意義的型樣及關聯性，第四節透過關聯法則之FP-growth演算法發現高頻率出現的缺失態樣及相關性，以期作為各單

位內部稽核之參考依據，第五節對WEKA軟體工具簡介，並透過該軟體執行FP-growth演算法強化驗證論述結果，第六節探討相關研究文獻使用WEKA軟體工具運用範例，加以驗證本研究實用性。

一、內部控制及內部稽核

汪家淦（2014）研究本部內部控制作業機制及推展，說明就監察院及審計部審核事項納入年度強化內部控制重點，並落實要求自行評估作業，發揮自我監督功能，確保控制作業有效性，在「制度、執行、監督」各面向，能有效控制環境，並進一步產生預警防範機制，另應持續強化各單位主官對於內部控制之重視及表彰其責任履行情形，並發揮縱向督導職能及橫向聯繫回饋，在稽核單位主動發掘執行窒礙，適時回饋風險資訊，做為政策制定策進參考。

林宛瑩、陳錦烽（2015）探討先進國家辦理內部控制之法規面作法，並建議我國政府內部控制制度法制化可採行訂定特法或增修訂條文，就施政角度而言，政策機關應賦予政府治理及績效管理應有之法律位階，併同內部控制制度等相關納入法制，俾有效促使機關首長重視及強化施政績效與風險管理。

鄭瑞成（2015）以新北市政府推動內部控制實務經驗，探討公部門執行內部控制及稽核作業上各面向，其研究發現內部控制制度得以生命週期的概念來檢視，區分為導入期、成長期、成熟期及衰退期共4期，分別為規劃與準備、啟動內控計畫、內控制度建置、監督機制的推動等；另外發現內部控制效能彰顯在於主管的認同及承辦單位同仁的使命感，是影響的重要因素，因為在現今各項業務推展及人力精簡下，維持足夠的作業能量著實不易，額外增加的稽核制度易流於形式，肇生人或事物之排

斥性致效能不彰，而在該單位積極設法內化內部控制制度及內部稽核作業，降低同仁負面或抗拒心態，逐步逐案促使各部門配合辦理各項控制工作，由於機關首長的重視及幕僚單位的積極處理，使得該府在推動上逐漸步上軌道。

石淑媛（2016）探討主計資訊系統導入持續性稽核技術研究成果的運用，以持續性風險評估、持續性控制監督模組及持續性異常偵測模組共三面向評估其成果，並推導出關鍵風險指標（Key Risk Indicators, KRIs）是一個預測風險變化的關鍵，得做為機關組織風險判斷之指標，當訊號出現時，組織應具備有效的反應能力去回應及降低，甚至迴避風險，並說明持續性稽核的特性是以科技為導向之查核作業及持續性的風險評估，從資料庫由來源端經萃取、轉換、匯入至稽核目標，伴隨取得關鍵風險指標的限制，進行持續動態地進行稽核作業，唯獨其作業須配合能夠彙總不同平台資料及擁有分析大量交易資料的能量及人力，在導入持續性稽核技術，必須是資訊科技作業平台有足夠能量予以協助，惟其資訊技術仍有其應用限制，需由具有專業知識背景人員才予以正確判讀。

譚家蘭（2015）探討風險為基礎之持續性稽核與監控，為減輕組織內稽核人員工作負擔，在提升稽核作業效率價值及降低查核風險下，持續性稽核（Continuous Auditing, CA）及持續性監控（Continuous Monitoring, CM）議題受到重視，優良的組織結構係良好企業文化能內化及生根，使得營運績效獲得確保，持續性稽核是作業資訊化，事後查核變革為即時、預警及事前的機制，避免或減少企業營運或決策上錯誤的損失。持續性稽核亦是連續性審計概念的延伸，藉由持續性稽核方式，針對企業內部流程控制作監控，對資訊系統及人為各項

交易資料作蒐集，並建立證據。

綜合以上論述，運用傳統稽核方式隨著資訊軟硬體設備的更迭及複雜化後，可能已不敷使用，而持續性稽核及循序漸進分析則能強化企業營運及決策流程，透過有系統的關聯演算法去找出缺失態樣之裙帶關係，可即時發現異常現象，預先提出因應措施，以供內外部組織的溝通，並加以防範有能的風險危害。

二、國軍主財資訊雲端服務網之薪餉次系統

「國軍主財資訊雲端服務網」係以國軍雲端服務如電子郵件、國軍智慧卡為基礎，整合主計、財務及帳務等各類資訊系統，提供各預算單位、財務單位與業務單位及支薪單位雲端化的作業介面與資訊服務，並利用電子郵件系統即時回饋國軍官兵個人給與撥戶資訊（國防部主計局，2015）。

薪餉次系統包含月支薪給、各項補助費（結婚、生育、喪葬、殮葬、教育及休假）、年終工作獎金、考績獎金等給與發放核結作業，另有薪資所得稅、軍人保險費、退休撫卹基金及強制執行等代扣款項計算及系統檔案維護作業，提供全國各財務單位每月薪餉發放作業使用，各單位作業承辦人均能依實際業務在電腦上即時辦理資料異動（含新增、更正、註銷、查詢、統計、印表，藉此產生發放資料及產製相關結帳、統計、查核分析表單）。

三、資料探勘

「資料探勘」原文為 Data Mining，有學者另譯為「資料採礦」、「資料發掘」、「資料挖礦」。近年來，許多文獻對資料探勘這個名詞給予了不少定義，茲整理其中較具代表性的部份，條列如下：

- (一)「資料探勘的目的是要從龐大的資料庫中，挖掘出隱含的、未知的、且可能有用的資訊」（Frawley & Piatetsky-Shapiro & Matheus, 1992）。
- (二)「資料探勘乃使用自動或半自動的方式，對龐大的資料作分析，為了從資料庫中找出有意義的關係或規則」（Berry & Linoff, 1997）。
- (三)「資料探勘是一種新的且不斷循環的決策支援分析過程，它能夠從組合在一起的資料中，發現出隱藏價值的知識，以提供給企業專業人員參考」（Kleissner, 1998）。
- (四)「資料探勘為提煉未知且可能有用的知識來幫助決策的一系列流程」（Kim, 2000）。
- (五)「資料探勘乃是從儲存在資料庫、資料倉儲或其他資訊儲存媒體中的大量資料發現有意義與效益的型樣」（Han & Kamber, 2001）。

資料探勘的技術時有進展而日趨成熟，許多實用的探勘技術和系統方法被學者專家提出來，這些探勘技術和系統方法是針對不同的領域與需求而設計的。不同的企業體，有不同的需求，所使用的資料探勘方式也就不盡相同。甚至資料型態的不同，都會影響到資料探勘技術的選擇。使用的資料探勘技術是否適宜，將會影響到處理效能與資料探勘後其資訊的可用性。

四、關聯法則之FP-growth演算法

黃楷鈞（2016）在龐大的資料庫中某一些物件間存在有彼此的關聯性，起源在於分析市場購物籃資料中的商品之間的關聯程度，因此也被稱為市場購物籃分析（Market Basket

Analysis），關聯規則通常應用在市場購物籃分析上，藉由資料庫所記錄的消費者購買紀錄，分析商品之間的關聯性。

FP-growth（Frequent Pattern Growth）演算法是一種不產生候選項目集而採用頻繁形樣數增長的方法挖掘頻繁項的演算法。算法只需要掃描2次資料庫：第一次掃描資料庫，得到頻繁1-項目集（Frequent 1-itemsets）；第二次掃描資料庫，利用頻繁1-項目集過濾資料庫中的非頻繁項目，同時生成FP-Tree包含了所有的頻繁1-項目集及它們間的關係資訊，其後的頻繁項集的挖掘只需要在FP-Tree上進行即可。

本篇採關聯法則中的FP-growth演算法查找存在於項目集合或對象集合之間的頻繁模式、關聯、相關性或因果結構。FP-growth演算法是韓家煒（Han jia wei）等人在2004年提出的關聯分析演算法，它採取如下分治策略：將提供頻繁項集的資料庫壓縮到一棵頻繁模式樹（FpTree），但仍保留項集關聯信息。

五、WEKA軟體介紹

WEKA是懷卡托智慧分析環境（Waikato Environment for Knowledge Analysis）的英文字首縮寫，在該網站可以免費下載可執行軟體和原始碼，還可以獲得說明文件、常見問題解答、資料集和其他文獻等資源。WEKA是紐西蘭懷卡托大學用Java開發的資料探勘著名開源軟體，該系統自1993年開始由紐西蘭政府資助，至今已經歷了20年的發展，其功能已經十分強大和成熟。

WEKA作為一個公開的資料探勘工作平臺，集合了大量能承擔資料探勘任務的機器學習演算法，包括對資料進行預處理、分類、迴歸、聚類、關聯規則以及在新的互動式介面上的視覺化。

六、參考相關WEKA軟體使用情形及驗證

葉聰憲（2015）以關聯式規則探勘零組件市場需求分析為例，利用6,496筆測試資料（其交易內容皆與工廠自動化用之機械標準零組件相關），並訂定零組件關聯規則所需的最小支持度及信賴度後，再來將整理過後的訂單（每項訂單上商品的組合及頻繁度）建立探勘所需之交易資料庫，在WEKA軟體內對此交易資料庫進行關聯規則探勘，最後再進行分析，相互驗證以得出結論。

研究中利用了 WEKA 軟體來對台灣三家公司於2014 年所產生的工廠自動化用之機械標準零組件交易資料進行關聯式規則的探勘（收集公司 A、B 及 C 之工廠自動化用之機械標準零組件訂單後，將其整合成為測試的交易資料庫，其整合的原因在於此三家均為代理商，其代理的產品大同小異），透過設定最小支持度、最小信賴度以及探勘規則數目以呈現這三家公司在 2014 年裡販售工廠自動化用之機械標準零組件間的關聯性，透過所探勘出以下訊息：

- (一)供應商可以依據此關聯規則以得到備料存貨產品項目之預測參考。
- (二)供應商可以針對在同一關聯規則內的零組件進行相關的組合促銷，以增加潛在的下游廠商客戶預測參考。
- (三)供應商可以根據這樣的資訊，給予下游廠商客戶欲在自動化機械標準零組件產品上之佈局建議。

而根據其在零組件廠服務多年之經驗，所探勘出來的關聯規則，的確在零組件銷售市場中是常見之購買及銷售組合，因此，期研究動機利用關聯式規則進行探

勘之方式的確可用於實際之產品應用上，並用以提供決策者更佳之決策資訊；綜上，我們可以套用此研究在內部稽核時利用資料庫缺失內容之頻繁項或組合等數據來快速預測出可利用之參考依據與建議。

參、作業現況

一、薪餉資料查核比對稽核機制

國防部主計局財務中心現為國軍官兵薪餉發放作業主要承辦機關，擔負各財務單位薪餉發放處理與會計事務指導工作，為貫徹人員經費覈實政策，藉由國軍人事與薪給資料查核管制作業及產生之資料，降低「人薪差異」異常情事，詳實產生量化數據，期達成人事與薪給資料合一目標，發揮本中心財務管理應具效能。

國軍主財業務作業配合資訊系統的強化，依業務職能劃分為薪餉次系統、軍費次系統、退俸次系統、主計次系統、帳審次系統等各項，利用雲端資料庫的作業特性，使財務單位創造即時回饋薪餉作業訊息，利用異常訊號產生及判別，藉以反映人事單位人員薪給現況，據以掌握官兵個人權益。

二、稽核步驟

在運用自動化科技及電腦輔助系統下，巨量資料的彙整時效性得以增強，較傳統稽核作業精確且快速執行，其作業步驟引用FP-growth演算法，FP-growth演算法通過構造一個樹結構來壓縮資料記錄，使得挖掘頻繁項集只需要掃描兩次資料記錄，而且該演算法不需要生成候選集合，所以效率會比較高。另以休補費退件單資料集為例：

休補費 Items

休補費1 {A07,B03}

休補費2 {B03,B06,B14,C02}

休補費3 {A07,B06,B14,C10}

休補費4 {B03,A07,B06,B14}

休補費5 {B03,A07,B06,C10}

上述休補費缺失代碼一覽表：

(一)申請名冊部分：A01名章不符、A02階級不符、A03身分證字號不符、A04申請金額不符、A05休假日數不符、A06檢附單據張數不符、A07檢附單據金額不符、A08申請年度不符、A09合計大寫金額不符、A10各級簽證人員漏蓋職章、A11塗改漏蓋主官章、A12造列之表冊份數不足、A13非為在職期間發生之事實、A16支薪單位全銜或代號錯誤、A22其他。

(二)申請表部分：B01名章不符、B02身分證字號不符、B03申請日期不符、B04休假日數不符、B05塗改漏蓋主官章、B06合計大寫金額不符、B07服役年資請重新計算、B08各級簽證人員漏蓋職章、B09申領破月人員未註明原因及加蓋私章、B10年度內已超支、B11支薪單位全銜錯誤、B12申請年度、日期不符、B13階級職務不符、B14休假日為國定休假日、B15休假日數合計不符、B16當年度已領休補費日數、金額不符、B18申領人未簽章、B19檢附單據張數錯誤、B21申領金額錯誤、B24其他。

(三)憑證部分：C01收據買受人、買受人地址漏填、C02申領人單據未簽名、C03單據數量、單價與總價不符、C04 單據日期非休假日、C05單據金額未達申領數二分之一、C06統一發票日期不清晰、C07發票未附收執聯、C08一次開立多聯發票有缺漏、C09未依規定開立發票、C10收據未註明營利事

業統一編號、C11收據未註明負責人姓名、C12所附單據依規定不得作為申領單據、C13發票日期非當期之月份、C14電子發票不得有統一編號、C15買受人非當事人、C16 未註明中文品名（可附明細表）、C24已逾申領期。

三、建立FpTree

FpTree (Frequent Pattern Tree) 是一種樹結構，樹結構定義如下：

```
1 public class FpNode {
2     String idName ; // 收文號
3     List < FpNode > children ; // 子結點
4     FpNode parent ; // 父結點
5     FpNode next ; // 下一個收文號相同的結點
6     long count ; // 出現次數
7 }
```

Step 1：掃描資料結果，生成一級頻繁項集，並按出現次數由多到少排序，假設我們的最小絕對支援度是3，如下所示：

Item Count

A07 : 4

B03 : 4

B06 : 4

B14 : 3

可以看到，C10和C02沒有出現在上表中，因為C10只出現2次，C02只出現1次，小於最小支援度，因此不是頻繁項集，所以C10和C02不需要再考慮。

Step 2：再次掃描資料記錄，對每條記錄中出現在Step 1產生的表中的項，按表中的順序排序。初始時，新建一個根結點，標記為null：

第一條記錄：{A07, B03}，按Step 1表過濾排序得到依然為{A07, B03}，新建一個結點，

idName為{A07}，將其插入到根節點下，並設定count為1，然後新建一個{B03}結點，插入到

{A07}結點下面，插入後如下所示：

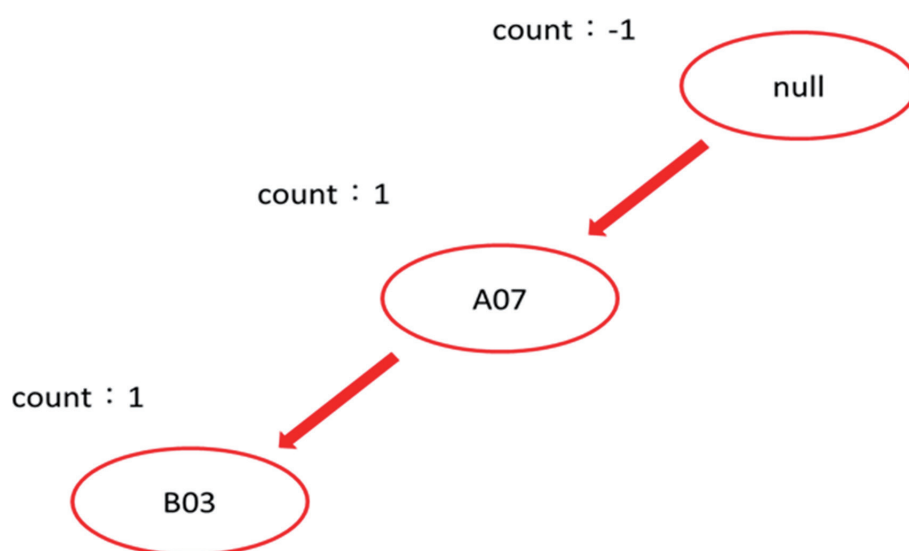


圖1 第一條記錄樹結構圖

第二條記錄：{B03, B06, B14, C02}，過濾並排序後為：{B03, B06, B14}，發現根結點沒有包含{B03}的結點（有一個{B03}次結點但不是結點），因此新建一個{B03}結點，插

在根結點下面，這樣根結點就有了兩個結點，隨後新建{B06}結點插在{B03}結點下面，新建{B14}結點插在{B06}下面，插入後如下所示：

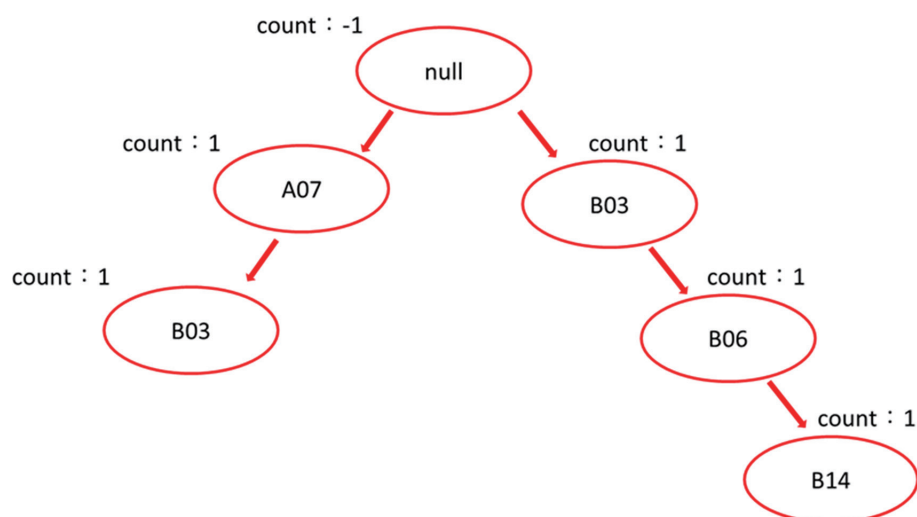


圖2 第二條記錄樹結構圖

第三條記錄：{A07, B06, B14, C10}，過濾並排序後為：{A07, B06, B14}，這時候發現根結點有結點{A07}，因此不需要新建結點，只需將原來的{A07}結點的count加1即可，往

下發現{A07}結點有一個結點{B06}，於是新建{B06}結點，並插入到{A07}結點下面，隨後新建{B14}結點插入到{B06}結點後面。插入後如下圖所示：

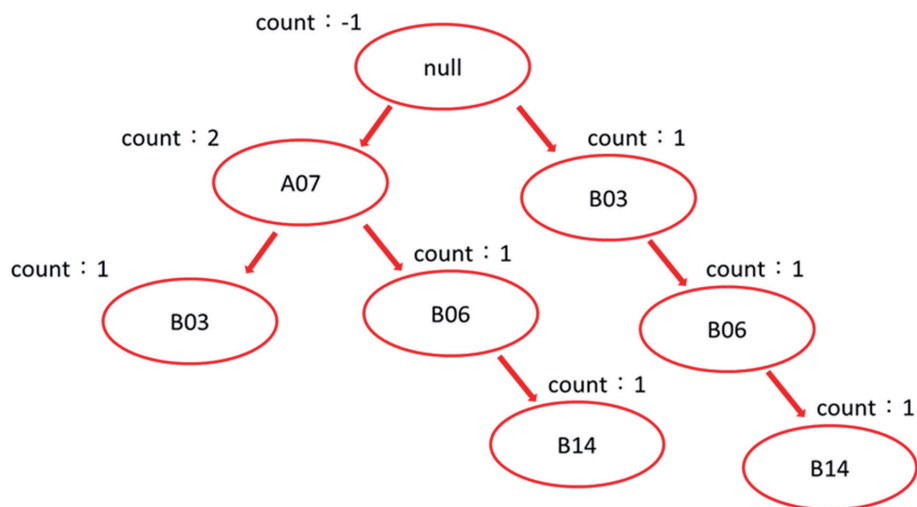


圖3 第三條記錄樹結構圖

第四條記錄：{B03, A07, B06, B14}，過濾並排序後為：{A07, B03, B06, B14}，這時候發現根結點有結點{A07}，因此不需要新建結點，只需將原來的{A07}結點的count加1即可，往下發現{A07}結點有一個結點{B03}，於是也

不需要新建{B03}結點，只需將原來{B03}結點的count加1，由於這個{B03}結點沒有結點，此時需新建{B06}結點，插在{B03}結點下面，隨後新建{B14}結點，插在{B06}結點下面，插入後如下圖所示：

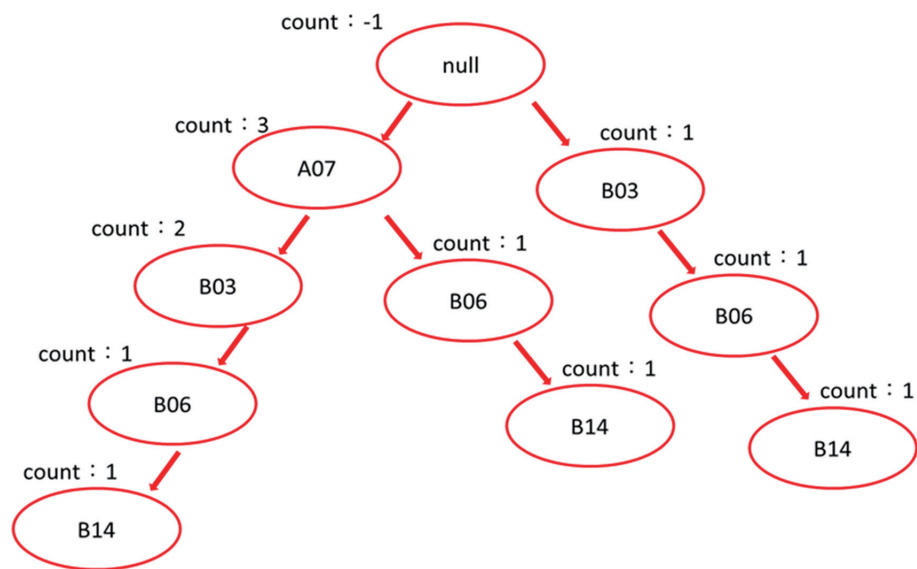


圖4 第四條記錄樹結構圖

第五條記錄：{B03, A07, B06, C10}，過濾並排序後為：{A07, B03, B06}，檢查發現根結點有{A07}結點，{A07}結點有{B03}結點，

{B03}結點有{B06}結點，本次插入不需要新建結點只需更新count即可，示意圖如下：

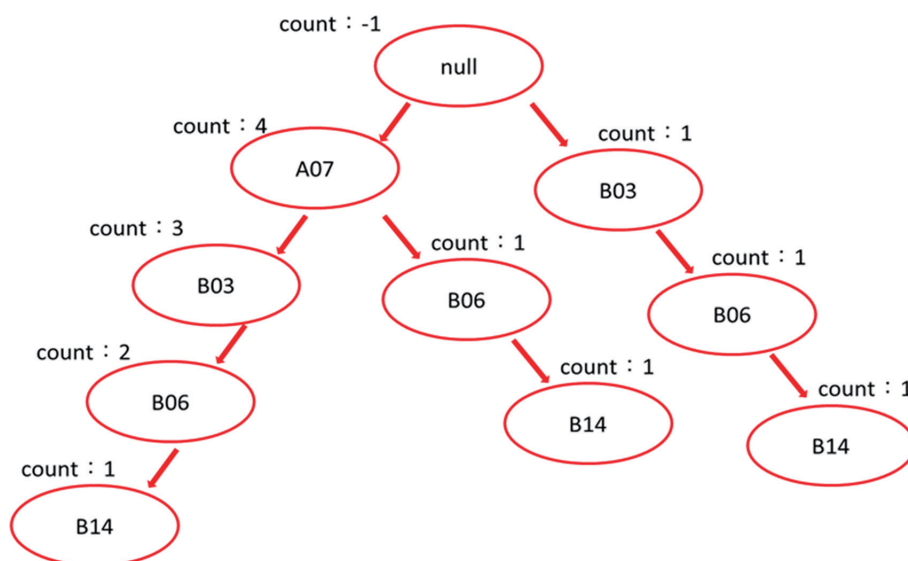


圖5 第五條記錄樹結構圖

按照上面的步驟，我們已經基本構造了一棵FpTree，樹中每天路徑代表一個項集，因為許多項集有公共項，而且出現次數越多的項越可能是公共項，因此按出現次數由多到少的順序可以節省空間，實現壓縮儲存，另外我們需要

一個表頭和對每一個idName相同的結點做一個線索，方便後面使用，線索的構造也是在建樹過程形成的，但為了簡化FpTree的生成過程，在上述我們沒有一一列舉程式碼，並轉而以圖形化方式呈現線索和表頭的FpTree如下：

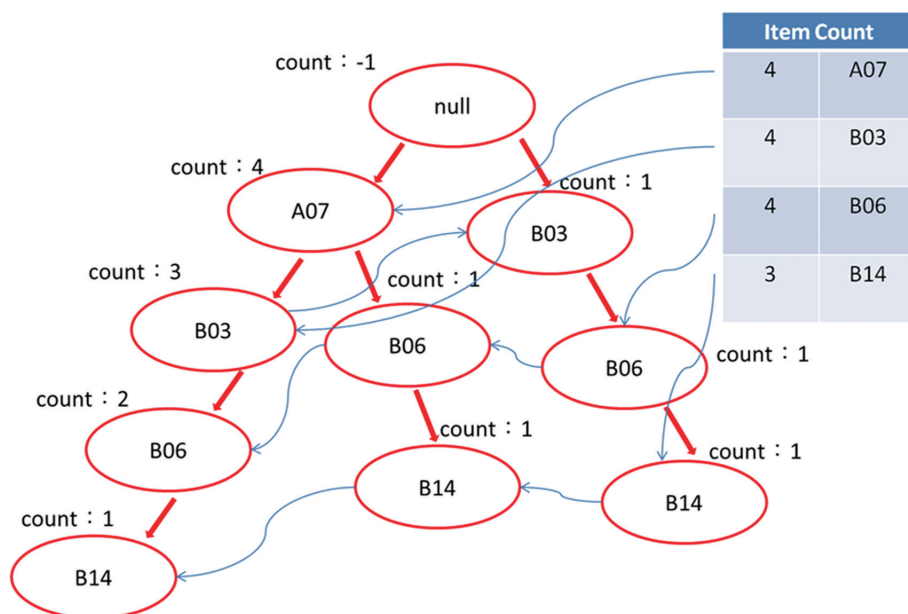


圖6 最終完整樹結構圖

至此，整個FpTree就構造好了，在下面的挖掘過程中我們會看到表頭和線索的作用。

四、利用FpTree挖掘頻繁項集

FpTree建好後，就可以進行頻繁項集的挖掘，挖掘演算法稱為Fp-growth演算法，挖掘從表頭header的最後一個項開始。

步驟一：此處即從{B14}開始，根據{B14}

的線索鏈找到所有{B14}結點，然後找出每個{B14}結點的分支：{A07, B03, B06, B14: 1}, {A07, B06, B14:1}, {B03, B06, B14:1}, 其中的「1」表示出現1次，注意，雖然{A07}出現4次，但{A07, B03, B06, B14}只同時出現1次，因此分支的count是由字尾結點{B14}的count決定的，除去{B14}，我們得到對應的字首路徑{A07, B03, B06: 1}, {A07, B06:1}, {B03, B06:1}，根據字首路徑我們可以生成一棵

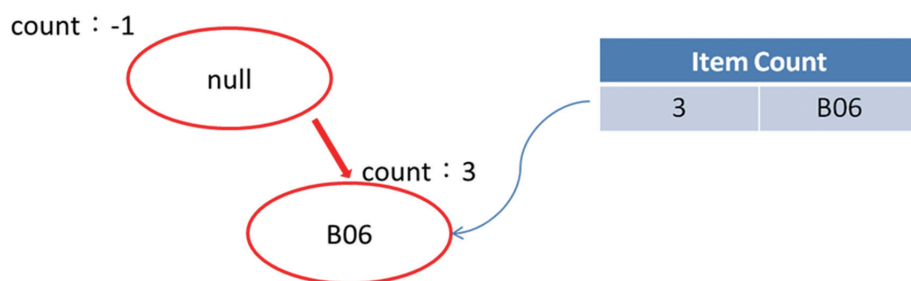


圖7 FpTree示意圖

構造好條件樹後，對條件樹進行遞迴挖掘，當條件樹只有一條路徑時，路徑的所有組合即為條件頻繁集，假設{B14}的條件頻繁集為{S1, S2, S3}，則{B14}的頻繁集為{S1+{B14}, S2+{B14}, S3+{B14}}，即{B14}的頻繁集一定有相同的字尾{B14}，此處的條件頻繁集為：{{}, {B06}}，於是{B14}的頻繁集為{{B14}, {B06, B14}}。

步驟二：接下來找header表頭的倒數第二個項{B06}的頻繁集，同上可以得到{B06}的字首路徑為：{B03: 1}, {A07: 1}, {A07, B03: 2}，條件FpTree的資料集為：

休補費	Items
休補費1	{B03}
休補費2	{A07}
休補費3	{A07, B03}
休補費4	{A07, B03}

注意{A07, B03: 2}，即{A07, B03}的count為2，所以在{A07, B03}重複了兩次，這樣做的目的是可以利用之前構造FpTree的演算法來構造

條件FpTree，構造方式跟之前一樣，此處的資料記錄變為：

休補費	Items
休補費1	{A07, B03, B06}
休補費2	{A07, B06}
休補費3	{B03, B06}

絕對支援度依然是3，構造得到的FpTree為：

條件FpTree，不過這樣效率會降低，試想如果{A07, B03}的count為20000，那麼就需要展開成20000條記錄，然後進行20000次count更新，而事實上只需要對count更新一次到20000即可，另綜合上述觀點，而此構造的條件FpTree為：

這棵條件樹已經是單一路徑，路徑上的所有組合即為條件頻繁集：{{}, {A07}, {B03}, {A07, B03}}，加上{B06}後，又得到一組頻繁項集{{B06}, {A07, B06}, {B03, B06}, {A07, B03, B06}}，這組頻繁項集一定包含一個相同的字尾：{B06}，並且不包含{B14}，因此這一組頻繁項集與上一組不會重複。

重複以上步驟，對header表頭的每個項進行挖掘，即可得到整個頻繁項集，可以證明頻繁項集即不重複也不遺漏，而依前面步驟二所驗證之結論為A07及B03互為正相關且B06及B03亦互為正相關，當稽查人員檢核出A07（檢附單據金額不符）或B06（合計大寫金額不符）後可研判出B03（申請日期不符）的犯錯可能性，可縮短爾後尋找相同缺失的時間，進而增加稽核的效率。

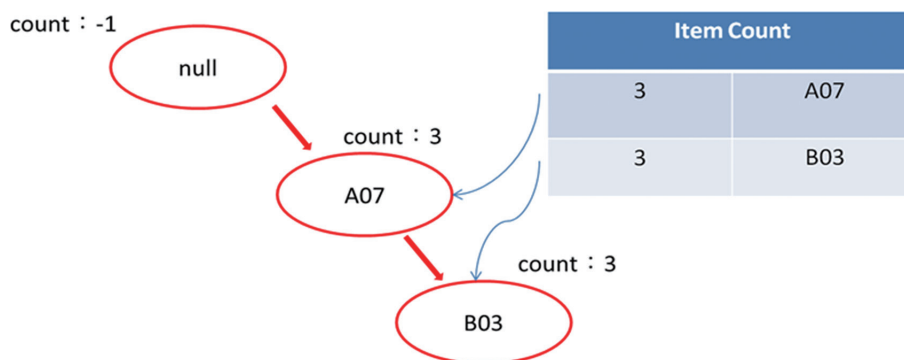


圖8 FpTree示意圖

五、利用WEKA軟體驗證各項集關聯性

(一)首先讀取我們所需驗證的資料：

	ITEM	UPFS_REAS_CD1	UPFS_REAS_CD2	UPFS_REAS_CD3	UPFS_REAS_CD4
1	ITEM1	A07	B03		
2	ITEM2	B03	B06	B14	C02
3	ITEM3	A07	B06	B14	C10
4	ITEM4	B03	A07	B06	B14
5	ITEM5	B03	A07	B06	C10

圖9 驗證資料示範圖

(二)確認WEKA軟體完整寫入檔案資料

Selected attribute

No.	Label	Count	Weight
1	ITEM1	1	1.0
2	ITEM2	1	1.0
3	ITEM3	1	1.0
4	ITEM4	1	1.0
5	ITEM5	1	1.0

圖10 WEKA操作示意圖

(三)以FP-growth演算法驗證結論

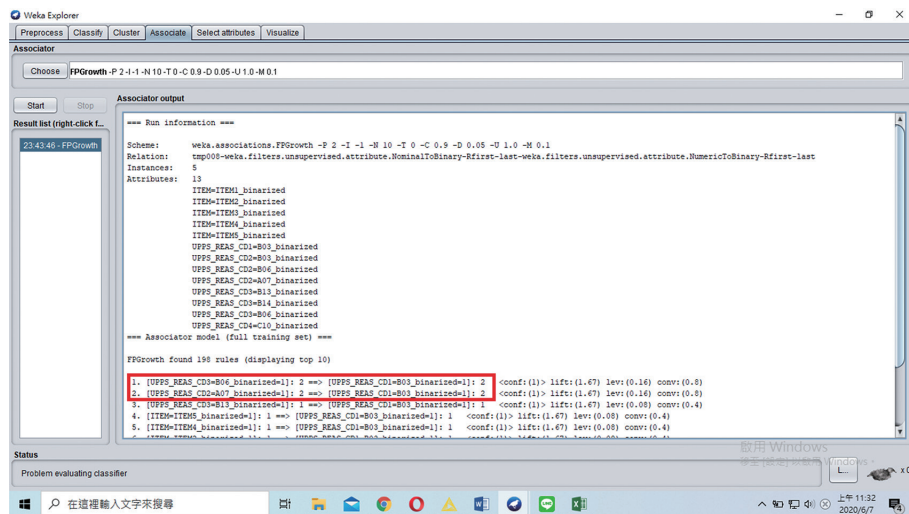


圖11 WEKA操作示意圖

而最後結果的最右邊，分別有conf（信賴度）、lift（提升度）、lev（槓桿度）、conv（確信度）這四種不同的數值，用來代表各欄位屬性之間的相關性為何，其代表之意義將分別說明如下：

1. Confidence, conf（信賴度） $\rightarrow P(B|A)$ ，即在發生A的情況下，B也會發生的條件機率。
2. Lift（提升度） $\rightarrow P(A \cap B) / (P(A) P(B))$ ，當Lift=1的時候，代表A和B相互獨立，因此Lift的值越大（>1），A和B之間的關係越密切。
3. Leverage, lev（槓桿度） $\rightarrow P(A \cap B) - P(A) P(B)$ ，當Leverage=0的時候，代表A和B相互獨立，因此Leverage的值越大，A和B之間的關係越密切。
4. Conviction, conv（確信度） $\rightarrow P(A) P(B') / P(A \cap B')$ （B'表示B沒有發生），Conviction也是用來衡量A和B之間的獨立性。從它和Lift的關係來看，其對B取反，並代入Lift的公式之後求倒數，可以得知Conviction的值越大，A和B之間

的關係也越密切。

最後由各欄位屬性之間的相關性判斷，並可以由槓桿度及確信度再次驗證（A07 檢附單據金額不符、B03 申請日期不符）及（B06 合計大寫金額不符、B03 申請日期不符）皆互為高度正相關，亦即發現休補費申請名冊檢附單據金額不符之缺失後，找出申請表申請日期不符之缺失機率偏高，而也同時高機率發現申請表合計大寫金額不符之缺失與申請表申請日期不符之組合。

綜整以上論述，我們可以在內部稽核時運用資料探勘中的FP-growth演算法去找尋各項缺失態樣、錯誤樣本及可能風險等關聯性，並針對此關聯去快速找出其他可能伴隨的缺失，另期後續嘗試各式資料探勘演算法及機器語言來蒐整不同數據之關聯性，藉以迅速研擬受檢單位改進方針也減少督導單位所花費在尋找缺失的時間，亦可運用此方法由國軍主財雲端服務網資料庫上規劃各項稽核子功能來減少可能的作業缺失及疏漏。

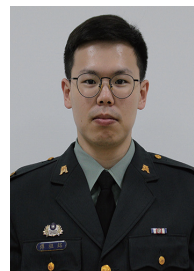
肆、結論與建議

現今國軍預算支用單位及支薪單位分布各地，透過資訊系統執行各項財務業務任務，其資料處理作業量更是數以萬筆計算，假以透過數據分析運用，可以了解各地區財務業務之作業實況，或藉由各項資料之關聯性，進行交叉比對，據以財務單位辦理人事資料及各項加給異動審查及發放等作業，俾利各項管控任務之遂行。

未來期透過國軍主財資訊雲端服務網陸續建置及整合，透過自動化的稽核程序及不斷強化財務收支及會計帳務資訊集中管控，配合電腦運用以輔助稽核人員查察，以期在持續性的稽核作業下，能有效預警執行單位，避免違失案件產生。

參考文獻

1. 石淑媛(2016).主計資訊系統導入持續性稽核技術研究成果之運用.主計月刊,第726期,94-99.
2. 林宛瑩、陳錦烽(2015).政府推動內部控制法制化初探-美、英、加、澳等國之借鏡.主計月刊,第711期,20-25.
3. 汪家淦(2014).強化本部內控機制,有效落實風險管制.主計季刊,第347期第55卷第四期,25-48.
4. 黃楷鈞(2016).於Hadoop Yarn平台上建置平行化的FP-Growth.研究論文,105年7月.
5. 國防部主計局(2015).國軍主財資訊管理之創新與精進,主計月刊,第 714 期:104-107.
6. 葉聰憲(2015).以關聯式規則探勘零組件市場需求分析.研究論文,104年8月.
7. 鄭瑞成(2015).新北市政府推動內部控制經驗分享.主計月刊,第719期,30-34.
8. 譚家蘭(2015).風險為基礎之持續性稽核與監控.主計月刊,第719期,36-40.
9. Berry, Linoff (1997) Organization of Knowledge Discovery and Customer Insight Activities.
10. C Kleissner (1998) Data mining for the enterprise.
11. Han jia wei, Pei Jan等(2004) Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach.
12. J Han, M Kamber (2001) Spatial clustering methods in data mining.
13. Kim (2000) On the security analysis of Lee,Hwang & Lee (2004) and Song & Kim(2000) key exchange / agreement protocols.
14. WJ Frawley, G Piatetsky-Shapiro, CJ Matheus (1992) Knowledge discovery in databases : An overview.



傅祖銘

⇒現職

國防部主計局財務中心上尉預財官

⇒學歷

國防大學國防管理學院資管系57期103年班

財務正規班107期（106年班）

⇒經歷

預財官、資處官