

行政院國家科學委員會專題研究計畫 成果報告

漸進式郵件類別判定模式(I) 研究成果報告(精簡版)

計畫類別：個別型
計畫編號：NSC 100-2221-E-343-001-
執行期間：100年08月01日至101年07月31日
執行單位：南華大學資訊管理學系

計畫主持人：楊士霆

計畫參與人員：碩士班研究生-兼任助理人員：龔鈺婷
碩士班研究生-兼任助理人員：黃家偉
碩士班研究生-兼任助理人員：吳朝宏

報告附件：出席國際會議研究心得報告及發表論文

公開資訊：本計畫可公開查詢

中華民國 101 年 10 月 25 日

中文摘要：目前處理郵件管理分類之方式，通常是採用人工閱讀內容、判斷其信件性質，再進行分類與管理，但此種做法不僅需要聘請大量人工，亦需要花費許多時間等待，才可以整理分類好所有資料，且分類正確率往往不佳。本研究提之「漸進式郵件分類模式」乃以郵件中所包含之郵件內容為基礎，先行將這些資料進行資料層級之歸類（包含為「數字型」、「數據型」、「文字型」、「附件檔案型」及「郵件聯絡人」等資料型別）；當中，「數字型」與「數據型」為第一層級資料、「文字型」與「附件檔案型」為第二層級資料、「郵件聯絡人」則為第三層級資料；待資料彙整完成後，即可以本研究建構之「郵件數據型、數字型資料解析模組」分析郵件數據型及數字型資料，先行過濾垃圾郵件，以判定適合分類之郵件群；之後，利用「郵件文字型資料解析」模組，計算目標郵件與各類別之隸屬關係；最後，當目標郵件之類別判定結果不甚理想，即可利用「郵件聯絡人資料解析模組」重新計算目標郵件與各類別之隸屬關係，進而形成一套漸進式郵件類別判定模式，期望以具效率之方式進行郵件類別判定任務。此外，本研究於最終將建立一套漸進式郵件分類系統，並以一個案例評估此模式與技術之有效性與可行性。

中文關鍵詞：電子郵件、郵件分類、關鍵字擷取、知識管理

英文摘要：Traditionally, the e-mail management of users is to read the content to identify the e-mail properties, and then classify and manage the e-mails by themselves. However, it is time-consuming and the effect is unsound. To efficiently provide appropriate e-mail management and classification model for e-mail users, this research develops an Incremental E-mail Classification Model (IECM) based on e-mail data structure. Before developing the IECM, this research examines all the critical data such as e-mail sender, e-mail subject, e-mail content, Message-ID, Received and e-mail contact, etc. contained in e-mails and then classifies these data into 'number type', 'data type', 'text type' and 'e-mail contact type'. The proposed IECM can be divided into three main modules namely of Number/Data Type Data Analysis (NTDA), Text Type Data Analysis (TTDA) and E-mail Contact Type Data Analysis (CTDA). In NTDA module, number/data type data such

as Message-ID or IP is used to filter the spam. Then, in the TTDA module, the correlations between the target e-mails (i.e., remained e-mails after spam filtering) and categories can be determined based on text type data such as e-mail subject and contents. Finally, in CTDA module, the e-mail contact data is used to re-determine the e-mail/category correlations. In conclusion, by using the proposed Incremental E-mail Classification Model, the e-mail/category correlations can be efficiently and effectively determine to e-mail users to enhance their e-mail management performance. In addition to the Incremental E-mail classification algorithm, a Web-based Incremental E-mail classification system is also developed and a demonstration case is applied to verify the performance of the proposed approach. The attempt of this research is to enhance the accuracy and efficiency of Incremental E-mail classification technology and to enhance e-mail management performance.

英文關鍵詞： E-mail, E-mail Classification, Keyword Extraction, Knowledge Management

中文摘要

目前處理郵件管理分類之方式，通常是採用人工閱讀內容、判斷其信件性質，再進行分類與管理，但此種做法不僅需要聘請大量人工，亦需要花費許多時間等待，才可以整理分類好所有資料，且分類正確率往往不佳。本研究所提之「漸進式郵件分類模式」乃以郵件中所包含之郵件內容為基礎，先行將這些資料進行資料層級之歸類（包含為「數字型」、「數據型」、「文字型」、「附件檔案型」及「郵件聯絡人」等資料型別）；當中，「數字型」與「數據型」為第一層級資料、「文字型」與「附件檔案型」為第二層級資料、「郵件聯絡人」則為第三層級資料；待資料彙整完成後，即可以本研究所建構之「郵件數據型、數字型資料解析模組」分析郵件數據型及數字型資料，先行過濾垃圾郵件，以判定適合分類之郵件群；之後，利用「郵件文字型資料解析」模組，計算目標郵件與各類別之隸屬關係；最後，當目標郵件之類別判定結果不甚理想，即可利用「郵件聯絡人資料解析模組」重新計算目標郵件與各類別之隸屬關係，進而形成一套漸進式郵件類別判定模式，期望以具效率之方式進行郵件類別判定任務。此外，本研究於最終將建立一套漸進式郵件分類系統，並以一個案例評估此模式與技術之有效性與可行性。

關鍵字：電子郵件、郵件分類、關鍵字擷取、知識管理

Abstract

Traditionally, the e-mail management of users is to read the content to identify the e-mail properties, and then classify and manage the e-mails by themselves. However, it is time-consuming and the effect is unsound. To efficiently provide appropriate e-mail management and classification model for e-mail users, this research develops an Incremental E-mail Classification Model (IECM) based on e-mail data structure. Before developing the IECM, this research examines all the critical data such as e-mail sender, e-mail subject, e-mail content, Message-ID, Received and e-mail contact, etc. contained in e-mails and then classifies these data into “number type”, “data type”, “text type” and “e-mail contact type”. The proposed IECM can be divided into three main modules namely of Number/Data Type Data Analysis (NTDA), Text Type Data Analysis (TTDA) and E-mail Contact Type Data Analysis (CTDA). In NTDA module, number/data type data such as Message-ID or IP is used to filter the spam. Then, in the TTDA module, the correlations between the target e-mails (i.e., remained e-mails after spam filtering) and categories can be determined based on text type data such as e-mail subject and contents. Finally, in CTDA module, the e-mail contact data is used to re-determine the e-mail/category correlations. In conclusion, by using the proposed Incremental E-mail Classification Model, the e-mail/category correlations can be efficiently and effectively determine to e-mail users to enhance their e-mail management performance.

In addition to the Incremental E-mail classification algorithm, a Web-based Incremental E-mail classification system is also developed and a demonstration case is applied to verify the performance of the proposed approach. The attempt of this research is to enhance the accuracy and efficiency of Incremental E-mail classification technology and to enhance e-mail management performance.

Keywords: E-mail, E-mail Classification, Keyword Extraction, Knowledge Management

1. 研究動機與目的

隨著全球人們漸漸習慣網路之通訊方式，利用網路收發電子郵件已成為人們生活中重要部分。由於電子郵件之多樣性（如廣告信件、教育信件、會議信件等），使用者常需要花費時間去管理分類，並從中擷取、儲存其所需要之知識。目前管理郵件類別之方式，通常是採用人工閱讀內容、判斷其信件性質，再進行分類與管理，但此種做法需要花費許多時間，才可整理分類所有資料，且分類效果往往不佳。有鑑於此，目前多數研究乃著重於解析電子郵件標頭欄位內「Date」、「From」、「Subject」以及本文內容之文字擷取，並且發展各種技術及運算法以解析郵件內容，然而當中僅解析電子郵件之郵件主旨、郵件內容等特徵值，並非分析郵件中所有內容，因此可能遺漏電子郵件中之重要分類訊息。

是故，一套有效率之郵件管理分類系統，其所著重之重點應協助郵件使用者將所需資訊有效地擷取、儲存，並且分類利用，為了達到以上所期望之目的，於本研究中需探討以下之重要議題：

1. 如何有效率之擷取郵件中使用者所需之資訊。
2. 如何對郵件進行有效率且正確之管理與分類。

綜上所述，有鑑於目前郵件分類仍使用人工分類、近期所發展之以關鍵字為基礎或其他演算法之郵件分類機制，可能衍生分類時間過長、分類結果易受主觀影響或系統自動分類效果不佳等問題；因此，本研究將解析與以往郵件分類模式所解析之電子郵件格式內容尚有差異，除了一般電子郵件之寄件者、郵件主旨、郵件內容和檔案內文等特徵值，另外亦解析件標頭欄位之Message-ID、Received、聯絡人及本文內容之附加檔案之檔名及副檔名，以及本文之寫作格式等內容。基於此些郵件需解析之內容，本研究乃先行將郵件中所包含之資料進行層級歸類（包含「數字型」、「數據型」、「文字型」、「附檔案型」及「郵件聯絡人」等資料型別），之後逐一解析同層級之資料（如文字型內容、數據型內容等），進而形成一套「漸進式郵件類別判定模式」；當中，此模式乃郵件資料進行資料層級之歸類（包含為「數字型」、「數據型」、「文字型」、「附檔案型」及「郵件聯絡人」等資料型別）；當中，「數字型」與「數據型」為第一層級資料、「文字型」與「附檔案型」為第二層級資料、「郵件聯絡人」則為第三層級資料；待資料彙整完成後，即可以本研究所建構之「郵件數據型、數字型資料解析模組」分析郵件數據型及數字型資料，先行過濾垃圾郵件，以判定適合分類之郵件群；之後，利用「郵件文字型資料解析」模組，計算目標郵件與各類別之隸屬關係；最後，當目標郵件之類別判定結果不甚理想，即可利用「郵件聯絡人資料解析模組」重新計算目標郵件與各類別之隸屬關係，是故，本研究所提出之漸進式郵件類別判定模式，除能考量郵件中所有包含之資料外（亦即郵件中所有資料），並能以「漸進式」之概念逐層解析並判定類別；因此，本研究之漸進式郵件類別判定模式除可具備郵件分類較佳之分類效果外，亦可兼具較佳之分類效率。本研究所提漸進式郵件類別判定之期望模式（TO-BE Model）如圖1.1所示。



圖1.1、郵件分類管理之期望模式（TO-BE Model）

2. 文獻回顧

本研究所涉及之研究主題包含「郵件分類管理與應用」及「郵件分類技術」等兩大研究方向，以下即針對此兩項主題之相關研究進行文獻回顧及探討。

2.1 郵件分類管理與應用

於此郵件管理應用中，包含有「郵件分類管理」及「產業應用」等課題，其中透過此些研究之成果，即可了解目前郵件管理系統應用方式與方向。

(A) 郵件分類管理

現今多數人以電子郵件進行郵件溝通，致使郵件數量過於龐大，而超乎個人之處理能力，產生郵件超載問題。為解決此問題，Schuff 等人(2006)建構一套自動類聚管理系統(Automatic Clustering E-Mail Management System)，該系統乃運用多屬性(Multi-Attribute)、多重權重分配(Multi-Weight)類聚之方法，於電子郵件進入此系統後進行郵件解析任務。此外，Andreas 及 Marcin (2006) 架設一套以自組織網路(Self-Organizing Maps)為基礎之郵件管理工具。該工具乃提供一個直覺化與簡單的未傳送郵件清單，並且提供直覺式導覽功能，使內容相似之郵件先行歸類於同一區塊，並以不同分類夾進行畫分，因此用戶可輕易掃描相似內容電子郵件。

甚者，由於電子郵件之盛行，加拿大政府乃導入電子郵件分類系統，並整合現有檔案文件資訊管理系統(Records, Document, and Information Management System; RDIMS)，然而郵件保密之機制尚未完善。因此，Eun 及 Natasha (2008) 提出電子郵件管理與分類方式，可依循「資訊公開法」、「保密法」及「圖書館與檔案法」制定郵件分類與其傳送方式，並給予 RDIMS 權限以新建郵件檔案、管理和增加郵件資訊與文件，即可加強機密郵件之傳輸安全。電子郵件之使用促使郵件軟體之迅速發展，然而郵件軟體於發送與接收郵件上尚存速度過於緩慢之問題。因此，Laura 及 Maria(2001) 提出以 SMTP 與 POP3 協定模式建構於 SPECmail2001 軟體中，以改善郵件收發效率。而根據社會心理學與經濟學所架構得「利社會行為模組 (Models of Prosocial Behavior)」指出，接收到一個有幫助回應之可能性與收件人的數量呈反函數存在，因此為檢驗此理論，Barron 和 Yechiam (2002) 提出一項實驗來調查電子郵件對資料之請求，研究結果顯示，對於單一收件者而言，有著更多有幫助並詳述的回應，此些結果亦可解釋「社會穿插理論 (Social Cueing Theory)」應用程式之意義，以及電子郵件通訊議題之討論。

(B) 產業應用

對於目前企業而言，網路郵件乃為目前企業間、企業內與員工間、或者企業對顧客間進行溝通之重要管道之一。其中包含客服領域、教育領域等，如以下說明之。

➤ 客戶關係管理領域

Joung 及 Yang (2009) 提出一套個人化電子郵件篩選機制，當中用戶可依照本身所設立郵件內容與條件，允許郵件進入用戶信箱，並儲存至對應之類別中。雖然現今電子郵件系統的功能已相當完善，然尚缺乏較為直覺且簡單之操作模式。此外，有鑑於先前之郵件回覆系統乃先行提供各類建議範本，並使客服人員先對範本設定編號，以取得所需範例，進而減少搜索相關回覆資訊時所消耗時間，然而此文件分類技術仍以單一概念為基礎，未能真正回覆使用者之問題 (Kosseim 等人，2001)。是故，Sung 及 Chih (2004) 乃使用多樣概念，並結合概念與分類兩者關聯，進行簡單郵件分類。該系統乃透過不同概念及分類結合成一套動態統一電子郵件機制，以建議不同內容範本，進而提升建議範本準確性，即可減輕客服人員回覆郵件往返次數，增加回覆郵件效率。

此外，隨著電子郵件的使用，有許多組織意識到電子郵件所伴隨而來的問題，如電子郵件收發數量上的限制，以及撰寫內文的文筆方面，都是目前的組織迫切渴望解決的缺陷 (Burgess 等人；2004)。此外，Merisavo 和 Raulas (2004) 考核使用行銷式電子郵件以維持與加強品牌忠誠度之成效，以及消費者族群對於郵件內容之評價 (Gimenez, 2000)。

➤ 教育應用領域

由於電子郵件上盛行，電子郵件已用來學生與教師之間教育與溝通上之應用。Hassini (2006) 認為電子郵件清單可提供學生與教師之間一個良好的溝通平台，亦即可作為教學內容作之補救措施。最後根據實際在班級上應用的結果，明顯增加了學生與教師之間的互動性，學生們也利用此平台對教師們的教學方式給予了大量的評價。此外，Hu 等人 (2009) 藉由調查 2998 位具備 24 至 48 個月教學經驗之教師，試圖解析學校教師們對於電子郵件的使用模式。該研究之調查結果顯示以下兩種情形：(1)當教師們任教於較高等級的學校之時，教師們主要將電子郵件使用在與同事的溝通上面、(2)當教師們任教於較低等級的學校之時(特別是小學教師)，則主要用於與家長們的溝通中；此結果可應用於討論教師們的職前教育以及專業發展中。Couzenza (2009) 提出以電子郵件為媒介，將相關資料層次性地提供給學生，以協助學生完成作業；該研究結果顯示，學生們參與課堂上的活動與討論情形有大幅度的改善。

此外，Cavus 和 Bicen (2009) 為研究在科技學院中，學生們對於電子郵件的運用方式，以及

找出大多數學生愛用的免費電子郵件服務，根據分析結果顯示，大多數科技學院的學生普遍愛好使用 HOTMAIL 免費電子郵件服務。Itakura (2004) 利用在香港理工大學之日本留學生，以及在日本鹿兒島大學的香港留學生，探討不同文化、環境下電子郵件所使用之方式，進而解析文化觀念是如何形成、修改、撤銷或加強之問題，同時也評估電子郵件在其中所扮演的協助角色作用。此外，Liaw 和 Johnson (2001) 探討台灣與美國跨文化之間電子郵件通信溝通的型態，該研究之數據分析偏重於文化信息傳播的類型、文化理念的影響以及價值觀上的溝通效果，研究結果發現，好奇心對於了解其他文化有著激勵之作用，而透過跨文化的電子郵件能更好的去認識到彼此之間的文化差異。

現有的資料顯示，手機電子郵件在日本青少年與其社會關係之間，扮演著一種樞紐、橋樑與打破舊有關係之存在；是故，Boase 和 Kobayashi (2008) 研究結果可用於一般用戶以及重度用戶上。此外，為研究電子郵件的通訊與傳統面對面之間的溝通之相互影響的關係，以及電子郵件在工作場所中，員工如何有效的管理電子郵件的交流，是故，O'Kane 和 Hargie (2007) 透過問卷的方式來解析員工們對電子郵件的看法，研究結果顯示，隨著以電腦為媒介進行通訊的使用率增加，電子郵件與傳統面對面之間的溝通有著正面與反面之影響，並逐漸具有越來越複雜之特性，而管理電子郵件方面，主要的關鍵是在於提高員工對於電子郵件通訊的意識，以發揮最大效益，並防止消極後果的出現，此研究結果將協助建立一個電子郵件通訊的解釋與調查之概念。

➤ 員工訓練績效評量領域

為調查電子郵件如何影響到公司內部員工之工作績效，Mano 和 Mesch (2010) 使用次級層次分析 (Secondary Level Analysis) 對普網和美國生活樣本 (The Pew and American Life sample) 進行分析；該研究分析結果顯示：(A)管理階級的人員對於電子郵件的收發數量較非管理階級之人員多、(B)電子郵件的影響與用戶的年齡、性別、婚姻狀況以及教育程度有關、(C)電子郵件的收發數量與工作績效成正相關之關係。Baruch (2005) 調查在大型國際公司中，使用電子郵件進行的網路霸凌對工作所造成的影響。該研就結果顯示受到網路霸凌影響的員工，在工作的滿意度與效率上有著負面的影響，並帶有強烈的焦慮感，意圖離開組織之行為，是故根據這些結果可得知，一個在心理上受到威脅與衝擊以及資訊的誤用，將會對個人或組織帶來難以想像的傷害。

➤ 醫療應用領域

由於供應商的關注程度、病人的不便以及技術上的低效率利用，導致電子郵件在醫學上難以推廣的主要因素，是故，White 等人 (2004) 為克服上述問題，從 3007 封病人與醫師之間的電子郵件隨機抽取樣本，並另隨機抽取 10% 的電子郵件作為對照組，經過分析得知，要推廣電子郵件系統以作出適當的初級醫療保健，患者們必須遵守以下幾點準則：(1)內容重點、(2)限制每封郵件請求的數量、以及(3)避免緊急要求或高度敏感的內容以達到電子郵件的有效利用。此外，Cornwall 等人 (2008) 為研究電子郵件在專科護士與肺癌病人及其家屬之間的溝通效果，對已使用六個月以電子郵件為溝通方式的兩位專科護士以及 16 位病人與家屬進行問卷調查，實驗結果顯現，對病人及其家屬而言，此種通訊方式較以往來的更快速、簡易，並對其反應速度感到滿意，另外專科護士們也對此種通訊方式給予正面的肯定。最後，Hobbs 等人 (2003) 乃評估電子郵件於整合傳遞系統中，醫生與病人間之使用情形，該研究認為透過充足的事前篩選、分流以及償還機制的話，將會有效的提升醫師與病人間以電子郵件做為通訊形式的發展。

2.2 郵件分類技術

於郵件分類技術之議題中，過去相關研究乃以「郵件分類演算法」及「郵件資料解析」兩方面進行探討，本研究乃期望觀察其中郵件分類技術與郵件資料特性之相異處，以更深一層瞭解郵件分類技術與模式。

(A) 郵件分類演算法

現今存在的主要分類演算法技術中，主要包含 K-NN、決策樹、自然貝式、隨機森林、支持向量機等資料探勘演算技術；此外，由於垃圾郵件之氾濫及郵件安全性逐漸受重視，故現有研究亦著重於垃圾郵件過濾與保密協定演算法之發展，以下即分別探討此兩類演算法之研究發展，如以下說明之。

➤ K-NN、決策樹、自然貝式等資料探勘演算法

Zhou 等人 (2007) 結合圖片信息測量 (Picture Information Measurement; PIM) 及關鍵字擷取

技術，建構一套智慧型電子郵件分類系統；該系統乃使用簡易貝式分類機 (Native Bayes Classifier) 先行過濾郵件 (以郵件內文之關鍵字以作為判斷特徵)，之後，以圖片信息測量分析郵件附錄中所包含之圖像訊息，並將各郵件進行分類；是故，藉由關鍵字過濾與圖片資訊分析後，即可針對目標郵件進行有效地分類。

此外，**Poon 及 Chang (2003)** 利用電子郵件之詞彙相關性與 K 值鄰近演算法以進行電子郵件分類。該研究係依照郵件事先所建立之詞彙相關性，及各關鍵詞彙與類別之關係進行第一次分類，之後再利用 K 值鄰近演算法進行郵件最後之分類。此兩階段之分類方式即可獲得較佳之郵件分類效果。甚者，**Yu 與 Zhu (2009)** 提出結合倒傳遞類神經網路 MBPNN (Modified Back-Propagation Neural Network) 和語義特徵空間 SFS (Semantic Feature Space) 以進行郵件之分類。該研究乃修改倒傳遞類神經網路 MBPNN，增加引進網路起點和使用適應性學習法以調整學習比率，即可有效改善上述之缺點。此外，於在郵件分類中，電子郵件之內容解析易產生詞意誤判之情形，故該研究乃以語義特徵空間 SFS (Semantic Feature Space) 技術，以減少相似的特徵值和詞意，進而提高分類的精度和效率，以達到分類郵件效果。

Segal 及 Kephart (2000) 利用 TF-IDF 分類器，建構一套智慧型郵件分類機制 (SwiftFile)，以取代目前使用者需自行分類郵件之任務。該系統係先預設三個分類資料夾，使用者可先自行分類郵件至各資料夾，系統即透過分類器分析資料夾內各郵件內容定義與資料夾類型之關係，以學習使用者分類方式，使郵件寄送至使用者時自行分類至此三個資料夾，以達到郵件自動分類效果。

在郵件分類方面，**Crawford 等人 (2004)** 利用分類器中之向量空間、k-NN、決策樹、自然貝式與最大似法 ML (Maximum-Likelihood)，以及詞語分析中詞彙庫 BOW (Bag of Words) 與特徵選擇進行交叉組合，以比較當中各組合對於電子郵件之分類具備有最佳效益，實驗結果乃顯示最大似法與特徵選擇之組合具備電子郵件分類之最佳績效。此外，**Islam 等人 (2009)** 乃以統計學習機制 (Statistical Learning Algorithms) 建構一套郵件多層分類器 (Multi-classifier)，能夠有效地改善現今郵件分類系統中，所存在的詞彙誤判 (即具備灰色地帶之詞彙集) 以及準確度之問題。

此外，為瞭解郵件內文之字詞對郵件分類之影響，**Chang 及 Poon (2009)** 於電子郵件分類器中，利用郵件字詞作為分類特徵進行分類效能之評估，該研究係透過天真貝式分類器 (Naive Bayes Classifier)、第 K 鄰近點的相似性分類器 (K-NN Classifiers Using Resemblance) 與第 K 鄰近點之 TF-IDF (詞頻計算與加權技術) 等分類器，以及郵件集合以分別探討「片語長度」、「涵蓋區域大小」及「最鄰近點長度」對分類之影響，並計算各方法可改善分類準確率之效益；然而實驗結果發現，僅使用字詞為分類特徵之分類效果上，對於各分類器皆無明顯之顯著差異；此外，於此實驗中亦發現公共郵件較私人郵件易於分類。

➤ 垃圾郵件過濾與保密協定演算法

另於過濾垃圾郵件方面，**Shih 等人 (2005)** 整合現有的郵件分類方法 (貝式、貝式網路及決策樹) 之優點，以建構一套檢測惡意郵件分類器。此分類器乃架設於電子郵件之伺服器中，並形成網路郵件過濾器，以自動掃描與過濾所接受到之郵件，並阻擋病毒式電子郵件與封鎖潛在之惡意郵件，進而達杜絕惡性郵件之成效。此外，**Ying 等人 (2010)** 結合決策樹、支持向量機和倒傳遞類神經網路，提出一套整合性模式以分類垃圾郵件。該方法先將郵件內容、標題、寄件等資料歸納得 14 種郵件特徵，再以綜合分類法分析此 14 種特徵，以判定是否為垃圾郵件，進而達到過濾效果。此外，為了有效改善郵件過濾之效能，**Gonzalez (2006)** 整合先前研究所提之黑名單 (Blacklist) 與白名單 (Whitelist) 解決方案之特性而成，經由實驗結果可得知，該程式可提升過濾垃圾郵件之績效。此外，**Herzberg (2009)** 結合路由器和 DKIM (Domain Keys Identified Mail) 以協助電子郵件過濾任務，此機制之任務主要著重在不受歡迎電子郵件訊息之過濾，以改善目前垃圾郵件過濾效果不彰之情況。更甚者，**Duan 等人 (2007)** 發展一套區別轉寄郵件協定，該協定乃允許收件者可控制由不同寄件者於網路上之遞送郵件，此外，該研究亦發展簡單的數學模式控制垃圾郵件過濾機制。而透過實驗驗證，證明 DMTP 能有效減少垃圾郵件接收，相較於目前簡易信件傳輸協定所建立之電子郵件系統，**Duan 等人 (2007)** 所建構之電子郵件系統能使垃圾郵件於網路上緩衝較長之時間，以使垃圾郵件過濾機制可即時修正郵件黑名單。

最後，**Kwon 等人 (2009)** 提出一項基於密碼認證方式之電子郵件保密協定，以提供保密的功能，根據實驗結果顯示，該協定可於不要求公共密鑰認證的情形之下，能充分有效的執行在資源受

到限制的移動裝置上。此外，Phan (2008) 為驗證最近聲稱提供了最完善保密機制的兩項電子郵件協定，是否具有其功效進行實驗，此兩項電子郵件協定乃使用驗證以及密鑰交換技術 (Key-Exchange Techniques)，並包含了完美轉送私密 (Perfect Forward Secrecy)、已知金鑰安全 (Known-Key Security)、金鑰及時性 (Key Freshness) 和未知金鑰分享安全 (Unknown Key-Share Resilience) 等技術，經過實驗證實，此兩項協定無法抵制重放攻擊 (Replay Attacks)，並進一步指出，在第一次攻擊時便突破了未知金鑰分享安全技術，而第二次攻擊時就已突破完美轉送私密技術，此結果明顯違背了開發人員的要求，故得知，目前較新的技術並不一定意味著更安全。

(B) 郵件資料解析

就郵件資料分析之議題而論，資料挖礦與分析技術不斷發展與進步，其主要分析郵件中「郵件社群資料」、「郵件垃圾訊息」、「郵件文字型資料」及「郵件使用者資料」等範疇，如以下分別說明之。

➤ 郵件社群資料分析

針對社群探勘之議題中，過去研究乃試圖由通聯紀錄中(如電子郵件)找出熟識的使用者社群，此等社群關係包含家人、鄰居、同事、以及同學等。故 Duan 等人 (2007) 乃利用電子郵件分類以協助社群探勘任務，該研究中乃包含三階層式過濾模組，以進行電子郵件之分類，該系統係透過分析時間屬性以過濾大量的電子郵件資料，先將其中文本資訊以及位址進行分層，以有效地減少搜索範圍，之後系統乃自動的將電子郵件之文本以及主體資料進行分類，再根據隱藏在郵件資料中位址信息以建構社群結構，以建構得電子郵件社群網絡，進而將同屬性之使用者歸類於同一社群中，以使專屬社群資訊提供作業可迅速達成。此外，Chundi 等人 (2009) 提出以時間為觀點以分析嵌入時間序列之分割 (Time Series Segmentation)，應用於發掘時間點跨越模式及隨時間變化之電子郵件溝通模式，並計算用戶端郵件溝通模式中時間序列項目集 (Item-Set)，尋找目標使用者之個人化溝通模式；此外，亦計算所有使用者的電子郵件資料中時間序列項目集，用以建立以社會為中心之溝通模式。最後以安然公司的電子郵件數據集 (Enron Email Data Set) 進行實證，亦獲得具正面效益之驗證。

➤ 郵件垃圾訊息分析

由於電子郵件濫用的增長，調查員需要高效率的電子郵件自動化分析工具。Appavu 等人 (2009) 提出以決策樹 (Decision Tree) 演算法為基礎，建構一套智慧型過濾恐怖信息郵件系統，以協助調查員於收集郵件線索和證據之任務。此外，因電子郵件具備高度之便利性，利用電子郵件以進行消費者詐騙之情況日趨嚴重。因此 Neese 等人 (2005) 分析電子郵件內容之產品推銷、產品、價錢、行銷通路方式之關鍵詞彙以判別該郵件是否為詐欺郵件，以降低消費者受騙機率。而隨著網路技術發展，電子郵件已經成為重要通訊工具，許多職業依靠者電子郵件傳達訊息(如：商業貿易或者教育機構)，當中為避免錯過重要之郵件訊息，Kadoya 等人 (2004) 提出確認郵件回覆時間，並依照多重屬性規則檢測時間語意表達，判斷使用者郵件重要訊息以及分類，此系統過濾郵件可讓使用者直接讀取重要郵件，忽略不必要之郵件。

➤ 郵件使用者資料、文字型資料解析

Sallis 和 Kassabova (2000) 為瞭解以電腦為媒介溝通中電子郵件所扮演的角色，該研究使用乃針對電子郵件進行定量和定性的分析；當中實驗資料係取自於新聞群組大型數據組，並透過分析電子郵件中文本屬性(如文字的數目、句子的長度與其他文體之特點)，並以分數加權機制評估該電子郵件之易讀性，該研究結果顯示若電子郵件有著缺乏易讀性之問題，將導致資訊傳遞上尚存相當模糊的區域。而 Nagabhushan 等人 (2009) 則提出應用軟性演算模型 (Soft Computing Model)，使郵件地址可準確映射至確定之目的地，甚者若存不完整或相似之郵件地址，則運用可讀性之通訊地址，以進行象徵性分析，並依據內容相似以判別送達地點，進而提高郵件寄出正確目的之準確率。最後，Sakurai 及 Suyama (2005) 使用文字探勘技術分析顧客之郵件資料，該研究之分析資料包含郵件主題、郵件內容、及郵件主體以決定郵件類型，並從電子郵件中取得關鍵字和統計資料(如關鍵字數等資訊)，藉由上述之郵件類型與關鍵資訊加以判定此顧客之類型，進而協助顧客中心操作人員主動地將各類資訊予顧客得知。

3. 漸進式郵件分類模式

本研究所提之「漸進式郵件分類模式」乃以郵件中所包含之郵件內容為基礎，先行將此些資料進行資料層級之歸類（包含為「數字型」、「數據型」、「文字型」、「附件檔案型」及「郵件聯絡人」等資料型別）；當中，「數字型」與「數據型」為第一層級資料、「文字型」與「附件檔案型」為第二層級資料、「郵件聯絡人」則為第三層級資料；待資料歸類完成後，即可以本研究所建構之「郵件數據型、數字型資料解析模組」分析郵件數據型及數字型資料，先行過濾垃圾郵件，以判定適合分類之郵件群；之後，利用「郵件文字型資料解析」模組，計算目標郵件與各類別之隸屬關係；最後，當目標郵件之類別判定結果不甚理想，即可利用「郵件聯絡人資料解析模組」重新計算目標郵件與各類別之隸屬關係，進而形成一套漸進式郵件類別判定模式，期望以具效率之方式進行郵件類別判定任務。因此本研究之主要流程可分為三大部份，分別為「郵件數據型、數字型資料解析模組」、「郵件文字型資料解析」及「郵件聯絡人資料解析模組」，如以下之說明。

3.1 郵件數據型、數字型資料解析模組

本研究所提之「郵件數據型、數字型資料解析法則」，乃以郵件於郵件資料中所儲存數據型、數字型資料為分析基礎，以先行過濾此目標郵件是否為「垃圾郵件」。首先，郵件數據型資料乃包含「Message-ID」、「Received」及「IP」等資料，此些資料乃記錄郵件之發送 Gateway 主機名稱、郵件伺服器主機名稱及發送之 IP 位址；其次，數字型資料則包含郵件送發日期等（當中數據型、數字型之紀錄細項詳見表 3.1）。此些資料皆可用以辨識此目標郵件是否歸類為「垃圾郵件」（包含廣告、病毒、色情等郵件），若為垃圾郵件者，即不需進行分類之任務；反之，則將此郵件進行解析（即郵件文字型資料解析），並完成郵件類別判定作業。

表 3.1、數據型、數字型資料說明

資料類型	資料項目	資料說明
數據型	Message-ID	根據 RFC 規範，Message-ID 代表之資訊係為該電子信件之唯一識別代號，並藉由 SMTP Gateway 之送發信作業，Message-ID 馬上會被註記 Gateway 主機名稱（如 xxx@aaa.com）
	Received	Received 乃屬於「追蹤 (Trace)」欄位，於郵件伺服器接受一封電子郵件處理完後，並往下傳送時，郵件伺服器需於該電子郵件之開頭紀錄寫下 Received 資訊，因此可用以追蹤郵件伺服器資訊。
	IP	記錄郵件送發位址。

數字型	Date	記錄郵件送發日期。

- CA_k 第 k 種數據型屬性， $k=1, \dots, q$
- $CA_{k,l}$ 第 k 種數據型屬性之第 l 個限制值， $l=1, \dots, t$
- $CASet(CA_k, P.)$ 對所有郵件而言，第 k 種數據型屬性之限制值所成集合
- IA_j 第 j 種數字型屬性， $j=1, \dots, p$
- $LL(IA_j, P.)$ 對所有郵件而言，第 j 種數字型屬性之篩選區間下界值
- LLM_j 對所有郵件而言，第 j 種數字型屬性之篩選區間下界係數
- $Mean(IA_j, P.)$ 所有郵件於第 j 種數字型屬性之平均表現值
- P_T 目標郵件
- $P(IA_j, P_T)$ 於第 j 種數字型屬性下，目標郵件之分類合適性得分值
- $P(CA_k, P_T)$ 於第 k 種數據型屬性下，目標郵件之分類合適性得分值
- PV_T 目標郵件於後續分類作業之合適值 PV_T
- $P_CA(CA_k, OP_T)$ 於第 k 種數據型屬性下，目標郵件之數據型資料
- $P_Sum(IA_j, WR_m, OP_i)$ 於第 j 種數字型屬性下，目標郵件之表現值

$UL(I A_j, P_T)$	對所有郵件而言，第 j 種數字型屬性之篩選區間上界值
ULM_j	對所有郵件而言，第 j 種數字型屬性之篩選區間上界係數
$Var(I A_j, P.)$	對所有郵件而言，第 j 種數字型屬性之表現值變異數
AP_Set	判定為可進行後續分類之郵件群

Step (A1)：界定郵件篩選條件

依據郵件之往返及郵件收信者使用 Outlook 軟體收信之經驗，各郵件收信者可依主觀經驗定義各屬性於目標郵件 P_T 之「數字型屬性 ($I A_j$)」與「數據型屬性 ($C A_k$)」篩選條件，以作為判斷郵件分類合適性之準則。當中，本研究乃定義「目標郵件 P_T 於數字型屬性 $I A_j$ 之篩選條件」時，郵件收信者可依目標郵件發送時間（即病毒、廣告或色情等郵件之發送時間多數為非正常時間，如 2012 年等）自行定義篩選區間之範圍。其計算方式乃利用所得之平均表現值 $Mean(I A_j, P.)$ 作為篩選區間之中心點、以各收發郵件表現值變異數 $Var(I A_j, P.)$ 作為篩選區間之單位寬度，提供郵件收信者自行定義篩選區間之寬度大小（ ULM_j 、 LLM_j 係數值表示篩選區間之寬度倍數），所定義之篩選區間為 $(LL(I A_j, P.), UL(I A_j, P.))$ 。其數學計算如公式(3.1)與公式(3.2)所示。

$$LL(I A_j, P.) = Mean(I A_j, P.) - LLM_j * Var(I A_j, P.) \quad (3.1)$$

$$UL(I A_j, P.) = Mean(I A_j, P.) + ULM_j * Var(I A_j, P.) \quad (3.2)$$

而定義「目標郵件 P_T 於數據型屬性 $C A_k$ 之篩選條件」時，乃不同於數字型屬性 $I A_j$ 之篩選條件，數據型屬性 $C A_k$ 之篩選條件，乃由郵件收信者主觀定義影響作業項目之離散性限制值，而所有限制值所形成之集合即為數據型屬性之篩選條件。郵件收信者定義目標郵件 P_T 於數據型屬性 $C A_k$ 之所有篩選條件限制值 $C A_{k,l}$ ，並將所有篩選條件限制值以集合型式 $CASet(C A_k, P_T)$ 表示，其篩選條件集合如公式(3.3)所示。

$$CASet(C A_k, P.) = \{C A_{k,1}, \dots, C A_{k,l}, \dots, C A_{k,t}\} \quad (3.3)$$

Step (A2)：判斷目標郵件於後續分類作業之合適性

此步驟乃取得目標郵件 P_T 於各類型屬性之相關表現值；首先，若目標郵件於數字型屬性 $I A_j$ 下之表現值 $P_Sum(I A_j, P_T)$ 落於定義之篩選區間 $(LL(I A_j, P.), UL(I A_j, P.))$ ，則其合適性得分 $P(I A_j, P_T) = 1$ ，代表於該屬性下目標郵件適合於後續分類作業；反之，合適性得分 $P(I A_j, P_T) = 0$ ，代表於該屬性下目標郵件不適合後續分類作業（如公式(3.4)所示）。

$$P(I A_j, P_T) = \begin{cases} 1, & \text{IF } LL(I A_j, P.) \leq P_Sum(I A_j, P_T) \leq UL(I A_j, P.) \\ 0 & \text{Otherwise} \end{cases} \quad (3.4)$$

若目標郵件 P_T 於數據型屬性 $C A_k$ 下，其數據型 $P_CA(C A_k, P_T) \notin CASet(C A_k, P.)$ ，則其合適性得分 $P(C A_k, P_T) = 1$ ，代表於該屬性下目標郵件適合於後續分類作業；反之，其合適性得分 $P(C A_k, P_T) = 0$ ，代表於該屬性下目標郵件不適合於後續分類作業；其數學計算如公式(3.5)所示。

$$P(C A_k, P_T) = \begin{cases} 1, & \text{IF } P_CA(C A_k, P_T) \notin CASet(C A_k, P.) \\ 0 & \text{IF } P_CA(C A_k, P_T) \in CASet(C A_k, P.) \end{cases} \quad (3.5)$$

Step (A3)：判定適合分類之郵件群

基於 Step (A3) 所計算之目標郵件於不同屬性合適性得分後，此步驟乃計算目標郵件於後續分類作業之合適值 PV_T （即目標郵件於數據型、數字型中，各屬性合適性得分之連乘積），如公式(3.6)所示。當中，若目標郵件之合適值 PV_T 不為 0 時，則代表經郵件數據型、數字型資料解析後，該目標郵件適合於後續分類作業，並將此目標郵件歸類於可進行後續分類之郵件群中 AP_Set（如公式(3.6)所示）。

$$PV_T = \prod_{all\ j,k} P(I A_j, P_T) \cdot P(C A_k, P_T) \text{ and IF } PV_T \neq 0 \text{ THEN } P_T \in AP_Set \quad (3.6)$$

3.2 郵件文字型資料解析模組

本研究所提之「郵件文字型資料解析法則」，乃以郵件於郵件資料中所儲存文字串型別資料為分析基礎（包含「郵件主旨」、「郵件內文」、「附加檔案名稱」等資料），先行篩選儲存內文中已存在之非關鍵字集（如「的」及「與」等字詞），以擷取當中候選關鍵字詞（即上述非關鍵字集外，尚餘未歸類完成之字集，本研究乃將此類字集定義為「候選關鍵字集」），再藉由領域專家所建構之電子郵件關鍵字/類別之隸屬關係，將上階段所獲得之關鍵字計算其所屬類別，進而得知該目標郵件之類別。以下乃先彙整「郵件文字型資料解析法則」之推論模式所使用之符號，之後說明「郵件文字型資料解析法則」推論模式之詳細步驟。

UI	電子郵件所包含之文字型資料集合
UI _i	電子郵件中第 <i>i</i> 項文字型資料， <i>i</i> =1,2,3；UI ₁ 為「郵件主旨」、UI ₂ 為「郵件內文」及 UI ₃ 為「附加檔名」
NKW	非關鍵字集
FUI _i	第 <i>i</i> 項重要文字型資料（即經 NKW 篩選過之第 <i>i</i> 項文字型資料）
IW _{i,j}	第 <i>i</i> 項重要文字型資料中第 <i>j</i> 筆重要詞彙
ST	領域專家所建置之領域類別關鍵字集合
ST _q	第 <i>q</i> 類別之關鍵字集 (<i>q</i> =1,2,..., <i>m</i>)
KW _k	關鍵字集中第 <i>k</i> 筆類別關鍵字
C(IW _{i,j})	IW _{i,j} 之字數
M _{i,j,k}	IW _{i,j} 與第 <i>k</i> 筆類別關鍵字之媒合比例
D _r	既有文件庫中之第 <i>r</i> 份文件
F(D _r , KW _k)	文件 <i>r</i> 中關鍵字 KW _k 之發生頻率
SF _{k,q}	KW _k 於第 <i>q</i> 類別之隸屬係數值
R _{i,j,k}	第 <i>i</i> 項文字型資料之第 <i>j</i> 筆重要詞彙與第 <i>k</i> 筆類別關鍵字之關聯係數
UR _{i,q}	第 <i>i</i> 筆文字型資料與第 <i>q</i> 類別之關聯係數值
AP _T	經郵件數據型、數字型資料解析後之可分類目標郵件
PR _{T,q}	目標郵件與第 <i>q</i> 類別之類別隸屬係數

Step (B1)：篩選非關鍵字集

為能於電子郵件文字型資料中獲得重要詞彙，以推論關鍵字。本模式於蒐集郵件內容中所包含之各項郵件文字型資料（UI_i，當中 *i*=1,2,3；UI₁ 為「郵件主旨」、UI₂ 為「郵件內文」及 UI₃ 為「附加檔名」）後，經由非關鍵字集（NKW）篩選，去除各項郵件文字型資料之非關鍵字，即成為各項重要郵件文字型資料集合（FUI_i；即經 NKW 篩選過之郵件內文集合），如公式(3.7)所示。

$$FUI_i = UI_i - NKW \quad \text{當中 } i=1,2,3 \quad (3.7)$$

Step (B2)：儲存重要詞彙

將各項重要郵件文字型資料（FUI_i），利用二至六字詞之解析方式將其拆解成若干字詞（即 $FUI_i = \{IW_{i,1}, IW_{i,2}, \dots, IW_{i,j}, \dots\}$ ），以視為各項重要郵件文字型資料之重要詞彙（IW_{i,j}），彙整如表 3.2。

表 3.2、重要詞彙統整表

	字詞 1	字詞 2	...	字詞 <i>j</i>	...
FUI ₁	IW _{1,1}	IW _{1,2}	...	IW _{1,j}	...
...
FUI _i	IW _{i,1}	IW _{i,j}	...

Step (B3)：計算重要詞彙與類別關鍵字之媒合率

本步驟乃根據各筆重要詞彙（IW_{i,j}）之字數（C(IW_{i,j}））與其比對相符之第 *k* 筆關鍵字（KW_k）之字數（C(KW_k））之媒合比例（M_{i,j,k}），如公式(3.8)所示。

$$M_{i,j,k} = \frac{C(IW_{i,j} \cap KW_k)}{C(KW_k)} \quad (3.8)$$

Step (B4)：彙整類別關鍵字與類別之隸屬關係

彙整關鍵字與類別之隸屬關係可分數個階段進行之。首先，於建構模式前需由系統管理者或領域專家建構一龐大關鍵字資料庫，此任務可運用孫銘聰、侯建良（2003）擷取訓練文件之字串，歸納出現頻率較高且屬於關鍵字詞者為關鍵字，進而得到所有文件之關鍵字集；這些關鍵字於不同文件之發生頻率可整理如表 3.3。其次，利用領域文件（如電子郵件等資料）與類別之相關特性，將已知內容與類別之文件認定為訓練文件（即表 3.3 之 D_r ），並根據侯建良、林峰興與畢威寧（2003）之方法論，將類別與關鍵字隸屬係數加以計算及精確化後，即可針對所有類別分別得到關鍵字 KW_k ，與類別 ST_q 之隸屬係數 $SF_{k,q}$ ，其結果整理如表 3.4。

表 3.3、各文件之關鍵字整併頻率摘要表

$F(D, KW)$	KW_1	KW_2	...	KW_k	...
D_1	$F(D_1, KW_1)$	$F(D_1, KW_2)$...	$F(D_1, KW_k)$...
...
D_r	$F(D_r, KW_1)$	$F(D_r, KW_2)$...	$F(D_r, KW_k)$...

表 3.4、關鍵字與類別之類別隸屬係數表

SF	ST_1	ST_2	...	ST_q	...	ST_m
KW_1	$SF_{1,1}$	$SF_{1,2}$...	$SF_{1,q}$...	$SF_{1,m}$
...
KW_k	$SF_{k,1}$	$SF_{k,2}$...	$SF_{k,q}$...	$SF_{k,m}$

Step (B5)：彙整重要詞彙與類別之隸屬關係

本步驟乃將各項郵件文字型資料中各重要詞彙與各類別關鍵字之媒合率（ $M_{i,j,k}$ ）乘上與其相對相符之各筆類別關鍵字其類別隸屬係數（ $SF_{k,q}$ ）（參見公式(3.9)），即可求得各項郵件文字型資料中每筆重要詞彙（ $IW_{i,j}$ ）與各類別之隸屬關係（ $R_{i,j,q}$ ），合併整理後得表 3.5。

$$R_{i,j,q} = \text{Max}(M_{i,j,k}) \times SF_{k,q} \quad \text{for all } i, j \text{ and } q = 1, 2, \dots, m \quad (3.9)$$

表 3.5、重要詞彙與類別之隸屬關係彙整表

R	ST_1	ST_2	...	ST_q	...	ST_m
$IW_{i,1}$	$R_{i,1,1}$	$R_{i,1,2}$...	$R_{i,1,q}$...	$R_{i,1,m}$
...
$IW_{i,j}$	$R_{i,j,1}$	$R_{i,j,2}$...	$R_{i,j,q}$...	$R_{i,j,m}$

Step (B6)：計算郵件內文與類別之關聯係數

本步驟乃先將第 i 項郵件內文所解析之各筆重要詞彙其類別隸屬係數（ $R_{i,j,q}$ ）予以加總，然後計算各筆重要詞彙之類別隸屬係數（ $R_{i,j,q}$ ）所占百分比（參見公式(3.10)），即可得到第 i 項郵件內文與其各類別之關聯係數值（ $UR_{i,q}$ ）。對於郵件內文各項郵件內文反覆進行前步驟之計算後，則各項郵件內文皆可求得其與各類別之類別關聯係數值（ $UR_{i,q}$ ），合併整理後如表 3.6 所示。

$$UR_{i,q} = \frac{\sum_{all\ j} R_{i,j,q}}{\sum_{all\ j,q} R_{i,j,q}} \times 100\% \quad (3.10)$$

表 3.6、郵件內文與類別之關聯係數彙整表

UR	ST ₁	ST ₂	...	ST _q	...	ST _m
UI ₁	UR _{1,1}	UR _{1,2}	...	UR _{1,q}	...	UR _{1,m}
...
UI _i	UR _{i,1}	UR _{i,1}	...	UR _{i,q}	...	UR _{i,m}

Step (B7)：計算郵件與類別之類別隸屬係數

本步驟乃以步驟(B6)所得之類別隸屬係數表為基礎，計算各項郵件內文所屬目標郵件 (AP_T) 與各類別 (ST_q) 之類別關聯係數。其詳細作法乃將目標郵件所包含之第 i 項文字型資料與第 q 類別之隸屬係數 (UR_{i,q}) 予以加總，並乘上對應之權重值 (α_i，假設第 1 項文字型資料 (即 UI₁「郵件主旨」之代表性較高，即可賦予較大之權重值)，並且計算該類別所佔所有類別關聯係數之比例 (參見公式(3.11))，即可得到目標郵件與其各類別之類別隸屬係數值 (PR_{T,q})。

$$PR_{T,q} = \frac{\sum_{all\ i} UR_{i,q} \times \alpha_i}{\sum_{all\ i} \sum_{all\ q} UR_{i,q} \times \alpha_i} \cdot 100\% \quad \text{WHERE} \quad \sum_{all\ i} \alpha_i = 1 \quad (3.11)$$

Step (B8)：郵件與類別之類別隸屬係數總表建立

所有郵件反覆進行前步驟之計算後，即可獲得各郵件與各類別之類別隸屬係數，如表 3.7 所示。

表 3.7、郵件與類別之類別隸屬係數表

PR	ST ₁	ST ₂	...	ST _q	...	ST _m
AP ₁	PR _{1,1}	PR _{1,2}	...	PR _{1,q}	...	PR _{1,m}
...
AP _T	PR _{T,1}	PR _{T,1}	...	PR _{T,q}	...	PR _{T,m}

3.3 郵件聯絡人資料解析模組

利用聯絡者關聯程度修正目標郵件類別隸屬係數可分為五大步驟進行。運用使用者與其他聯絡者之電郵發送頻率關係，計算其與各聯絡者之電郵係數，進而評定各聯絡人與使用者間關聯程度高低，建立不同關聯等級之區隔，以選定關鍵等級之聯絡人，之後於關鍵聯絡者所有聯繫信件中，選取關鍵郵件與各類別之關係，並修訂使用者類別。以下即為聯絡者關聯程度之推導過程。

Step (C1)：計算各聯絡者之電郵係數

首先，定義目標使用者 U_T 與各聯絡人之電郵係數為其與對應聯絡人之電子郵件收發頻率 (含回信、轉信、寄信及收信) 所計算而得。因此，記錄目標使用者 U_T 與其第 t 位聯絡者 L_t 之回信、轉信、寄信及收信等活動之發生頻率，並配合給定之權重值計算其與聯絡者 L_t 之電郵係數 M[L_t]，如公式(3.12)所示。

$$M[L_t] = W(R_t) \times R_t + W(T_t) \times T_t + W(S_t) \times S_t + W(M_t) \times M_t \quad (3.12)$$

其中，所有權重之總和為 1 (即 W(R_t)+W(T_t)+W(S_t)+W(M_t)=1)。

Step (C2)：制定各聯絡者預選等級

依據電郵係數 M[L_t] 之計算結果予以排序，依此排序結果可制定預選等級。如選定排序前 S 位聯絡者 (即為關鍵聯絡者 CL_j) 為修訂目標郵件之層級，即應賦予每位聯絡者對應之權重值 α_j (j=1,...,S)，以修訂目標郵件所屬類別。

Step (C3)：計算各聯絡郵件之郵件關聯係數

首先，先行彙整各關鍵聯絡者之聯絡郵件，並定義目標郵件與各聯絡郵件之「郵件關聯係數」為其與對應聯絡郵件中郵件主旨字元重複比例。因此，計算目標郵件 AP_T 與第 s 位聯絡者之第 t 個聯絡郵件 $OWD_{s,t}$ 之郵件主旨重複字數比例，即可獲得與各郵件之關聯係數 $M[OWD]$ ，如公式(3.13)所示。

$$M[OWD_{s,t}] = \frac{N(AP_T[TITLE] \cap OWD_{s,t}[TITLE])}{N(P_T[TITLE])} \quad (3.13)$$

當中， $N(AP_T[TITLE] \cap OWD_{s,t}[TITLE])$ 係代表目標郵件與第 t 個聯絡郵件之郵件主旨重複字數。

Step (C4)：制定各聯絡郵件預選等級

於各關鍵聯絡者中將各郵件關聯係數 $OWD_{s,t}$ 之計算結果予以排序，依此排序結果可制定預選等級。如選定排序前 V 個聯絡郵件為修訂目標郵件類別之層級，即應賦予各聯絡郵件對應之權重值 β_j ($j=1, \dots, V$)，以修訂目標郵件所屬類別。

Step (C5)：修正目標郵件與類別之類別隸屬係數

選定前 S 個關鍵聯絡者之前 V 個聯絡郵件後，並配合所賦予之權重值（如以聯絡郵件為例，由郵件關聯係數最高之聯絡郵件指定權重為 β_1 、其次 β_2 ，依此類推直至 β_s ，即排序為前之權重值大於等於後者）。本研究乃依此權重修正目標郵件與各領域類別之類別隸屬係數，以求得修正後類別隸屬係數 $PR'_{T,q}$ ，如公式(3.14)所示。此外，定義公式(3.14)中之 $M_{j,h}[PR_q]$ 為選擇之前 S 個關鍵聯絡人之前 V 個聯絡郵件中，排序為第 j 個關鍵聯絡人之第 h 個關鍵聯絡郵件與類別 ST_q 之類別隸屬係數。

$$PR'_{T,q} = \frac{PR_{T,q} + \sum_{j=1}^S \sum_{h=1}^V \alpha_j \cdot \beta_h \cdot M_{j,h}[PR_q]}{1 + \sum_{j=1}^S \sum_{h=1}^V \alpha_j \cdot \beta_h} \quad (3.14)$$

$$\text{where } 0 \leq \alpha_s \leq \alpha_{s-1} \leq \dots \leq \alpha_1 \leq 1 \text{ and } 0 \leq \beta_V \leq \beta_{V-1} \leq \dots \leq \beta_1 \leq 1$$

此公式可用以修訂表 3.7，以提升計算目標郵件與類別之隸屬係數正確性；因此，修正後目標郵件之類別隸屬係數如表 3.8 所示。

表 3.8、修正後目標郵件與類別隸屬係數表

PR	ST_1	ST_2	...	ST_q	...	ST_m
AP_1	$PR'_{1,1}$	$PR'_{1,2}$...	$PR'_{1,q}$...	$PR'_{1,m}$
...
AP_T	$PR'_{T,1}$	$PR'_{T,2}$...	$PR'_{T,q}$...	$PR'_{T,m}$

4. 系統應用流程

根據第三章所提出之漸進式郵件分類模式，本研究乃發展一套漸進式郵件分類系統。本研究乃分別就「系統管理者」及「一般使用者」進行各功能模組之介紹，以說明此系統於實際應用之運作方式（如圖 4.1 所示）。

➤ 系統管理者設定系統參數及關鍵字庫建立

系統管理者必須先設定系統內之各項參數，以保持系統主功能判斷之正確性。首先，管理者必須要決定郵件類別名稱，並點選郵件類別維護模組之郵件類別新增功能以設定郵件類別名稱，管理者再點選系統參數設定模組，以設定系統參數之門檻值及權重值（如圖 4.1 所示），之後點選關鍵字/非關鍵字維護模組，以設定關鍵字，最後管理者乃透過郵件資料維護模組之郵件新增功能，上傳具有分類代表性之訓練郵件樣本至系統中，並新增與郵件相關之關鍵字及設定其類別隸屬係數（如圖 4.2 所示），完成系統參數設定及關鍵字庫建立之步驟。

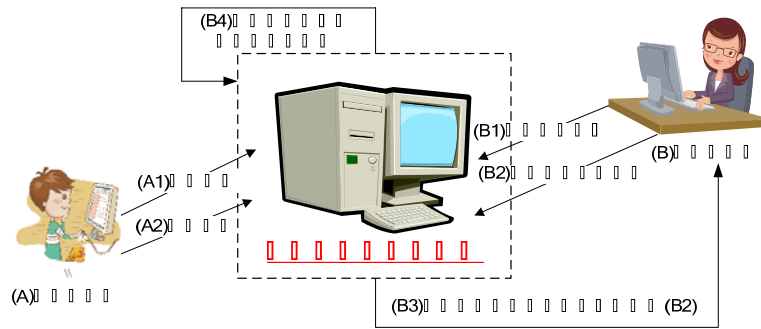


圖 4.1、漸進式電子郵件類別判定系統之應用流程



圖 4.2、系統參數設定模組之各模組門檻值設定



圖 4.3、設定郵件類別隸屬係數

➤ 一般使用者蒐集電子郵件

一般使用者於使用本系統前，蒐集許多未分類之電子郵件，之後乃利用本系統幫助其進行類別判定。使用者於使用本系統之前，蒐集了從 Pchome 個人電子報中所提供之各類電子報，透過網路郵件工具「outlook」收信，並使用 outlook 管理功能將所訂閱之電子報（如圖 4.4 所示）以「*.eml」格式儲存並彙整（如圖 4.5 所示），之後乃利用本系統分析使用者欲分析之電子郵件。

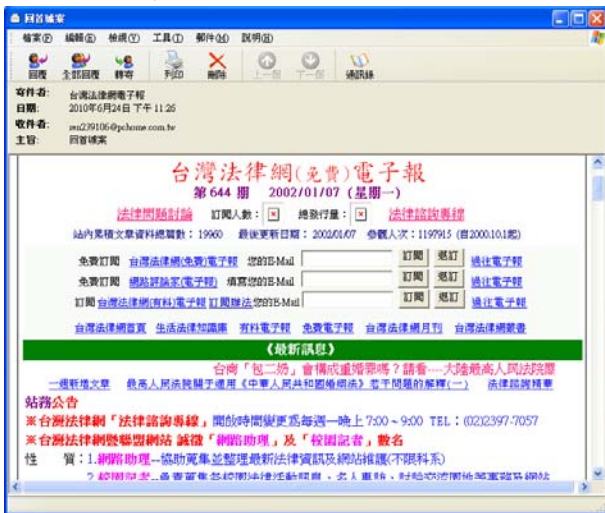


圖 4.4、以 outlook 顯示之電子報內容

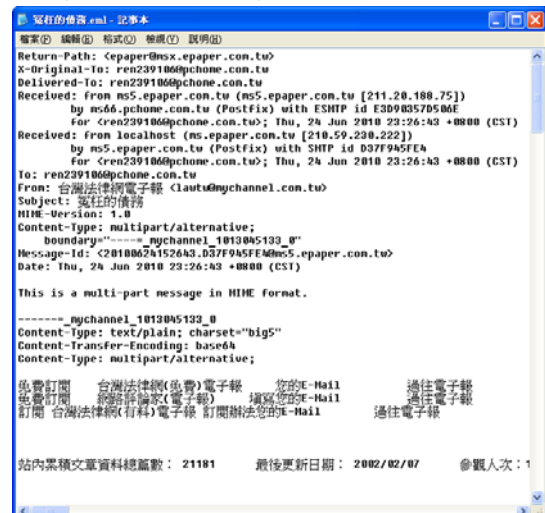


圖 4.5、「*.eml」格式之電子郵件內容

➤ 一般使用者上傳未分類之電子郵件

首先使用者要點選郵件資料維護模組之郵件新增功能，並選擇一份未分類之電子郵件名稱為「CHD、糖尿病與惡化缺血性中風女性功能性結果有關.eml」（如圖 4.6 所示），並輸入「CHD、糖尿病與惡化缺血性中風」之郵件基本資料，即可將「CHD、糖尿病與惡化缺血性中風」之郵件上傳至系統中（如圖 4.7 所示），以提供後續主功能漸進式郵件類別判定模組使用。

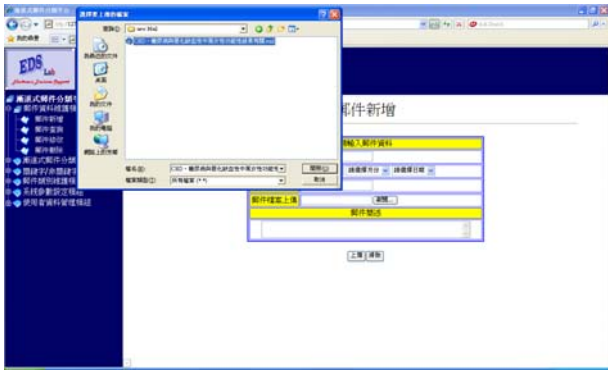


圖 4.6、選擇欲分析類別之郵件



圖 4.7、郵件新增完成

➤ **系統管理者使用主推論模組判斷郵件類別**

系統管理者須先點選漸進式郵件類別判定模組，該模組有三層針對不同郵件資料類型的分析模組，如：數字/數據型、文字型、聯絡人資料解析模組，管理者點選數字/數據型資料解析模組以開始判斷使用者所上傳之郵件類別，點選查詢按鈕以進行郵件類別判定，系統即會顯示數字/數據型資料分類結果（如圖 4.8 所示），如判斷出垃圾信則標記為垃圾信，剩下之郵件系統則會建議繼續往下一層文字型資料解析模組分析，以取得正確類別判定結果，如果門檻值低於系統門檻值 70%，系統則會建議繼續往下一層聯絡人資料解析模組分析（如圖 4.9 所示），並如上述步驟點擊系統訊息提供之超連結，最後，管理者即可點選分類結果儲存，以將本次分析紀錄儲存至系統中。



圖 4.8、數字/數據型資料解析模組之分類結果



圖 4.9、文字型資料解析模組之郵件分類結果

➤ **一般使用者查詢郵件及郵件類別關係**

一般使用者可查詢其所上傳之郵件及該郵件類別判定結果。使用者可透過點選郵件資料維護模組之郵件查詢功能，查詢其所上傳之郵件「CHD、糖尿病與惡化缺血性中風」是否成功完成分類並儲存至系統中，首先須先輸入查詢該郵件相關資料並選擇好查詢條件再點選查詢，即可顯示搜尋結果（如圖 4.10 所示），並可點選詳細資料，以查詢郵件判定結果為醫療保健（如圖 4.11 所示）。



圖 4.10、搜尋結果顯示



圖 4.11、郵件判定結果顯示(1)

5. 案例驗證

為驗證本系統之電子郵件類別自動推論績效，本研究乃以Pchome個人電子報中所提供之各類電子報透過網路訂閱，並利用郵件管理功能將所訂閱之電子報（如圖5.1）以「*.eml」格式收信並彙整（如圖5.2），並以其為探討案例。而驗證過程可分為訓練與測試兩大階段，以下即針對驗證資

料取得、系統驗證方式說明、評估指標定義與驗證結果分析而依序進行說明。



圖 5.1、以郵件系統顯示電子報內容

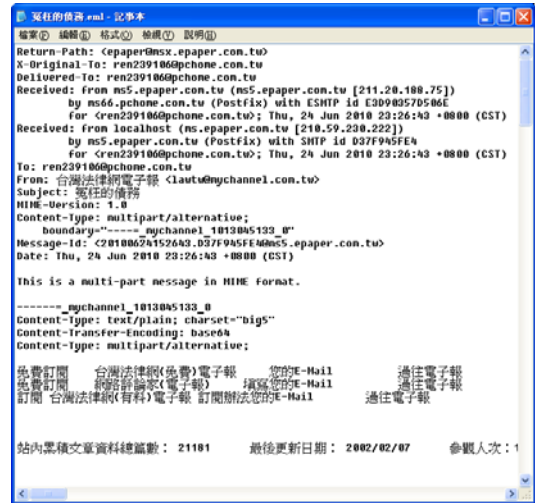


圖 5.2、以記事本顯示「*.eml」之內容

5.1 驗證資料取得

於系統訓練階段中，本研究乃以Pchome個人電子報中所提供之各類電子報為基礎，蒐集與影視娛樂 (M1)、資訊類 (M2)、工商金融經濟 (M3)、生活休閒人文 (M4)、知識教育 (M5)、政治法律 (M6) 及醫療保健 (M7) 等7個類別之電子郵件，總計700篇電子郵件。

5.2 系統驗證方式說明

為驗證本論文所提出之方法論與系統績效，首先於系統驗證第一階段（即系統訓練階段）乃自700份與「影視娛樂」、「資訊類」、「工商金融經濟」、「生活休閒人文」、「知識教育」、「政治法律」及「醫療保健」內容相關之電子郵件中，共挑選210份電子郵件作為訓練資料（各類別30份訓練資料），並逐一匯入系統中，以建構關鍵字與領域類別之關係。之後，於此所選定之七大類別中，各類隨機挑選3份文件（共計21份，整理如表5.1），以作為系統測試資料，並利用訓練階段所取得之關鍵字/類別隸屬關係，推論此21份測試電子郵件之隸屬類別，並藉由觀察系統所推論之電子郵件類別是否符合該電子郵件之實際類別，以確認本研究提方法論之正確性。待完成上述之第一階段系統績效驗證後，於第二階段（即系統測試階段）乃分為7個週期持續匯入訓練用之電子郵件資料，每週期皆再匯入70份電子郵件（各類別10份電子郵件），總計再匯入490份電子郵件（各類別70份電子郵件）；於各週期中乃利用前述21份測試電子郵件重新進行電子郵件類別推論，以分析系統於不同訓練電子郵件數量下之長期學期趨勢。

表 5.1、測試資料（僅列示兩類別資料，共六筆測試資料）

類別代號	電子郵件名稱	內容
M1	五月天唯一單曲「我」！	五月天大變了！？這首未發行的單曲，在網路上引起了各方愛樂人的討論！
	侯湘婷《17、18、19...侯湘婷》	湘婷 17 到 19 歲最完整的成長紀錄！
	亞洲電視劇搜查線 108 期	歡迎 F4 外借 言承旭為她「叫床」 兩人好得很
M2	星空舊電腦報第 25 期_登陸檔的認識	登陸檔(Registry),是儲存電腦裡各種資料的地方。
	星空舊電腦報第 34 期_計算機概論	電腦起源概略介紹電腦的前身就是計算機
	星空舊電腦報第二十期_燒錄軟體的使用 NERO5_X 2	NERO 5.X 軟體基本使用方法 2 觀看空白片容量和超燒空白片抹除 CD-RW 資料

5.3 評估指標定義

為評估電子郵件經漸進式分類模式之分類，對於電子郵件之推論是否具備精確性與可行性，本研究乃利用「郵件分類召回率」、「郵件分類正確率」與「郵件類別隸屬係數」此三項量化指標進行系統績效驗證，當中(1)郵件分類召回率乃為一相對比例值，為「實際類別與推論類別相符之類別個數」與「實際類別個數」之比例；(2)郵件分類正確率乃為一相對比例值，為「實際類別與推論

類別相符之類別個數」與「推論類別個數」之比例；(3)郵件類別隸屬係數乃代表「系統推論目標郵件於實際類別隸屬係數，期望藉由此些指標評估系統推論之類別與實際類別間的差異程度。

5.4 系統驗證結果分析

本研究將系統驗證結果分為「第一階段驗證結果分析」與「第二階段驗證結果分析」兩大項目。

5.4.1 第一階段驗證結果分析

於第一階段系統驗證中，本研究先由 700 封與「影視娛樂」、「資訊類」、「工商金融經濟」、「生活休閒人文」、「知識教育」、「政治法律」及「醫療保健」內容相關之電子郵件中，共挑選 210 封電子郵件作為訓練資料（各類別 30 封訓練資料），並逐一匯入系統中，以作為第一階段驗證之基礎訓練資料。以下即針對各項指標說明系統驗證過程，並分析系統驗證之結果。

(A) 漸進式郵件類別判定之「文字型指標郵件結果」(21 封指標郵件)

在 210 封電子郵件作為訓練資料之基礎下，系統針對 21 封指標郵件之文字型平均召回率與平均正確率均為 90%（標準差為 0.3007），而文字型隸屬係數為 42%（標準差為 0.1526）；其中，指標郵件之推論結果與指標郵件之理想結果完全符合之郵件 21 件共 19 封，佔總測試郵件之 90.47%。而驗證結果彙整如表 5.2。整體而言，於此階段中，文字型平均召回率與平均準確率之分佈趨勢趨於兩極化，而文字型平均類別隸屬係數之分佈趨勢則趨向常態分布，即表示在本階段中，大多數指標郵件經本系統判定後，即可獲得準確地郵件類別判斷結果。

表 5.2、第一階段文字型指標績效彙整

郵件名稱	實際類別	推論類別	實際類別數	推論類別數	正確類別數	指定最大值		類別隸屬係數
						召回率	正確率	
五月天單曲	M1	M1	1	1	1	100%	100%	33.00%
17、18、19 侯湘婷	M1	M1	1	1	1	100%	100%	79.00%
亞洲電視劇 108 期	M1	M1	1	1	1	100%	100%	49.00%
星空電腦報第 25 期	M2	M2	1	1	1	100%	100%	33.00%
星空電腦報第 34 期	M2	M2	1	1	1	100%	100%	29.00%
星空電腦報第二十期	M2	M2	1	1	1	100%	100%	46.00%
...
平均值						90%	90%	42%

(B) 漸進式郵件類別判定之「聯絡人指標郵件結果」(21 封指標郵件)

在 210 封電子郵件作為訓練資料之基礎下，系統針對 21 封指標郵件之聯絡人平均召回率與平均正確率均為 95.23%（標準差為 0.2182），而聯絡人平均類別隸屬係數為 82.10%（標準差為 0.2003）；其中，指標郵件之推論結果與指標郵件之理想結果完全符合之郵件 21 件共 20 封，佔總測試郵件之 95.23%。而驗證結果之績效彙整如表 5.3。

表 5.3、第一階段聯絡人指標績效彙整

郵件名稱	實際類別	推論類別	實際類別數	推論類別數	正確類別數	指定最大值		類別隸屬係數
						召回率	正確率	
五月天單曲	M1	M1	1	1	1	100%	100%	94.00%
17、18、19 侯湘婷	M1	M1	1	1	1	100%	100%	98.00%
亞洲電視劇 108 期	M1	M1	1	1	1	100%	100%	96.00%
星空電腦報第 25 期	M2	M2	1	1	1	100%	100%	72.00%
星空電腦報第 34 期	M2	M2	1	1	1	100%	100%	72.00%
星空電腦報第二十期	M2	M2	1	1	1	100%	100%	74.00%
...
平均值						95%	95%	82.10%

5.4.2 第二階段驗證結果分析

第二階段系統驗證之作法乃以第一階段驗證中，匯入系統資料庫之 210 封電子郵件為基礎，並將其作為第二階段之第一週期，再將剩餘之 490 封訓練用電子郵件分為七個週期依序匯入系統後端之資料庫中，並以第一階段驗證時所選取之 21 封指標郵件重新進行系統績效測試，以瞭解系統於不同訓練資料數量基礎下進行漸進式郵件類別判定之「文字型指標郵件結果」及漸進式郵件類別判定之「聯絡人指標郵件結果」之績效變化趨勢，進而分析本系統是否具備學習成長能力。

(A) 第二階段漸進式郵件類別判定之「文字型指標郵件結果」

由表 5.4 及圖 5.3 可知，以每週期增加 70 封訓練用電子郵件為單位，平均每週期文字型召回率與準確率之整體平均成長率皆為 1.59%；而文字型類別隸屬係數之整體平均成長率為 6.51%，故可知本系統分類能力良好，且具學習能力。

表 5.4、文字型指標績效彙整

漸進式郵件類別判定 文字型	各週期訓練用漸進式郵件類別判定—文字型郵件數量								平均	
	第一週期	第二週期	第三週期	第四週期	第五週期	第六週期	第七週期	第八週期		
	210 封	280 封	350 封	420 封	490 封	560 封	630 封	700 封		
召回率	平均值	90.48%	90.48%	95.24%	95.24%	100.00%	100.00%	100.00%	100.00%	96.43%
	成長率	-	0.00%	4.76%	0.00%	4.76%	0.00%	0.00%	0.00%	1.36%
	標準差	0.3007	0.3007	0.2182	0.2182	0	0	0	0	0.1297
正確率	平均值	90.48%	90.48%	95.23%	95.23%	100.00%	100.00%	100.00%	100.00%	96.43%
	成長率	-	0.00%	4.76%	0.00%	4.76%	0.00%	0.00%	0.00%	1.36%
	標準差	0.3007	0.3007	0.2182	0.2182	0	0	0	0	0.1297
類別隸屬係數	平均值	42.00%	40.76%	44.19%	45.57%	55.85%	72.76%	78.57%	80.09%	57.47%
	成長率	-	-1.24%	3.34%	1.38%	10.29%	16.90%	5.81%	1.35%	5.40%
	標準差	0.1526	0.1413	0.1276	0.1225	0.1036	0.0294	0.0461	0.0417	0.0956

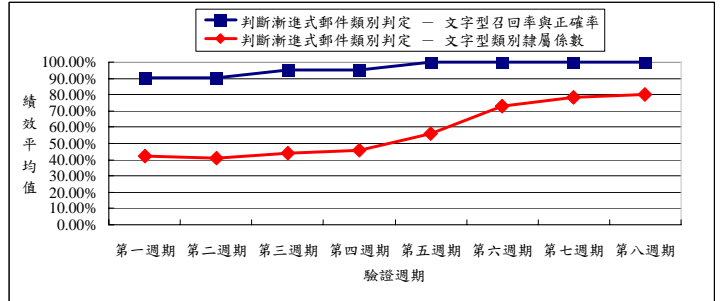


圖 5.3、「文字型資料解析」績效分佈趨勢

(B) 第二階段漸進式郵件類別判定之「聯絡人指標郵件結果」

由表 5.5 及圖 5.4 可知，以每週期增加 70 封訓練用電子郵件為單位，平均每週期聯絡人召回率與準確率之整體平均成長率皆為 0.79%；而聯絡人類別隸屬係數之整體平均成長率為 2.16%，故可知本系統分類能力良好。

表 5.5、聯絡人指標績效彙整

漸進式郵件類別判定—聯絡人型	各週期訓練用漸進式郵件類別判定—聯絡人型郵件數量								平均	
	第一週期	第二週期	第三週期	第四週期	第五週期	第六週期	第七週期	第八週期		
	210 封	280 封	350 封	420 封	490 封	560 封	630 封	700 封		
召回率	平均值	95.24%	95.24%	95.24%	95.24%	100.00%	100.00%	100.00%	100.00%	97.62%
	成長率	-	0.00%	0.00%	0.00%	4.76%	0.00%	0.00%	0.00%	0.68%
	標準差	0.2182	0.2182	0.2182	0.2182	0	0	0	0	0.1091
正確率	平均值	95.24%	95.24%	95.24%	95.24%	100.00%	100.00%	100.00%	100.00%	97.62%
	成長率	-	0.00%	0.00%	0.00%	4.76%	0.00%	0.00%	0.00%	0.68%
	標準差	0.2182	0.2182	0.2182	0.2182	0	0	0	0	0.1091
類別隸屬係數	平均值	82.10%	84.33%	84.71%	86.24%	89.14%	89.43%	94.05%	96.80%	88.30%
	成長率	-	2.23%	0.83%	1.53%	2.90%	0.29%	4.62%	2.81%	2.17%
	標準差	0.1955	0.1962	0.1745	0.1642	0.1308	0.131	0.0482	0.0397	0.1350

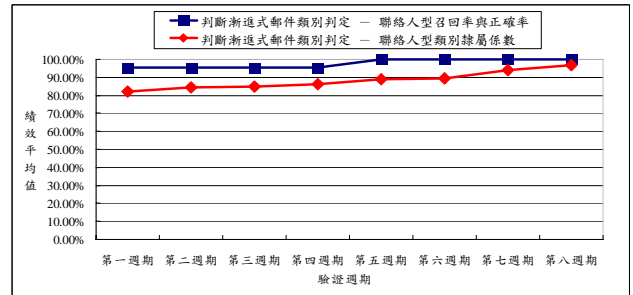


圖 5.4、「聯絡人資料解析」績效分佈趨勢

5.5 驗證結果整體分析

綜合兩階段之驗證成效後，可將各項驗證指標之相關結果整理如表 5.6。由表 5.6 可知，各項驗證指標之收斂前，文字型召回率及正確率初始只有 91%，但至第五週期便以達到 100%，呈現收斂狀態，而文字型之類別隸屬係數雖然初始只有 41%，但至第五週期呈現收斂狀態，且至第八週期仍然有小幅成長並以成長至 80%，以整體平均值而言，文字型召回率與正確率為 97%，而文字型類別隸屬係數雖然只有 60%，但最後達到 80% 以上之水準。

整體而言，雖然文字型判定功能判斷郵件類別正確性已達到 80% 以上，但是本模式又另外建構一個聯絡人判定功能來輔助文字型，其令召回率及準確率初始便提高至 95%（原本文字型為 91%），且第五週期便已呈現收斂狀態，令類別隸屬係數初始便提高至 84%（原本文字型為 41%），並且第五週期便已呈現收斂狀態（原本第六週期才呈現），以整體平均值而言，聯絡人召回率與準確率提高至 98%（原為 97%），而聯絡人類別隸屬係數則提高至 89%（原為 60%），並且最後達到 97% 之水準（原本只有 80%）。雖然單獨使用文字型判定功能便能有 80% 以上的正確性，但如果再用聯絡人判定功能輔助文字型，便能將正確性提高至 96% 以上的高水準，不過本研究提供了可以單獨使用文字型判定功能，或者兩者互相搭配使用之選擇。

各項驗證指標皆於七個週期內呈現收斂狀態，因此，以本研究所選驗證個案（即 Pchome 個人電子報）為例，當本系統使用約 490 封至 560 封訓練用電子郵件時，可讓系統之各項推論績效提升至 80% 以上之水準，故本研究所建置之漸進式類別判定模式從一開始數字/數據型判定功能便可以

有效刪除垃圾信，並將有效信件往下繼續透過文字型判定功能來判斷類別，且如果判斷結果不理想，更可以繼續往下透過聯絡人判定功能來輔助文字型之判定結果，令判斷之正確性更進一步從80%提高至96%以上，因此得知，本系統可有效應用於電子郵件分類判定，並準確地依據判定結果給予使用者分類建議。

表 5.6、各項驗證指標成長率之彙整表

驗證指標	整體平均值	收斂週期	收斂前每週期平均成長率	整體每週期平均成長率
文字型召回率	96.43%	第五週期	1.59%	1.36%
文字型正確率	96.43%	第五週期	1.59%	1.36%
文字型類別隸屬係數	57.47%	第六週期	6.13%	5.40%
聯絡人型召回率	97.62%	第五週期	1.19%	0.68%
聯絡人型正確率	97.62%	第五週期	1.19%	0.68%
聯絡人型類別隸屬係數	88.36%	第五週期	1.87%	2.17%

6. 結論

電子郵件使用者使用電子郵件之目的係為透過電子郵件取得本身所需之資訊，由於電子郵件資訊種類過於繁複，導致部分郵件使用者不易找到其所需之電子郵件；或是電子郵件使用者因郵件收信軟體介面過於複雜、整理分類郵件過程過於冗長，導致電子郵件使用者無動力執行，當然旁人更無法確切得知電子郵件使用者真正需要何種類型之信件類別。因此，本研究透過擷取以及分析電子郵件中所包含之「數字/數據型」、「文字型」及「聯絡人」等資料型別，將原本由電子郵件使用者自行手動整理郵件信箱內之郵件之流程方式，修正為由本系統使用者直接上傳新增欲分類之郵件，以此作為後續本系統判定郵件類別之依據。爾後乃藉由本系統運算判定後，便能得知該郵件是否為垃圾郵件，如結果為否，則可往下繼續進行推論以得知該郵件推論類別為何，並將此結果推薦予使用者，如判定結果未達門檻值，系統亦會推薦使用者可以繼續往下推論，以取得更精確之郵件類別判定結果，如此使用者便能縮短整理作業時間以及節省人力資源。

參考文獻

1. Andreas, N. and Marcin, D., 2006, "Externally growing self-organizing maps and its application to e-mail database visualization and exploration," *Applied Soft Computing*, Vol. 6, No. 4, pp. 357-371.
2. Appavu, S., Rajaram, R., Muthupandian, M., Athiappan, G. and Kashmeera, K. S., 2009, "Data mining based intelligent analysis of threatening e-mail," *Knowledge-Based Systems*, Vol. 22, No. 5, pp. 392-393.
3. Barron, G. and Yechiam, E., 2002, "Private e-mail requests and the diffusion of responsibility," *Computers in Human Behavior*, Vol. 18, pp. 507-520.
4. Baruch, Y., 2005, "Bullying on the net: adverse behavior on e-mail and its impact," *Information & Management*, Vol.42, pp. 361-371.
5. Belkin, N. J. and Croft, W. B., 1992, "Information filtering and information retrieval: two side of the same coin?," *Communication ACM*, Vol. 35, No.12, pp.29-38.
6. Boase, J. and Kobayashi, T., 2008, "Kei-Tying teens: Using mobile phone e-mail to bond, bridge, and break with social ties- A study of Japanese adolescents," *International Journal of Human-Computer Studies*, Vol. 66, pp. 930-943.
7. Burgess, A., Jackson T. and Edwards J., 2004, "Email training significantly reduces email defects," *International Journal of Information Management*, Vol. 25, pp. 71-83.
8. Cavus, N. and Bicen, H., 2009, "A study to find out the preferred free e-mail services used by university students," *Procedia - Social and Behavioral Sciences*, Vol. 1, pp. 419-425.
9. Chang, M. and Poon, C. K., 2009, "Using phrases as features in email classification," *Journal of Systems and Software*, Vol. 82, pp. 1036-1045.
10. Chundi, P., Subramaniam, M. and Vasireddy, D. K., 2009, "An approach for temporal analysis of email data based on segmentation," *Data & Knowledge Engineering*, Vol. 68, No. 11, pp. 1253-1270.
11. Cornwall, A., Moore, S. and Plant, H., 2008, "Embracing technology: Patients', family members'

- and nurse specialists' experience of communicating using e-mail," *European Journal of Oncology Nursing*, Vol. 12, pp. 198-208.
12. Couzenza, T., 2009, "Using electronic mail to motivate students," *Teaching and Learning in Nursing*, Vol. 4, pp. 76-78.
 13. Crawford, E., Koprinska, I. and Webster, P. J., 2004, "Phrases and feature selection in e-mail classification," *Proceedings of the 9th Australasian Document Computing Symposium, Melbourne, Australia*.
 14. Duan, D., Guo, S. Z., Li, Z. B. and Liu, S., 2007, "Community mining based on email classification," *Journal of Computer Applications*, Vol. 27, No. 12, pp. 3039-3041.
 15. Duan, Z., Dong, Y. and Gopalan, K., 2007, "DMTP: Controlling spam through message delivery differentiation," *Computer Networks*, Vol. 51, No. 10, pp. 2616-2630.
 16. Eun, G. P. and Natasha, Z., 2008, "Canadian government agencies develop e-mail management policies," *International Journal of Information Management*, Vol. 28, No. 6, pp. 468-473.
 17. Gimenez, J. C., 2000, "Business e-mail communication: some emerging tendencies in register," *English for Specific Purposes*, Vol. 19, pp. 237-251.
 18. Gonzalez, T. G., 2006, "A simple, configurable SMTP anti-spam filter: Greylists," *Computers & Security*, Vol. 25, No. 3, 2006, pp. 229-236.
 19. Hassini, E., 2006, "Student-instructor communication: The role of email," *Computers & Education*, Vol. 47, pp. 29-40.
 20. Herzberg, A., 2009, "DNS-based email sender authentication mechanisms: A critical review," *Computers & Security*, Vol. 28, No. 8, pp. 731-742.
 21. Hobbs, J., Wald, J., Jagannath, Y. S., Kittler, A. Pizziferri, L. Volk, L. A., Middleton, B. and Bates, D. W., "Opportunities to enhance patient and physician e-mail contact," *International Journal of Medical Informatics*, Vol. 70, pp. 1-9.
 22. Hou, J. L. and Lin, F. H., 2004, "A document and user matching model via document keyword analysis," *Journal of Computer Information Systems*, Vol. 44, No. 4, pp. 1-15.
 23. Hu, C., Wong, A. F. L., Cheah, H. M. and Wong, P., 2009, "Patterns of email use by teachers and implications: A Singapore experience," *Computers & Education*, Vol. 53, pp. 623-631.
 24. Islam, M. R., Zhou, W., Guo, M. and Xiang, Y., 2009, "An innovative analyser for multi-classifier e-mail classification based on grey list analysis," *Journal of Network and Computer Applications*, Vol. 32, No.2, pp. 357-366.
 25. Itakura, H., 2004, "Changing cultural stereotypes through e-mail assisted foreign language learning," *System*, Vol. 32, pp. 37-51.
 26. Joung, Y. J. and Yang, C. J., 2009, "Email licensing," *Journal of Network and Computer Applications*, Vol. 32, No. 3, pp. 538-549.
 27. Kadoya, Y., Fuketa, M., Atlam, E., Morita, K., Kashiji, S. and Aoe, J. I., 2004, "An efficient e-mail filtering using time priority measurement," *Information Sciences*, Vol. 166, No. 1-4, pp. 213-229.
 28. Kosseim, L., Beaugard, S. and Lapalme, G., 2001, "Using information extraction and natural language generation to answer e-mail," *Data & Knowledge Engineering*, Vol. 38, No. 1, pp. 85-100.
 29. Kwon, J. O., Jeong, I. R. and Lee, D. H., 2009, "A forward-secure e-mail protocol without certificated public keys" *Information Sciences*, Vol. 179, pp. 4227-4231.
 30. Laura, B. and Maria, C. C., 2001, "Models of mail server workloads," *Performance Evaluation*, Vol. 46, No. 2-3, 2001, pp. 65-76.
 31. Liaw, M. L. and Johnson, R. J., 2001, "E-mail writing as a cross-cultural learning experience," *System*, Vol. 29, pp. 235-251.
 32. Mano, R. S. and Mesch, G. S., 2010, "E-mail characteristics, work performance and distress," *Computers in Human Behavior*, Vol. 26, pp. 61-69.
 33. Merisavo, M. and Raulas, M., 2004, "The impact of e-mail marketing on brand loyalty," *Emerald Group Publishing Limited*, Vol. 13, pp. 498-505.
 34. Nagabhushan, P., Angadi, S. A. and Anami, B. S., 2009, "A soft computing model for mapping incomplete/approximate postal addresses to mail delivery points," *Applied Soft Computing*, Vol. 9, No. 2, pp. 806-816.
 35. Neese, W. T., Ferrell, L. and Ferrell, O. C., 2005, "An analysis of federal mail and wire fraud cases related to marketing," *Journal of Business Research*, Vol. 58, No. 7, pp. 910-918.

36. O'Kane, P. and Hargie, O., 2007, "Intentional and unintentional consequences of substituting face-to-face interaction with e-mail: An employee-based perspective," *Interacting with Computers*, Vol. 19, pp. 20-31.
37. Phan, R. C.-W., 2008, "Cryptanalysis of e-mail protocols providing perfect forward secrecy," *Computer Standards & Interfaces*, Vol. 30, pp. 101-105.
38. Poon, C. K. and Chang, M., 2003., "An email classifier based on resemblance," *Proceedings of the 14th International Symposium on Methodologies for Intelligent Systems*, Vol. 4, No. 1, pp. 334-338.
39. Sakurai, S. and Suyama, A., 2005, "An e-mail analysis method based on text mining techniques," *Applied Soft Computing*, Vol. 6, No. 1, pp. 62-71.
40. Sallis, P. and Kassabova, D., 2000, "Computer-mediated communication: experiments with e-mail readability," *Information Sciences*, Vol. 123, pp. 43-53.
41. Schuff, D., Turetken, O. and D'Arcy, J., 2006, "A multi-attribute, multi-weight clustering approach to managing 'e-mail overload'," *Decision Support Systems*, Vol. 42, No. 3, pp. 1350-1365.
42. Segal, R. B. and Kephart, J. O., 2000, "Swiftfile: An intelligent assistant for organizing e-mail," *Proceedings of AAAI 2000 Spring Symposium on Adaptive User Interfaces*, pp.107-112.
43. Shih, D. H., Chiang, H. S. and Yen, C. H., 2005, "Classification methods in the detection of new malicious emails," *Information Sciences*, Vol. 172, No. 6, pp. 241-261.
44. Shih, T. H. and Fan, X., 2009, "Comparing response rates in e-mail and paper surveys: A meta-analysis," *Educational Research Review*, Vol. 4, No. 1, pp. 26-40.
45. Sun, M. T. and Hou, J. L., 2003, "The architecture and models for security reasoning in an EDMS," *Journal of the Chinese Society of Industrial Engineers*, Vol. 20, No. 4, pp. 305-316.
46. Sung, S. W. and Chih, K. L., 2004, "Using text classification and multiple concepts to answer e-mails," *Expert Systems with Applications*, Vol.26, No.4, pp. 529-543.
47. White, C. B., Moyer C. A., Stern D. T. and Katz, S. J., 2004, "A content analysis of e-mail communication between patients and their providers: patients get the message," *Journal of the American Medical Informatics Association*, Vol. 11, pp. 260-267.
48. Ying, K. C., Lin, S. W., Lee, Z. J. and Lin, Y. T., 2010, "An ensemble approach applied to classify spam e-mails," *Expert Systems with Applications*, Vol. 37, No. 3, pp. 2197-2201.
49. Yu, B. and Zhu, D. H., 2009, "Combining neural networks and semantic feature space for email classification," *Knowledge-Based Systems*, Vol. 22, No. 5, pp. 376-381.
50. Zhou, G. W., Cheng, J. and Ping, X. J., 2007, "Intelligent Email Classification System based on PIM and Keywords," *Jisuanji Gongcheng / Computer Engineering*, Vol. 33, No. 15, pp. 199-201.

出席國際學術會議心得報告

101 年 07 月 17 日

計畫編號	NSC 100-2221-E-343 -001
計畫名稱	漸進式郵件類別判定模式(I)
出國人員姓名 服務機關及職稱	楊士霆 南華大學資訊管理學系 助理教授
會議時間地點	2012/7/8~2012/7/11, Lithuania (Vilnius)
會議名稱	EURO 2012: 25th European Conference on Operational Research
發表論文題目	A Qualitative Model for Collective Intelligence Analysis

一、參加會議經過與心得

此次 25th European Conference on Operational Research 研討會乃安排於立陶宛(Lithuania)之維爾紐斯(Vilnius)舉辦，配合研討會主辦單位之行程規劃與可行機位安排，個人與國內學者(包含清華大學王小璠教授、清華大學侯建良教授、銘傳大學林進財教授、東華大學陳啟斌教授、義守大學江育民教授、淡江大學阮聘如教授等人)於 7/6 晚上 7 點，即出發前往桃園國際機場，進行登記作業且於晚間 09:05 由桃園國際機場起飛，並於晚間 22:40 抵達香港(HKG)香港國際機場，抵達後於 7/7 上午 00:25 出發轉機續飛往芬蘭(赫爾新基)，並於當地時間 7/7 上午 06:05 抵達該地，台北與芬蘭時差為 5 小時；此程搭機時間總計 13 小時左右，故辦理入境手續並稍做休息後，即先於芬蘭及拉脫維亞逗留一天，瞭解當地風俗民情與生活習慣，即於 7/8 搭乘渡輪至研討會舉辦國立陶宛之首都維爾紐斯。

7/8，由於研討會晚間才得以報到，因此利用本日之上半天至立陶宛之維爾紐斯周遭環境參觀，並參觀教堂、古城，以瞭解立陶宛歷史背景，之後前往當地市場，瞭解當地風俗民情與飲食習慣，最後於晚上五點則前往 Cooperation College，完成報到手續(如圖 1 及圖 2 所示)。

此次研討會舉辦地相當成功與盛大，由於僅需以摘要投稿，總計發表篇數約為 1918 篇左右(來自 28 個國家)，議程數為 524 個，當中來至台灣共計 53 篇論文，而德國論文總數 245 為投稿數最多之國家。大會正式行程日期為 7/8~7/11 三日，正式發表日期乃為 7/9~7/11 三日。個人乃於 7/8 下午五點已至會場，並完成報到手續(如圖 1 及圖 2 所示)。本次研討會內容乃安排與此次會議主題(作業研究)相關之數值分析、線性規劃等專題演講與論文發表，再依不同論文主題每天分 4 個時段，以及 45 個左右平行 Session 進行發表。個人的論文被安排於發表日第二天(7/9)的上午場次(編號 Session MB-40)「MADM Applications VI」發表，於發表後其他學者亦表示對此研究的高度興趣，詢問本研究之網頁論壇智慧推論技術與其他研究之差異，個人並作完整回答，互動甚佳。此外個人亦參加多場與研究興趣較相關之發表場

次，並對於其他學者發表內容提出詢問，對於知識管理、作業研究等課題觸發新的研究靈感（如圖 3 至圖 8 所示）。

除會議發表時間外，在其他交流活動時，個人與國際/國內學者（如清華大學王小璠教授、銘傳大學林進財教授、東華大學陳啟斌教授、淡江大學阮聘如教授等人）亦有良好交流，於此次研討會所認識之多數先進乃屬國內「台灣作業研究學會」及「中華決策科學學會」，因此可瞭解許多國際/國內工業工程、資訊管理學者之研究方向，並規劃未來合作之可能作法，收穫極大，此對於個人學術經歷尚屬資淺而言，乃一大助益。



圖 1、抵達 EURO 2012 會場並註冊(1)



圖 2、抵達 EURO 2012 會場並註冊(2)



圖 3、論文報告與研討(1)



圖 4、論文報告與研討(2)



圖 5、論文報告與研討(3)

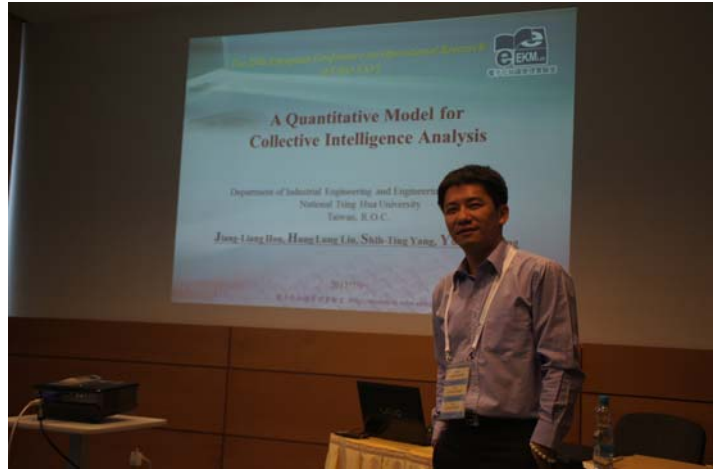


圖 6、論文報告與研討(4)



圖 7、論文報告與研討(5)



圖 8、論文報告與研討(6)

待研討會圓滿結束後（7/11），個人與國內學者一行人則利用些許時間前往包含主辦國在內之波羅的海東岸的愛沙尼亞、拉脫維亞和立陶宛三國參觀，以瞭解波羅的海三小國當地之相似性地理位置和歷史。之後乃搭車、搭船前往機場並搭機飛返回桃園國際機場，結束此次 EURO 2012:學術研討活動。

三、建議

此次會議中的各項活動安排都可發現主辦單位頗為用心，對於遠道造訪之學者給予多項貼心之服務，為國內學校爭取主辦國際型研討會可加以參考之長處。然而，雖然主辦單位之用心可見，但由於此次主辦單位乃將研討會場地劃分為 4 個地點（包含 Radisson Blu Hotel Lietuva、Cooperation College、Best Western Hotel、Holiday Inn Hotel）舉行，雖然各場所皆可以步行到達，但由於研討會期間，維爾紐斯之天氣不甚穩定，時有大雨時而天晴，對各參與

者皆為考驗之一。此外，主辦單位非以該學者所處之議程地點為餐地點，更使多數學者需至另一地點用餐，實為不便。最後，研討會於飯店舉辦（如 Holiday Inn Hotel），因此需受限於飯店之場地限制（如各議程場地較為狹小及休息區之規劃皆不甚完美），此可提供國內學者於辦此類大型學術研討會之借鏡。

整體而言，本次大會舉辦頗為用心，個人於此行收穫豐富，且結識多位國際學者，希望能於未來建立更長遠的交流與合作。

四、攜回資料名稱及內容

1. 研討會論文集：含議程集 1 本及論文摘要集 1 本。
2. 國內外學者學術交流名片。

A qualitative model for collective intelligence analysis

Jiang-Liang Hou, Dept of Ind. Eng. & Eng. Mgt., National Tsing Hua Univ., No. 101, Section 2,
Kuang-Fu Road, 300,

Hsinchu, Taiwan, adamhou02@gmail.com, Hung Lung Lin, Shih-Ting Yang, Yu-Min Chiang

Based on the concept of CI, a three-stage tendency determination model of community opinions is proposed to develop a platform for gathering public opinions, determining tendency of the community opinions and supporting decision makers with decision conclusions. The three stages include "analysis of opinions contents", "representation of formatted opinion contents", and "analysis of opinions tendency and representation of conclusions". The proposed model is to extract text-based opinions of individuals into a generally accepted decision making reference for the decision maker.

Investigation of feature dimension reduction based GLCM/SVM for color filter defect classification

Yu-Min Chiang, Industrial Management, I-Shou University,

No.1, Sec. 1, Syuecheng Rd., Dashu District., 84001, Kaohsiung

City, Taiwan, Taiwan, ymchiang@isu.edu.tw, Yao-Chang

Lin, Jiang-Liang Hou, Shih-Ting Yang

This paper presents an efficient color filter (CF) defect recognition system, based on gray-level co-occurrence matrix (GLCM) and support vector machines (SVM). GLCM is used to extract textural information first. The extracted features are further selected by F-score, information gain and individual feature accuracy selection methods. The idea is to reduce dimensionality of CF defect space. The selected features are used as features for the SVM classification process. Experimental results indicate the proposed method achieves outstanding performance for CF defect classification.



Professor
Marielle Christiansen
Telephone no.: +47 73 59 36 02
E-mail: Marielle.Christiansen@iot.ntnu.no

Dated:
2012-03-14

Dear Yang, S.-T.,

On behalf of the Organizing and Programme committees of the 25th European Conference on Operational Research (EURO XXV), we would like to invite you to participate at the conference that will take place in Vilnius, Lithuania, July 8-11, 2012.

Your abstract entitled "A qualitative model for collective intelligence analysis" has been accepted for presentation at the conference.

We are looking forward to seeing you in Vilnius.

Sincerely yours

Marielle Christiansen
Programme Committee Chair
EURO XXV



Professor

Marielle Christiansen

Telephone no.: +47 73 59 36 02

E-mail: Marielle.Christiansen@iot.ntnu.no

Dated:
2012-03-14

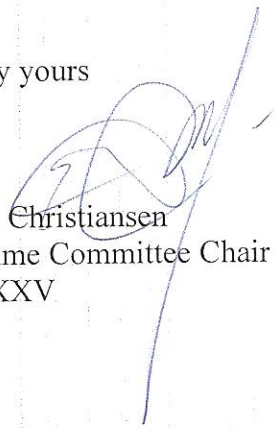
Dear Shih-Ting Yang,

On behalf of the Organizing and Programme committees of the 25th European Conference on Operational Research (EURO XXV), we would like to invite you to participate at the conference that will take place in Vilnius, Lithuania, July 8-11, 2012.

Your abstract entitled "Investigation of feature dimension reduction based GLCM/SVM for color filter defect classification" has been accepted for presentation at the conference.

We are looking forward to seeing you in Vilnius.

Sincerely yours



Marielle Christiansen
Programme Committee Chair
EURO XXV

present a mathematical model for resource optimization which is based on Takt time concept (average time a patient spends in a hospital for treatment) for each department in all the timeslots and considers the dependency constraints. The Linear programming solution provides the resource allocation pattern for each department during every time slot within the resource availability to balance the patient flow and achieve higher patient satisfaction levels.

2 - Optimizing the emergency distribution network in the case of epidemic outbreaks response

Tom Dasaklis, Department of Industrial Management, University of Piraeus, 80 Karaoli & Dimitriou str., 18534, Piraeus, Attica, Greece, dasaklis@unipi.gr, *Nikolaos Rachaniotis*, *Costas Pappis*

Effective control of an epidemic's outbreak calls not only for a rapid response but also for the establishment and management of an emergency supply chain. During the control effort huge amounts of medical supplies should be distributed from central warehouses to local areas. In this paper an epidemic diffusion model based on differential equations is coupled with an emergency logistics distribution model. An integrated solution approach is proposed combining epidemics control rules and logistics network design. A numerical example is finally presented illustrating the methodology proposed.

3 - Modelling ambulance location and deployment in Wales

Leanne Smith, School of Mathematics, Cardiff University, Senghennydd Road, CF24 4AG, Cardiff, United Kingdom, SmithL13@cf.ac.uk, *Paul Harper*, *Vincent Knight*, *Israel Vieira*, *Janet Williams*

Response time targets for the Welsh Ambulance Service Trust (WAST) are not currently being met, particularly for high priority emergencies in rural areas. A discrete event simulation makes use of initial vehicle allocations given by a developed location model, in order to identify the affect demand, number of available vehicles and handover times have on regional response. The aim of these mathematical models is to help WAST make better decisions on locations, capacities and deployments, so that they may provide a more efficient and effective service and achieve the Government set targets.

4 - MetSim: a simulation support tool using meteorological information to improve the planning and management of hospital services

Paul Harper, School of Mathematics, Cardiff University, CF24 4AG, Cardiff, United Kingdom, harper@cardiff.ac.uk, *R John Minty*

Improved short-term predictions of hospital admissions and bed occupancy offer the potential to plan resource needs more accurately and effectively. The MetSim project explores the relationship between weather and health, building novel Bayesian models that are more sensitive to the weather. Short-term forecasts of the numbers of admissions categorised. In turn, MetSim uses hazard ratios embedded within a simulation framework to provide forecasts of short-term bed needs. The web-based tool has been piloted in UK hospitals and is supported by the Met Office.

■ TB-38

Tuesday, 10:30-12h00

HH-Colombus

Hybrid Evolutionary Multiobjective Optimization Algorithms

Stream: Multiobjective Optimization

Invited session

Chair: *Karthik Sindhya*, of Mathematical Information Technology, University of Jyväskylä, Finland, karthik.sindhya@jyu.fi

1 - A Preference-based Evolutionary Algorithm for Multi-objective Optimization: The Weighting Achievement

Scalarizing Function Genetic Algorithm

Ana Belen Ruiz, Applied Economics (Mathematics), University of Malaga, C/ Ejido, 6, 29071, Malaga, Spain, abruiz@uma.es, *Rubén Saborido Infantes*, *Mariano Luque*

We suggest a preference-based EMO algorithm which asks the DM for a reference point (RP) and which is based on the EMO algorithm NSGA-II and on achievement scalarizing functions (ASFs). In practice, an ASF finds the closest efficient solution to the RP, and this solution also depends on a weight vector used in the ASF. The proposed approach modifies the dominance criterion of NSGA-II in order to highlight solutions closer to the RP. The classification of the points into the nondominated fronts is done according to the values that each solution takes on an ASF, using different weight vectors.

2 - Preference-based Evolutionary Algorithm for Multi-objective Bilevel Optimization

Pekka Malo, Information and Service Economy, Aalto University School of Economics, Runeberginkatu 22-24, 00100, Helsinki, Finland, pekka.malo@aalto.fi

We present an evolutionary framework for solving multiobjective bilevel problems, where the upper-level has multiple objectives and lower-level has a single objective. The implementation of the algorithm is based on a new technique for approximating the optimal lower-level solution mapping by a class of smooth and continuous single-valued functions. The algorithm is well-motivated by the recent results obtained for general parametric optimization problems and set-valued analysis. The performance of the algorithm is demonstrated with respect to a collection of scalable test-problems.

3 - A Hybrid Evolutionary Multi-objective Optimization Algorithm for Enhanced Convergence and Diversity

Karthik Sindhya, of Mathematical Information Technology, University of Jyväskylä, Finland, karthik.sindhya@jyu.fi, *Kaisa Miettinen*, *Kalyanmoy Deb*

Evolutionary multi-objective optimization (EMO) algorithms are used to find approximate Pareto optimal solutions, though they have slow convergence speed. We present a modular hybrid EMO algorithm including a local search module (involving an achievement scalarizing function) to increase the convergence speed, a diversity enhancement module, etc. It can be connected with various EMO methods like MOEA/D. Numerical tests on test problems indicate the efficacy of our hybrid EMO algorithm, which can also be used to handle dynamic process simulation based multi-objective optimization problems.

■ TB-39

Tuesday, 10:30-12h00

HH-Cousteau

MADM Applications VI

Stream: MCDA: New Approaches and Applications

Invited session

Chair: *Chin-Tsai Lin*, Department of Business Administration, Ming Chuan University, 5F, No.130, Jihe Rd., Shihlin District,, 11111, Taipei, Taiwan, ctlin@mail.mcu.edu.tw

Chair: *Chie-bein Chen*, International Business, National Dong Hwa University, 1, Sec. 2, Da-hsueh Rd. Shou-feng, 974, Hualien, Taiwan, cbchen@mail.ndhu.edu.tw

1 - A qualitative model for collective intelligence analysis

Jiang-Liang Hou, Dept of Ind. Eng. & Eng. Mgt., National Tsing Hua Univ., No. 101, Section 2, Kuang-Fu Road, 300, Hsinchu, Taiwan, adamhou02@gmail.com, *Hung Lung Lin*, *Shih-Ting Yang*, *Yu-Min Chiang*

Based on the concept of CI, a three-stage tendency determination model of community opinions is proposed to develop a platform for gathering public opinions, determining tendency of the community opinions and supporting decision makers with decision conclusions. The three stages include "analysis of opinions contents", "representation of formatted opinion contents", and "analysis of opinions tendency

and representation of conclusions'. The proposed model is to extract text-based opinions of individuals into a generally accepted decision-making reference for the decision maker.

2 - Taiwan Health Communication Research: Application and Impact of New Media

Rose Hung, Department of Communications Management, Shih Hsin University, Taiwan, rose@jetgo.com.tw, *Pi-Fang Hsu*

With the rise of social networks and mobile media, health communication models also come with each passing day, compared to traditional media, new media more effectively to achieve the dissemination of health knowledge and help people or patients to implement lifestyle adjustment? This study analysis of the Taiwan government and private health communication activities in recent years, through the combination of interviews and questionnaires, the results are analyzed to find out the health communication effect and construct a health dissemination pattern.

3 - Hybrid MCDM Approach for Company Growth Strategy Selection in Marketing

Tuncay Gürbüz, Industrial Engineering, Galatasaray University, Ciragan cad. No.36, Ortakoy, 34357, Istanbul, Turkey, tuncaygurbuz09@gmail.com, *Y. Esra Albayrak*

In today's competitive markets, one can easily see that the importance is not producing goods and services but being able to sell what one produces. Nowadays, the concept of marketing is customer-oriented and aims to meet the needs and expectations of customers to increase their satisfaction. While creating a marketing strategy, an enterprise must consider many factors and decide to perform either target determination or company growth strategies. In this study, company growth strategies are analyzed with a hybrid MCDM method consisting of Analytic Network Process and Choquet Integral.

■ TB-40

Tuesday, 10:30-12h00

HH-Livingstone

Preference Learning 1

Stream: Preference Learning

Invited session

Chair: *Willem Waegeman*, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure links 653, 9000, Ghent, Belgium, willem.waegeman@ugent.be

Chair: *Krzysztof Dembczynski*, Institute of Computing Science, Poznan University of Technology, Piotrowo 2, 60-965, Poznan, Poland, kdembczynski@cs.put.poznan.pl

1 - Active learning for preference learning

Nir Ailon, Computer Science, Technion Israel Institute of Technology, Taub Building, Technion City, 32000, Haifa, Israel, nailon@cs.technion.ac.il

Analyzing preferences lies in the intersection of many fields, most notably econometrics (discrete choice theory), combinatorial optimization and machine learning theory and practice. In a typical setting, the ultimate goal is to be able to predict the order of a given set of alternatives given a set of observed, noisy pairwise comparisons. In this work I will discuss the adversarial noise case, and present a new active learning ERM (Empirical Risk Minimization) algorithm with an almost optimal preference query complexity and provable guarantees.

2 - Preference elicitation for interactive learning of Optimization Modulo Theory problems

Andrea Passerini, disi, University of Trento, via Sommarive 5, 38100, Povo di Trento, Italy, passerini@disi.unitn.it, *Paolo Campigotto*, *Roberto Battiti*

We consider unknown decision maker preferences as weighted combinations of predicates in a certain theory of interest. We developed an iterative approach alternating a search and a learning stage: (i) the

current approximation of the preference function is optimized (Optimization Modulo Theory), providing a set of candidate solutions; (ii) the preference function is refined according to the DM feedback on these candidates. The learning stage relies on a sparse ranking algorithm aimed at automatically selecting the relevant predicates and their combination, among a catalogue of candidate ones.

3 - Optimal recommendation sets

Paolo Viappiani, Computer Science Dept, Aalborg University, selma lagerfos vej 300, 9000, Aalborg, Denmark, paolo.viappiani@gmail.com

Utility elicitation is an important component of many applications, such as decision support systems. Such systems query users about their preferences and offer recommendations based on the system's belief about the user's utility function. We analyze the connection between the problem of generating optimal recommendation sets and the problem of generating optimal choice queries, considering both Bayesian and regret-based elicitation. Our results show that, somewhat surprisingly, under very general circumstances, the optimal recommendation set coincides with the optimal query.

■ TB-41

Tuesday, 10:30-12h00

HH-Heyerdahl

AHP/ANP 6

Stream: Analytic Hierarchy Process, Analytic Network Process

Invited session

Chair: *Pekka Korhonen*, Economics and Management Science, Helsinki School of Economics, Runeberginkatu 22-24, 100, Helsinki, Finland, korhonen@hkkk.fi

1 - R&D Personnel Selection with Fuzzy Analytic Hierarchy

Erhan Berk, Defence Science, Hava Duragi Lojmanlari 4.Blok Daire:1, 06790, Ankara, Turkey, erhanberk@hotmail.com

In the 21st century where the competitive environment becomes very tough for companies in parallel with the drastic development in technology, only innovative companies became prominent. The R&D personnel in particular have played the major role in these emerging innovations. In this study a Fuzzy Analytical Hierarchy Process (AHP) method for the selection of R&D personnel has been proposed. A case study has been conducted by using the weights obtained from the Fuzzy AHP method and the results have been compared.

2 - Determining effective criteria of Customers' satisfaction in kitchen worktops by using AHP

Majid Azizi, Department of Wood and Paper, Faculty of Natural Resources, University of Tehran, Daneshkadeh St., 31585-4314, Karaj, Tehran, Iran, Islamic Republic Of, mazizi@ut.ac.ir, *Vahid Nabavi*, *Mehdi Faezipour*

Determination of effective criteria in customer satisfaction of worktops is very important for any company that produces this product. To identify effective criteria for customers' satisfaction five major groups which are selling condition and price, delivery condition, guaranty, raw material quality and technological specification were selected. These major groups have 19 sub-criteria. The priority rate of these criteria and sub-criteria was obtained by AHP. Sub-criteria of price, payment condition, top veneer quality, MDF quality and fast delivery had the highest priorities, respectively.

3 - Analysing numerically three different methods to estimate a priority vector from inconsistent pairwise comparison matrices in AHP

Attila Poesz, Corvinus University of Budapest, Fovam ter 8, 1093, Budapest, Hungary, attila.poesz@uni-corvinus.hu, *Pekka Korhonen*

In our paper we present a simulation framework in which the performance of the estimating methods used to determine priority vector

crashes in railway transport is presented. Immune algorithm is proposed for optimization problem solution. Algorithm is tested using computer model of the transport system.

■ MD-39

Monday, 14:30-16:00

HH-Cousteau

MADM Applications IV

Stream: MCDA: New Approaches and Applications

Invited session

Chair: *Chin-Tsai Lin*, Department of Business Administration, Ming Chuan University, 5F, No.130, Jihe Rd., Shihlin District,, 11111, Taipei, Taiwan, cotlin@mail.mcu.edu.tw

Chair: *Chie-bein Chen*, International Business, National Dong Hwa University, 1, Sec. 2, Da-hsueh Rd. Shou-feng, 974, Hualien, Taiwan, cbchen@mail.ndhu.edu.tw

1 - The study of the university students consumer behavior tea-shops as example

Yuan-Du Hsiao, Department of Business Administration, Chungyu Institute of Technology, No.40, Yi7th Rd., Keelung, 20103 Taiwan, R.O.C., 20103, Keelung, Taiwan, R.O.C., Taiwan, yuduhsia@cit.edu.tw, *Jen-Chia Chang*, *Mu-Hui Lai*, *Liang-Yuan Hsiung*

This study main purpose are 1. To discuss the subjects who get drinks from consumer behavior. 2. Testing the difference from the consumer of tea-shops. The study finding 1. More than half consumer is choosing the shop criterion "important" of the project. 2. Consumer general doesn't costs at fixed shop. 3. The ones often go shopping at the noon. 4. Almost ones always go to "Dou Di Tea-shops". From findings, the suggestions for the tea-shop boss or studies experts reference related.

2 - The Study on Teachers' Willingness to Adopt Computer-Aided Instruction in Public Vocational High Schools

Jen-Chia Chang, Institute of Technology & Vocational Education, National Taipei University of Technology, 10608, Taipei, Taiwan, jc5839@ntut.edu.tw, *Yuan-Du Hsiao*, *Mu-Hui Lai*, *Liang-Yuan Hsiung*

The study with 126 sampled subject in mid-Taiwan. The reliability and validity of the questionnaire have been rendered positive, upon verification. The mean value of teachers' willingness to adopt CAI is 3.97, ranged as moderate. Teachers are generally dissatisfied with the CAI software, hardware facilities support and promote CAI in school; however, these three factors have no significant connection with teachers' willingness to apply CAI. The researcher is to offer various suggestions, based on this study, to schools and the education administrative departments.

3 - A study on relationship model among travel motivation, travel type, travel involvement, and travel benefit of Facebook users: A case of Taiwan domestic tourism

Pei-Ting Chen, Travel Management, National Kaohsiung University of Hospitality and Tourism, 81267, Kaohsiung, Taiwan, tober770831@hotmail.com, *Tang-Chung Kan*, *Kuo Tzu-hsuan*

This study aims to discuss the relation among travel motivation, travel involvement, and tourism benefit of Facebook users. It also uses travel type as the intervening variable to further explore its influence in relation to travel pattern, travel involvement and tourism benefit. We expect this research result will help understand Facebook users to understand characteristics of Taiwan domestic tourism. This study is expected to build an appropriate social networking users' travel behavior model to develop distinctive and customized community tourism products.

4 - Multi-Criteria Comparison of Catering Service Companies Using Grey Relational Analysis: The Case of Turkey

Alper Hamzadayi, Dokuz Eylul University, Afghanistan, alper.hamzadayi@deu.edu.tr, *Simge Yelkenci Kose*, *Sener Akpinar*, *Atabak Elmi*, *Hanifi Okan Isguder*

Catering is the business of providing food and beverage at a remote site or a site like business meetings and other social occasions. Selecting catering is an important multi-criteria decision making problem. In this study grey relational analysis (GRA) is used for the aim of selecting the best catering firm providing the most firm satisfaction. Due to the importance of criteria weights in decision making, Analytic Hierarchy Process is used to find best values of criteria weights in GRA process. A case study is demonstrated and evaluated using the developed multi-criteria methodology.

■ MD-40

Monday, 14:30-16:00

HH-Livingstone

MADM Applications VII

Stream: MCDA: New Approaches and Applications

Invited session

Chair: *Chin-Tsai Lin*, Department of Business Administration, Ming Chuan University, 5F, No.130, Jihe Rd., Shihlin District,, 11111, Taipei, Taiwan, cotlin@mail.mcu.edu.tw

Chair: *Chie-bein Chen*, International Business, National Dong Hwa University, 1, Sec. 2, Da-hsueh Rd. Shou-feng, 974, Hualien, Taiwan, cbchen@mail.ndhu.edu.tw

1 - Efficiency and Productivity of Human Resources in Science and Technology in Innovation and Efficiency Driven Nations

Ying-Chyi Chou, Department of Business Administration, Tunghai University, 181 Taichung-kang Rd., Sec. 3, Taichung, Taiwan, R.O.C., 407, Taichung,, Taiwan, rosechyi@yahoo.com.tw, *Hsin-Yi Yen*

Data Envelopment Analysis (DEA) was applied to assess the relative efficiency of HRST and determine productivity variation of each nation using the Malmquist model. The empirical results demonstrate that the efficiency of HRST differs among nations, depending on the level of economic development. From Malmquist model estimates, productivity indices were calculated and decomposed into two categories: technological change and efficiency change of productivity variation. A benchmark for nations aiming to improve their HRST efficiency and productivity has been established.

2 - An Assessment of Taiwan's Solar Water Heater Subsidy Policy Using Logistic Diffusion Curves

Pao-Long Chang, Department of Business Administration, Feng Chia University, 100, Wenhwa Road, Seatwen, Taichung, Taiwan, R.O.C, 407, Taichung, Taiwan, paolong_chang@yahoo.com.tw, *Chiung-Wen Hsu*

As many countries are implementing policies to achieve carbon reduction, assessing the effectiveness of measures is important. Using installation area data from the non-subsidized period (1992—1999) and subsidized period (2000—2010), logistic curves are used to model the adoption of solar water heating and assess the effectiveness of Taiwan's solar water heater subsidy policy. The result showed that the government's average annual investment of NT \$1,052 for solar water heaters led to a 1-ton carbon emission reduction in 10 years. Thus, only 57% of the goal by 2020 can be achieved.

3 - Investigation of feature dimension reduction based GLCM/SVM for color filter defect classification

Yu-Min Chiang, Industrial Management, I-Shou University, No.1, Sec. 1, Syuecheng Rd., Dashu District,, 84001, Kaohsiung City, Taiwan, Taiwan, ymchiang@isu.edu.tw, *Yao-Chang Lin*, *Jiang-Liang Hou*, *Shih-Ting Yang*

This paper presents an efficient color filter (CF) defect recognition system, based on gray-level co-occurrence matrix (GLCM) and support vector machines (SVM). GLCM is used to extract textural information first. The extracted features are further selected by F-score, information gain and individual feature accuracy selection methods. The idea is to reduce dimensionality of CF defect space. The selected features are used as features for the SVM classification process. Experimental results indicate the proposed method achieves outstanding performance for CF defect classification.

■ MD-41

Monday, 14:30-16:00

HH-Heyerdahl

AHP/ANP 4

Stream: Analytic Hierarchy Process, Analytic Network Process

Invited session

Chair: *Y. Ilker Topcu*, Industrial Engineering, Istanbul Technical University, Istanbul Teknik Üniversitesi, İslatme Fakültesi, Macka, 34357, Istanbul, Turkey, ilker.topcu@itu.edu.tr

1 - Applying ANP to Select Strategic Alliances for International Tourist Hotels in Taiwan

Su-Chuan Shih, Business Administration, Providence University, 200 Chung Chi Road, Business Administration, 43301, Taichung City, Taiwan, Taiwan, scshih@pu.edu.tw

International tourist hotels often adopt strategic alliance to increase their competition advantage. This study employs Analytic Network Process to select strategic alliance types. The result shows the most important key criterion is the performance goal of the alliance. Under the key criterion, the index of profit rate is weighted more than the others. The preference alliance types are ordered as: the complex activities, the electronic channels, and the physical shops.

2 - An Integrated Multi-Criteria Approach for the Evaluation and Control of Strategic Options

Ralf Kaspar, Chair of Managerial Accounting, University of Osnabrueck, 49069, Osnabrueck, Germany, ralf.kaspar@uni-osnabrueck.de

The evaluation and control of strategic options within the strategic management process is fundamentally complicated by multiple and mutually influencing objectives, internal and external factors, lack of transparency, complexity and conflicts of interest. To cope with these difficulties a suitable multi-criteria decision support approach is required. Research has led to the development of an ANP based framework. Supported by a case study it is a contribution to an improved evaluation of strategic options incorporating advanced SWOT-Analysis, BOCR modeling and group decision making.

3 - A Multi-Criteria Based Evaluation of 'Innovation Strategy Selection'

Ozge Surer, Industrial Engineering, Istanbul Technical University, ITU İslatme Fakültesi, Macka, 34357, Istanbul, Turkey, surer@itu.edu.tr, *Sezi Cevik Onar*, *Y. Ilker Topcu*

Innovation is a crucial tool for firms to compete, thus it is necessary to define the factors which affect this process. Containing interrelations among the related factors, a multi criteria approach utilizing network models should be used. In accordance with the iterative steps of Analytic Network Process, innovation strategy types and related factors are identified by opinions of experts and the literature survey. Relations among these factors are revealed and their relative importance and the priorities of the strategies are assessed, based on the pairwise comparison answers of the experts.

4 - On-line modified AHP algorithm for the graduation papers group estimating.

Vsevolod Kuzmitskiy, Far Eastern State Transport University, 47, Seryshev St., 680021, Khabarovsk, Russian Federation, sevan_sevan@bk.ru, *Boris Davydov*

The expert group estimating of the graduation papers is the multifactor problem with the previously unknown scale. We use the modified AHP-method for search solution. The adaptive property of the algorithm is used by the on-line estimating of the middle grade of the expert's judgments. Moreover, the complex pairwise comparison process is simplified by the replacement of ordinal scale comparisons. The objectivity of the final grade is increased due to the liquidating conflict interest's graduation committee members.

■ MD-42

Monday, 14:30-16:00

BW-Amber

EEPA 2012 - 2

Stream: EURO Excellence in Practice Award 2012

Invited session

Chair: *Michel Bierlaire*, ENAC INTER TRANSP-OR, École Polytechnique Fédérale de Lausanne (EPFL), GC B3 454, Station 18, CH-1015, Lausanne, Switzerland, michel.bierlaire@epfl.ch

1 - SPRINT: Optimization of Staff Management for Desk Customer Relations Services

Daniele Vigo, DEIS, University of Bologna, Via Venezia 52, 47023, Cesena, Italy, daniele.vigo@unibo.it, *Claudio Caremi*, *Angelo Gordini*, *Sandro Bosso*, *Giuseppe D'Aleo*, *Beatrice Beleggia*

Staff Management (SM) is a critical issue in the design and management of modern Desk Customer Relation Services (DCRSs) due to the large amount of resources involved and the increasing pressure towards maintaining a very high level of service. We illustrate a Decision Support System for SM, called SPRINT, developed for a large multi-utility company in northern Italy. After one year of service SPRINT has introduced a considerable improvement in the process and achieved relevant savings in terms of resource consumption and very high level of service.

2 - Logistic planning using DSS FlowOpt

Mikael Rönnqvist, Département de génie mécanique, Université Laval, G1V 0A6, Québec, Canada, mikael.ronnqvist@nhh.no, *Patrik Flisberg*, *Mikael Frisk*

Skogforsk is the central research body for the Swedish forestry sector. Skogforsk has developed the decision support system FlowOpt. A central part is a powerful optimization engine which has been a key contribution to the success of the system. It makes use of a flexible network model with extensions including backhauling and cost sharing possibilities. The system has been used in many case studies and implementations during the last 10 years. It has been used to evaluate the performance and suggest improvements at single companies as well as several companies making collaborative schemes.

3 - Flood Prevention by Optimal Dike Heightening

Kees Roos, EWI, TU Delft, Mekelweg 4, 2628CD, Delft, ZH, Netherlands, c.roos@tudelft.nl, *Dick den Hertog*, *Ruud Brekelmans*, *Carel Eijgenraam*

The Dutch government is very much aware of the flooding risks that The Netherlands is exposed to. This research tries to answer the fundamental questions of when and how much to invest in which parts of the constituent segments of a dike ring protecting a certain area of land. This problem is solved by using Mixed-Integer Nonlinear Programming techniques. For homogeneous dike-rings explicit formulas for periodic solutions have been obtained. Dutch government agencies use the model to analyze the main dike rings and to propose new safety standards to be incorporated in the Dutch Water Act.

國科會補助計畫衍生研發成果推廣資料表

日期:2012/10/24

國科會補助計畫	計畫名稱: 漸進式郵件類別判定模式(I)
	計畫主持人: 楊士霆
	計畫編號: 100-2221-E-343-001- 學門領域: 資訊系統
無研發成果推廣資料	

100 年度專題研究計畫研究成果彙整表

計畫主持人：楊士霆		計畫編號：100-2221-E-343-001-					
計畫名稱：漸進式郵件類別判定模式(I)							
成果項目		量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）	
		實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比			
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	1	1	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力 (本國籍)	碩士生	3	3	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	1	2	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	1	1	100%		
		專書	0	0	100%		章/本
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力 (外國籍)	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

<p>其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)</p>	<p>無</p>
--	----------

	成果項目	量化	名稱或內容性質簡述
科 教 處 計 畫 加 填 項 目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與(閱聽)人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文： 已發表 未發表之文稿 撰寫中 無

專利： 已獲得 申請中 無

技轉： 已技轉 洽談中 無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

電子郵件使用者使用電子郵件之目的係為透過電子郵件取得本身所需之資訊，由於電子郵件資訊種類過於繁複，導致部分郵件使用者不易找到其所需之電子郵件；或是電子郵件使用者因郵件收信軟體介面過於複雜、整理分類郵件過程過於冗長，導致電子郵件使用者無動力執行，當然旁人更無法確切得知電子郵件使用者真正需要何種類型之信件類別。因此，本研究透過擷取以及分析電子郵件中所包含之「數字/數據型」、「文字型」及「聯絡人」等資料型別，將原本由電子郵件使用者自行手動整理郵件信箱內之郵件之流程方式，修正為由本系統使用者直接上傳新增欲分類之郵件，以此作為後續本系統判定郵件類別之依據。爾後乃藉由本系統運算判定後，便能得知該郵件是否為垃圾郵件，如結果為否，則可往下繼續進行推論以得知該郵件推論類別為何，並將此結果推薦予使用者，如判定結果未達門檻值，系統亦會推薦使用者可以繼續往下推論，以取得更精確之郵件類別判定結果，如此使用者便能縮短整理作業時間以及節省人力資源。

根據上述之研究課題，本計畫已完成之工作項目分述如下：

1. 建立「漸進式郵件類別判定」模式

分析電子郵件使用者所上傳之電子郵件，並建構「數字/數據型」、「文字型」與「聯絡人」等資料型別之方法論，用以建構漸進式郵件類別判定系統，進而提供系統使用者郵件類別判定之推論結果建議。

2. 漸進式郵件類別判定系統之建置

開發「郵件資料維護模組」、「漸進式郵件類別判定模組」、「關鍵字/非關鍵字維護模組」、

「郵件類別維護模組」、「訓練郵件維護模組」及「系統參數設定模組」等六大功能模組，並整合此六大功能模組於一網際網路環境下執行之漸進式郵件類別判定系統。之後，電子郵件使用者即可透過漸進式郵件類別判定系統，獲得各類電子郵件之正確類別建議，以作為電子郵件使用者整理電子郵件時之參考。

3. 案例驗證與成果分析

根據前述之推論模式，發展漸進式郵件類別判定系統，並以 Pchome 電子報為來源所收集之郵件樣本進行驗證，以確認本研究提模式與技術之正確性、合理性。待模式驗證完成後，最後檢討實際成效與預期成果間之符合程度，並由分析評估瞭解本論文之未來發展與應用方向。

整體而言，本研究期望整合此些模式，解決電子郵件使用者使用電子郵件時之既有問題(即作業流程、整理時間冗長，以及閱讀電子郵件時未能真正符合所有電子郵件使用者所需)，並透過本方法論之發展，提供此類業者更為客觀、系統化、有效率之整理電子郵件建議，進而更有效地服務更多電子郵件使用者。

綜合而言，本計畫於理論方法、技術開發與實務應用等層面皆具相當之成果與貢獻，以下即分別說明本計畫於此三層面之具體成效。

A. 理論方法層面：本計畫提出一套漸進式郵件類別判定模式，包含「數字/數據型」、「文字型」與「聯絡人」等資料型別之方法論，以提供系統使用者判定電子郵件類別之建議。

B. 技術開發層面：本計畫完成漸進式郵件類別判定系統平台之技術開發，整合郵件資料維護模組及漸進式郵件類別判定模組等模組於一系統中，透過此些模組以提高郵件分類之正確率，並提供推論之郵件類別結果建議，以減少作業時間及節省人力資源，進而提升電子郵件整理時之效率與取得正確資訊的正確率。

C. 實務應用層面：透過線上網站 Pchome 電子報所取得之郵件樣本為來源，經過實際實驗測試以確認系統技術之可行性。