

運用關聯分析探勘民眾關注議題與發展方向:以環保議題為例

王界人 Chieh-Jen Wang
工業技術研究院巨資中心
Computational Intelligence Technology Center
Industrial Technology Research Institute
chiehjen@itri.org.tw

沈民新 Min-Hsin Shen
工業技術研究院巨資中心
Computational Intelligence Technology Center
Industrial Technology Research Institute
mshen@itri.org.tw

摘要

關聯分析近年來被應用於許多不同研究領域，在巨量資料中，探勘資料間之相互關係與規則。本研究目標在於運用關聯分析，從巨量非結構化資料中，探勘民眾關注之議題，並分析未來可能發展方向。我們以環保領域資料為研究樣本，建構環保議題偵測模型，探勘民眾目前關注之議題與未來可能討論風向。分析結果顯示，探勘 PTT 電子佈告欄之文章，可有效瞭解民眾關注議題，以及準確預測未來議題發展方向。透過此環保議題偵測系統，不但可以讓環保機關精準掌握輿情焦點，提高施政品質，也可強化政策宣導內容，大幅涵蓋與解答民眾所關注之議題。此分析系統也可以應用在不同領域，例如：產品市場調查與口碑分析。

Abstract

Association analysis has attracted considerable attention recently in many research fields, mining data relations and rules from huge volume of data especially. This study aims at mining issues of public concern and analyzing its relations from massive of unstructured data. The main resource of this study is environmental related documents from PTT bulletin board system. A model is constructed via the collected environmental documents for predictions of issues of public concerns and possible future directions. The experimental results show that mining information from documents of PTT bulletin board system can effectively understand the public concerns and predict possible future directions. The reports from the prediction system may be used as a reference for environmental authorities. The prediction model we propose not only precisely masters of opinions from public to improve the administrative quality of environmental authorities, but also strengthens the content of press release to cover and answer the significant important issues of public concerns. The prediction system can be also applied to different applications, such as market investigation and opinion analysis.

關鍵詞：關聯分析，意見探勘，議題偵測

Keywords: Association Analysis, Opinion Mining, Topic Detection

一、緒論

近年來社群網路的使用者來越多，根據 eMarketer 調查指出，2015 年全球使用社群網站的人數預估將突破 21 億 8000 萬人¹，不同種類的討論話題都可能會出現在各大社群網站中。網路將人與人之間的距離拉近，不同來源的資訊也隨著網路的便利性以及社群網站的發達，快速地將資訊傳播開來。而網路上的資訊來源不只來自新聞媒體，民眾個人經驗、小道消息更是有別於新聞媒體儘量保持客觀的態度，主觀的陳述與透過社群網路的公開討論，使得訊息的面向更加豐富多元，因此，社群網路使用者討論資料已經成為文字探勘與議題分析之重要素材來源。

批踢踢實業坊(PTT)是一個電子佈告欄系統(Bulletin Board System, BBS)，於 1995 年創立，目前在批踢踢實業坊與其分站批踢踢免註冊總人數約 125 萬人，兩站尖峰時段超過 15 萬名使用者同時上線，擁有超過 2 萬個不同主題的看板，每日約 4 萬篇新文章被發表，是非常好的文字探勘與議題偵測素材來源。

文字探勘是近年來隨著人工智慧和自然語言處理技術發展的一門新興技術。主要從大量文字資料中自動化辨識與挖掘有用的資訊，萃取出隱含的或過去不為人知，但可信與有效的訊息。並且依據使用者文字表達特徵，在一群未經處理的資料中找到使用者可能感興趣的資訊。其中關聯法則分析就是文字探勘中一種重要的分析模型，利用關聯分析探勘出有用的資訊，可做為政府決策的參考依據[1] [2]。

政府單位發布政策資訊給民眾，常使用新聞稿的形式向民眾宣達，因此，如何加強政策溝通對政府機關十分重要。一般而言，公部門擬稿人是憑藉個人經驗與蒐集過去歷史資訊來撰寫文稿。然而，擬稿人可能會受限於個人迷思或因特定領域知識不足，對於議題焦點之掌握程度，有參差不齊的現象。現今社群網路的出現，使得任何人都可以透過社群網路取得資訊，並且透過巨量網民的討論資訊，發掘目前大眾所關注的議題與輿情焦點。透過對巨量社群網路資料進行文字探勘即可以達到上述的目標。建立議題預測模型，協助公部門擬稿人撰寫新聞稿之方向；運用巨量資料，分析民眾關注之議題，作為政策研擬與新聞發布後之輿情蒐集，可以即時回應或加強政策溝通，應是可行之有效方式。

本研究架構共分為五章，包含「緒論」、「文獻探討」、「實驗資料」、「預測模型」與「結論」，內容分別說明如下：

第一章「緒論」說明研究背景與動機以及論文架構。第二章「文獻探討」則回顧整理過去文字探勘方法與應用、關聯分析演算法理論及我國政府對於文字探勘與巨量資料分析之策略。第三章「實驗資料」說明資料取得來源與統計資訊。第四章「預測模型」介紹本研究預測模型設計與參數設定，以及系統分析流程。第五章「實驗結果與分析」探討與分析實驗結果。第六章「結論」總結所有分析資訊、研究貢獻以及未來可能研究方向。

二、文獻探討

文字探勘(Text Mining)主要是針對半結構化(Semi-structured)或非結構化(Unstructured)儲存格式的文件資料進行探勘，這些非結構化資料隱藏著許多重要的資訊，是近年來重要的研究領域之一[3]。透過分析文本中的文字特徵，從中萃取出隱含性資訊，轉換成

¹ <http://www.emarketer.com/Article/Social-Networking-Reaches-Nearly-One-Four-Around-World/1009976>

為有價值的訊息。近幾年文字探勘技術越臻成熟，除了可以將文本資料有效擷取，且可以利用語意分析（Semantic Analysis）技術偵測資料內容的屬性。目前已經廣泛應在不同研究主題，包含文件分類（Document Classification）[4]、文件分群（Document Clustering）[5]和網頁探勘（Web Mining）[6]。

隨著行動通訊及網路的發展，社群網路平台如 Facebook、Plurk 與 PTT 等相繼興起，加上智慧手機的普及，使用者可以隨時隨地透過網路發表自己的意見，網路使用者主動將各種不同的意見與評論資訊上傳網路，如果能利用自動化文字分析技術，就可以把這些巨量的社群網路資料，轉換成有用的資訊，協助政府單位與各集團公司之決策參考[7][8]。近年來我國政府也感受到社群資料的重要性，著手研究利用「開放資料」、「巨量資料」與「群眾外包」，協助政府運用網路媒體與科技技術創造「有感施政」²。

巨量資料(Big Data)的議題在我國從 2012 年開始持續發燒，許多應用也隨之而生，其中巨量資料是指對海量資料進行分析與探勘，而獲取深入、有用且有價值的訊息。巨量資料的特性包括：資料多、速度快、變化多，以及真實性，所以巨量資料的分析演算法是決定最終產出資訊是否具有價值的重要因素³。關聯分析（Association Analysis）是巨量資料中重要的分析演算法之一，特徵為可以產生法則，用來描述資料間的關聯性。這種方法具有簡潔、易懂的分析結果，常常使用在購物籃分析(Market Basket Analysis)、交叉比對、預測、分群、分類等研究領域[9]。最早的關聯分析是 1994 年 IBM Almaden Research Center 的學者 Agrawal[10][11]所提出 Apriori 演算法，主要針對市場購物籃問題加以探討，其演算法透過反覆產生候選項目集合(Candidate Item Set)，以找出所有高頻項目集合，並藉由最小支持度與最小信心度之篩選後，推導出所有的關聯法則。但 Apriori 演算法有一個先天上的問題，就是需要產生大量候選項集和需要重複地檢視資料。Han 等學者提出的 FP-growth 演算法[12]，有效地克服了這方面的問題。FP-growth 將大量的資料壓縮成一種緊密的樹狀結構 FP-tree，這種做法可以大量減少候選項目集合的個數，並且只需檢視兩次資料庫，可以改善 Apriori 需要大量時間計算的問題，顯著提升執行效能。

關聯分析已經被廣泛應用在不同領域，在一般商業應用上，藉過去客戶購買行為之關聯，分析客戶的消費習性，進而變更規劃產品銷售、擺設、推出有競爭力的商品促銷方案與評估搭配銷售模式[13]。在信用卡市場中，可有效且準確地分析持卡人的信用狀況，預防呆帳與惡意倒帳的行為，減少發卡銀行損失[14]。在電子商務應用上，分析使用者進站瀏覽及購物行為，其關聯性可提供網站經營者很好的銷售決策與商品推薦之參考[15]。在證券投資應用上，股票投資的領域中，常使用技術指標分析來評估股票交易策略。技術指標很多元，利用關聯分析產生許多可供判斷之投資進出場規則，協助制訂股票投資交易策略[16]。在數位學習應用上，利用教學網站，從學生的學習特徵中，探勘學生的學習行為，藉以掌握學生的學習狀態[17]。在災害防治應用上，經由分析溼度、溫度、風力等各種不同氣象特徵，可更詳細準確預測氣象[18]。在醫學應用上，分析醫療資料庫，進行疾病類別與病人特徵之預測，顯示疾病與病人特徵之關聯性，如性別、年齡與血型等，可以作為醫師診斷時輔助參考[19]。

三、實驗資料

本研究主要使用兩種不同來源資料，分別為社群文章資料與環保關鍵字資料，資料內容

² <http://technews.tw/2014/12/24/taiwan-new-the-prime-minister-talks-about-tech-policy/>

³ <http://www.npf.org.tw/post/2/14788>

說明如下：

社群文章資料集：

此資料集是利用網路爬蟲蒐集 2002 到 2014 年在 PTT/Ecophilia 環境板共 12,412 篇與環保議題有關之文章，文章內容包含：作者 ID、標題、時間、文章內容、圖片等，此 12,412 篇將用來建立社群文章資料集(PttForumDB)。此資料集是由網路爬蟲 (Web crawler) 自動抓取，網路爬蟲是一種「自動化瀏覽網路」的程式，也是一種網路機器人。網路爬蟲被廣泛用於網際網路搜尋引擎或資料蒐集網站，用以取得或更新網站的內容。網路爬蟲可以自動收集所有頁面內容，供文字探勘演算法做探勘 (分析處理下載的頁面)，然後進一步得到隱含在網頁內容中之資訊。現今的網路爬蟲包提供者有開源軟體與商用服務廠商，開源軟體包含：Apache Nutch、Heritrix、Aperture、Grub 等；國內商用服務廠商包含：意藍資訊、碩網資訊、磐古數位、威知資訊、i-buzz 亞洲指標數位行銷等；國外商用服務廠商則有 80legs 等。此外由 PTT 抓取文章原始格式為 HTML，無法直接進行文字探勘與分析，所有抓取文章都經過剖析與清潔處理，將 HTML 轉換成純文字檔的格式，如： 轉為空格、"轉為雙引號""、刪除所有的 STYLE 設置等步驟。

環保關鍵字資料集：

透過與環境保護領域專家學者多次討論，分析新聞資料與環保署新聞稿，擬定重要環保關鍵字共 97 組。

四、預測模型

環保議題預測模型，主要係透過關聯法則學習法，分析長期的網路社群 (PTT/Ecophilia 環境板) 資料，據此建立網民討論環保關鍵字的關聯法則，以了解民眾關注之環保議題。資料分析流程如圖 1 所示，主要分為三個階段：資料蒐集、關鍵字比對、關鍵法則集合建立。此環保議題預測模型適用之兩種情境：(一) 政策研擬階段之輿情蒐集和 (二) 新聞發布後之即時回應。

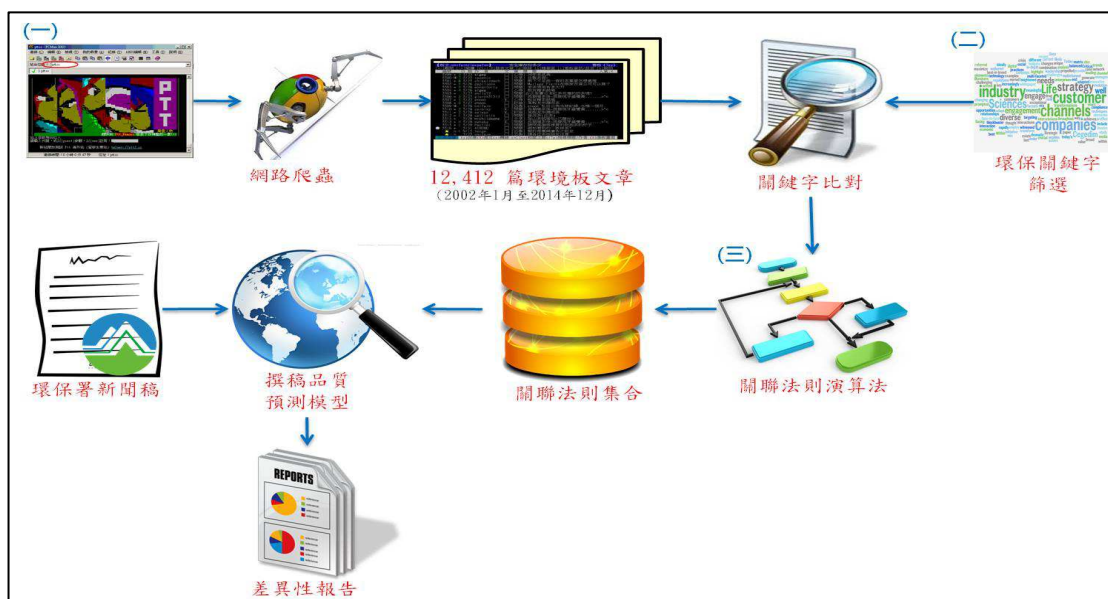


圖 1、議題預測模型分析流程

（一）政策研擬階段輿情蒐集

新聞稿為環保相關單位對外發布政策重要管道。當各級環保單位因應政策須發布新聞稿時，可自動分析新聞稿內容包含的關鍵字，並經由關鍵字關聯法則，由歷史社群資料中分析出網民會同時討論的關鍵字進行比對，列出此新聞稿中未包含在關聯法則中的關鍵字列表，代表須關注之焦點。

透過關聯法則學習法，分析長期的網路社群(PTT/Ecophilia 環境板)資料，據此建立網民討論環保關鍵詞的關聯法則。當環保相關單位將發布新聞稿時，自動分析新聞稿內容所包含的關鍵詞，並經由關鍵詞關聯法則分析與比對，列出此新聞稿中未包含在關聯法則中的關鍵詞列表，這些是由歷史社群資料中分析出網民會同時討論的關鍵詞，因此可以預先提示新聞稿中所欠缺網民關切的關鍵詞。

分析架構如下述步驟：

- (1) 以欲發布新聞稿之時間為啟始時間，從社群資料集(PttForumDB)中取出前一個月的文章，建立社群資料集子集合(PttForumDB’)
- (2) 比對社群資料集子集合(PttForumDB’)與手動設定欲發布新聞稿之關鍵字資料集(EpaKeyword)，找出社群資料集子集合(PttForumDB’)中所包含的關鍵字集合(EpaKeyword’)。
- (3) 分析包含的關鍵字集合(EpaKeyword’)所隱含的詞彙關聯，建立關聯法則集合(AssociationRuleSet)，並依據法則信心值排序。
- (4) 比對新聞稿與關鍵字資料集，找出新聞稿中所包含的關鍵字集合(EpaPressKeyword)。
- (5) 比對新聞稿中所包含的關鍵字集合(EpaPressKeyword)與社群關鍵字關聯法則集合(AssociationRuleSet)的差異程度，依照法則信心值，排序列出所研擬新聞稿中未關注到之議題。

（二）新聞發布後之即時回應

環保相關單位發布新聞稿之後，透過關聯法則學習，分析網路社群（PTT/Ecophilia 環境板）在新聞發布後的時間區段內，網民討論關鍵字的關聯性。進而比對新聞內容所包含的關鍵字與此段時間內的關鍵字關聯法則，追蹤掌握環保相關單位發布的新聞稿與網民關切的關鍵字之間是否出現差異，以便即時因應處理或加強政策溝通。

分析架構如下述步驟：

- (1) 以新聞稿發布時間為起始時間，從社群資料集(PttForumDB)中取出後一個月的文章，建立社群資料集子集合(PttForumDB’)
- (2) 比對社群資料集子集合(PttForumDB’)與環保關鍵詞資料集(EpaKeyword)，找出社群資料集子集合(PttForumDB’)中所包含的關鍵詞集合(EpaKeyword’)。
- (3) 分析該筆關鍵詞集合(EpaKeyword’)所隱含的詞彙關聯，建立關聯法則集合(AssociationRuleSet)，並依據法則信心值排序。

- (4) 比對新聞稿與關鍵詞資料集，找出新聞稿中所包含的關鍵詞集合(EpaPressKeyword)。
- (5) 比對該筆關鍵詞集合(EpaPressKeyword)與社群關鍵詞彙關聯法則集合(AssociationRuleSet)的差異程度，依照法則信心值，排序列出已發布新聞稿中未包含的關鍵詞彙。

五、實驗結果與分析

(一) 政策研擬階段之輿情蒐集

2014 年 11 月，高雄市爆發嚴重的登革熱疫情⁴。假設環保相關單位將發布一篇與登革熱疫情相關的新聞稿，撰稿輔助系統執行與分析的過程如下：

- (1) 假設環保相關單位欲發布新聞稿包含了 2 個環保關鍵字，分別為：廢水、廢棄物。
- (2) 利用議題偵測系統，分析關鍵詞集合所隱含的詞彙關聯，建立關聯法則集合如下，信心值越高代表法則越強健，也就是說產生的關鍵字關聯度非常高。關聯法則集合如表 1，以第 1 條法則為例，可以發現「水污染」會跟「廢水」的關聯度非常高。表 1 中斜體字表示原本已知之關鍵字，粗體字表示新發現之關鍵字。
- (3) 經過比對欲發布新聞稿關鍵字集合與社群關鍵詞彙關聯法則集合，可以得到環保關鍵字差集並為「水污染」、「環保局」、「資源回收」、「戴奧辛」、「水質」、「環保署」、「地下水污染」、「廢清法」和「飲用水」。此方式可以協助擬稿者，當撰寫有關登革熱防疫的新聞稿時，可以知道網民通常也會關注「水污染」、「環保局」、「資源回收」、「戴奧辛」、「水質」、「環保署」、「地下水污染」、「廢清法」和「飲用水」等議題，協助擬稿人撰寫新聞稿時，可納入參考，提高新聞稿內容廣度與關注議題涵蓋率，以此評估是否跨單位合作發布新聞稿，以及拉近與網民的距離。

(二) 新聞發布後之即時回應

假設以環保相關單位在 2014 年 11 月 5 日所發表的新聞稿為例⁵，標題為「廢食用油回收管理工作推動情形及未來規劃」，根據上述分析架構得到分析結果如下：

- (1) 環保相關單位發布新聞稿擷取關鍵字共有 4 個，分別為：環保局、廢食用油、廢棄物、環保署。
- (2) 設定分析資料來源時間區間：本研究使用新聞發布後 1.5 個月(2013.11.06~2014.12.21)，此為動態設定參數，可依需求調整。
- (3) 利用議題偵測系統，分析關鍵詞集合所隱含的詞彙關聯，建立關聯法則集合如表 2，以第 1 條法則為例，可以發現「廢棄物」與「水污染」跟「水質」的關聯度非常高。此外，因為 2014 年 11 月 6 日到 2014 年 12 月 5 日在 PTT/Ecophilia

⁴ <http://www.chinatimes.com/newspapers/20141109000258-260102>

⁵ http://enews.epa.gov.tw/enews/fact_Newsdetail.asp?InputTime=1031105164404

環境板之文章，都沒有提到關鍵字「廢食用油」，所以沒有產生與關鍵字「廢食用油」有關之法則。表 2 中斜體字表示原本已知之關鍵字，粗體字表示新發現之關鍵字。

- (4) 經過比對新聞稿關鍵字集合，與社群關鍵字關聯法則集合，可以得到關鍵字差集為「水質」、「水污染」、「地下水污染」和「戴奧辛」。故當廢食用油的議題在網路傳開後，PTT 網民閱讀的文章除了包含與環保相關單位新聞稿相同的關鍵字外，通常也會關注或討論與「水質」、「水污染」、「地下水污染」和「戴奧辛」等相關議題之文章，可提供相關單位參考因應。另一方面新聞稿關鍵字「廢食用油」並沒有被網民討論，也可思考是否加強政策宣傳。

表 1 政策研擬階段關聯法則分析結果

廢水← 水污染
廢水← 環保局
廢水← 環保局 水污染
水污染 ← 環保局 廢水
水污染← 環保局
環保局 ← 水污染 廢水
環保局← 水污染
水污染← 廢水
環保局← 廢水
廢棄物← 環保局
廢棄物← 環保局 廢水
廢棄物← 環保局 水污染
廢棄物← 環保局 水污染 廢水
資源回收 ← 環保局
資源回收← 環保局 廢水
資源回收← 環保局 水污染 廢水
資源回收← 環保局 水污染
戴奧辛 ← 環保局 廢水
戴奧辛← 環保局 水污染 廢水
水質 ← 環保局 廢水
水質← 環保局 水污染 廢水
環保署 ← 環保局
地下水污染 ← 環保局 廢水
地下水污染← 環保局 水污染
地下水污染← 環保局
廢清法 ← 環保局 廢水
廢清法← 環保局 水污染
廢清法← 環保局
飲用水 ← 環保局 廢水
飲用水← 環保局 水污染 廢水

表 2 新聞發布即時回應關聯法則分析結果

水質 ← 廢棄物 水污染
水污染 ← 廢棄物 水質
水質 ← 環保署
環保署 ← 水質
水質 ← 環保局
水質 ← 環保局 環保署
環保署 ← 環保局 水質
環保署 ← 環保局
水質 ← 水污染
水質 ← 水污染 環保署
環保署 ← 水污染 水質
環保署 ← 水污染
水質 ← 水污染 環保局
環保局 ← 水污染 水質
水質 ← 水污染 環保局 環保署
水質 ← 廢棄物
水質 ← 廢棄物 環保署
環保署 ← 廢棄物 水質
環保署 ← 廢棄物
水質 ← 廢棄物 環保局
環保局 ← 廢棄物 水質
地下水污染 ← 水污染 水質
環保局 ← 環保署 水質
戴奧辛 ← 環保局 環保署 水質
戴奧辛 ← 環保局
地下水污染 ← 環保局 環保署 水質
廢棄物 ← 環保局 水質

六、結論與未來研究方向

本研究以環保領域資料為研究樣本，建構環保議題偵測模型，探勘民眾目前關注議題與未來討論方向。分析結果顯示，採用 PTT 電子佈告欄之文章，經過實驗證明，能有效瞭解民眾過去關注議題及準確預測未來議題發展方向，可作為環保機關之參考資訊。然而，經由分析結果可發現，環保相關單位所發布之新聞稿與網民所關注的議題焦點仍有些微差異。若能及早處理與補強新聞稿未提到之議題，能降低民怨，提升人民對政府施政效率與滿意度。

未來研究方向可分為以下 3 點：

- (一) 本研究使用 PTT 為主要分析資要來源，網路上還有許多不同態樣資料來源可當成未來分析資料目標，例如 Facebook、Plurk 與政府單位的市民信箱等，透過多種不同來源資料交叉比對，應可有效提升分析資料的廣度與面向。
- (二) 資料所使用的時間區段可能會直接影響分析結果，如使用較多的資料（時間區段設定拉長）可以分析較多的資料樣本，但是有可能會提高雜訊包含率；使用較少的資料（時間區段設定減短），可降低雜訊被包含的機率，但是可能會影響到分析資料豐富性。如何拿捏最佳的分析區段，需要時間經驗累積來調整。
- (三) 本研究使用關聯法則演算法，可嘗試不同的機器學習演算法，並將不同的預測結果加以合併整理，應可再提升分析系統之效能。

參考文獻

- [1] X. Hang, J. N. K. Liu, Y. Ren, and H. Dai, “An incremental FP-growth web content mining and its application in preference identification,” in *Proceedings of the 9th international conference on Knowledge-Based Intelligent Information and Engineering Systems - Volume Part III*, Melbourne, Australia, 2005, pp. 121–127.
- [2] L. Peipeng and R. T. T. Sim, “Research experience of big data analytics: the tools for government: a case using social network in mining preferences of tourists,” in *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, Guimaraes, Portugal, 2014, pp. 312–315.
- [3] M. W. Berry, *Survey of Text Mining*. Springer-Verlag New York, Inc., 2003.
- [4] H. Guo and L. Zhou, “Segmented document classification: problem and solution,” in *Proceedings of the 17th international conference on Database and Expert Systems Applications*, Kraków, Poland, 2006, pp. 538–548.
- [5] Q. Luo, “Dynamic Fluzzy Clustering Algorithm for Web Documents Mining,” in *Proceedings of the 2010 International Conference on Computational Intelligence and Security*, 2010, pp. 64–67.
- [6] W. Chung, “An automatic text mining framework for knowledge discovery on the web,” The University of Arizona, 2004.
- [7] H. Alvarez, S. Sebastián A., F. Aguilera, E. Merlo, and L. Guerrero, “Enhancing social network analysis with a concept-based text mining approach to discover key members on a virtual community of practice,” in *Proceedings of the 14th international conference on Knowledge-based and intelligent information and engineering systems: Part II*, Cardiff, UK, 2010, pp. 591–600.
- [8] M. M. Mostafa, “More than words: Social networks’ text mining for consumer brand sentiments,” *Expert Syst Appl*, vol. 40, no. 10, pp. 4241–4251, 2013.
- [9] R. Kosala and H. Blockeel, “Web mining research: a survey,” *SIGKDD Explor Newsl*, vol. 2, no. 1, pp. 1–15, 2000.

- [10] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," in *Proceedings of the 20th International Conference on Very Large Data Bases*, 1994, pp. 487–499.
- [11] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, Washington, D.C., USA, 1993, pp. 207–216.
- [12] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, Dallas, Texas, USA, 2000, pp. 1–12.
- [13] Z. Huang, X. Lu, and H. Duan, "Mining association rules to support resource allocation in business process management," *Expert Syst Appl*, vol. 38, no. 8, pp. 9483–9490, 2011.
- [14] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, "Association rules applied to credit card fraud detection," *Expert Syst Appl*, vol. 36, no. 2, pp. 3630–3640, 2009.
- [15] Z. Xizheng, "Building Personalized Recommendation System in E-commerce Using Association Rule-based Mining and Classification," in *Proceedings of the Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing - Volume 03*, 2007, pp. 853–857.
- [16] P. Paranjape-Voditel and U. Deshpande, "A stock market portfolio recommender system based on association rule mining," *Appl Soft Comput*, vol. 13, no. 2, pp. 1055–1063, 2013.
- [17] S. B. Aher, *Recommendation System in Education: An Association Rule based Approach*. LAP Lambert Academic Publishing, 2012.
- [18] Z. Zhang, W. Wu, and Y. Huang, "Mining Dynamic Interdimension Association Rules for Local-Scale Weather Prediction," in *Proceedings of the 28th Annual International Computer Software and Applications Conference - Workshops and Fast Abstracts - Volume 02*, 2004, pp. 146–149.
- [19] K. R. Kumar, *Association Rule Mining - A Research: In Medical Perspective*. LAP Lambert Academic Publishing, 2012.