

無線點對點資訊分享網路中對搭便車資訊 享用者之無滲透研究

苑守慈

政治大學資訊管理系

黃子恆

輔仁大學資訊管理系

摘要

Gnutella、Napster 的點對點 (Peer to Peer, P2P) 資訊分享系統之間非利他主義的影響之下面臨了非常嚴重的挑戰，此挑戰即為資訊分享不對稱的問題。在本論文中將之稱為「資訊便車者 (free rider)」議題。隨著點對點無線隨意網路 (P2P wireless ad-hoc network, WP2P) 亦是具有高度發展潛力的領域，本論文因此提出點對點無線網路環境下衡量「資訊貢獻度」的模型，以解決在無線點對點資訊分享網路中資訊便車者的問題，進而提供點對點分散式網路架構一位良好的資訊分享基礎環境。此模型稱之為「情境化資訊貢獻度模型」(CICM)，在點對點資訊分享的過程之中，其考慮到資訊品質、資訊價值、時效性、分享策略、情境因素、網路頻寬及載具的效能等因子，並且利用上述因子，以 P-Grid 分散式儲存架構為基礎，評估個體對於此資訊分享網路的貢獻度，來消除在點對點無線隨意網路資訊分享中資訊便車者的問題，以期達到經濟效率及公平原則，避免點對點網路中無效率、不對稱的資訊分享，或是造成個體之間資訊分享成本上無謂的損失。

關鍵字：點對點、分散式處理、資訊分享、資訊貢獻度、便車資訊享用者、行動隨意式網

A Study on Free-Rider Distributed Impermeability for WP2P Content Sharing

Soe-Tsyu Yuan

MIS Department, National Chengchi University

Tzu-Heng Huang

IM Department, Fu-Jen University

Abstract

Gnutella and Napster, once most popular P2P file sharing systems, have encountered a serious problem for their significantly unbalanced content sharing between peers. Most peers in the sharing networks benefit from the sharing of others without contributing. This problem is so-called the free-rider problem. It is a very important issue to avoid the free-rider problem so that all of the peers in the networks can benefit substantially from collective actions so as to sustain the operations of the sharing networks. For the foreseeable wireless P2P contents sharing, this paper presents a correspondent solution model (utilizing the P-Grid storage model) named as “Contextualized Information Contribution Model, CICM”. This model enables peers in the network to share content fairly and efficiently by considering information contribution of peers and instituting peers the incentives to account for the global benefits of a private act.

Keyword: P2P, distributed computing, content sharing, information contribution, free rider, ad-hoc network

壹、緒論

近來，「點對點分散式網路架構」被 Intel 喻為第三代網路革命[楊朝凱, 2001] [A. Oram, 2001]。在網際網路 (Internet) 上的存在許多點對點資訊分享方案，如 Gnutella、Napster 之資訊分享。然而其中 90% 以上的個體 (peer) 通常是不會分享資訊卻只要下載資訊的 [E. Adar et al., 2000]，這種資訊分享不對稱的問題嚴重影響資訊服務的持久性。在本文將此挑戰統稱之為「資訊便車者 (free rider)」。

另外，隨著有線到無線的科技之快速演進，使得未來的資訊科技產品將以移動能力為最新的訴求，無線的應用將更貼近個人生活，甚至融合為生活一的部分。在無線應用服務中的一枝，適域商務 (ubiquitous commerce)，這個概念將促使行動商務 (mobile commerce) 邁向另一個全新的境界，走出主從式 (client/server) 的架構 [S. Schapp et al., 2002]。資訊應用服務不再只是消費性服務為主，將會轉向個人化的資訊應用，並且把點對點無線網路的特性納入設計的考慮，例如：自我組織 (self-organization)、動態 (dynamic) 聚合、情境察覺 (context awareness) 等特性 [N. Daswani et al., 2002]。是故，無線點對點網路的架構上將亦能演化出嶄新的資訊應用服務，在資訊服務設計的架構上也將是全新的挑戰，例如：自我組織的能力讓使用者在使用資訊服務時擁有更大的彈性 (scalability)，動態聚合的能力使其具有即時化組合方案，情境察覺的能力則提供了最適合當下的服務或應有的反應。

而在無線點對點網路中，資訊提供者的角色會有所改變，傳統的主從架構，已經宣告了資訊服務提供與使用的角色是固定不變的，資訊提供者與使用者之間的角色是可以互換，非固定不變，亦即網路中的每個參與者都有可能提供服務或使用服務，配合動態的無線網路環境，整個資訊網路儼然形成一個動態市集，充滿各種形式的資訊服務在市集上流動，所以點對點資訊分享的商業模式會變得更有趣、更有變化性，可以發展更複雜的策略與多元化的資訊服務。

然而，如同網際網路的點對點資訊分享，在無線點對點網路環境中（例如購物中心內的點對點無線網路環境）一樣會遇到資訊分享不對稱的困境，舉如：(1) 資訊分享集中在特定個體，分享不對稱的情境。(2) 購物資訊因為資訊便車者無法有效散佈到商場的每位角落，無法有效的刺激消費。(3) 分享網路的瓦解或停擺。

因此在如此高速變動的點對點無線網路環境之下，確保個體之間資訊分享對稱的機制是重要的。因為確保資訊分享網路中資訊的流通性是點對點資訊分享網路的基礎理論，也唯有確保點對點資訊分享網路中資訊的流通性與個體 (peer) 的分享率，才能有效經營一個點對點資訊分享的平台。然而資訊分享對稱的機制是需要一個具備「經濟效率」及「公平客觀」之資訊貢獻度模型設計。

本文因此提出「情境化資訊貢獻度模型 (Contextualized Information Contribution Model, CICM)」，其在點對點資訊分享的過程之中，考慮到資訊品質、資訊價值、時效性、分享策略、情境因素 (context factor)、網路頻寬及載具 (device) 的效能

(performance)等因子，以期達到具經濟效率及公平客觀特性之資訊分享對稱的機制。

本文以大型購物中心 (shopping mall) 為實驗情境 (如圖 1 所示)，在點對點資訊分享無線網路之下，消費者或商家可以分享即時的購物資訊，例如：商品資訊、廣告資訊、促銷活動或折價券等，且商店之間亦可舉行策略聯盟的合作活動 (例如：跨商家的認同卡，聯名消費的折扣)。本研究將每一個消費者均視為個體 (peer)，個體所扮演的角色是資訊提供者也是資訊下載者，商品資訊透過區域性的 provision servers 發送，再由每位資訊個體找到目前購物中心內鄰近存在的資訊 (服務) 提供者進行資訊交換。本研究的實驗平台架構屬混合式點對點無線網路。

在購物中心的環境中，消費資訊分享是一種不間斷的活動，又消費資訊分享必須能動態的組織週遭的個體形成分享網路來進行分享活動，而集中式架構有單一點癱瘓的系統風險以及高訊息溝通與資訊處理能力的需求，且網路無法自我組織，是故本研究以彈性較大的分散式架構做為此實驗環境的基礎。藉由貢獻度 (contribution) 評估個體，以消除在點對點隨意網路資訊分享中資訊便車者的問題，期望達到經濟效率及公平原則的「情境化資訊貢獻度模型」，避免點對點無線網路中無效率、不對稱的資訊分享，或是造成個體之間資訊分享成本上無謂的損失，並提高了購物中心裡個體之間的資訊流動性，且降低業主以推式廣告造成的成本與無效率的推薦結果，或給予消費者適地化的即時消費資訊。

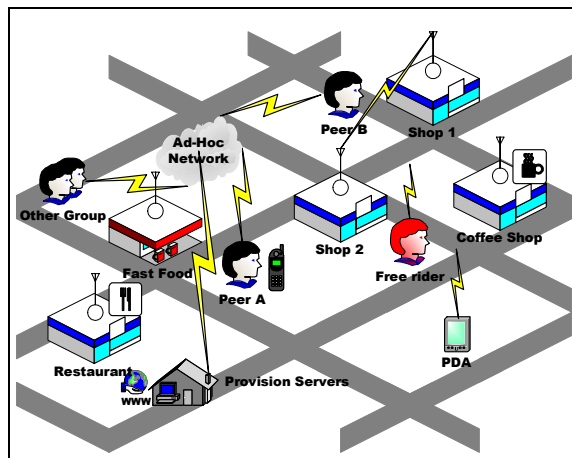


圖 1.點對點無線網路中，購物中心的即時資訊分享

本文主要分為三部份。第一部份說明目前存在之資訊貢獻度模型機制等相關研究。第二部份則呈現 CICI 之研究方法與架構。第三部份列出了本研究之 CICI 系統平台製作與實驗設計，並藉由實驗數據驗證 CICI 之價值。最後為本文結論，並提出後續研究的建議。

貳、文獻探討

在本節中我們針對點對點資訊分享發展現況及過去有過的相關貢獻度模型進行探討，學習過去的知識與經驗並找出過去研究不足之處。

Websense Inc. (NASDAQ: WBSN) 所作的調查報告指出，由 Napster 帶起的點對點檔案分享網絡，所交換的檔案已經由 MP3 音樂擴展至韓劇、流行電視遊戲等，包羅萬有。變化多端的交換內容，不單令新的 P2P 應用程式及網站如雨後春筍般日益增多，亦為全球企業帶來了各種關於頻寬、法律及保安的問題。事實上，P2P 檔案分享網站於過去 12 個月內增長逾 300%，達至 89,000 個之多。據員工上網管理(EIM)方案全球領導供應商 Websense 指出，現時互聯網上有超過 130 個不同的 P2P 應用程式供用戶使用，包括 KaZaa、Grokster 等。研究機構 Yankee Group 估計去年自 P2P 網絡下載的音樂檔案已逾 50 億個，而遊戲開發商 Trymedia 估計去年的電視遊戲下載數目亦超過 500 萬個。此外，顧問機構 Viant 亦估計每日的電影檔案下載率亦達 40 萬至 60 萬個；而每日約有 300 萬名用戶於 KaZaa 下載”Buffy the Vampire Slayer”等受歡迎的電視連續劇。

在 [G. Kortuem, 2001] [N. Daswani et al., 2003] 的研究指出，點對點資訊系統是一個自治的分散式系統，每位個體都是獨立的經濟體¹，擁有自己的決策，且其分散式的架構可以分享個體的一切電腦資源²，而點對點最大的特色就是分散式的處理，沒有中央管理（centralized control），方便了資源的共用，但也造成了系統效能評估不易、資訊搜尋與資訊系統安全的問題，其中關於資訊搜尋與資訊安全的討論整理如下：

一、搜尋機制（search mechanism）

一個好的搜尋機制，在點對點的架構中扮演了靈魂的角色，因為有了良好的搜尋引擎才有辦法正確又有效率的找到個體所需到的資訊。而一個搜尋機制將定義了點對點網路設計的架構以及個體行為模式，我們從下面三個維度討論：

- 拓撲（topology），個體連接的方式，即整個分散式網路組成的架構，會因為不同的系統有不同的設計。
- 資料置放（data replacement），在分散式網路中，個體對於資料儲存與獲取的方式。
- 訊息路由（message routing），個體轉送網路封包的路由工作，定義了個體的行為，舉如如何查詢、回應或是傳遞訊息的規範。

¹所謂的經濟體指的是每位個體會以尋求自身利益最大化為出發點，進行點對點資訊分享或是交易，是故分享網路是一個小型的經濟圈。

²電腦資源，泛指一切檔案、應用程式、計算能力、儲存空間、路由轉存等能力。

二、安全 (Security)

對於惡意個體 (malicious peer) 的侵入或破壞，可能影響到分散式系統的穩健度，所以安全的議題不容忽視。而對於安全的設計可考慮的因子如下：

- 可得性 (availability)，每個點對點網路中的個體，必須能接受其他個體的訊息，並且分享網路或是電腦資源，但可能被惡意個體濫用，如發動阻斷式攻擊 (denial-of-service, DoS)。
- 檔案認證 (file authenticity)，對於個體所查詢的檔案資料，並需能作到真切性的驗證，確定標的物存在於點對點資訊分享網路之中。
- 匿名 (anonymity)，對於資訊分享者或是請求者必須能作到保護個體個人資訊的安全控制。
- 存取控制 (access control)，對於系統資源或是權限的管理，以及智慧財產權的問題。

目前點對點資訊分享所遭遇到較大的問題是資訊便車者 (free rider) 與信任 (trust) 的議題。在資訊便車者的部分，Eytan Adar 等人的研究 [E. Adar et al., 2000] 中以 Gnutella 為例指出，在 Gnutella 資訊分享系統中，將近百分之七十的使用者是不分享資料的，而且將近百分之五十的回應訊息是來自百分之一的使用者，也就是說幾乎九成的使用者對這個資訊網路是沒有貢獻的。這個影響直接造成了資訊分享網路的停滯或是崩解，是目前點對點資訊分享網路急需解決的議題之一。在信任的部分，Karl Aberer [K. Aberer et al., 2001] 指出信任是點對點資訊分享重要的一環(如圖 2 所示)。因為在點對點資訊分享網路中，一般而言，使用者所接觸到的個體都是不認識的個體，所以，個體間信任的管理將影響點對點系統的成敗，而管理信任的方法可以利用個體的信譽 (reputation) 來衡量信任度。

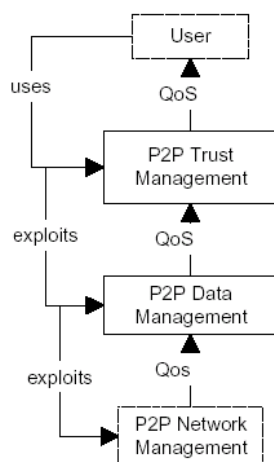


圖 2.點對點系統階層管理 [K. Aberer et al., 2001]

目前具代表性的貢獻度模型來解決資訊便車者的方法就是 Sepandar D. Kamvar 等人提出以信譽 (reputation) 為基礎的信任度管理 [S. D. Kamvar et al., 2003] 與 C. Courcoubetis 提出的市場成本模式為基礎的評量系統 [C. Courcoubetis et al., 2002]。而無論信譽的衡量準則或是演算法如何變化衡量，其背後有一個相同的理論基礎，那就是經濟市場，每位研究者都視點對點資訊分享網路的資訊分享交易為一項經濟市場上的交易，且每個個體都是一個經濟的個體，以最大化個體利益為決策交易的原則，有了代價的概念，舉凡每次交易都有一個付費的概念。從信譽 (reputation)、效度 (utility)、成本 (cost) 出發，並有一般經濟市場的競爭，在此模式下達到經濟平衡，以獲得整體經濟市場最大的效益。因此，市場模式可作為點對點資訊分享一個重要的基礎環境，平移上述的概念，若是能達到完全效率的經濟市場，那麼亦能間接解決部分的資訊分享不對稱的議題。

不過目前這些貢獻度模型都還是建立在主從式的架構之下，無論是信任度的管理或是付費機制計算都需要透過中央的伺服器來搜尋個體之間的資訊以進行資訊分享活動。另外這些貢獻度模型忽略了情境屬性 (contextualized attribute) 這個因子。因此，雖然能計算出資訊分享本身是否能為需求者或群體 (group) 創造出利益，進而決定是否資訊分享，但是卻無法計算出因為情境屬性、使用者剖象 (user profile) 所造成的資訊分享成本，以及利用分散式計算為基礎的優點與特性，因此值得加以改善。

而本研究提出的 CICM 模型則是選擇以分散式為基礎架構，個體之間的資訊交換與貢獻度的計算是分散式且在動態的環境之下所組合而成。而兩種架構最大的差異在於：

- 分散式的架構沒有集中式單點癱瘓導致網路停擺的問題。
- 分散式可以更動態且彈性的由週遭的個體來決定分享的群體，義即動態聚合 (slice of moment) 之能力。
- 分散式可以分散資訊處理的資源與網路流量，避免集中在中央，而需要大型主機來處理龐大的溝通與其他資訊處理的成本。

是故，分散式且動態聚合個體是與現存的研究最大的不同之處，是本研究的貢獻所在—提供的一個新分享評估架構。

參、研究方法

為避免點對點無線網路中無效率、不對稱的資訊分享，或是造成個體之間資訊分享成本上無謂的損失，本論文提出一個「情境化資訊貢獻度模型」(CICM)。在點對點資訊分享的過程之中，考慮到資訊品質、資訊價值、時效性、分享策略、情境因素、網路頻寬及載具的效能等因子，並且利用上述因子，以分散式計算為基礎，評估個體對於此資訊分享網路的貢獻度，並利用此資訊貢獻度消除在點對點隨意網路資訊分享中資訊便車者的問題，以期達到經濟效率及公平原則。

另外，由於無線網路的動態性，故 CICM 提供的並非是一個考慮整體的貢獻度 (global contribution) 模型，而是一個動態聚合 (slice of moment) 式的評估模型來除去資訊分享網路中資訊便車者 (free rider) 的問題。此外，CICM 資訊分享平台記錄動態聚合間個體的交易經驗，並且平移經驗到下次交易活動，以提昇資訊分享服務的個人化程度。

在本節中，我們將先描述 CICM 資訊分享平台之架構及是如何達到預期目標，並且在陸續的小節中詳細的介紹 CICM 內部各元件的功能與函數模型的定義及演算法。

一、CICM 資訊分享平台之架構

於 CICM 資訊分享平台架構 (於圖 3 所示) 中，點對點資訊分享評估考量了個體之分享價值 (sharing value)、個體之拒絕率及個體之被拒絕率三個向量以計算出資訊貢獻度 (information contribution)。並且藉由資訊貢獻度的計算進而避免無線點對點資訊分享網路中資訊便車者的問題。在評估資訊貢獻度的同時，每位使用者擁有自己的資訊分享策略，而資訊分享策略的形成則是透過 CICM 平台中 Strategy Evaluator 的元件來產生，Strategy Evaluator 會根據使用者 Profile 中的屬性或是使用者預設的特定情境分享策略，以及 Context Agent 所偵測到的環境因子，如網路頻寬、流量、連線數、終端設備處理能力，研擬出最適合使用者當下條件的分享策略，因此，當進行資訊分享的時候，CICM 平台會考慮到當下週遭的環境狀態、網路狀態以及使用者對不同場合時機的分享模式。

此外，在資訊分享的過程之中，CICM 平台會平移每次資訊分享交易的經驗到下次資訊分享活動身上，依據使用者的交易活動，CICM 透過 Preference Learning Agent (PLA) 來調整活動類別 Ontology [N. Guarino, 1997] [R. Jasper et al., 1999] 的類別權重，亦即學習使用者的交易經驗，學習使用者在選擇資訊檔案的類別偏好與特色，並在下次資訊分享活動產生時，而後根據使用者每次資訊分享交易對標的資訊的選擇，提供篩選資訊清單時給予適當的調整，例如將使用者喜好的類別資訊挑選出來並置於清單前列，是故，能達到平移每次交易經驗的目的。

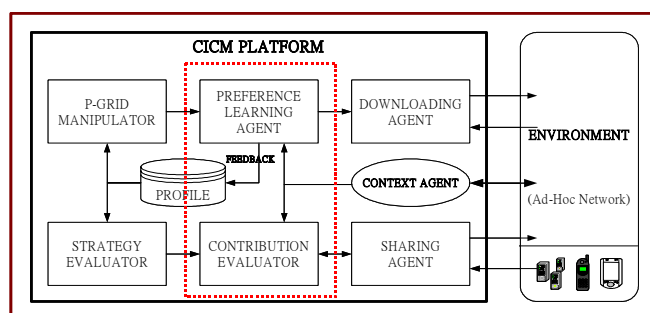


圖 3. CICM 資訊分享平台架構圖

茲將 CICM 資訊分享平台架構內部個別元件（component）的功能描述與概念如下：

- **Profile**：紀錄使用者的個人資料，舉凡個人資料、偏好、類別權重 Ontology、資訊分享經驗、可分享資源、分享策略等資訊都囊括其中，使用者可以自訂自我的 Profile，並且可以決定在什麼時間或是地點，將這些個人的資訊分享出去，以供其他個體運用。Profile 的設計可以讓使用者可以得到更貼近個人的資訊服務，因為在點對點網路存取資訊服務時，可透過 Profile 把有關個人的資訊，如：偏好、習慣、會員帳號、密碼...等定義特殊的分享策略，讓資訊服務提供者可以提供更貼近使用者需要的資訊服務，使用者具有控制 Profile 裡個人資料的權力，而且這些私密的個人資料只存放在個人的終端設備身上。
- **Context Agent (CA)**：CICM 平台中的 Context Agent (CA) 專門負責蒐集分享環境中與資訊分享相關的情境因子 (context factor)，例如：所在的時間、所在地點、特殊情境活動、網路頻寬、載具效能、下載連線數等因子，這些經由分享網路上取得的環境變數可以作為 Preference Learning Agent (PLA) 學習與 Contribution Evaluator (CE) 評估的參數值，讓學習和評估的結果能依據分享網路中環境因子的變化與發展，做出適當的判斷與回應。所以，情境化 (contextualized) 的資訊服務提供了更多的環境資訊，使得資訊分享或是提供資訊服務的過程中，能更精準了解當下使用者的需求與判斷環境的變化。
- **P-Grid Manipulator (PGM)**：由於本研究使用了 P-Grid 點對點分散式的資料儲存架構，當存取每位個體的資料時，需要透過每位個體 CICM 平台上的 P-Grid Manipulator (PGM) 來搜尋及存取標的資料，並且 PGM 會從 Profile 中取得使用者活動類別 Ontology 的權重，依據這些權重再篩選出合適的資訊清單給使用者作選擇。在 P-Grid 的儲存架構下，每位個體存有部分的搜尋樹 (p-tree)，搜尋樹中的節點記錄了每位個體的識別碼 (identifier) 以及該節點擁有的資料項目，當 P-Grid Manipulator (PGM) 在個體本身的搜尋樹上找不到標的物時，PGM 就將訊息轉送 (forward) 出去給其他的個體的 P-Grid Manipulator (PGM) 協助搜尋標的物，直到最後發現標的物或回傳無法搜尋到該標的物，而轉送的對象則是根據當時個體路由表 (routing table) 當中所儲存的節點為對象，且每位個體的路由表中至少會

儲存一個通往另一顆搜尋樹節點的路徑，以及同一顆搜尋樹下的節點路徑。是故，利用 P-Grid 的分散式儲存架構，能以分散式的存取有效的從事點對點的資訊分享，同時也借助其他個體的終端設備處理能力來搜尋擁有標的物的個體，所以，亦達到了 Grid Computing 的境界，即分散式的點對點網路中個體之間共享系統處理資源的模式。

- Preference Learning Agent (PLA)：CICM 平台中的 Preference Learning Agent 會根據使用者每次資訊選擇的交易結果，來調整 Profile 中活動類別 Ontology 的類別權重，並將調整學習後的結果再回饋到 Profile 之中，進而學習使用者的資訊分享偏好、與使用習慣。透過 Preference Learning Agent 學習的這些交易經驗，能在下次請求下載的搜尋結果清單中平移已存在的資訊分享經驗，最後於清單中做出符合使用者習性及喜好的調整。例如：若是活動類別的 Ontology 分為食、衣、住、行、育、樂五個維度，當使用者在食的維度中選擇了「時尚」的類別特性，則可以透過 Preference Learning Agent 的學習調整將「時尚」類別的權重提高，並且在下次需要其他下載維度的資訊服務時，將「時尚」類別及其他權重較高類別的相關的資訊篩選出來，以供使用者選擇，達到平移交易經驗的目的。因此，CICM 平台會累積使用者的資訊分享經驗，並提供個人化的資訊分享服務供使用者使用。
- Downloading Agent (DA)：當使用者從 Preference Learning Agent 學習篩選過後的資訊清單中選擇的檔案標的物後，Downloading Agent 會負責發送請求分享的訊息給資訊提供者，並且負責和資訊提供者的 Sharing Agent (SA) 作資訊下載的談判溝通，最後 Downloading Agent 會回覆給使用者目前資訊提供者檔案的下載狀態，以及管理被許可的下載連線。
- Strategy Evaluator (SE)：根據使用者 Profile 內的個人資訊或是預設的分享模式，透過 Strategy Evaluator 擬出使用者個人的資訊分享策略，例如：分享策略可劃分成完全開放、部分開放、好友分享、特殊地點分享、限速限量等模式，然而資訊分享策略將隨著 Profile 設定以及時間地點或是分享對象改變而演化出不同的分享策略給不同的情境使用，當選定了分享的策略後，再把策略模式傳送給 Contribution Evaluator (CE) 和 Sharing Agent (SA) 作分享檔案時的指標。如此，當個體分享資訊時，即可依據個人習性和意願分享資料檔案。
- Contribution Evaluator (CE)：在接收到請求者要求分享的訊息與確定了資訊分享策略後，CICM 平台的 Contribution Evaluator 將開始評估請求者的資格，進而決定是否分享資訊檔案給請求者。其中，資訊分享的標準則是根據請求者對分享網路的資訊貢獻度而決定，而貢獻度的計算則是根據使用者本身的交易經驗及網路中相似個體的交易經驗來評估請求者。由於點對點無線網路的動態性，所以網路中的成員並非固定的個體且沒有固定的網路識別，是故，此貢獻度的評估結果可以說是一個動態聚合 (slice of moment) 的評估結果。又因為受限於無線網路的動態性，個體流動率高，所以無法評估個體的整體貢獻度 (global contribution)，但透過此動態聚合的評估，可以得到最具代表性、符合當下情境的評估結果。

- Sharing Agent (SA)：當資訊提供者的 Sharing Agent 接收到資訊請求者要求分享的訊息後，交由 Contribution Evaluator 評估請求者的網路資訊貢獻度，藉由此資訊貢獻度的結果和資訊請求者的 Downloading Agent 作資訊下載的談判溝通，最後 Sharing Agent 會回覆給使用者目前資訊請求者的下載狀態，以及管理、分享符合資訊貢獻度的下載連線。

二、P-Grid 儲存結構

在資料存取 (data access) 的部分，本研究採用結構化的分散式儲存資料結構 (P-Grid)。P-Grid 以結構化的方式動態建立搜尋樹 (P-Grid Tree)，每個個體都只存有部分的搜尋樹。首先，每個個體的路由表 (Routing table) 中至少存有一組到另一顆樹的節點 (node)，當個體要循尋找某一特定的資料時，以查詢 (query) 的方式取得資料，且當查詢的主鍵 (ID) 不屬於收到訊息的節點時，收到訊息的節點會在轉送 (forward) 出去，給最靠近目標節點的節點，此步驟反覆執行直到找到目標節點為止。其次，有新的資料要加入此搜尋樹或是修改資料時，則是分別以新增 (insert) 和修改 (modify) 的函數完成。圖 4 圖與 5 則為圖解的說明。舉例說明，資訊個體 6 存有索引 00 的資訊項目，當網路中其他的資訊個體 (如：資訊個體 2) 想要搜尋索引是 00 的資訊項目，此時資訊個體 2 就會查看其路由表認識的資訊個體中是否有儲存索引為 00 的資訊個體，以圖 5 來看，資訊個體 2 中的路由表中有資訊個體 4、6，其中資訊個體 6 為索引 00 的個體，是故資訊個體 2 可以查詢到索引 00 的資訊。再另一個例子，若是資訊個體 2 想查詢的索引為 11 開頭的資訊項目，又在資訊個體 2 沒有認識索引 11 的個體，此時資訊個體 2 會將查詢傳給它所認識的個體，看看這些資訊個體是否有認識索引 11，如此規則以此類推，圖例中最後資訊個體 2 則會透過資訊個體 4 查詢到資訊個體 5 的項目。

換言之，有了 P-Grid 樹狀的架構後，接下來透過每位資訊個體身上的路由表資訊傳遞查詢的訊息，在 P-Grid 搜尋樹中的任何一個節點都可以發出查詢，每位個體會有自己所認識的資訊個體，如圖 4 中個體之間連結線即是，因此形成一種分散式、遞移的資訊網路。P-Grid 是屬於一個完全點對點 (pure P2P) 的系統架構，不需要集中式的架構來處理通訊。

採用 P-Grid 主要原因則是為了避免像 Gnutella 這種分散式分享平台對網路造成洪流 (flooding) 以及僅能對個體的鄰居作有限範圍的廣播等問題，且有相關研究報告的實驗數據指出 P-Grid 對於個體的搜尋較能準確的掌握且在資訊分享的過程中效率多了 [K. Aberer et al., 2003]。

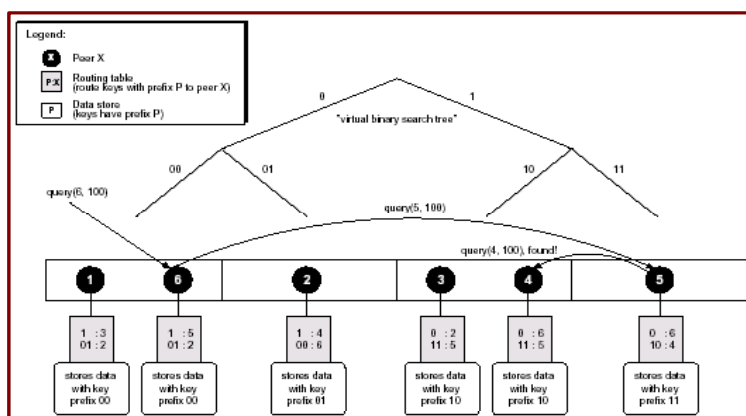


圖 4.P-Grid 結構說明圖[K. Aberer et al., 2003]

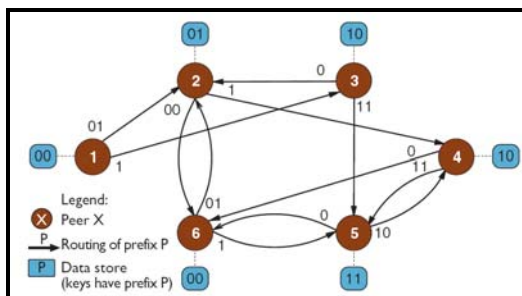


圖 5.對應圖 4 之 P-Grid Routing Table Exchange [K. Aberer et al., 2003]

P-Grid 分散式的資料儲存在點對點的環境中，已經漸漸地廣泛被使用，其中 P-Grid 主要的特色可歸納如下：

- 完全分散式的環境架構，沒有集中的伺服器處理任何分享過程中所需要的資料訊息。
- 隨機化的搜尋演算法，當建構資料存取結構時，沒有特定的集合對象 (peer set)，此方法可以提高資料搜尋的範圍以及資料搜尋的命中率，並且可以提升系統穩定性 (robust)，因為當資訊網路中的某些資訊個體失敗 (failure) 或是離開無法提供服務時，搜尋的演算法不會被固定在某些特定個體之上。
- 由於上一點的特性，使得此分散式的資料結構涵蓋力 (scalability) 強，能有效的搜尋整個資訊網路。

三、Contribution Evaluator

本研究將資訊分享網路中個體 (peer) 每次的資訊分享經驗都視為 Instance-Based Learning (IBL) [T. Mitchell, 1997] 中的一個樣本 (sample instance)，並且儲存在各別個體的終端設備身上，等到資訊請求者發出要求分享的訊息，在由資訊提供者透過

Contribution Evaluator (CE) 評估資訊請求者的資訊貢獻度，分辨出對分享網路有貢獻的個體與沒有貢獻的資訊便車者 (free rider)。當進行資訊貢獻度的評估時，需要透過 P-Grid Manipulator (PGM) 來搜尋 P-tree，整個 P-Grid 是透過分散式儲存架構，存取點對點網路中其他個體與資訊請求者曾經發生過的資訊分享經驗，亦即仿效 Instance-Based Learning (IBL) 的學習模式下，透過 P-Grid 來存取樣本，進行資訊貢獻度的評估。最後，若評估出來的資訊貢獻度達到系統預設的標準量 (threshold) 則與該請求者進行資訊分享，並記錄下該次資訊分享經驗，反之則拒絕請求者的要求，並紀錄拒絕與該個體交易。

舉一個實際案例說明(如圖 6 所示)，說明當個體 A 在接受到個體 C 的請求訊息時，如何透過 P-Grid Manipulator (PGM) 分散式儲存結構 (P-Grid)，存取其他個體的資訊分享經驗 (sharing experience)，並在評估個體 C 的資訊貢獻度後，決定是否與個體 C 進行資訊分享交易。首先，個體 C 將獲得其他個體的資訊清單 (information list)，此資訊清單中紀錄著擁有個體 C 所需資料檔案的個體名單，在個體 C 選定了其中的一個分享個體後，個體 C 選擇了個體 A，是故個體 C 的 Downloading Agent (DA) 會發出資訊分享的請求訊息給個體 A 的 Sharing Agent (SA)，在個體 A 接收到個體 C 的請求訊息後，個體 A 會先啟動 Context Agent (CA) 蒐集有關情境因子的資訊，例如：所在的時間、所在地點、特殊情境活動、網路頻寬、載具效能、下載連線數等因子。然後，根據目前的環境狀態與使用者的分享喜好習慣，由 Strategy Evaluator (SE) 擬出適合當下與符合使用者的資訊分享策略，例如：分享策略可劃分成完全開放、部分開放、好友分享、特殊地點分享、限速限量等模式。

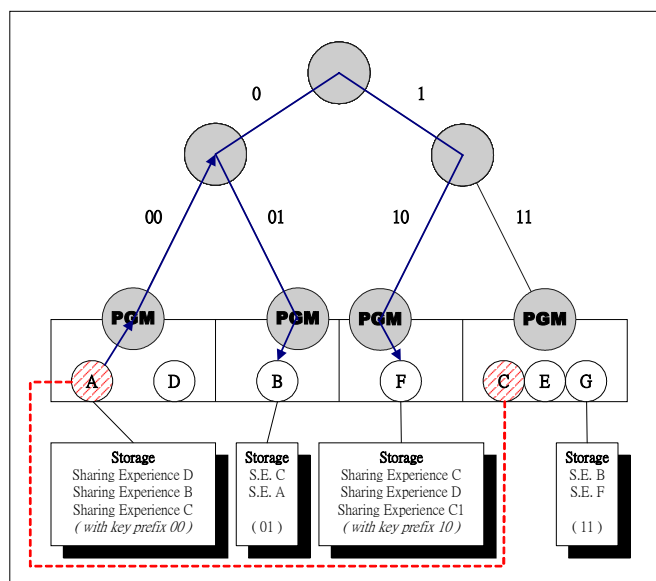


圖 6. 資訊貢獻度評估流程示意圖

在確認分享模式後，個體 A 就進入了評估個體 C 資訊貢獻度的階段，個體 A 需要資訊分享網路中其他個體與個體 C 的資訊分享經驗，或是本身與個體 C 曾經有過的

資訊分享經驗。在此一階段中，個體 A 會透過 P-Grid Manipulator (PGM) 分散式儲存結構 (P-Grid)，存取其他個體的資訊分享經驗 (sharing experience)，而存取的方式則是藉由每個個體身上的路由表 (routing table)，每張路由表除了紀錄到其他個體的路徑之外，還會記錄個別的個體儲存了哪些資訊分享經驗 (如圖 4 與圖 5 之說明)。如此的設計有助於每位個體能透過 P-Grid Manipulator (PGM) 尋找特定個體之間的資訊分享經驗，進而作為自身評估其他個體資訊貢獻度的依據。有了其他個體的資訊分享經驗後，個體 A 即可啟動 Contribution Evaluator (CE) 進行評估個體 C 的資訊貢獻度，最後決定與個體 C 分享資訊或是拒絕分享交易的請求。

由於蒐集其他個體分享經驗的過程是尋找目前分享網路中存在的個體，並且以分散式的架構存取分享經驗，所以，本研究模型得到的資訊貢獻度是一個動態聚合 (slice of moment) 的評估結果，不同於固定式的網路架構所評估的結果，分散式架構能夠提供即時隨地的評估，評估的過程不需要定置的設備來處理此評估結果，提供了更具動態性與移動性的服務。而 Contribution Evaluator (CE) 元件定義與 Contribution Evaluator (CE) 演算法將列表並詳細說明如下：

表 1. Contribution Evaluator Algorithm 變數說明

Requester X : a peer who requests content sharing in the p2p ad-hoc network.

Peer i : peers who currently are alive in the p2p ad-hoc network.

Definition :

Ic (Requester X, Peer i): the information contribution of Requester X evaluated by Peer i .

Fwi : the weights of features that constitute the information contribution construct, e.g. value of sharing, rate of rejection and rate of refusal.

Contextual attributes

En (Requester X) : detected location, device and time from the environment.

Nt (Requester X) : detected bandwidth, connection and speed from the network.

St (Peer i) : the selected sharing strategy from the strategy pool by Strategy Evaluator.

Vs (Requester X) : Value of sharing delivered by X who triggers the computation of $En()$, $Nt()$, and $St()$.

$\sigma(y)$: Sigmoid function, the output is a continuous function of its input which is between 0 and 1.

$$\text{i.e., } 0 \leq \sigma(y) = \frac{1}{1 + e^{-y}} \leq 1$$

Rrj (Requester X, Peer i) : rate of rejection representing the ratio of content-sharing requests (from other peers) rejected by X (the records of these rejections are retained at the side of the requestors – Peer i).

Rrf (Requester X, Peer i) : rate of refusal denoting the ratio of peer X's requests of content sharing refused by other peers – Peer i (the records of these refusals are retained at the side of the providers – Peer i).

Contribution Evaluator Algorithm:

Function $Ic(Request\ X, Peer\ i)$ **returns** the information contribution of peer X evaluated by Peer i in a slice of moment.

Input : X, a peer who is requesting for information

sub function $Vs(Requester\ X)$ **returns** the value of sharing delivered by X

1. Start the *Context Agent* to detect the contextual values between *Requester X* and Peer i in the p2p ad-hoc network.
2. Start the *Strategy Evaluator* to choose a strategy of content sharing based on the contextual values.
3. Choosing a model of content sharing based on the strategy.
4. Return $\sigma(En(Requester\ X) + Nt(Requester\ X))$

sub function $Rrj(Requester\ X, Peer\ i)$ **returns** the ratio of content-sharing requests (from other peers- Peer i) rejected by X

1. Start the *P-Grid Manipulator* to search the transaction experience about X in the p2p network
2. Return $\frac{\sum_{i=1}^n Trj(X, Peer\ i)}{Tt(X, Peer\ i)}$

sub function $Rrf(Requester\ X, Peer\ i)$ **returns** ratio of peer X's requests of content sharing refused by other peers – Peer i.

1. Start the *P-Grid Manipulator* to search the transaction experience about X in the p2p network
2. Return $\frac{\sum_{i=1}^n Trf(X, Peer\ i)}{Tt(X, Peer\ i)}$

While receiving the request MSG. from Requester X

1. $Vs =$ Call function $Vs(Requester\ X)$
2. $Rrj =$ Call function $Rrj(Requester\ X)$
3. $Rrf =$ Call function $Rrf(Requester\ X)$

在 Contribution Evaluator (CE) 計算資訊貢獻度 (information contribution) 的演算法中:

$\frac{\sum_{i=1}^n Trj(X, Peer\ i)}{Tt(X, Peer\ i)}$ (Formula 1) 所計算的是拒絕率 (Rate of Rejection) 是指當個體 i 發出資訊分享的請求後, 標的個體 (target peer) X 拒絕分享資料的比率, 而式子 $\frac{\sum_{i=1}^n Trf(X, Peer\ i)}{Tt(X, Peer\ i)}$ (Formula 2) 所計算的是被拒絕率 (Rate of Refusal), 則是指每次個體

X 發出資訊分享的請求後, 遭受其他個體 i 拒絕分享的比率。是故, Formula 1 是用來衡量標的個體 X 分享意願的指標, 用來計算標的個體 X 不願意分享資料的比率, 當比

率愈高，表示標的個體 X 資訊分享的貢獻度愈差。相對而言，Formula 2 則是一個衡量個體 X 人氣指數的指標，用來計算個體 X 請求分享卻遭受拒絕的比率，當比率愈高，表示 X 資訊分享的歷史交易記錄太差，導致貢獻度低落，所以每每遭受拒絕。藉由 Formula 1&2，本研究可以推論一件事實，即當某一個體的拒絕率（Rate of Rejection）與被拒絕率（Rate of Refusal）都超過系統的預設值（threshold）時候，則表示此個體為資訊分享網路中的一名資訊便車者（free rider）。如此，便能更有效率地找出對資訊分享網路沒有貢獻的資訊便車者，解決點對點網路中，資訊分享無效率與不對稱的問題。此外，在轉換分享價值（Vs, value of sharing）時，本研究運用 s 形函數（sigmoid function） $\sigma(y) = \frac{1}{1+e^{-y}}$ （Formula 3）轉換從環境中偵測到的情境因子及網路變數，環境

變數轉換到介於零到一之間的常數³。最後，結合分享價值、拒絕率與被拒絕率，計算資訊貢獻度，並且依據情境因子與 User Profile 產生分享策略，例如：當使用者在特定的科技展覽會場參觀時，可以設定資訊分享只與公司團隊成員分享，此時情境因子會偵測目前的環境地點，結合使用者透過 Strategy Evaluator 所做的 Profile 的設定（i.e. 好友分享模式），就可以產生屬於使用者的資訊分享策略模式，屬於好友的成員 *Stw* 值會被設成 1，反之，非好友的使用者 *Stw* 值會被設成 0。

四、Strategy Evaluator

根據使用者 Profile 內的個人資訊或是預設的分享模式，透過 Strategy Evaluator（SE）擬出使用者個人的資訊分享策略，例如：分享策略可劃分成完全開放、部分開放、好友分享、特殊地點分享、限速限量等模式，然而資訊分享策略將隨著 Profile 設定以及時間地點或是分享對象改變而演化出不同的分享策略給不同的情境使用，當選定了分享的策略後，再把策略模式傳送給 Contribution Evaluator（CE）和 Sharing Agent（SA）作分享檔案時的指標，作為選擇資訊分享的模式的依據。如此，當個體分享資訊時，即可依據個人習性和意願分享資料檔案。而分享策略模式的權重表如下：

3 如將網路上傳貨下傳的頻寬速率帶入，轉換出 0 到 1 之間的常數。

表 2.分享策略模式權重表

分享策略模式	權重值
完全開放	1
部分開放	0.8
不開放分享	0
好友分享	{Yes : 1, No : 0}
特殊地點分享	{Yes : 1, No : 0}
限速限量	1, 限制流量與大小
對等分享	{Yes : 1, No : 0}

當計算資訊貢獻度時，Contribution Evaluator(CE)即可依據依據 Strategy Evaluator(SE)產生的資訊分享策略，選擇上表的權重權重值，做最後資訊貢獻度的處理，表 2 中的權重即式子 $Stw * (Fwv * Vs + Fwj * Rrj + Fwf * Rrf)$ (Formula 4) 中的 Stw 。

五、Preference Learning Agent

Preference Learning Agent (PLA) 在每次使用者完成資訊選擇的交易之後，會根據活動類別 Ontology 的交易結果轉換出對應的類別向量矩陣，之後再根據此向量矩陣去調整平台上 Profile 中類別權重 (category weight)，使用者偏好的類別權重將會漸漸被提高，因此，透過此模式可以學習到使用者在從事資訊分享時資訊選擇的偏好、與使用習慣。是故，本研究利用 Preference Learning Agent (PLA) 學習及累計這些資訊分享的交易經驗，並且，在下次提供使用者請求下載的搜尋結果清單時，依據計算出的類別權重，對分享的資訊清單二次篩選，達到平移目前已存在的資訊分享交易經驗的目的，亦即依據活動類別 Ontology 所計算出來的類別權重，提供符合使用者習性及喜好的資訊分享搜尋結果清單。

首先，Preference Learning Agent (PLA) 會接受來自 P-Grid Manipulator (PGM) 以及 Context Agent (CA) 兩個元件輸入的參數值。在 P-Grid Manipulator (PGM) 的部分，將接收到使用者對於資訊清單項目的選擇結果，在 Context Agent (CA) 部分，則是接收到網路與情境因子的參數，本研究則依據使用者的選擇結果以及環境變數值來調整類別權重，而調整的方式則是將這些參數值轉換成類別選擇向量與類別向量矩陣 (category vector matrix, CVM)，透過此類別向量矩陣計算出新的權重並調整類別權重，最後再回饋到使用者的 Profile 之中。而類別選擇向量的屬性與類別向量矩陣的定義如表 3：

表 3.類別選擇向量屬性表

類別屬性	屬性值
類型	時尚、科技、復古、一般、運動、保健、居家
活動型態	動態、靜態

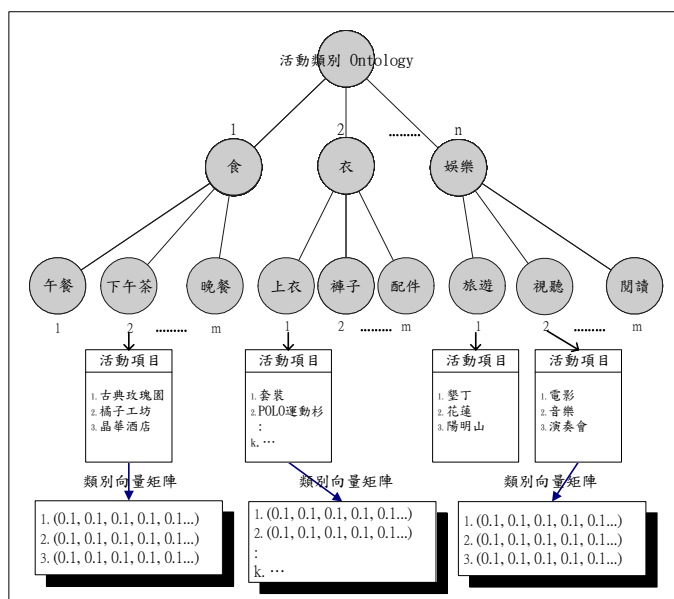


圖 7.活動類別 Ontology

本研究模型中的類別權重是一個累積的結果，Profile 紀錄使用者原有的類別權重，若此個體在資訊分享網路中沒有從事過資訊分享的交易活動時，則給予預設值 0.1 的初始權重值，亦即將所有類別的活動類別視為同等重要，如圖 7。爾後，當有新的分享交易活動產生時，經由轉換成類別選擇向量與類別向量矩陣，並且依據此新活動計算出新的類別權重，最後再與原有的類別權重作平滑均數，透過平滑指數取兩者的平滑平均值作為個體新的類別權重，平滑均數在這裡的目的是用來平滑因為特殊情況所造成的極端值之影響。所以，本研究模型中的類別權重（category weight）是一個遞移的累進結果，其中包含以紀錄類別權重值的方式反映使用者選擇資訊時偏好及習慣的歷史紀錄。

於圖 7 中，將活動類別的 Ontology 分為食、衣、娛樂三個維度，當使用者在食的維度中的資訊項目中其含有「時尚」的類別特性，此時則可以透過 Preference Learning Agent (PLA) 的學習調整將「時尚」特徵值的權重提高，並且在下次需要其他下載維度的資訊服務時，將「時尚」類別及其他權重較高類別的相關的資訊篩選出來，以供使用者選擇，達到透過 PLA 平移過去食的資訊項目所學到的交易經驗的目的。因此，

CICM 平台會累積使用者的資訊分享經驗，並提供個人化的資訊分享服務供使用者使用。而 Preference Learning Agent (PLA) 權重學習以及平移歷史經驗的元素定義和演算法如下：

表 4.Preference Learning Agent Algorithm 變數說明

CVL (*Srt*, *Cf*) : Category vector learning machine that calculates

the category weights based on the behavior of peerX.

Srt is a set of selections (rendered by peer*i*) from the option menu.

Cf is the contextualized factor created by Context Agent.

CVM () : Category Vector Matrix storing the category vector for peerX.

CVM (*Srt*, *Cf*) = def $\exists n, m, k$ is (*CIT1I1*, *CIT1I2*, ..., *CIT1Ik*,
CIT2I1, *CIT2I2*, ..., *CITmIk*, *C2T1I1*, *C2T1I2*, ..., *CnTmIk*) , $n, m, k \in N$.

CnTmIk = def is $x_1, x_2, x_3, \dots$, or $x_{14} \in CVM()$, $x_1, x_2, x_3, \dots, x_{14} \in \{0, 1\}$.

Itv (*s*) : Item to vector (transforming the selected options and the contextualized factors into a category vector matrix). i.e., basing on the file feature set match with the feature of category selected vector.

CWold : An old category weight stored in the user profile.

CWnew : An new category weight calculated with user's file sharing behavior.

α : Learning rate, $\alpha \in [0, 1]$, $\alpha \in \mathbb{R}$.

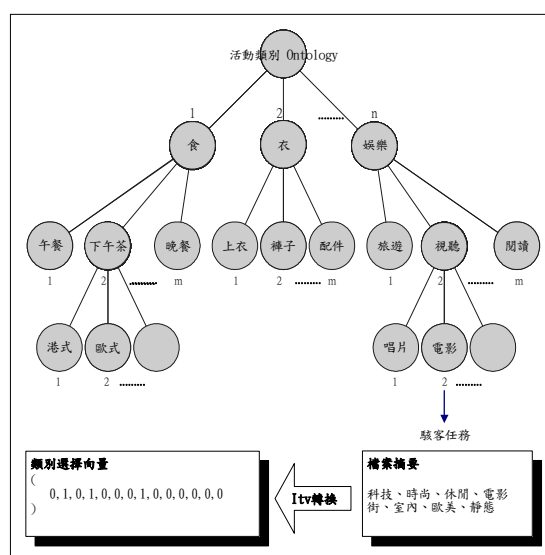


圖 8.活動類別 Ontology 圖例說明

表 4 定義了類別向量矩陣的組成元素，其中類別選擇向量 $C_n T_m I_k$ 定義了在第 n 個類別中第 m 個種類的第 k 個資訊分享項目之向量值。而每次使用者選擇了資訊分享項目後，將依據其項目的屬性特徵轉換成類別選擇向量，又交易經驗與向量之間的屬性轉換標準，則是依據表 4 所定義的類別向量矩陣屬性表為主，藉由轉換後所得到的類別向量矩陣，進而可以計算出新的類別權重。如圖 8 活動類別 Ontology 圖例說明：使用者選擇下載駭客任務電影的影片資訊與電影折價卷，其檔案摘要中說明此資訊項目在娛樂、視聽、電影的類別維度中擁有科技、時尚、靜態的特徵⁴以及一般特徵關鍵字中的休閒關鍵字，因此透過 Itv 的分析轉換將此資訊分享交易轉換成類別選擇向量得到 $(0,1,0,1,0,0,0,1,0,0,0,0,0,0)$ ⁵，而轉換的方式則是依據表 4 類別選擇向量屬性表，Itv 則依據分享檔案本身的屬性摘要將此次交易對應到所屬的類別向量屬性，得到類別選擇向量後，結合每一筆選擇交易的類別選擇向量 $C_n T_m I_k$ ，即可得到類別向量矩陣。最後 Preference Learning Agent (PLA) 依據新的類別向量矩陣產生新的類別權重，再將此新的類別權重與使用者原有的類別權重作平滑均化的處理，如此，即可獲得累積並學習使用者交易經驗的類別權重，亦可避免因極端值所造成的影響，如 Preference Learning Agent Algorithm 所示。

Preference Learning Agent (PLA) 計算類別權重 (category weight) 的演算法。其

$$\text{中, } \frac{\sum_{n=1}^N \sum_{m=1}^M \sum_{k=1}^K C_n T_m I_k(x_i)}{N * M * K} \quad (\text{Formula 5})$$

即是將類別向量矩陣 (category vector matrix,

CVM) 中的類別向量轉換成類別權重 (category weight)，其演算邏輯是去計算每位類別屬性被選用的次數最後取平均值，拿來代表此類別屬性的重要性，即使用者的喜好程度。例如使用者選擇的十個資訊分享項目中科技類的屬性佔了六個，由 Itv 將此十個資訊項目轉換成資訊選擇向量再透過 Compute CVM，將每個由 itv 轉換出來的解果合併起來所得到的向量矩陣，那麼在類別向量矩陣中則可計算出科技屬性的類別權重為 0.6 (6/10)，又每種資訊分享項目會符合多個類別屬性，最後累計出來的向量即可代表使用者的使用習慣以及使用偏好，經由上述過程反覆計算，方可算出使用者個人累計的類別權重，並可在下次提供資訊清單的時候，透過類別權重篩選資料清單，達到平移使用者資訊選擇經驗的目的。

4 其中特徵值具有一個靜態的語意樹 (ontology)，此 ontology 紀錄每位特徵的關鍵字用來對應檔案摘要中的資訊，亦即轉換檔案摘要成為選擇向量的標準。

5 依表 4 特徵擁有時尚、科技、復古、一般、運動、保健、居家七種特徵又每種特徵細分為動態與靜態兩種，是故 Itv 由特徵 ontology 中的特徵名稱或關鍵字 (如圖 7) 轉換的類別選擇向量，其中類別選擇向量依表之七種特徵依序排列，又每種特徵依序分為動態與靜態的類別，得 14 選項的類別選擇向量。

Preference Learning Agent Algorithm

Function CVL (*Srt, Cf*) **returns** the category weights of peer X learned evolved by Preference Learning Agent.

Input : *Srt, Cf*

1. Attain *CW_{old}* from the profile of the peerX.
2. Compute $CVM = Itv (Srt) \cup Itv (Cf)$.

$$\frac{C_n T_m I_k (xi)}{N * M * K}$$

3. Calculate $CW_{new} = \frac{\sum_{n=1}^N \sum_{m=1}^M \sum_{k=1}^K C_n T_m I_k (xi)}{N * M * K}$.

4. Return $\alpha * CW_{old} + (1 - \alpha) * CW_{new}$.

5. Replace *Cw_{old}* with $\alpha * CW_{old} + (1 - \alpha) * CW_{new}$ in the user profile.

肆、研究方法實驗結果

CICM 資訊分享平台乃是透過 JXTA [L. Gong, 2001] 與 P-Grid 分散式的資料儲結構來開發完成，並將之安裝在每一個 Peer 之 行動裝置上。JXTA 是一套開放原始碼版本的點對點協定，可讓同一網路上的任何裝置（如：手機、PDA、PC、伺服器 etc）作資料的交換與聯繫。（附錄一將簡略說明 CICM 之各模組互動關連。）

為驗證 CICM 平台之價值，我們進行了一連串的實驗。這些實驗的目的有兩個方向：

- (1) 驗證 CICM 資訊貢獻度模型對於資訊分享網路中資訊便車者的消除能力。
- (2) 驗證 Preference Learning Agent (PLA) 學習使用者資訊選擇偏好來過濾資訊項目搜尋結果的能力。

另外，在績效評估指標方面：

- (1) 於 CICM 資訊便車者的消除能力實驗中，將以 acceptance rate 作為評估 CICM 的指標 (metric)。其中，acceptance rate 用來了解整個資訊分享網路的均衡程度，並且透過資訊便車者的 acceptance rate 來檢驗是否 CICM 能夠辨別出資訊分享網路中沒有貢獻度的資訊個體。
- (2) 在 PLA 使用者資訊選擇偏好實驗中，以 ranking difference 作為評估 Preference Learning Agent 的尺度 (metric)。且 ranking difference 是用來衡量 Preference Learning Agent 所學習到的資訊使用選擇習慣是否符合使用者的習慣。

最後的實驗結果則是透過這些尺度所設計的評估函數 (Evaluative Function) 來驗證 CICM 模型在點對點資訊分享的過程之中是否能提供一個資訊效率流通性與平衡性的資訊分享環境。而 Evaluative Function 的公式設計與概念說明如下列所示：

$$\text{Evaluative Function 1} = (1 - \text{acceptance rate}) / \text{malicious peer rate}$$

Evaluative Function 1 藉由分享率可以了解整個資訊分享網路的分享程度是否活絡。此外，以整體的拒絕率、malicious peer 在網路中分佈的比率，藉由比率原則的統計推論，推測是否有達到客觀的資訊貢獻度評估。另外，本研究亦針對個別的資訊便車者來檢視是否在 CICM 模型的分享網路中會被拒絕分享，並期望能由宏觀與微觀兩個角度來檢視 CICM 模型。整體的 malicious peer rate 比率高時，當 Evaluative Function 1 愈接近 1（亦即拒絕率、malicious peer rate 接近一比一），以比率推論，兩者比例相近表示有客觀的區分出具資訊貢獻的個體。反之，假設整體的 malicious peer rate 比率低時，Evaluative Function 1 愈接近 0。是故，當 Evaluative Function 1 有能力區分出具資訊貢獻的個體時，本研究將可以進一步推論，在 CICM 模型的資訊分享之下，能達到對等的資訊分享。

$$\text{Evaluative Function 2} = \frac{1}{n} \sum_{i=1}^n (\text{peerItemRank}_i - \text{PLAItemRank}_i)$$

peerItemRank *i*, the information item rank provided by the peer.
PLAItemRank *i*, the information item rank provided by PLA.

Evaluative Function 2 藉由使用者心中值得等第（ranking）順序，與 preference learning agent 的結果來比較，看看是否符合使用者資訊選擇的使用習慣。而 ranking difference 尺度的好壞的評估如 Evaluative Function 3，依據實驗時的實際等第差狀況與最差狀況的比率，決定透過 Preference Learning Agent 所得到的排序結果的錯誤率，是故，當 Evaluative Function 3⁶的結果越高時表示 preference learning agent 學習的結果越差。

$$\text{Evaluative Function 3} = \text{real case of difference} / \text{worst case of difference}$$

關於 Evaluative Function 2、3 舉例說明如下，當 preference learning agent 排出的資訊標的等第為 $A > B > C$ ，而使用者的回饋的資訊等第為 $B > A > C$ ，藉由 Evaluative Function 2 我們可以知道等第差為 2，其計算過程為 $B:(3-2) - A:(2-3) + C:(1-1)$ 。其中，被低估的標的物等第以正數表示，被高估的標的物等第以負數來表示，因此當等第差越大時，則表示 Preference Learning Agent 的學習結果與使用者實際的選擇差異越大。而已此案例來說，最差狀況為 4，根據 Evaluative Function 3 可算出錯誤率為 50%（2/4）。

6 公式中的 real case of difference 表示經測試實驗後得到的等級差，worst case of difference 表示依據實驗設計的資訊項目個數計算出最差狀況下的等級差，其中最差的狀況為使用者的選擇與 PLA 學習的結果頭尾對調。

一、CICM 對資訊便車者的消除能力之實驗

在實驗環境的設定中，圖 9 顯示此組實驗的實驗流程概念。本實驗模擬資訊分享網路中兩邊資訊個體之間進行資訊分享的活動，一邊負責資訊項目查詢，一邊負責資訊項目分享，並且在實驗的過程之中查詢訊息的設定以約 100/200/300/400/500/600 次的實驗次數累進進行結果的分析，最後以約 600 次資訊查詢量的實驗次數來呈現 CICM 實驗的結果。其中在每次的實驗本研究假設將資訊便車者的產生率設定在 0.2 的比率⁷，作為實驗的基礎環境進行資訊分享的實驗。模擬實驗的總次數共為 12 次，每次的實驗時間約略為 2.5 小時。另外為了驗證 CICM 能運用在各種資訊便車者個體分佈的環境之上，本研究再將資訊便車者的產生率設定為 0.6，並且同上述實驗環境進行實驗，其結果如後述實驗結果所示。

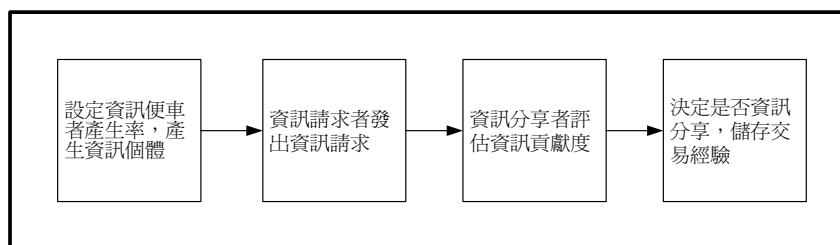


圖 9.CICM 實驗流程設計

在資訊個體的設計上，由於在本研究每位資訊個體所使用的網路環境參數（網路速率、流量限制）與情境參數（分享策略模式）其數值特性都屬於離散型的有限集合值，如表 5。是故，本研究決定在模擬資訊個體時，以隨機亂數的方式產生每次實驗中資訊個體 profile 中的網路環境與情境參數，如此便可模擬出不同個體的環境與行為模式，以供實驗使用，彌補只以兩邊（資訊項目查詢、資訊項目分享）來模擬點對點資訊分享網路的不足。

⁷ 此數值為實驗的假設，做此架設是為了要做實驗測試，與真實資訊分享網路的環境設定無關，並且依據此架設最後在透過實驗結果來檢視這些資訊便車者的個體是否能被 CICM 所杜絕。

表 5.資訊個體參數屬性表

類別	參數屬性
分享策略模式 ⁸	{"All"、"Part"、"Deny"、"Friend"、"Equal"、"LocationYes"、 "LocationNo"、"SpeedFast"、"SpeedNormal"、"SpeedSlow"}
網路速率、流量限制	上傳：{64、128、256、512、1500}
	下載：{512、1500、2000}

表 6 為系統實驗參數的設計與描述，用來表示實驗環境參數的使用與設定。

表 6.系統實驗參數

變數名稱	值域	描述	預設值
Contributive Score	natural number	貢獻度分數,CICM 的評估結果	0
Alpha	0..1	平滑指數,決定 CICM 學習調整的速率	0.5
Threshold	natural number	標準值,決定是否達到分享標準	15
Acceptance Rate	0..1	分享資訊成功的機率	0
Rate of Malicious Peers	0..1	不分享資訊的個體在分享網路中的比率	0.2

首先，第一個實驗(Free Rider Rate = 0.2, 資訊查詢量 630 次)的結果是以 acceptance rate 來呈現整體資訊網路的資訊分享率，亦即觀查整個資訊分享網路是否活絡。其實驗數據分析如圖 10。

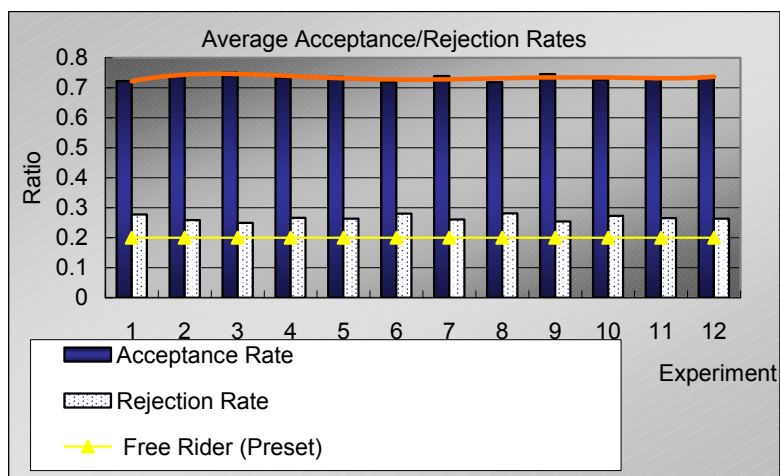


圖 10.整體分享率平均數

⁸ 分享策略模式說明，All：不限定任何模式以最高的策略權重分享，Part：分流流量分享，Deny：不分享，Friend：好友分享，Equal：對等資訊量分享，LocationYes/No：依據 SpeedFast/Normal/Slo 設定流量速度分享。

由圖 10 中的 X 軸來看可以看到每一次實驗平均分享率的都達到七成以上，資訊分享的活動在網路中能順利進行，並未因多了 CICM 的評估系統而導致過度貢獻度評量，使得資訊個體之間無法分享資訊資源的現象，又從每次實驗初期到實驗末期的過程之中 CICM 都能有效的評估資訊個體。而且整體的拒絕率都高於實驗設計的資訊便車者的比率 0.2，以整體的觀點來看，接受率在三次回歸趨勢線的趨勢下，平均分享率能夠穩定的維持在 75%左右的比率，且每次實驗之中無極端值的產生。是故，本研究初步推論透過 CICM 的評估，資訊分享活動在網路中仍可順利、活絡的進行，且有拒絕資訊便車者在網路中進行資訊分享的能力。此外，本研究在以 Evaluative Function 1 的計算來分析資訊分享率與資訊便車者比率之間的關係，由 Evaluative Function 1 帶入 12 次實驗平均的分享率 (0.734003724) 與實驗設計的資訊便車者比率 (0.2) 得到的數值為 1.329981 ($(1-0.734003724)/0.2$)，整體而言，CICM 所拒絕的資訊個體是高於資訊便車者的比率。以宏觀的角度來看 CICM 的資訊分享活動，資訊便車者是可以被 CICM 過濾掉的。

接下來，本研究同樣使用 acceptance rate 此尺度來衡量，再透過微觀的角度，由個別資訊便車者的資訊分享情形來進一步驗證 CICM 是否真的杜絕了資訊分享網路中的資訊便車者，其實驗數據與分析圖如圖 11 所示：

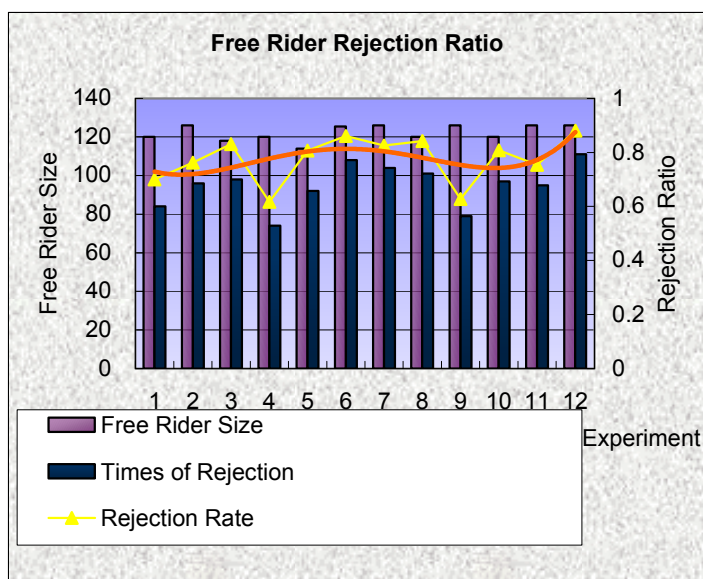


圖 11. 資訊便車者平均被拒絕率

由圖 11 的 X 軸來看可以看到每一次實驗平均被拒絕率的都達到七成左右，圖中 Y 軸左邊為每次實驗資訊便車者的數量對應在單次實驗的長條圖，而 Y 軸右邊為每次實驗資訊便車者被拒絕的比率對應在單次實驗的折線圖。是故，資訊分享的活動除了在網路中能順利進行，而且對於實驗中所設計的 20% 的資訊便車者都能有效的被去除，12 次的實驗平均拒絕率為 77.6%。而且資訊便車者被拒絕的比率在三次回歸趨勢

線的趨勢下，資訊便車者的被拒絕率有足漸攀升到 90%左右的趨勢，且每次實驗之中無極端值的產生。是故，不論以整體或是個別的觀點來看，本研究都能推論透過 CICM 的評估，資訊分享活動不但在網路中仍可順利、活絡的進行，而且能有效的拒絕掉資訊便車者在網路中進行資訊分享活動。

而為了驗證 CICM 能運用在各種資訊便車者個體分佈的資訊分享網路環境之上，本研究再將資訊便車者的產生率設定為 0.6，進行實驗測試，其實驗數據如分析圖 12 所示。由圖 12 的 X 軸來看可以看到資訊便車者的累積數量，圖中 Y 軸為資訊便車者的累積資訊分享率與累積拒絕率，此次實驗經歷了 630 次的資訊分享，一共有 352 位資訊便車者，其中有 89 位資訊便車者被其他資訊個體接受分享，是故整各資訊便車者的群體有 74.7% ($(352-89)/352$) 的個體被拒絕掉，又在累積拒絕分享次數的過程中，平均的拒絕率約在 83.5%。是故，以上述兩組實驗設計 (free rider rate = 0.2, 0.6) 與實驗數據結果來看，本研究能驗證透過 CICM 進行資訊分享能有效的拒絕掉分享網路中各類型分布的資訊便車者在網路中進行資訊分享活動。

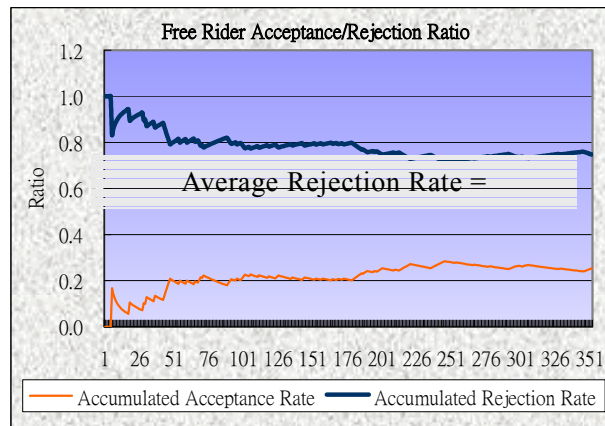


圖 12.資訊便車者累積被接受率/被拒絕率

二、PLA 學習符合使用者資訊選擇習慣之實驗

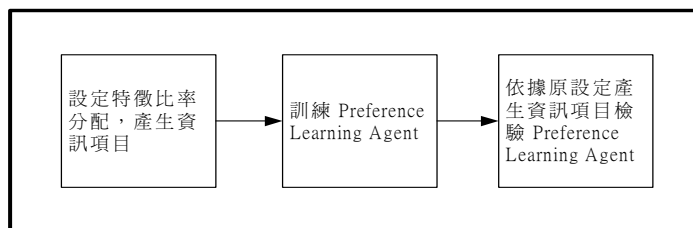


圖 13.Preference Learning Agent 實驗流程設計

圖 13 為 Preference Learning Agent 實驗的實驗流程概念，目的在於實驗模擬資訊分享網路中的資訊個體在選擇 P-Grid 搜尋資訊項目的結果時，Preference Learning Agent 學習使用者選擇偏好與篩選資訊項目結果的活動。在每次的 Preference Learning Agent 模擬實驗分別進行了約 1000/5000/10000 次的資訊選擇活動，用來訓練 Preference Learning Agent 來預先訓練 Preference Learning Agent 調整 Category Ontology 之中的特徵屬性權重值 (feature weight)，且學習的平滑指數設定為 0.5⁹，最後以預先訓練 10000 次的 Preference Learning Agent 來呈現進行實驗呈現結果，並且依據此訓練結果進行 1000 次的實驗測試，驗證學習的結果是否符合當初模擬使用者資訊選擇的資訊項目分配，最後的實驗結果以 10000 次的活動測試作為資料分析的數據。

其中，模型訓練活動設定了兩種資訊項目的分配。第一類型的資料其資訊項目分配以兩個資訊特徵 (fashion、technology) 為主要特徵。第二類型的資料其分配以五個資訊特徵 (classic、general、sport、health、home) 為主要特徵。模擬實驗的次數共為 12 次，Preference Learning Agent 的訓練與模型的測試每次產生 10 個資訊項目來測試，每次的實驗時間約略為 1.5 小時。

首先，實驗的第一部份必須先進行 Preference Learning Agent 喜好模型的訓練。訓練的資料有兩種類型，分別為以兩個 (類型一)、五個 (類型二) 資訊項目特徵為主的資訊項目分配產生資訊項目來訓練 Preference Learning Agent 來調整 category ontology 中的七個特徵權重 (feature weight)，並且進行 12 次的實驗，依據實驗的數據本研究預期能驗證當 Preference Learning Agent 面對不同的特徵喜好時能有效的學習出使用者的特徵選擇喜好，其實驗分析圖表如圖 14、15 所示。

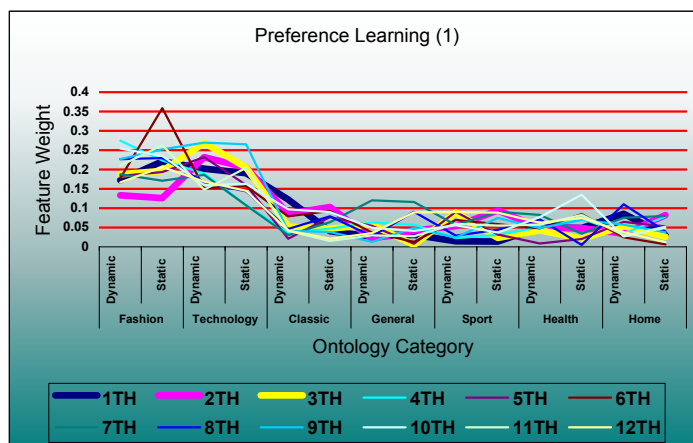


圖 14.平均 category ontology 特徵權重折線圖 (類型一)

9 此處的平滑指數表 PLA 累積舊的經驗與學習新的偏好經驗之間的比率，亦可表示 PLA 學習的速率，本研究認定新舊的經驗之間重要性依樣，故在此設定 0.5 的數值為本研究的實驗假設作為實驗所用，但在實際環境中此學習速率是可以依據個體的人格特質而有所不同。

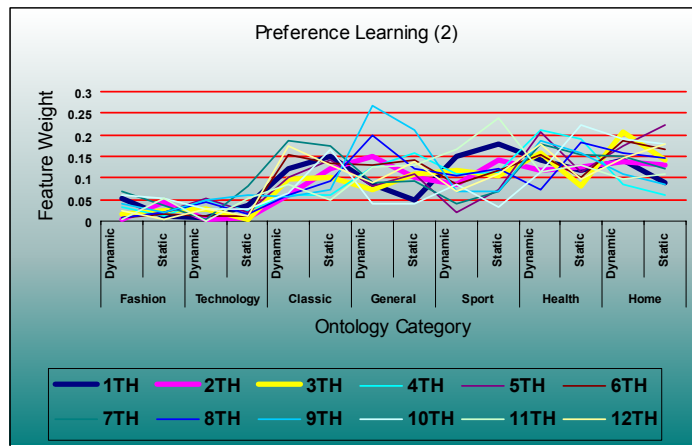


圖 15.平均 category ontology 特徵權重折線圖（類型二）

由圖 14 與圖 15 的 X 軸來看可以看到每一次實驗平均資訊項目特徵權重，圖中 Y 軸為每次實驗資訊項目特徵權重值對應在單次實驗的折點之上。圖 14 為類型一資料，其結果分布趨向 fashion、technology 兩個特徵值，圖 15 為類型二資料，其結果分布趨向 Classic、general、sport、health、home 五個特徵值，在圖上的折線可以清楚的看出無論是類型一或是類型二的線圖，大致上的分布趨勢都與原始的資料類型相符，且每次實驗之中無極端值的產生。因此，透過比率數據的說明，本研究驗證了透過 Preference Learning Agent 的資訊選擇偏好學習效果誠如實驗設計所預期。

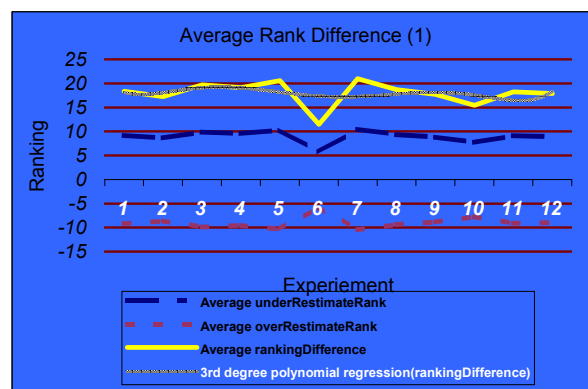


圖 16.Preference learning 平均等第差（類型一）

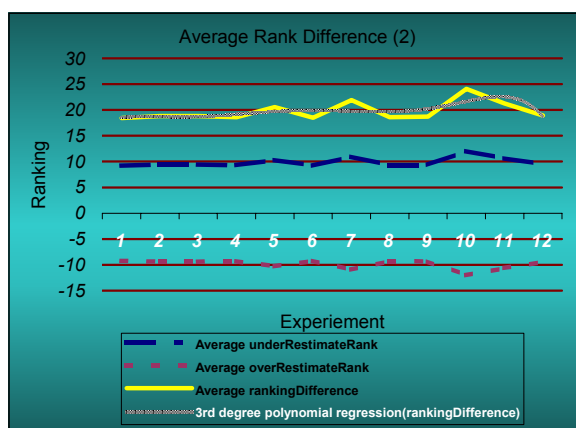


圖 17. Preference learning 平均等第差（類型二）

最後，根據 Evaluative Function 2、3 的 ranking difference 尺度，驗證 Preference Learning Agent 是否能正確的排序（sorting）兩種資料型態的資訊選擇樣式。由圖 16 與圖 17 的 X 軸來看可以看到每一次實驗平均等第差，圖中 Y 軸為每次實驗等第差分數對應在單次實驗的折點之上。圖 16 為類型一的資料，圖 17 為類型二資料，其種兩種資料類型的結果以三次回歸趨勢線分析，其等第差保持穩定，其值約略在 18~21 左右遊走，且每次實驗之中無極端值的產生。

最後，根據 Evaluative Function 3 此一評估函數來評量等第差的結果是否能驗證 PLA 的學習具有信度。其中，由於每次實驗所產生的資訊項目數量為 10 個項目，所以其最差狀況的等第差值為 50，又由 12 次的實驗中本研究可以得到等第差平均值約略為 19.77366667，是故 Evaluative Function 3 為 0.3954733334 ($19.77366667/50$)，亦即 PLA 有 61.5% 的正確率。故本研究驗證 Preference Learning Agent 所學習使用者的資訊選擇偏好具有 6 成以上的信心能符合實際的偏好選擇。

伍、結論

在不確定的無線點對點資訊分享環境之下，藉由本研究提出「情境化資訊貢獻度模型 (Contextualized Information Contribution Model, CICM)」，利用 P-Grid 分散式儲存架構儲存交易經驗，以計算資訊個體在資訊分享網路中的資訊貢獻度，嘗試解決資訊分享中的資訊便車者 (free rider) 的議題。同時，在「情境化資訊貢獻度模型 (Contextualized Information Contribution Model, CICM)」的點對點資訊分享的過程之中，考慮到資訊品質、資訊價值、時效性、分享策略、情境因素 (context factor)、網路頻寬及載具 (device) 的效能 (performance) 等因子，並且利用上述因子，以分散式計算 (distributed computing) 為基礎，評估個體對於此資訊分享網路的資訊貢獻度 (information contribution)。並且透過 Preference Learning Agent (PLA) 學習個體資訊選擇的習慣與偏好，再藉由平移交易選擇經驗的機制，提供個體個人化的資訊選擇。

根據 CICM 的實驗結果顯示，本論文所提之研究方法能維持整體分享網路的資訊分享活絡，並且能相當地杜絕資訊便車者的角色存在在網路中。另外，其亦有學習使用者資訊選擇偏好來過濾資訊項目搜尋結果的能力。換言之，本論文提供了一個能有效經營無線點對點資訊分享網路、確保個體之間資訊分享對稱的機制，進而創造出不同於傳統的商業模式。

未來研究將繼續改善 PLA 學習的方法，以提高準確度。可能方法包括調整學習速率、改善新舊兩種學習經驗用於學習結果的比例、設定不同的關鍵字重要性來得到更精緻的結果、提昇關鍵字分類的方法、建立靜態或是動態的字庫、依據特定的文件群或資訊項目群進行內容的解析以得到關鍵字 (terms) 建立關鍵字庫等。

參考文獻

1. [楊朝凱, 2001] 楊朝凱, P2P, 網路的大同之路?, 數位觀察者, <http://www.digitalobserver.com/71-80/78/kevin.htm>, 第七十八期 2001 年 7 月 4 日。
2. A. Oram, editor. Peer-to-Peer: Harnessing the Power of Disruptive Technologies. O'Reilly & Associates, March 2001.
3. [T. Mitchell, 1997] T. Mitchell, Machine Learning, McGraw Hill, 1997.
4. [C. Courcoubetis et al., 2002] C. Courcoubetis and P. Antoniadis, Market Models for P2P Content Distribution, International Workshop on Agents and Peer-to-Peer Computing, Bologna, Italy, 2002.
5. [E. Adar et al., 2000] E. Adar and B. A. Huberman, Free Riding on Gnutella, Internet Ecologies Area Xerox Palo Alto Research Center, 2000.
6. [G. Kortuem et al., 2001] G. Kortuem, J. Schneider, D. Preuit, T. G. C. Thompson, S.

- Fickas, Z. Segall, When Peer-to-Peer comes Face-to-Face: Collaborative Peer-to-Peer Computing in Mobile Ad-hoc Networks, 2001 International Conference on Peer-to-Peer Computing, Linköpings, Sweden, 2001.
7. [K. Aberer et al., 2001] K. Aberer, Z. Despotovic, Managing Trust in a Peer-2-Peer Information System, Proceedings of the Tenth International Conference on Information and Knowledge Management, McLean, VA, 2001.
 8. [K. Aberer et al., 2003] K. Aberer, P. Cudre-Mauroux, A. Datta, Z. Despotovic, M. Hauswirth, M. Ponceva, R. Schmidt, P-Grid: A Self-organizing Structured P2P System, ACM SIGMOD Record, 32(2), September 2003.
 9. L. Gong, JXTA: A Network Programming Environment. IEEE Internet Computing, 5(3), pp. 88-95, May/June 2001.
 10. [N. Daswani et al., 2002] N. Daswani, H. Garcia-Molina, and B. Yang, Open problem in file sharing of P2P, Stanford University, Stanford CA 94305, 2002.
 11. [N. Daswani et al., 2003] N. Daswani, H. Garcia-Molina, and B. Yang, Open Problems in Data-Sharing Peer-to-Peer Systems, Stanford University, Stanford CA 94305, <http://www-db.stanford.edu>, 2003.
 12. [N. Guarino, 1997] N. Guarino, Understanding, Building and Using Ontologies, International Journal of Human-Computer Studies, 46(2-3), pp. 293-310, February 1997.
 13. [R. Jasper et al., 1999] R. Jasper and M. Uschold, A Framework for Understanding and Classifying Ontology Applications, Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management, Voyager Inn, Banff, Alberta, Canada, 1999.
 14. [S. Schapp et al., 2002] S. Schapp and R. D. Cornelius, U-Commerce, Accenture, 2002.
 15. [S. D. Kamvar et al., 2003] S. D. Kamvar, M. T. Schlosser, H. Garcia-Molina, EigenRep: Reputation Management in P2P Networks, WWW2003, Budapest, Hungary, ACM, 2003.

附錄一：CICM 之各模組之互動關連

本論文以大型購物中心（shopping mall）為情境設計，在點對點資訊分享無線網路之下，消費者或商家可以分享即時的購物資訊，例如：商品資訊、廣告資訊、促銷活動或折價券等，且商店之間亦可舉行策略聯盟的合作活動（例如：跨商家的認同卡，聯名消費的折扣）。本論文將每一個消費者均視為個體（peer），個體所扮演的角色是資訊提供者也是資訊下載者，並透過 P-Grid 伺服器尋找到目前購物中心內存在的資訊（服務）提供者。於以下說明之圖 A1（CICM 系統流程圖），假設個體 X 為資訊分享者、個體 Y 為資訊請求者來作講解。

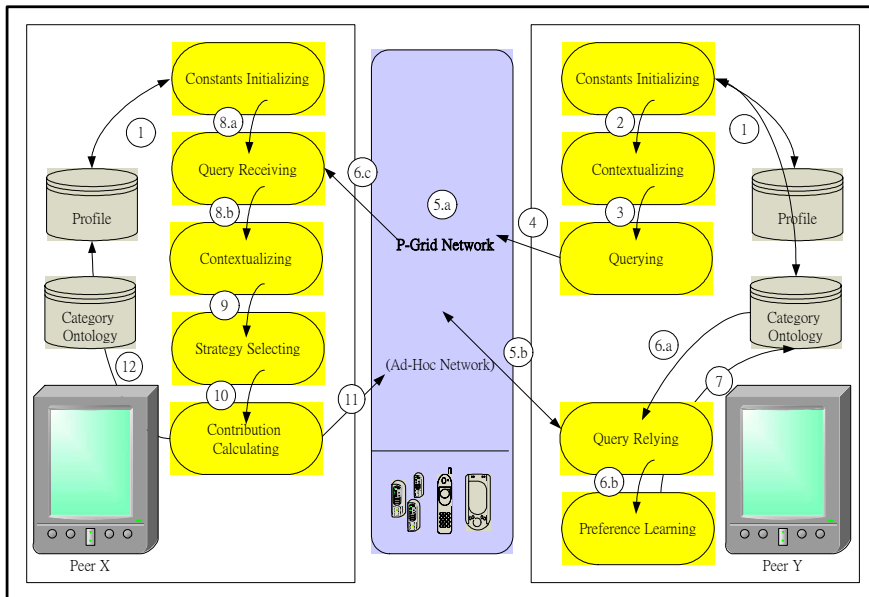


圖 A1. CICM 系統流程圖

CICM 平台之建立乃是依據此 CICM 系統流程圖與情境。其中與各系統元件之間的系統流程說明如下：

當個體 Y 搜尋資訊項目並向個體 X 發出資訊分享的請求，流程說明如下：

1. 當個體 X（分享者）、Y（請求者）啟動 CICM 時，為系統變數初始化。如：profile 中的個人設定、資訊分享的歷史經驗、category ontology 中的類別權重等設定值。
2. 個體 Y 偵測有關情境因子¹⁰與網路環境因子¹¹的環境變數，並依據個體 Y 個人需

10 系統使用的情境因子屬性如下述項目。此處的情境因子指的是了解請求者當下的資訊搜尋策略，如是否有個體群組搜尋、資訊項目特徵值（feature）等情境搜尋設定，以求在分享的過程中更充分的表達請求者當下的態度與想法。

11 系統使用的網路因子屬性如下述項目，此處的網路因子指的是了解請求者對於資訊分享者其網路頻寬（上／下傳）的大小選擇或是限制。

- 求，結合環境變數定義查詢的資料項目。
3. 個體 Y 依據個人需求，透過 P-Grid 資訊分享網路發出特定資訊項目的資訊請求。
 4. (a) P-Grid 平台依據請求者的查詢，搜尋資訊分享網路中符合請求者要求的資訊項目。
(b) P-Grid 平台依據請求者的查詢，回覆目前資訊分享網路中符合請求者的所有資訊分享者的項目清單。
 5. (a) 個體 Y 存取 categoryOntology.xml 中的類別權重，並依據此類別權重重新排列資訊項目的順序。
(b) 個體 Y 得到由類別權重排出的資訊項目清單後，選擇喜好的資訊項目，並向個體 X 發出資訊下載的請求。
(c) 依據個體 Y 的請求，P-Grid 平台將下載的訊息傳遞到個體 X 身上，由個體 X 的訊息接收器接收訊息。
 6. 依據個體 Y 選擇的喜好順序，經由 Preference Agent 學習使用者的資訊選擇喜好，並寫入 categoryOntology.xml 修改其類別權重。
 7. (a) 在個體 X 的訊息接收器個體 Y 收到訊息後，取得系統與個人設定等系統變數。
(b) 啟動情境代理人，蒐集情境因子¹²與網路環境因子¹³等參數值。
 8. 依據步驟 7、8 所蒐集的參數值與變數，選擇資訊分享的策略。
 9. 個體 X 依據所計算出來的資訊分享策略，計算請求者的資訊貢獻度。
 10. 依據資訊貢獻度的計算結果，回覆給請求者 Y，是否進行資訊分享。是的話完成資訊分享的請求，反之則否。
 11. 最後，將資訊分享的交易經驗紀錄到 profile 檔之中，並結束交易。

12 系統使用的情境因子屬性如下述項目。此處的情境因子指的是了解分享者當下的資訊分享策略，如是否有地點分享、好友分享、拒絕分享、對等分享等情境限制，以求在分享的過程中更充分的表達分享者當下的態度與想法。

13 系統使用的網路因子屬性如下述項目，此處的網路因子指的是了解分享者對於流量控制的選擇或是限制。