

# 整合社群網路事件以進行多元化知識管理應用之研究

曾守正

國立高雄第一科技大學

[imfrank@ccms.nkfust.edu.tw](mailto:imfrank@ccms.nkfust.edu.tw)

江翰榕

國立高雄第一科技大學

[u9924831@ccms.nkfust.edu.tw](mailto:u9924831@ccms.nkfust.edu.tw)

## 摘要

目前有許多研究是針對運用社群網站的龐大資料去做相關的應用和服務，其中大多數的研究是運用該社群的本地社群資料去做分析和設計。但是這些方法並不適用於即時回饋性質的查詢服務和相關應用。本研究主要目的是要針對符合搜尋條件的即時性異質性社群網站使用者，藉由群聚或者配對的方式，達到快速吻合目標的服務，進而應用於更優質的服務推薦。例如，即時性的找尋某商城中團購人數或者是共乘服務的概念。其概念是藉由降低商品價格來提高購買意願，進而讓雙方取得互惠的目標。

關鍵詞：資料串流、社群網路、推薦服務

# 整合社群網路事件以進行多元化網路團購之研究

## 壹、緒論 (Introduction)

社群網路服務 (Social Networking Service) 是目前網路應用上的新興領域，並且逐漸成為眾多主流的應用服務之一，而參與線上社群的使用者也隨著科技產品的發展而日漸增加。從較早期的論壇、部落格到近期的 YouTube<sup>1</sup>、Flickr<sup>2</sup> 以及 Google+<sup>3</sup> 等多媒體社群網站，無非是希望能提供使用者在友誼圈內去進行分享與互動等服務，其所產生的資料和訊息具有即時的特性，而且隨著時間的流逝將會累積龐大的資料。有效運用這些資料往往能提升各種領域的應用與分析，而這正是吸引許多研究人員投入此議題的一大誘因。

在各種社群網站上的使用者藉由發表文章、主題的跟隨、應用軟體的使用或是朋友間訊息的互動等，產生出彼此間互動的關聯性和資料訊息，這些訊息會經由科技產品 (例如：智慧型手機、電腦等) 以類似資料流 (Data Stream) 的型態，即時傳遞到系統和使用者手上。因此，更進一步整合異質性 (Heterogeneous) 社群網站的資料，將會是下一階段必須要進行的重點，因為透過資料的互補，將會產生更多的應用價值，但是其所產生的資料流也將會相當龐大。

社群網站所產生的資料流與傳統資料串流的資料型態並不盡然相同。傳統資料串流所產生的資料 (例如：股市行情、網路傳輸量) 大多是跟隨著時間，持續不斷產生龐大的資料流，其資料大多以數值類型呈現；而社群網路所產生的事件會依照時間的累積而產生多種屬性的事件資料 (例如：地點、姓名、關連等)，其類型可能包含數值、文字、影像，或甚至多媒體等資料類型。

本論文主要的研究目的是在於整合異質性社群網站的資料，並彙整社群來源的使用者與所產生的事件資訊，將其資訊轉換成統一規格的事件串流 (Event Stream)，利用結構化查詢的方式快速提取符合某些服務條件的使用者或分享資訊，並且藉由服務篩選的門檻來群聚符合條件的使用者、資料或文件，以達到社群知識整合與管理的應用目的。

本研究所提出的社群事件串流 (Social Event Stream) 概念如圖 1 所示，此架構乃是修改自資料串流 (Data Stream) 的概念，將異質性社群網路服務平台上的使用者經由應用程式、分享和發表等動作產生的事件訊息進行彙整，然後將取得的資訊 (例如，時間、貨幣等可能會因為地區的不同而有部分差異的整合元素)，轉換成相同的資料規格並且以一致的格式輸出成資料串流的型態，再依照某些屬性去查詢串流內部的事件並且群聚，以達到優化分享資訊，進而提昇使用者社群之間達成分享與互惠的相關應用。最後我們將會展示兩種不同類型的實例來呈現此架構的可行性。

---

<sup>1</sup> [http:// www.youtube.com/](http://www.youtube.com/)

<sup>2</sup> [http:// www.flickr.com/](http://www.flickr.com/)

<sup>3</sup> [http:// www.plus.google.com/](http://www.plus.google.com/)

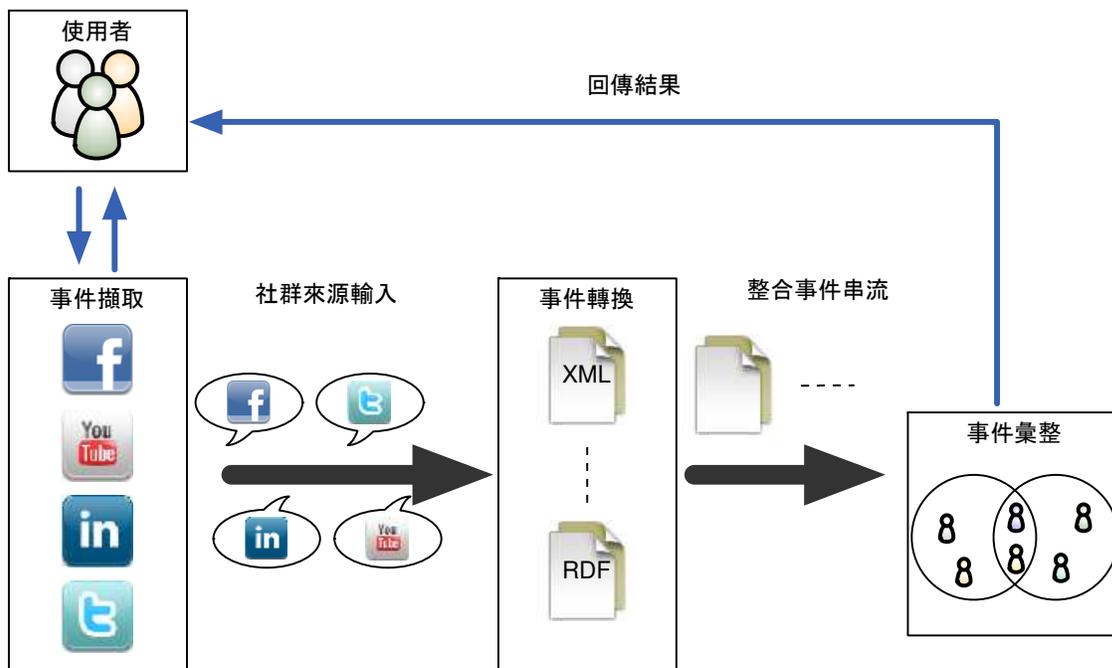


圖 1 社群事件彙整概念圖

## 貳、文獻探討

### 一、 社群網路服務 (Social Network Services)

社群網路服務主要是以網際網路為基礎，透過「社群網站」(Social Network) 提供具有相同興趣或是參與共同活動的使用者進行互相聯繫與交流的網路服務。這些網站通常透過周遭認識的人建立各自的社群，進而拓展成類似樹狀的架構。

依照 Anderson *et al.* (2010) 的研究，社群網站依照其應用可以分成以下兩種類型：

1. 普遍線上社群網站 (Popular Online Social Networks)，像是 Facebook<sup>4</sup>、Twitter<sup>5</sup> 等；
2. 廣泛收集協作工具 (Widely-Adopted Collaborative tools) 網站，像是 CiteULike<sup>6</sup> 和 Delicious<sup>7</sup> 等網站。

這些網站藉由標記、註解、應用程式、分享和文章的組織等，提供相當豐富的資訊 (Silberstein *et al.* 2011)。除此之外衛星定位系統 (GPS) 和藍牙功能 (Bluetooth-capable) 等行動裝置的附加功能，能夠產生不同層級的社群資料數據。

社群資訊可以被應用於各種不同面向的優化實作上，像是針對使用者的偏好推斷、調查使用者間的信賴程度 (Mislove *et al.* 2008) 等。例如，當某百貨發生災難時，可以提供使用者與其朋友或家人適當的共同逃生路線規劃；在可信賴程度之內找尋可以共乘的對象；分析與某議題具有相關性的組織或結構 (例如：恐怖犯罪組織)；或者提供彼此信賴的使用者之間的資源分享動作等應用。

<sup>4</sup> [http:// www.facebook.com/](http://www.facebook.com/)

<sup>5</sup> [http:// www.twitter.com/](http://www.twitter.com/)

<sup>6</sup> [http:// www.citeulike.com/](http://www.citeulike.com/)

<sup>7</sup> [http:// www.delicious.com/](http://www.delicious.com/)

在這些應用中，某些部份是利用使用者自身的社群文件或者是從另一個外部的社群作為資訊來源，去收集所需的文件資訊。但是只使用少數來源的社群資訊（像單純以 E-mail 溝通，或是單一社群網站的資料）去進行分析和推薦等應用時，往往會受限於其範圍與連結的對象，而產生資訊不足的狀況，而對使用者有所誤導。

目前已經有許多研究是結合不同社群網站的資料，以便讓社群資訊能夠更完整，進而設計或輔助具有社群意識的應用和服務。例如，Anderson *et al.* (2010) 結合多個社群資源的社群資訊來架構有向性和有權重值的社群多邊圖，去運用在較新穎且具有社群意識的應用和服務上面；而 Fard & Ester (2009) 提出一個應用聯合探勘和統計方法的架構自動從多重社群網路的資料分析和推論方法來轉換資料組。然而，由於異質性社群資料的整合必須是即時性的資料，過去在異質性資料庫上對大量資料所進行的靜態異質性整合運算（曾守正 & 周韻寰 2010；第十五章），雖然並不完全適合運用在即時性的動態服務應用上，但是我們認為加以修正與調整應該還是有某些可行的空間。

因此，本研究將以不同的社群服務做為資料的來源，經由應用程式或者是實體裝置將使用者資料擷取下來，再彙整各個相異的來源資料並且轉換成相同格式輸出，使得系統能夠藉由更豐富的資訊來優化或設計新穎的即時性社群應用服務。這些使用者即時提供的豐富事件資訊，會因為時間的進行和使用者的增加，產生具有即時性、大量性、持續性等類似資料串流的事件流，因此我們將在下一節針對資料串流的概念以及特性做進一步的探討，以便了解整合過程所面臨的問題。

## 二、 資料串流 (Data Stream)

社群網路每天所產生的大量資料可以看成是文字混和數字與多媒體的串流資料 (Streaming Data)。以往的串流資料研究大多著重在數值類型的資料串流管理，例如：網路的測量數據、電話通聯紀錄、網站的探訪資訊、地質監控數據等。近年來由於網際網路的應用廣泛，讓文字混和數字與多媒體的串流資料研究，也慢慢受到重視，像：股市、匯率等金融方面的應用、網際網路的安全問題監控、通訊資料管理等。例如 *Traderbot*<sup>8</sup> 即為一建構於全球資訊網上的財金搜尋引擎，可提供即時的財經新聞分析以預測未來的股價變化趨勢，並可透過即時的金融串流資料做評估查詢，像股市行情和新聞提要等。而 *iPolicy Networks*<sup>9</sup> 則提出了一個整合安全性平台，來提供服務網路封包串流的防火牆支援和入侵偵測的服務。另外還有大型網站監控網頁記錄的相關應用，例如：*Yahoo!*<sup>10</sup> 或 *Google*<sup>11</sup> 等搜尋引擎，往往需要搭配許多分散式點擊串流的資料分析，去尋找負載比較大的網頁存取並且做監控的動作，或是結合感測器監控 (Carney *et al.* 2002；Madden & Franklin 2002) 串流資料，以進行監控和分析等具有多樣性的，且隨著時間而有所變化的應用情況。

---

<sup>8</sup> [http:// www.tradebit.com/](http://www.tradebit.com/)

<sup>9</sup> [http:// www.ipolicynetworks.com/](http://www.ipolicynetworks.com/)

<sup>10</sup> [http:// www.yahoo.com/](http://www.yahoo.com/)

<sup>11</sup> <http://www.google.com>

資料串流的管理模式畢竟不同於傳統的資料管理模式，因此 Babcock *et al.* (2002) 提出下列差異來做比較：

1. 串流資料的輸入非常快速。
2. 資料的異動非常頻繁。
3. 資料元素在串流中傳遞抵達是即時線上運作 (On-Line Operations) 的，而非線外運作 (Off-Line Operations)。
4. 無論在資料串流內或者是整個資料串流，系統無法針對資料抵達的順序做處理。
5. 資料串流有可能是沒有大小限制的，也就是說資料量可能會非常的龐大。
6. 任何資料元素只能在資料串流中做一次性的丟棄或者儲存的處理動作，這些元素無法輕易的被重新檢索。

而社群網路所產生的資料串流也同樣會具有上列的某些相同性質 (例如：1~3 與 5)。每當事件產生時，往往會包含具有時間、人、地點、文件等不同的互動元素，這些元素之間會具有某種關聯性存在 (例如：十五號早上 (時間) 小明 (人) 在學校實驗室 (地點) 發佈了一篇關於美食的文章 (文件))，通常這些關聯性會以複雜度比較低的程度被使用者或者是分析者檢視 (例如：美食相關的文章、目前在學校的人等的分類群組)，但是大多數分析者會有興趣的是透過多維度層級的關係所表現出來的重要資訊 (例如：商機訊息、潮流資訊等)，如果要針對這些較高層級的資訊特性去做探討，就會需要利用執行即時線上的、多重關係層級的、多方面維度串流資料等的資料分析程序，以便讓資訊分析人員能夠有效取得所需要的相關資訊。

#### (一) 資料串流管理系統

Han & Kamber (2000) 將資料串流查詢程序的主要架構分成三個部分：目的用戶、查詢處理器和暫存空間。目的用戶對資料串流管理系統提出查詢的動作，查詢處理器接收到查詢要求之後在暫存空間執行資訊的運算，然後回傳結果給目的用戶。

查詢程序可以區分成一次性查詢 (One-Time Queries) 和連續性查詢 (Continuous Queries)(Terry *et al.* 1992)。一次性查詢是在某一個時間點快照的資料值組做一次的評估，然後將結果回傳給使用者；而連續性查詢則是針對不斷抵達的資料串流做持續性的查詢，連續性的查詢會依據時間回傳結果，從另一方面來看，連續性查詢可以看成是監控程序。而針對需要即時找尋社群網路上符合條件的使用者而言，我們能夠針對某些屬性經由持續性查詢去做事件串流監控的動作，以此來解決無法即時處理龐大異質性社群資料的問題。

此外，查詢可以是事先定義的查詢 (Pre-defined Queries) 或者特定的查詢 (Ad Hoc Queries)，事先定義查詢主要支援在任何相關的資料到達之前的資料串流管理系統，通常事先定義查詢是連續性查詢。特殊查詢是在資料串流已經開始之後，由線上發出的查詢，它可以是一次性查詢或者是連續性查詢。因為並無法事先知道查詢最佳化的目的，也無法辨別整個查詢的共用表達方式，使得特殊查詢會讓資料串流系統的設計複雜化。

#### (二) 資料串流的查詢

由於資料串流有可能是沒有限制的，對於要準確的查詢資料串流內的資訊而言，所需求的大量儲存空間也會呈現無限制的成長，因此在資料串流模組的運算中執行查詢程

序也是一項挑戰。儘管有些演算法 (Vitter 1999) 針對解決資料量大於主要記憶體的研究，但是這些演算法的即時反應太慢，而且無法支援連續查詢，因此不太適合用於資料串流的應用。

在這些問題方面，Arasu *et al.* (2004) 提出了一些初步的步驟來區分查詢的類型，一種是可以利用有限制的記憶體做精準的查詢回覆，另一種則是做近似值的查詢。而結論表示在不清楚輸入資料串流的大小情況之下，不可能配置有限的記憶體去做普遍的查詢 (包括聯結)，除非查詢的屬性範圍是有被限制住的。其中滑動窗框 (Sliding Windows)(Dani & Getta 2005 ; Jiao 2006 ; Zhu & Shasha 2002) 的技術是透過目前串流產生的資料去滑動窗框，而不是透過所有串流的歷史資料來做查詢。

我們將運用滑動窗框的概念解決事件串流中龐大而且即時的特性。運用事件窗框的概念套用到事件串流之中，將符合條件的事件使用者從事件串流中提取到窗框內，藉由設定窗框的大小 ( $n$  個消費者) 或者是時效性 (超過一個小時的需求者) 來掌控有限資源的運用。

本研究主要目的是運用異質性社群來源提供龐大的資料和使用者，藉由社群資源彙整將來源的資料轉換成事件串流的型態，在運用連續查詢動作監控事件串流內符合某特定屬性條件的事件，然後以滑動窗框的概念將其彙整，進而達到即時性團購意識的目的。下一章節將針對本次研究的主要概念架構作詳細討論。

## 參、研究架構

### 一、系統架構定義

本研究主要目的是要針對搜尋異質性社群網站的使用者，藉由提高來源的數量，進而快速找尋到具有類似需求的社群使用者，以達到即時群聚的目的，使得能進一步取得更優質的服務應用。例如，即時性的找尋某商城顧客之中有購買牛仔褲需求的消費者，將人數湊齊到滿足某店家牛仔褲產品的優惠數量等，然後藉由找尋相似性團體來達到優惠門檻。

為了針對某些條件元素所衍生的社群資料即時處理模式，本研究將以 Anita & Janusz (2005) 所提出的資料串流模組作為基礎，進而修改成為本研究的社群網路事件串流運算模組。

首先，概念的定義低階事件運算中的窗框模組，將滑動窗框定義為一個從社群網站所獲得的事件值組  $X$ ， $X = \langle x_1, x_2, x_3, \dots, x_n \rangle$ ，而事件從社群網站時間戳記為  $t_i$  所取得的值組為  $\langle x_i, t_i \rangle$ 。在模組當中如果有新的事件資料輸入的時候，將會觸發每個低階運算的動作，我們以低階運算的屬性改變類型來分類運算模式，其分類如下：

#### 1. 改變維度的運算：

事件過濾 (Filter) 運算：是將向量中的事件  $n$  依照使用者或管理者需求，將事件元素分割成有限數量的子事件組，如圖 2(a) 表示過濾器 (Filter) 允許複製向量中的元素到數個內部的子事件組。例如， $x_j, \dots, x_n$  要取消購買牛仔褲的條件，或者是小明提出購買需求的時間已經超出設定時效等，在此以過濾運算將這些資料從滑動窗框中移除。

事件合併 (Merge) 運算：將兩個向量重新建構成一個新的事件向量，其維度等於是加總所輸入的所有維度，如圖 2(b) 表示將兩個事件向量做合併的運算。我們可以假設將第二個向量附加到第一向量之中，這表示合併運算是無法替代的，此外合併運算也無法排除掉重複的部份。例如， $x_3$ 、 $x_4$  為新的需求者與舊有的需求者  $x_1$ 、 $x_2$  群聚成為新的群組。

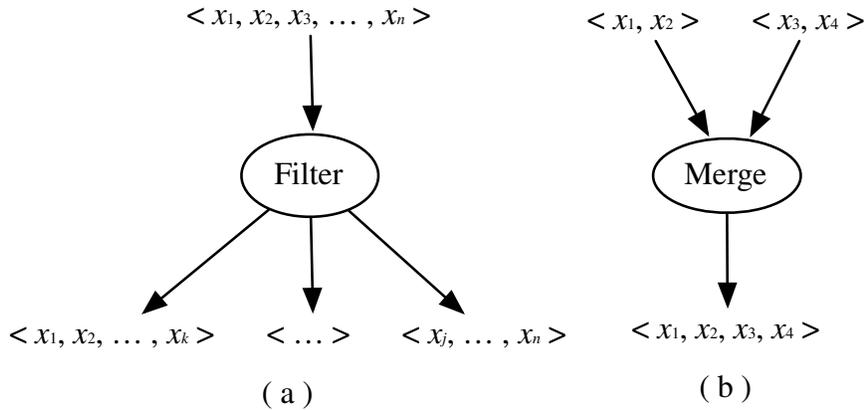


圖 2 過濾運算元 & 合併運算元

### 2. 改變數值的運算：

轉換 (Transition) 運算： $f_i: \mathfrak{R} \rightarrow \mathfrak{R}$  為一些基本的程式，主要是將各個取得的來源資料，去執行各種度量衡單位的轉換，例如溫度、貨幣換算等的元素。 $\forall i=1,2,\dots,n$ ，一般來說  $f_i \neq f_j$ 。轉換運算執行一個運算事件向量  $n$  時亦會形成一個方程式向量值組  $n$  像是  $\langle f_1, f_2, f_3, \dots, f_n \rangle$ 。上述的定義是滿足在當轉換向量維度大於等於窗框向量的維度之下。例如，將事件資料內的時間或者金錢等度量衡單位統一轉換算成固定的單位。

收集 (Aggregate) 運算：此運算主要是要找到或者是搜尋到一個事件或者是一組事件來滿足所給的條件。例如，從事件組  $x_1, \dots, x_n$  當中找尋符合系統需求的事件（或者消費者） $x_{k1}, \dots, x_{kn}$ 。

累加器 (Accumulator)：累加器運算  $V = \oplus x_i$ ，其  $\oplus$  為一個在所有事件  $x_i$  上的基本運算，其運算為從所給的事件向量  $n$  中輸出單一個數值。例如，彙總符合某條件的事件數量到同一群組之中。

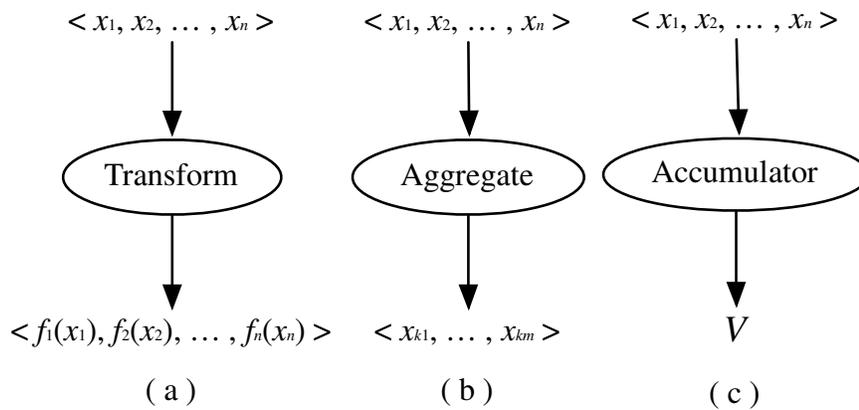


圖 3 轉換、收集和累加運算元

### 3. 改變排列位置的運算：

排列 (Permutation) 運算：此運算執行時會從不同來源輸入的事件經由此運算來改變事件在向量內的序列位置。例如， $x_{i1}, \dots, x_{in}$  為多個社群來源所擷取到的資料，藉由排列運算將資料依照時間順序重新排列。

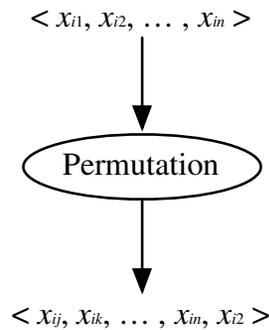


圖 4 排列運算元

#### 4. 串流運行的計算需求：

讀取 (Reader) 運算：假設 Stream 為原始社群資料值組或者新的事件的產生。則讀取運算即是從 Stream 中讀取新事件做為支流的輸出，讀取新的事件將觸發窗框的滑動以及接續的運算。

寫入 (Writer) 運算：寫入運算主要需求是為了儲存或者暫存狀態的結果，然後在滑動窗框應用中重複使用。其寫入可以區分為寫入永久儲存空間如圖 5a 所表示，將資料放置在本機或者雲端等空間來提供給後續分析或者執行；以及寫入至暫存空間如圖 5b，以提供後續即時所需的運算。

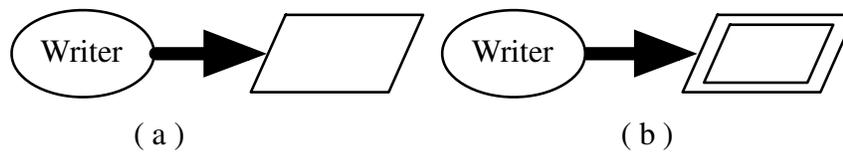


圖 5 寫入運算元

## 二、 系統流程

依照上述所提建立的數種運算元執行方式，進而提出本研究所架構的社群網路事件串流處理程序如圖 6 所表示。藉由社群網站的使用者允許應用程式存取的动作來取得所需的資訊，例如，使用者個人資料、使用者在該社群內與其他使用者的關聯資訊、使用者所產生的事件資訊等等，(在此不討論資料取得權限的問題和技術) 使得只要有使用者產生相關事件的動作將會觸發應用程式的運行。其系統的運作情況如下列步驟：

1. 最先開始系統會將資料從各個異質性的來源位置取回 (Stream)，其取得方式可能會運用社群網站的應用服務、行動裝置的應用程式或者是網頁等方式取得。
2. 針對取得的事件資訊依照所對應的原始檔案格式進行轉換 (Transform) 的动作，其轉換內容會包含資料原始格式的統一轉換 (例如，統一轉換成 XML 格式) 或者針對內部所需換算的度量衡單位的換算 (例如，各國的年份格式、各國國家幣值的匯率、溫度單位等等度量衡單位的轉換)。

3. 將處理過後的資料寫入到儲存空間 (Writer) 以提供其後續分析應用去使用，其儲存位置可以為本地端的資料庫或者是雲端位置等等。
4. 將資料寫入的同時將不同來源所收集到的資料依照時間的維度進行彙整並且重組 (Permutation)(由於相異來源輸入的時間排列順序會有差異)，最後依照時間維度排列依序輸出結果。
5. 把前面步驟所產生的結果，藉由連續性的查詢來監控事件串流，並且將符合條件的事件運用收集運算元 (Aggregate) 集合起來做為輸出。
6. 然後藉由事件過濾運算元 (Filter) 將已經超過時效性或者是不具有該查詢資格的使用者做過濾。例如，以查詢方式將兩三個小時以前有意願購買該產品的消費者或者是已經不再有需求該服務的使用者移除掉。
7. 接著將符合條件要素需求的事件或者使用者以累加運算 (Accumulator) 彙整起來，並且將彙整後的資料儲存到暫存空間之中以利於後續降低資料轉換的運算成本。
8. 同時將收集後的資料轉換成所要呈現的資料格式並且傳送至使用者端，然後經由應用程式或者是網頁呈現其服務給予使用者。
9. 往後的使用者產生事件行為之後，系統會將該事件單獨進行前述的轉換運算，再將轉換後的資料寫入至儲存空間以更新分析所需的資訊。
10. 接著將轉換後的事件資料結合 (Merge) 步驟 7 所儲存在暫存區的彙整事件，即更新完整個事件窗框成為一個新的群組。
11. 最後將依序前面 4、5、6、7、8 的流程將最新的訊息傳送至使用者端，以達到即時更新的服務。

藉由異質性社群網站服務的使用者來提供資料和其服務需求，然後經由系統做初步的即時性查詢分析，將符合企業或者是提供者條件的相異社群服務使用者，即時的回饋使用者的需求。以此概念來降低個人花費成本來提高交易達成率，進而提升企業與消費者之間的互利情形來達到本次研究架構的主要目的。

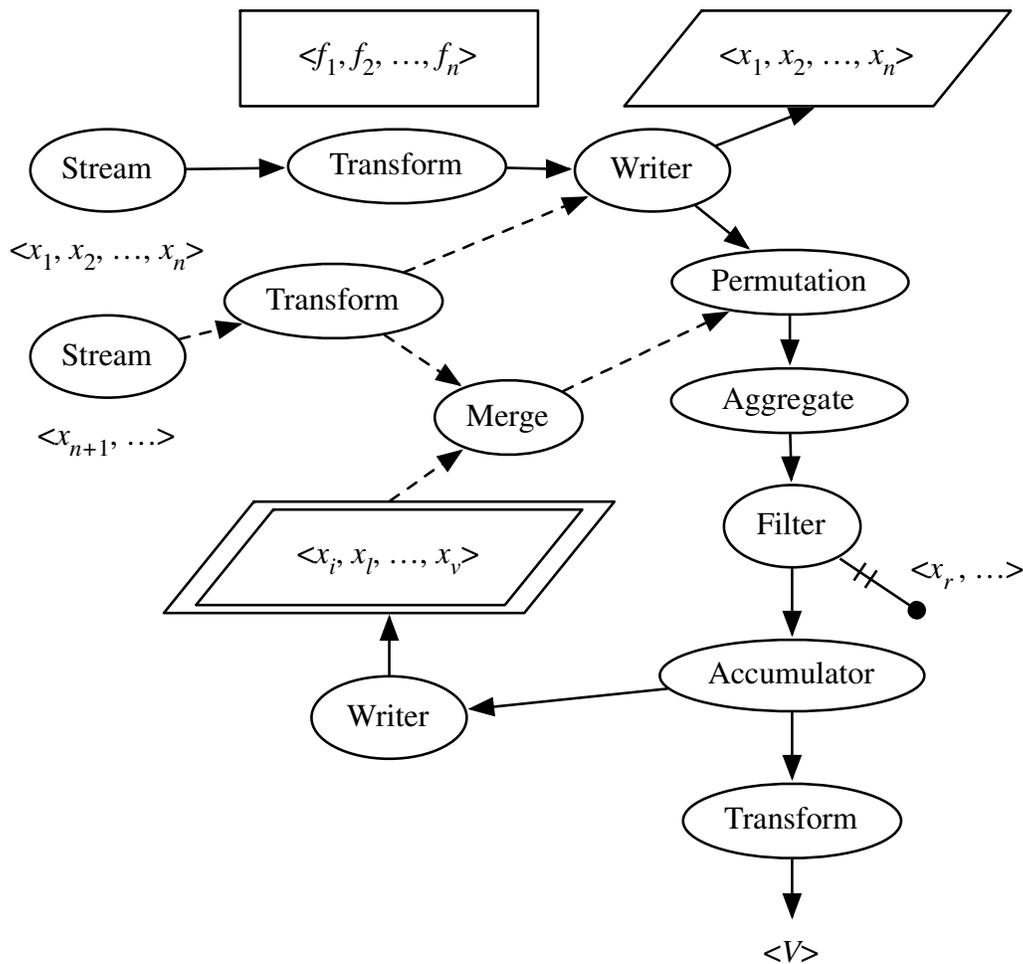


圖 6 系統運算流程圖

#### 肆、實際應用舉例

以下將針對此研究所建立之模組提出未來在社群網路與實際生活上可以應用的範例，以驗證其可行性。

##### 一、 實際範例一：即時性微團購

首先所要提出的應用範例是針對在各大百貨業者、特惠賣場或者是展覽場等環境為基礎，由商場店家或者企業這些提供者所推出類似團購促銷性質的優惠服務(例如，H牌牛仔褲一條 250 三條 599)作為購買條件，而對於大多消費者可能會有購買該商品(牛仔褲)的需求但是卻沒有大量購買的需求，導致大幅降低購買的意願。因此運用我們所建構的模式來提出即時性為團購的概念來解決上述問題。

首先商場(百貨)的消費者會經由習慣的社群網站(Google Plus、Facebook)所使用的應用程式即時的提出購買需求，藉由消費者的需求事件來觸發本系統的流程運作；

1. 最先將社群擷取的事件經由轉換運算將事件內的度量衡單位(貨幣、時間等)進行統一格式的換算；

2. 將新的事件寫入到儲存空間，同時結合暫存空間中所儲存的舊有事件，使其成為新的窗框資訊；
3. 然後將轉換後的事件窗框寫入到儲存空間做備份或者提供給其他後續應用；
4. 同時將轉換之後的事件組利用排列運算依照時間維度重新組合序列；
5. 接著透過查詢把具有購買 H 牌牛仔褲的使用者收集作為輸出結果；
6. 然後經由過濾運算元將已經超過系統設定延遲時間的消費者或者是取消購買 H 牌牛仔褲的消費者做移除；
7. 再來先將過濾之後新的事件組儲存回暫存空間去，再經由累加運算把符合門檻的消費者（三位有購賣一條牛仔褲的消費者）叢聚為一個購買群組；
8. 最後將購買資訊反轉換成這三個購買群組內的消費者所使用的社群環境格式，將推薦訊息或團體購買確認資訊呈現給該消費者。

經由系統使這些消費者們能夠快速的找尋到可共同購買該商品的其他消費者。藉由找尋群聚相同購買意願的消費者以提高購買意願，促進產品成交機率，以達到消費者取得商品優惠的價格，同時業者亦能提高銷售的業績，呈現雙方互惠的主要目的。

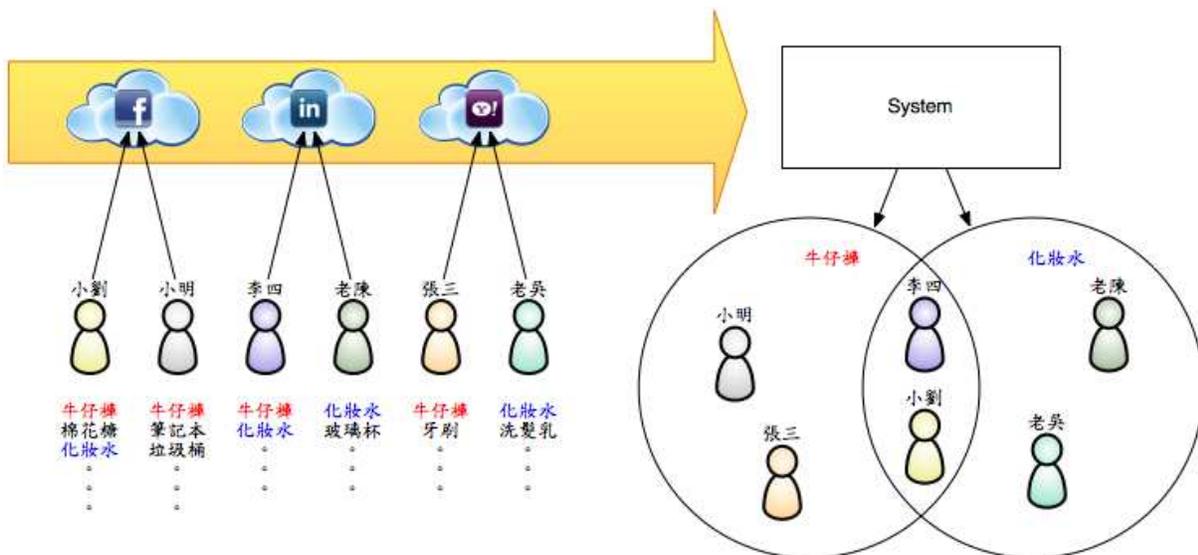


圖 7 範例一

如圖七中所表示，消費者藉由不同的社群網站提出需求，接著系統從所有異質性的社群網站擷取資訊並且做查詢收集，最後系統依照符合條件的消費者將其適合的資訊串送回消費者所使用的社群來源應用程式並且呈現給使用者相關資訊。

## 二、 實際範例二：共乘概念

第二個能夠實際所應用本研究概念的環境實例是結合地理資訊的共乘應用。在這一部分主要是運用系統內部針對各個不同社群資源位置所取得的關聯架構圖，並且針對使用者彼此之間的互動去計算關連的信賴權重值，然後系統可以利用使用者在各個社群來源的關連性，整合成更完整的社群關係圖並且附加上不同的權重值屬性，利用這些來建構出更完整而且有附加權重屬性的社群多向關係圖，以此能夠達到更完整、更準確的去輔助相關的應用程式。在最初，系統會從異質性使用者社群來源擷取關聯圖，運用現今計算個體間信賴程度的相關計算方式預先將彼此之間的信賴程度計算出來，以及整合各

個不同社群的使用者關係來建構出更完整的使用者關係圖，並且以此為基礎去輔助吉食性共乘服務的概念。

1. 一開始使用者可以藉由行動裝置的應用程式或者電腦網站等方式，將共乘的需求資訊（搭車時間、起始地點、金額額度等）遞送給系統並且觸發系統的運算；
2. 當事件產生先將事件內的資訊經由轉換運算將事件內的度量衡單位（貨幣、時間等）進行統一格式的換算；
3. 將轉換之後的事件寫入到儲存空間去做資訊更新的動作，同時結合暫存空間裡的舊有事件群組，成為新的一組窗框群組資料；
4. 接著將結合後的資料組依照時間的維度重新排列，使其窗框內的資料按照時間做排序輸出；
5. 然後以連續性查詢的方式監控窗框內事件，將有共乘服務需求的使用者收集並且輸出；
6. 運用過濾運算元將放棄共乘服務或者是超出設定時間的消費者從窗框內移除；
7. 利用系統預先計算的信賴程度和整合後的完整社群關聯圖來將具有相同性質的共乘消費者做分群（例如，共同目的地、相同的嗜好等）；
8. 最後將分群後的推薦資訊反轉繪成消費者的原來格式，並且展現給消費者。

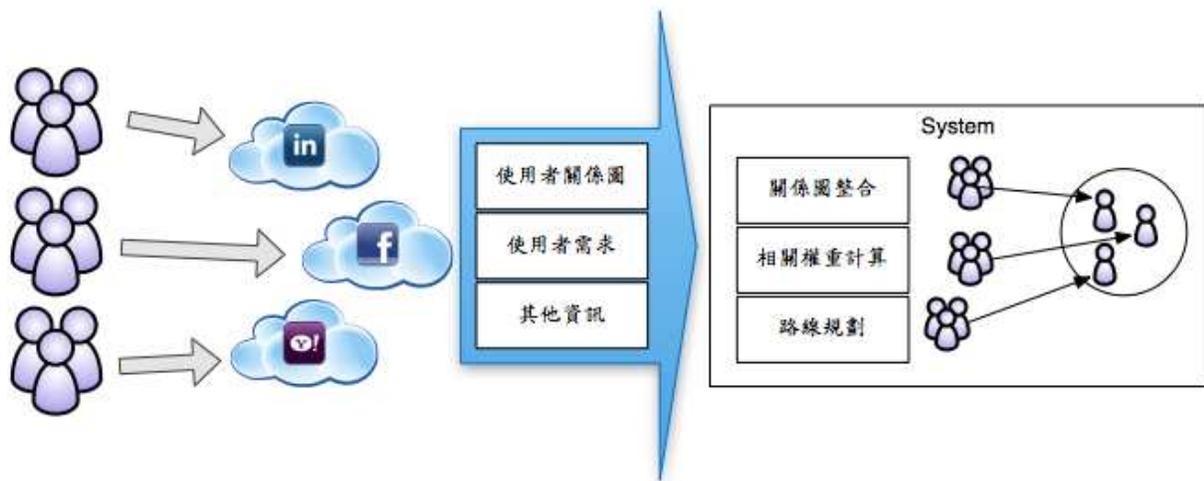


圖 8 範例二

如圖 8 所表示為其實際可以發展的應用範例，首先消費者將提出所需的乘車資訊提出請求，接著系統將針對收集使用者在異質性社群服務的相關資訊做彙整（例如，興趣、喜好、關係圖等等），然後經由眾多消費者所提出的請求依照時效性進行運算配對，最後將符合需求的消費者進行群聚，然後發送所規劃的共乘路徑和共乘人員，甚至於提出額外的服務推薦給予消費者。

## 伍、結論與未來研究

本論文的研究主要是要建構一個即時性結合異質性社群資料處理的概念，將彙整不同社群來源的消費者去整合各個單獨社群網站零散且具有共同需求的消費者問題，利用

異質性社群來源取得更豐富的資訊以及更大量的使用者，以此提高找尋到相同消費意願消費者的機率，進而達到以某目的為基礎的微團購群組叢聚，一方面能降低消費者購買所需的成本，另一方面能提高賣方成交的機會，藉此得到買賣雙方皆互惠的成果，來達到本次研究的目的。這概念可應用在輔助各種具有某些類似因素而有叢聚需求的環境下去做規劃設計，如第四章節所提出的應用實例。

在未來的研究方面可以進一步延伸到針對異質性社群網站所產生的龐大文件，即時的去解析字句內的關鍵字詞或者是語句正負向性質等等的值組，使的能夠附加到事件或者是使用者彼此之間相關權重數值的標記；另一方面可以針對所擷取的龐大異質性社群資源做預先的處理，將龐大且即時的社群資料流所形成的事件串流事先依照不同維度、階層關聯等，將整個事件串流依照社群資料倉儲的結構索引，自動將資料放到所屬的集合或群組中，以便後續的查詢能快速找到所要的資料集，甚至整個社群資料倉儲也能夠自動調整其維度結構。

### 參考文獻

1. 曾守正 & 周韻寰 (2010)。資料庫管理系統理論與實務 (SQL Server 2008 版)。台北市：華泰文化。
2. Citeulike home page. (<http://www.citeulike.org/>)
3. Delicious home page. (<http://delicious.com/>)
4. Facebook home page. (<http://www.facebook.com/>)
5. Ipolicy networks home page. (<http://www.ipolicynetworks.com/>)
6. Linkedin home page. (<http://linkedin.com/>)
7. Traderbot home page. (<http://www.tradebit.com/>)
8. Twitter home page. (<http://twitter.com/>)
9. Yahoo home page. (<http://yahoo.com/>)
10. YouTube home page. (<http://youtube.com/>)
11. Alon, N., Matias, Y. and Szegedy M. “The space complexity of approximating the frequency moments.” *J. Comput. Syst. Sci.*, 58:137--147, February 1999.
12. Anderson, P., Kourtellis, N., Finnis, J. and Iamnitchi, A. “On managing social data for enabling socially-aware applications and services.” In *Proceedings of the 3rd Workshop on Social Network Systems*, SNS '10, pages 7:1--7:6, New York, NY, USA, 2010. ACM.
13. Arasu, A., Babcock, B., Babu, Shivnath., McAlister, J. and Widom, J. “Characterizing memory requirements for queries over continuous data streams.” *ACM Trans. Database Syst.*, 29:162--194, March 2004.
14. Babcock, B., Babu, S., Datar, M., Motwani, R. and Widom, J. “Models and issues in data stream systems.” In *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, PODS '02, pages 1--16, New York, NY, USA, 2002. ACM.
15. Babcock, B., Datar, M. and Motwani, R. “Sampling from a moving window over streaming data.” In *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '02, pages 633--634, Philadelphia, PA, USA, 2002. Society for Industrial and Applied Mathematics.

16. Carney, D., Çetintemel, U., Cherniack, M., Convey, C., Lee, S., Seidman, G., Stonebraker, M., Tatbul, N. and Zdonik, S. "Monitoring streams: a new class of data management applications." In *Proceedings of the 28th international conference on Very Large Data Bases*, VLDB '02, pages 215--226. VLDB Endowment, 2002.
17. Chaudhuri, S. and Motwani, R. "On sampling and relational operators." *IEEE Data Eng. Bull.*, 22(4):41--46, 1999.
18. Chaudhuri, S., Motwani, R. and Narasayya, V. "On random sampling over joins." In *Proceedings of the 1999 ACM SIGMOD international conference on Management of data*, SIGMOD '99, pages 263--274, New York, NY, USA, 1999. ACM.
19. Dani, A. and Getta, J. "Conceptual modelling of computations on data streams." In *Proceedings of the 2nd Asia-Pacific conference on Conceptual modelling - Volume 43*, APCCM '05, pages 43--47, Darlinghurst, Australia, Australia, 2005. Australian Computer Society, Inc.
20. Fan, W., Watanabe, T. and Asakura, K. "A framework for flexible clustering of multiple evolving data streams." *Int. J. Adv. Intell. Paradigms*, 1:178--195, April 2008.
21. Fang, M., Shivakumar, N., Garcia-Molina, H., Motwani, R. and Ullman, J. D. "Computing iceberg queries efficiently." In *Proceedings of the 24rd International Conference on Very Large Data Bases*, VLDB '98, pages 299--310, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.
22. Greenwald, M. and Khanna, S. "Space-efficient online computation of quantile summaries." In *Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, SIGMOD '01, pages 58--66, New York, NY, USA, 2001. ACM.
23. Guha, S., Koudas, N. and Shim, K. "Data-streams and histograms." In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, STOC '01, pages 471--475, New York, NY, USA, 2001. ACM.
24. Han, J. and Kamber, M. *Data mining: concepts and techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2000.
25. Jiao, Y. "Maintaining stream statistics over multiscale sliding windows." *ACM Trans. Database Syst.*, 31:1305--1334, December 2006.
26. Koudas, N. and Srivastava, D. "Data stream query processing." In *Web Information Systems Engineering, 2003. WISE 2003. Proceedings of the Fourth International Conference on*, page 374, dec. 2003.
27. Madden, S. and Franklin, M. J. "Fjording the stream: An architecture for queries over streaming sensor data." In *Proceedings of the 18th International Conference on Data Engineering*, ICDE '02, pages 555--568, Washington, DC, USA, 2002. IEEE Computer Society.
28. Matias, Y., Vitter, J. S. and Wang, M. "Wavelet-based histograms for selectivity estimation." In *Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, SIGMOD '98, pages 448--459, New York, NY, USA, 1998. ACM.
29. Mislove, A., Post, A., Druschel, P. and Gummadi, K. P. "Ostra: leveraging trust to thwart unwanted communication." In *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*, NSDI '08, pages 15--30, Berkeley, CA, USA, 2008. USENIX Association.
30. Motwani, R. and Raghavan, P. "Randomized algorithms." *ACM Comput. Surv.*, 28:33--37, March 1996.

31. Raman, V., Raman, B. and Hellerstein, J. M. "Online dynamic reordering for interactive data processing." In *Proceedings of the 25th International Conference on Very Large Data Bases, VLDB '99*, pages 709-720, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc.
32. Seshadri, P., Livny, M. and Ramakrishnan, R. "Sequence query processing." In *Proceedings of the 1994 ACM SIGMOD international conference on Management of data, SIGMOD '94*, pages 430--441, New York, NY, USA, 1994. ACM.
33. Seshadri, P., Livny, M. and Ramakrishnan, R. "Seq: A model for sequence databases." In *Proceedings of the Eleventh International Conference on Data Engineering, ICDE '95*, pages 232--239, Washington, DC, USA, 1995. IEEE Computer Society.
34. Seshadri, P., Livny, M. and Ramakrishnan, R. "The design and implementation of a sequence database system." In *Proceedings of the 22th International Conference on Very Large Data Bases, VLDB '96*, pages 99-110, San Francisco, CA, USA, 1996. Morgan Kaufmann Publishers Inc.
35. Snodgrass, R. and Ahn, I. "A taxonomy of time databases." In *Proceedings of the 1985 ACM SIGMOD international conference on Management of data, SIGMOD '85*, pages 236--246, New York, NY, USA, 1985. ACM.
36. Terry, D., Goldberg, D., Nichols, D. and Oki, B. "Continuous queries over append-only databases." In *Proceedings of the 1992 ACM SIG-MOD international conference on Management of data, SIGMOD '92*, pages 321--330, New York, NY, USA, 1992. ACM.
37. Tucker, P. A., Maier, D., Sheard, T. and Fegaras, L. "Exploiting punctuation semantics in continuous data streams." *IEEE Trans. on Knowl. and Data Eng.*, 15:555--568, March 2003.
38. Vitter, J. S. *External memory algorithms*. chapter External memory algorithms and data structures, pages 1--38. American Mathematical Society, Boston, MA, USA, 1999.
39. Vitter, J. S. and Wang, M. "Approximate computation of multidimensional aggregates of sparse data using wavelets." In *Proceedings of the 1999 ACM SIGMOD international conference on Management of data, SIGMOD '99*, pages 193-204, New York, NY, USA, 1999. ACM.
40. Vitter, J. S., Wang, M. and Iyer, B. "Data cube approximation and histograms via wavelets." In *Proceedings of the seventh international conference on Information and knowledge management, CIKM '98*, pages 96--104, New York, NY, USA, 1998. ACM.
41. Zhu, Y. and Shasha, D. "Statstream: statistical monitoring of thousands of data streams in real time." In *Proceedings of the 28th international conference on Very Large Data Bases, VLDB '02*, pages 358--369. VLDB Endowment, 2002.
42. Silberstein, A., Machanavajjhala, A. and Ramakrishnan, R. "Feed following: The big data challenge in social application" In *Databases and Social Networks, DBSocial '11*, pages 1—6, New York, NY, USA, 2011. ACM.
43. Fard, A. M. and Ester, M. "Collaborative mining in multiple social networks data for criminal group discovery." In *Proceedings of the 2009 International Conference on Computational Science and Engineering – Volume 04, CSE '09*, pages 582—587, Washington, DC, USA, 2009. IEEE Computer Society.

# A Study on the Integration of Social Network Events for Versatile

## Knowledge Management Applications

Frank S.C. Tseng

National Kaohsiung First University of Science and Technology

imfrank@nkfust.edu.tw

Han-Jung Chiang

National Kaohsiung First University of Science and Technology

u9924831@nkfust.edu.tw

### Abstract

Although there are many studies focused on utilizing social network information for related applications and services, most of them only employ single site information for target analysis and application design. They are not practically applicable for real-time query services and related feedback applications. The objective of our study is to resolve the information heterogeneities collected from multiple social network sites to meet the requirements of instant matching by clustering consistent services, and provide a better quality of service recommendation. We will present two examples to illustrate our approach. The first one is to search the consumers who are willing to buy the same products under an expected discount. The other is on matching passengers for carpooling service. Hopefully, these concepts will reduce commodity prices and increase the willingness of shopping, and establish a win-win platform for social networking users.

Keywords: Data Stream, Social Network, Recommendation Service.