

Combine Naïve Bayesian with Collaborative Filtering for Mobile Community

彭愷翔

輔仁大學資訊管理研究所
台北縣新莊市中正路 510 號
rickpeng@im.fju.edu.tw

苑守慈

輔仁大學資訊管理研究所
台北縣新莊市中正路 510 號
yuans@tpts1.seed.net.tw

摘要

本研究著眼於蓬勃的行動商務市場，希望藉由結合被視為殺手級應用的適地性服務 (Location Service)，並且透過虛擬社群的機制，將現階段的資訊價值予以提升，更進一步的透過社群的機制提供有價資訊，成為行動商務中以使用者為主軸的 WCP(Wireless Content Provider)，而系統業者亦能夠從社群成員間的互動中分析使用者喜好，進而針對使用者喜好促銷商品，以期達到較有效率的目標行銷。本研究因此提出一套研究架構，並著手予以實作以期能夠真正建置出 context-aware voice-based mobile community。此研究架構利用 IVR 系統與定位系統 (Location Service) 並且再加上本研究提出之 EPN(Euclidean distance with Positive and Negative) Clustering 與 Naïve Bayesian Prediction 的預測模式及相關社群指標進行模擬驗證。

關鍵字：行動社群、Context-Aware、Naïve Bayesian、Collaborative Filtering

壹、緒論

行動通訊市場的高度成長 [9]，促使著新的應用技術不斷推陳出新，再加上行動通訊時代中，提供資訊和分享資訊一直是系統業者所致力於的目標，而如今 Context awareness application [4] [14] 亦成為眾人注目的焦點，但何謂 Context awareness application 呢？簡單來說就是針對使用者記錄並且分析其所在的『實體位置』、『過去的互動記錄』、『行動通訊設備』以及『其他使用者在相同地點所產生的互動』，而提供之動態資訊服務。正因結合了上述多項的應用，才會使的 Context awareness application 成為世人注目的焦點，本研究基於此一理念，遂提出 Context-Aware Voice-Based mobile community，將語音技術與行動通訊的適地性服務(location service)做一結合，以期建構一個高互動性之虛擬行動社群。

貳、題描述與動機

正當無線通訊網路快速成長的同時，如何才能夠整合此一強大的平台傳遞使用者所需的資訊，以現階段而言似乎面臨了許多的問題，接下來將針對這些問題與現況逐一說明：

(一) ICP(Internet Content Provider)的資訊不

易獲得

環顧目前國內的 ICP 市場，幾乎所有的服務都是定位在免費服務，舉凡音樂、美食、旅遊、新聞等等服務都有網站在提供，可是所能夠提供給使用者的資訊也十分有限，如何才能從傳統的 ICP 業者跨足到 WCP(Wireless Content Provider)，以目前行動電話系統業者而言，就必需要從現階段 ICP 業者當中尋找有能力提供足夠資訊的業者，可是這之間的利潤和費用該如何計算就是一個棘手的問題，而這個問題確實令許多的業者困惑不已，基於這個考量本研究便思索以社群方式來提供資訊的來源，將原本必需由 ICP 提供的資訊轉而從社群之中取得。

(二) 語音輸入受到限制

語音辨識的技術雖然已經探討許多年，但直到今天仍舊有一個很大的問題便是『背景噪音』，所謂的背景噪音便是當使用者在操作語音輸入時，週遭環境所產生的聲音，然而行動電話使用者本身就是具有高度移動性，並不會因為要執行語音輸入而尋找一個合適的地點，如此一來若要採用語音輸入法並且再將其辨識成文字的應用範圍便受到了限制，因此本研究便著眼於此，將不採用語音辨識的技術，系統只會將使用者留言以語音格式的檔案予以儲存，此舉兼顧了語音輸入的便利性亦避免了語音辨識錯誤

的問題，也就是說使用者無須考慮到背景噪音所帶給語音辨識的問題，而且在不久的將來語音識別技術更趨成熟時，亦能夠讓系統便於整合語音識別的功能。

(三)簡訊成為行動商務的殺手級應用

根據 GSM Association 的統計資料指出，在 2000 年 12 月全球透過 GSM 無線網路傳送的簡訊(Short Message Service)數量就高達 150 億則，比起去年同時期成長了五倍，可見其十分受到使用者的青睞，而最新一期的 GSM association 的資料指出，2001 年第一季全球透過 GSM 系統傳送的簡訊數量就高達 500 億則[8]。

四、日本 NTTDoCoMo 的 i-mode 成功經驗

根據 NTTDoCoMo 所發佈的數據顯示，截至 2001 年七月份 i-mode 的用戶數已經達到了二千五百多萬人[12]，這些使用者透過手機下載音聲鈴聲、打電玩、收發 e-mail 打破了以往行動電話單純的語音服務，帶給了行動商務另一個嶄新的視野，讓系統業者莫不戰戰兢兢深怕在這一次的通訊革命中失去了先機。

經過以上的動機分析之後本研究認為，將資訊與地點結合的觀念能夠強化資訊價值，若能夠再結合不同使用者之間對相同地點的認知，便可以達到經驗和想法的交流，再加上藉由『語音』輸入的方式來取代透過手機鍵盤輸入的動作，事實上說話不就是我們人類最基本也最自然的溝通方式嗎？，本研究便基於以上多方面考量，架構出了『Context-Aware Voice-Based Mobile Community』的觀念，但語音傳輸的資料量十分的可觀，以現行手機進行數據傳輸成本十分的高，而且目前 WAP 之所以沒有辦法普及的原因也正是其過高的通訊費用和不理想的傳輸速率，相反的語音系統是每一位使用者較習慣的使用介面，因此本研究便採取 IVR(Interactive Voice Response) 和 SMS(Short message system) 來建構我們的 Context-Aware Voice-Based Mobile Community。

以往的社群[15][16]都是建構在網際網路之上，而資訊的傳遞都只侷限在文字與圖形的描述，但有一些資訊是無法透過文字或

圖形描述，例如某一個地點的描述，若不是熟悉當地環境的人，即便是再多的說明也無法讓他人明確的了解，可是藉由 Context-Awareness 的觀念，我們便可以把資訊、地點、人做一個緊密的結合，讓資訊具有更高的價值。

然而使用者的喜好並非一陳不變，本研究針對此特性便採用使用者定義檔(user profile)的方式來記錄其所參與的討論區和其使用討論區的頻率等資訊，以期掌握每一位客戶喜好的變動，如此一來企業便能更加了解其客戶也就更能依照客戶的需求來提供所需的服務，最終目的便是能夠建立良好的客戶關係管理機制，並且透過客戶所參與的討論區得知其喜好，善用手機的通路來達到目標行銷，再加上以地點為基礎的服務，讓使用者在適當的時間獲取到適當的資訊，此外個人化的服務機制能夠促進社群互動環境快速的成長，這也就是為何本研究提出藉由使用者定義檔來記錄使用者行為模式。

此外資訊的取得和管理完全屬於使用者，而不用依賴資訊內容供應商(Information Content Provider)來提供，也正因如此系統業者毋需耗費龐大的人力和物力來維護資訊的來源和品質，但獲得的卻是廣大使用者所提供的寶貴經驗，如此加值性的服務更能夠為系統業者建構出另一個全新的 Business Model。

參、研究方法與架構

3.1 Context-Aware Voice-Based Mobile Community 的特色

傳統的社群是由一群具有相同喜好的使用者所組成，而在 Context-Aware Voice-Based Mobile Community 之中亦是如此，但相較於傳統的社群，行動社群除了具有 Mobile Positioning(定位化)[7]之外，更加上了所謂的協同過濾(Collaborative Filtering)[6][10][18]的機制，之後便將定位(Position)和 CF 的機制加以結合，也就是說不同的使用者在相同的地點可能會有不同的交互作用，如此才能夠藉由此一機制激盪出真正具有價值的資訊，此外再參考使用者互動的歷史記錄提供符合使用者習性的資訊。

當使用者啟動社群功能時，系統就會依照使用者定義檔中互動的歷史記錄提供適合該使用者的『適地資訊』，但由於手機的物理限制，我們將會採取 Top-N 的資訊呈現方式，讓使用者可以先從系統推薦的 Top-N 中選取較有興趣的留言來收聽，此外本研究期望提供給使用者的 Top-N 值為三則，使用者可以先行聽取此三則留言的部分內容，再決定是否收聽全部留言或是自行選定其他主題的留言。以本研究架構為藍圖所建構的 Context-Aware Voice-Based Mobile Community 具有以下特點：

一、 適地性(Location sensitive)：

傳統的網際網路社群都是建構在網站之上，所能提供的資訊一般都是透過圖片、影像、文字等方式來傳達，可是藉由這些方式的描述若是使用者並不清楚該地點，則使用者還是無法掌握到正確的位置，但本研究提出之架構則是將資訊與地點緊密的結合，只要使用者在任何地方啟動行動社群功能，便能夠輕易的獲得與所在地點相關的適地資訊。

二、 互動性(Interaction)：

行動電話的使用者皆為獨立個體，但是藉由本系統的社群功能我們可以將具有相同喜好的使用者結合在一起，透過使用者參與不同討論區的討論[3]，進而提升使用者的使用率及對系統業者的忠誠度。

三、 即時性(Real time)：

所有的訊息都是由使用者自動提供，當使用者在該地點傳送留言時系統便立即記錄，其他具有相同喜好的使用者只要一進入相同的區域啟動社群功能，便可以立即收到符合其喜好且適地性的訊息，提供了最為即時性的資訊。

四、 便利性(Convenience)：

本研究利用語音技術儲存使用者的留言，避免使用者必需藉由手機輸入訊息，減少透過手機輸入文字所造成的不便，使用者只需輸入簡單的數字即可操作本系統。

五、 智慧性(Intelligent)：

本系統提供智慧的分群(Clustering)功能並結合使用者定義檔，可以過濾出符合使用者喜好的訊息，讓使用者不用花費過多的時間在搜尋符合喜好的資訊。

六、 演化性(Evolution)：

本系統會依據使用者喜好的改變而自動調整其使用者定義檔中的記錄，並且推薦符合其需求的訊息，不會發生使用者的喜好改變，但系統仍舊發送使用者從前喜好的訊息，如此一來便可以由系統自行調適，達到最為便利的客戶服務目標。

七、 個人化(Personalization)：

每一位使用者皆擁有 user profile，其中記錄使用者所有的互動資訊，系統能夠參考使用者的互動記錄，快速找出符合使用者喜好的訊息，並且以 Top-N 的原則呈現給使用者，提供個人化的使用環境。

3.2 研究架構

本研究的目的是希望利用 IVR 系統與定位系統(Location Service)並且再加上本研究提出之 EPN(Euclidean distance with Positive and Negative) Clustering 與 Naïve Bayesian Prediction[18][19]的預測模式建構出 Context-Aware Voice-Based Mobile Community。因此本研究架構可以概分為以下幾個模組，將分述如下，首先利用本研究之 Give Mode(留言模式)來解說本系統架構，請參見圖 1：

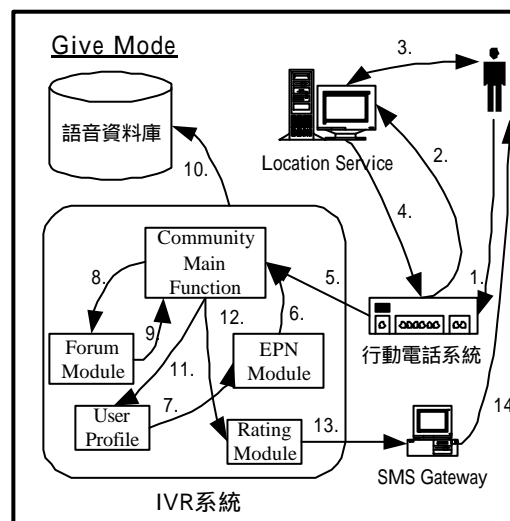


圖 1、留言模式系統流程圖

1. 使用者啟動虛擬行動社群功能
2. 行動電話系統將使用者的手機號碼傳送給定位服務(Location Service)
3. 定位服務系統 Pooling 使用者手機取得使用者所在位置

4. 定位服務系統將所取得之使用者位置回傳至行動電話系統
5. 系統將使用者導入 IVR(Interactive Voice Response)
6. Community Main Function 取得使用者的需求(Request), 便啟動 EPN Module 來計算使用者所可能喜好的討論區類別
7. EPN Module 推算使用者喜好時會參考使用者定義檔(User Profile)的分群記錄
8. 若使用者不接受系統建議, 即進入總討論區類別選擇功能, 由使用者自行選擇適當之討論區
9. 確定使用者所選擇的討論區類別並予以回傳至 main function
10. 將使用者的語音留言連同先前從 Location Service 所取得之位置資訊依其所選取之類別寫入資料庫中
11. 將使用者此次的互動記錄回寫至 user profile
12. 若是使用者啟動評比簡訊提示的服務, 系統便會將其記錄在 Rating Module 之中的提示排程中
13. 當使用者所定義的時間到達時, Rating Module 便將提示簡訊傳送給 SMS Gateway
14. SMS Gateway 便將簡訊透過簡訊系統傳送給使用者

透過圖 1 的系統架構說明之後, 就更能夠了解 context-aware application 其中所傳達的意念, 若是簡單的加以說明可以歸納如下:

1. 本研究透過 location Service 提供系統獲取使用者『實體位置』的資訊。
2. 藉由使用者定義檔來記錄使用者『過去的互動記錄』, 以做個人化服務的依據。
3. 結合留言與評比的機能讓使用者能夠在相同的地點產生互動, 達到使用者經驗交流的目標。
4. 行動通訊設備成為應用的重要媒介

系統模組功能說明

(1) EPN Module

此模組採用本研究所提出之 EPN 架構來進行使用者喜好度的推算, 藉由其他使用者對不同文章的喜好程度來推估我們所欲觀

察的使用者的喜好。而其中 EPN 模組所需的資料則來自於使用者定義與語音資料庫之的記錄, 來預測使用者的喜好。

(2) 討論區模組(Forum Module)

為了讓本研究所提出之系統更具有擴充性與彈性, 因此特別將討論區予以模組化, 讓各類別的討論區可以更容易的被新增進入社群之中。

(4) 社群主要模組(Community Main Function)

本模組為社群運作的核心系統, 提供使用者操作指引, 並且依照使用者需求啟動不同模組進行處理。

(5) 評比模組(Rating Module)

為了確保社群之中的留言具有一定的品質, 故採用使用者評比的方式, 但使用者並非在聽取完留言之後便立即可以進行評比, 若以本研究為例, 所提供的留言內容是屬於美食餐廳的類別, 但當使用者聽取完成留言之後並未實際進行消費, 基於此種考量本研究便提出結合簡訊提示的功能, 當使用者在聽取完成留言之後, 系統便會詢問使用者是否要進行評比的動作, 而使用者可以選擇立即進行評比或是利用簡訊提示服務, 在之後提示使用者進行評比的動作, 藉由此提示服務期望提高使用者參與評比的意願。

經過上述分析之後可以了解到本系統中各模組所負責的功能, 接下來便針對各模組所會使用到的關鍵技術 - EPN(Euclidean distance with Positive and Negative) Clustering 與 Naïve Bayesian Prediction 的預測模式-加以說明。

3.2.1 協同過濾(Collaborative Filtering)

傳統的 Collaborative Filtering 方式是藉由不同使用者的評比來計算使用者之間的相似度(相似之使用者則形成同一個群集)。但這必需具有一個前題便是評比的資料並非是稀疏的(sparse), 但是實際在應用的領域之中並非如此, 我們並沒有辦法可以要求所有的使用者都來針對某一個留言進行評比, 一般來說我們常常必需面對如表 1 的狀況。

所以在系統中存在著許多稀疏[11][13]的資料, 但這此資料要如何才能夠分析, 假使只單單的利用評比過的資訊便加以分

群，這麼一來所獲得的結果必定會和事實有很大的出入，所以本研究便提出結合 CF 概念和 Naive Bayesian 的方式來達到我們的目標。

然而現階段的 Naive Bayesian 都是應用在，從一個人的歷史行為來推估同一個使用者在相同的條件下是否會去做某件事的機率。此外使用者的喜好也有可能隨著時間而改變，為了掌握此一特性我們便必需定期的重新執行分群的動作，其目的就是為了讓所有的分群之後的資料都能保有一定程度的可靠性。

從表 1 知道每一位使用者皆可以針對每則留言加以評比，所以在討論區中的留言都會經過社群成員的評比，而我們評比所採用的分數是參考李氏克特量表 (Likert scale) 的評估方式，共分為五個等級，使用者可以透過手機的鍵盤輸入 1~5 的分數，接下來便利用評比之後的結果來區隔出留言的品質，因此留言的品質是由社群成員共同管理，這也就是所謂協同過濾的意義，表中的 U_i 表示 User，而 $i = 1 \sim n$ ，而每則留言別分予以編碼，以一個英文字加上五個數字，其中英文字與第一位數字表示討論區代碼，後四碼數字則討論區中的該則留言的代碼，例如 A13211 則表示是位於 A1 討論區中的第 3211 則留言

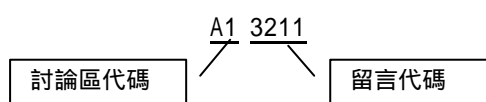


表 1 中 ' - ' 代表該使用者尚未針對該篇留言進行評比，而下表中只就十則留言為例，而非系統在進行分群時只針對十則留言加以處理。

表 1、原始 Rating Table

	A13211	A11123	B12321	A34121	B22112
U1	-	3	2	3	-
U2	1	3	-	-	2
U3	-	-	5	3	2
U4	1	-	-	2	-
U5	3	2	4	2	4
U6	-	-	4	4	4
	B14001	D21153	F13121	K12112	K21212
U1	2	4	3	-	2
U2	3	3	2	1	3
U3	3	3	5	-	1
U4	-	-	-	1	3
U5	-	4	5	3	2
U6	4	-	5	5	3

3.2.2 EPN-Based Clustering

以表 1 為例，存在系統之中的評比表會記錄使用者許多的評比記錄，但我們並不知道那些使用者具有相同的喜好，所以我們就必需將似的使用者加以分群，找出和某一位使用者有相同喜好的一群人。傳統的分群方式種類非常的繁多，但前題必需是要能夠擁有一個完整的評比表格 (rating table)，其中並不會存在所謂的稀疏 (sparse) 的情況，但在表 1 可以很明顯的發現其中存在許多稀疏的情況，基於這個考量本研究遂提出 *EPN Algorithm (Euclidean distance with Positive and Negative Strength)*，也就是結合 Euclidean distance 與正、負向強度來評估使用者之間的相關性。

本研究認為不只是正向強度可以表現出使用者之間的相關性，負向強度應該也包含許多的訊息，因此在本研究中所提及的負值就是尚未被評比的個數，如果以機率來分析的話這些尚未被評比的值，比已經都評比過的值，更有可能出現與我們想要預測的使用者還接近的可能性或是離我們所欲預測的使用者更遠的狀況，因此為了確保資料的正確性本系統會定期的重新執行評比，至於何謂定期的執行將視資料量的多寡來評估週期的長短。

為了讓您更容易瞭解其中的涵義，我們以下面這個例子來加以說明：

表、2 相似度比較

	A13211	A11123	B12321	A34121	B22112
U3	-	-	5	3	2
U4	1	-	-	2	-
U5	3	2	4	2	4
	B14001	D21153	F13121	K12112	K21212
U3	3	3	5	-	1
U4	-	-	-	1	3
U5	-	4	5	3	2

本研究認為在表 2 之中，U4 會較 U5 更有可能與 U3 相似，因為 U4 仍有許多的值尚未評比，相較於 U5 已經幾乎都 Rating 完成，故本研究認為 U4 有可能會較有機會與 U3 相似，但為避免 U4 後續的發展會朝

向與 U_3 相異的方向，所以本研究所提出的『定期分群』目的便是要克服此一問題，而此一方法所稱的『定期』我們將視資料量來評估此週期的長短，以求分群的正確性。

EPN Algorithm 提出一個計算兩個使用者總體關連強度值之演算方法，兩個使用者總體關連強度值為正負向強度與 Euclidean distance 中之加權總和，其中相關權重 w_1 、 w_2 、 w_3 (正負向與 Euclidean distance 的比重) 取決於已經 rating 的比率來決定個別所佔的比重。因為已經被 rating 的資料量少，避免某一個極值把 Euclidean distance 的值拉大，所以此時就是正、負向強度佔的權重需較大，而 Euclidean distance 的權重需小。相反情況當 rating 個數已經很多時，那麼 Euclidean distance 就較不易受某一個極值影響，所以其比重就加大。

以本研究為例如表 1 中的使用者比較，此時系統若欲從評比表(Rating table)中尋找出和 User3 相似的使用者，便會參考每一位使用者和 User3 這一位使用者之間已評比值的相似個數，和尚未評比的個數。

以 User1 和 User3 為例，我們先計算其正、負向強度，從表 1 中可以知道，User1 已經針對十則留言中的七則留言給予評比還有三則尚未評比，再跟 User3 比較得知七則中只有一則是相同的留言，所以我們便將 User1 和 User3 的正、負向相似度表示成

$$(U_1, U_3):(P, N) = (\frac{1}{7} \times \frac{7}{10}, \frac{6}{7} \times \frac{3}{10})$$

接下來再計算 User1 和 User3 的尤拉距離 (Euclidean distance) ,

$\sqrt{(5-2)^2 + (3-3)^2 + (3-2)^2 + (3-4)^2 + (5-3)^2 + (1-2)^2} = \sqrt{16}$
依上述的方式我們分別求算每位 User 和 User3 的相似度，即可得到下列數據。

$$(U_1, U_3):(P, N) = (\frac{1}{7} \times \frac{7}{10}, \frac{6}{7} \times \frac{3}{10})$$

$$E(U_1, U_3) = \sqrt{16}$$

$$(U_2, U_3):(P, N) = (\frac{3}{8} \times \frac{8}{10}, \frac{5}{8} \times \frac{2}{10})$$

$$E(U_2, U_3) = \sqrt{13}$$

$$(U_4, U_3):(P, N) = (\frac{0}{4} \times \frac{4}{10}, \frac{4}{4} \times \frac{6}{10})$$

$$E(U_4, U_3) = \sqrt{5}$$

$$(U_5, U_3):(P, N) = (\frac{1}{9} \times \frac{9}{10}, \frac{8}{9} \times \frac{1}{10})$$

$$E(U_5, U_3) = \sqrt{8}$$

$$(U_6, U_3):(P, N) = (\frac{1}{7} \times \frac{7}{10}, \frac{6}{7} \times \frac{3}{10})$$

$$E(U_6, U_3) = \sqrt{11}$$

依照分析的結果 U_2 較其他的四位使用者更接近 U_3 ，雖然其中 U_1 及 U_6 與 U_3 的正、負強度皆相同，但若再參考尤拉距離 (Euclidean distance) 的數值，就可以區分出來其中的差異性，所以藉由本法可以找出與 U_3 相似的 K 個 Users，以便於之後計算 Naï ve Bayesian 的機率值。

所以依照原則推論正值愈大表示愈相近，當正值相同時則採用負值比較，而負值愈大也表示愈相近，因此為確保分群的正確性系統必需定期執行分群的動作。所以經由以上的分群的機制之後，我們便得到 U_2 、 U_4 、 U_6 此三個使用者是較接近 U_3 。

EPN Algorithm

- 1、 所欲預測的使用者 (U_p)
- 2、 將其他使用者 (U_q) 的評比記錄與預測的使用者比對。
- 3、 計算正向強度 (P) 與負向強度 (N)

$$(U_p, U_q):(P, N) = (\frac{S}{R} \times \frac{R}{A}, \frac{R-S}{R} \times \frac{A-R}{A}) \quad (1)$$

U_p : 欲預測的使用者

U_q : 其餘使用者 (q=1~n)

S : 代表相同 U_p 與 U_q 相同評比的個數

R : 代表 U_q 已經評比過的個數

A : 代表所有的留言數

- 4、 計算出尤拉距離 (Euclidean Distance) (必需兩者皆有評比才計算)

$$E = \sqrt{\sum (r_{pj} - r_{qj})^2} \quad (2)$$

r : 評比

j : 為文章代碼

r_{pj} : P user 對第 j 篇文章的評比值

r_{qj} : Q user 對第 j 篇文章的評比值

- 5、 計算其他使用者 (U_q) 與預測的使用者 (U_p) 之總體 關連強度值 (Strength):

$$Strength = w_1 * P + w_2 * N + w_3 * E$$

w_1, w_2, w_3 : 相關權重

3.2.3 Naï ve- Bayesian-Based Prediction

而我們會對經過分群之後的評比表採用 Naï ve Bayesian 的方式來預測某一位使用者的評比值，藉由預測出來的機率值，就能夠知道使用者的喜好所在，現階段的 Naï ve Bayesian 都是從某一位使用者的歷史行為中去計算在發生某一個情境之下該使用者會去做某件事的機率值，並沒有應用在藉由相似的他人之歷史記錄來推估該使用者可能會去做某件事的機率值。

以本研究為例我們透過 EPN 演算法，先將與使用者有相同習性的使用者加以分群，找出和所欲預測的使用者屬於同一群的使用者，如表 3 所示即為系統針對表 1 進行 EPN 分群之後的結果。

表 3、經過 EPN 分群過的 Rating Table

	A13211	A11123	B12321	A34121	B22112
U2	1	3	-	-	2
U3	-	-	5	3	2
U4	1	-	-	2	-
U6	-	-	4	4	4
	B14001	D2115	F13121	K12112	K2121
U2	3	3	2	1	3
U3	3	3	5	-	1
U4	-	-	-	1	3
U6	4	-	5	5	3

接下來再利用 Naï ve Bayesian 的概念加以預測該使用者會針對某則留言，將其評比值為 1~5 值的機率。故本研究將採用 Naï ve-Bayesian-Based Prediction Algorithm 的方式來計算機率值，而我們將針對經過分群之後的評比表來進行 Naï ve Bayesian 的推算。

經過推算之後，我們便可以得到使用者對於每一則留言的喜好程度所代表的機率值，系統便是依據這些機率值與使用者定義檔來決定應該將何種組合的 Top-N 傳送給使用者，以達到適性化的資訊服務功能。

然而為何我們藉由他人之歷史記錄來推估該使用者可能會去做某件事的機率值。原因在於 Naï ve Bayesian 是建立在條件機率之上的一種預測模式，也就是某使用者由其歷史行為中過去的評比值為何的條件之下，該使用者會將某則留言 rating 為何的機率。

但當評比仍是稀疏狀況時，使用者之歷

史行為便無法計算出有效的機率值。而本研究認為相似的其他使用者其歷史行為之資訊可被用來彌補使用者本身不足之歷史行為資訊，進而計算出來一通用且有效的機率值，並且可以讓我們透過這個機率值立即瞭解到使用者對每一則留言喜好的可能性，以這個為基礎來實現我們的推薦服務[5]。

因為我們採用的是五個等級(scale)的方式來評比留言，若今天我們想推估 User3 對 k12112 留言的喜好程度，我們便必需分別求算出其可能評比值為 1~5 的各機率為何，在這個前提之下我們先計算可能評比值為 1 的機率為何。

已經完成分群過後的評比表中所有出現 1~5 的機率： $P(v_j)$

出現 1~5 的評比值又是屬於 k12112 留言的機率： $P(a_j|v_j)$

$$v_{NB} = \arg \max P(v_j) \prod P(a_i | v_j) \quad (3)$$

利用表 2 中的記錄，我們先計算 k12112 此則留言出現 1 的機率，計算之後得知已經被評比的值共有 25 個而其中值為 1 的共有 4 個，而在 k12112 留言的所有記錄之中有 1 個評比值為 1，針對 k12112 的評比記錄則有 2 個，因此我們可以得到下列的機率：

User3 針對 k12112 會評比值為 1 的機率為

$$\frac{5}{25} \times \frac{2}{5} = \frac{2}{25}$$

依照上述方式則可以分別求算出

$$\text{Rating 為 2 的機率：} \frac{4}{25} \times \frac{-}{4} \rightarrow d$$

$$\text{Rating 為 3 的機率：} \frac{9}{25} \times \frac{-}{9} \rightarrow d$$

$$\text{Rating 為 4 的機率：} \frac{4}{25} \times \frac{-}{4} \rightarrow d$$

$$\text{Rating 為 5 的機率：} \frac{4}{25} \times \frac{1}{4} = \frac{1}{25}$$

所以我們可以推斷 user3 對 k12112 的這一則留言最有可能的評比分數為 1，也就是代表他喜好這 k12112 的喜好程度為 $\frac{2}{25}$ ，接下

但我們以一個極微小的數 d 來代表 computation 內含「-」的值。

來便可以再計算出其他留言的喜好程度，做為 Top-N 推薦時的依據。

Naïve-Bayesian-Based Prediction Algorithm

- 1、計算經過 EPN 之後的評比表中所有已被評比的個數。
 T ：表示在分群過後的評比表中有多少值已被評比
- 2、計算經過 EPN 之後的評比表中所有已被評比為 i 的個數。
 R_i ：表示 1~5 每一個評比分數的個數， $i=1\sim5$
- 3、計算想要被預測的留言 m 評比為 i 的個數。
 R_m ：表示第 m 則中所評比為 i 的個數
- 4、計算每一則留言評比分數的機率為何。

$$P_i = \frac{R_i}{T} \times \frac{R_m}{R_i} \quad (4)$$
 P_i ：評比為 i 的機率， $i=1\sim5$
- 5、得到所有機率，便可找出機率最大值預測值

3.2.4 相關社群指標

社群品質指標(Quality of Community, QC)

在社群的發展之中，若能夠有效的掌握其發展狀態，就可以更進一步協助其採行不同的策略，達到品質監控的目的，故本研究便提出此一社群品質指標，希望藉由這個指標可以充份的反應出目前社群的討論品質，在這個社群品質指標中包含多個因素。傳統的社群機制之中都只注重每一個討論區的文章數，但文章的數目真的可以代表一個討論區的留言品質嗎？如果討論區中充斥著廣告的訊息，那麼文章的數目便無法有效的表現出討論區的品質，因此本研究提出 QC 來評估討論區的留言品質，而我們將討論區中所有文章數定義為 M ，已經有評比的文章數定義為 R ，被評為 3~5 的文章視為品質良好將其定義為 G ，而被評為 1~2 的文章則視為品質不良將其定義為 B 。

所有文章數：(M) message
 受評比文章：(R) rating 數目
 品質良好的文章：(G) Good, 被評為『好』的文章數目
 品質不良的文章：(B) Bad, 被評為不良的文章

$$QC = \frac{R}{M} \left(\frac{G-B}{R} \right) \quad (5)$$

社群活絡指標(Hot Community, HC)

傳統的社群經營都是由使用者自發性的管理，常在某一個討論區中會存有一些所謂的領袖成員(Key user)，但如果社群之中完全依賴這些領袖成員來經營，久而久之會不會導致這些領袖成員失去動力？然而我們不能夠就此讓領袖成員失去服務的動力，在相關的研究之中領袖成員的出現十分的不易，其必需具備有高度的專業知識和服務熱忱[1]。也因此基於此種考量本研究便提出社群活絡指標，讓建構社群的系統可以正確掌握到現階段社群討論區的發展狀況，並針對不同的狀況提供相對應的政策，以確保社群能維持在一定的討論品質之下。

目前一般網際網路上面絕大多數社群都會把回應文章數量當做是討論區熱門與否的指標，可是那隻可以視為是熱門話題的指標，而不能代表整個討論區目前的活絡程度。所以何謂活絡的社群呢？為了定義這個活絡的意義，本研究認為具有良好的討論品質與高人氣的瀏覽率才能算是活絡的社群，所以結合先前的 QC，再參考討論區瀏覽人數 v 佔整個社群瀏覽人數 V 的比率，即可以得到下列的指標：

社群品質指標：QC
 該討論區瀏覽人數： v
 該社群總瀏覽人數： V

$$HC = QC \times \frac{v}{V} \quad (6)$$

個人聲望指標(Personal Reputation)

在本研究中為了讓使用者能夠如同現實社會一樣擁有社會地位，便採行了『個人聲望』的機制，透過這個機制即能夠得知某一位使用者在社群中的貢獻度，當系統在針對使用者提供回饋服務時，此類高貢獻度的使用者便能夠獲得更多的回饋[2]，若能夠有效的掌握這些高貢獻使用者，藉由其來活絡整個社群，並豐富討論區中的內容，對於社群經營將會更加的容易，因此本研究便針對使用者所發表的留言被評比的分數和其發表的數量來加以分析，從中了解每一位使用者所發表留言被其他使用者閱讀之後的評價為何。

但為什麼會採用 PR 這個指標，深究其意就是在於虛擬的社群之中並沒有一個有力的依據可以來評斷一個使用者的地位，若

能夠將使用者在社群中的貢獻程度加以量化之後，其他使用者或是系統便能輕易的了解這位使用者在社群中的地位。舉例來說，系統會記錄每一位使用者留言的次數我們將其定義為 n ，而在討論區中存在的所有留言我們將其定義為 N ，但並非所有的留言都具有相同的價值，所以我們計算出每一則留言所獲得的平均分數為 R_i ，綜合上述 factor 就可以計算出每一位使用者在社群中所表現出的價值。

發表文章數： n

所有文章數： N

發表文章被 Rating 的分數： R_i

$$PR = \frac{n}{N} \times \sum \frac{R_i}{n} \quad (7)$$

肆、研究方法評估與討論

在未來研究方法評估之實驗設計上希望能夠評估由本研究所提出的 EPN 分群機制 CF 結合 Naïve Bayesian 的預測模式及社群中的數個相關指標，而為了能夠較為精確地模擬本研究的環境，故設計了不同類型的使用者，希望藉由模擬的環境來驗證本研究所提出來的各項假設或是指標是否具有顯著的效用。其中分別針對 EPN 分群機制與 Naïve Bayesian 機制採用 precision 與 recall 的方式來評估本研究所提出的機制是否能夠達到預期的目標。而其餘的各項社群指標亦採用效標的方式來測驗，藉以了解具備這些指標的社群是否較其他社群更能夠產生良好的社群品質。

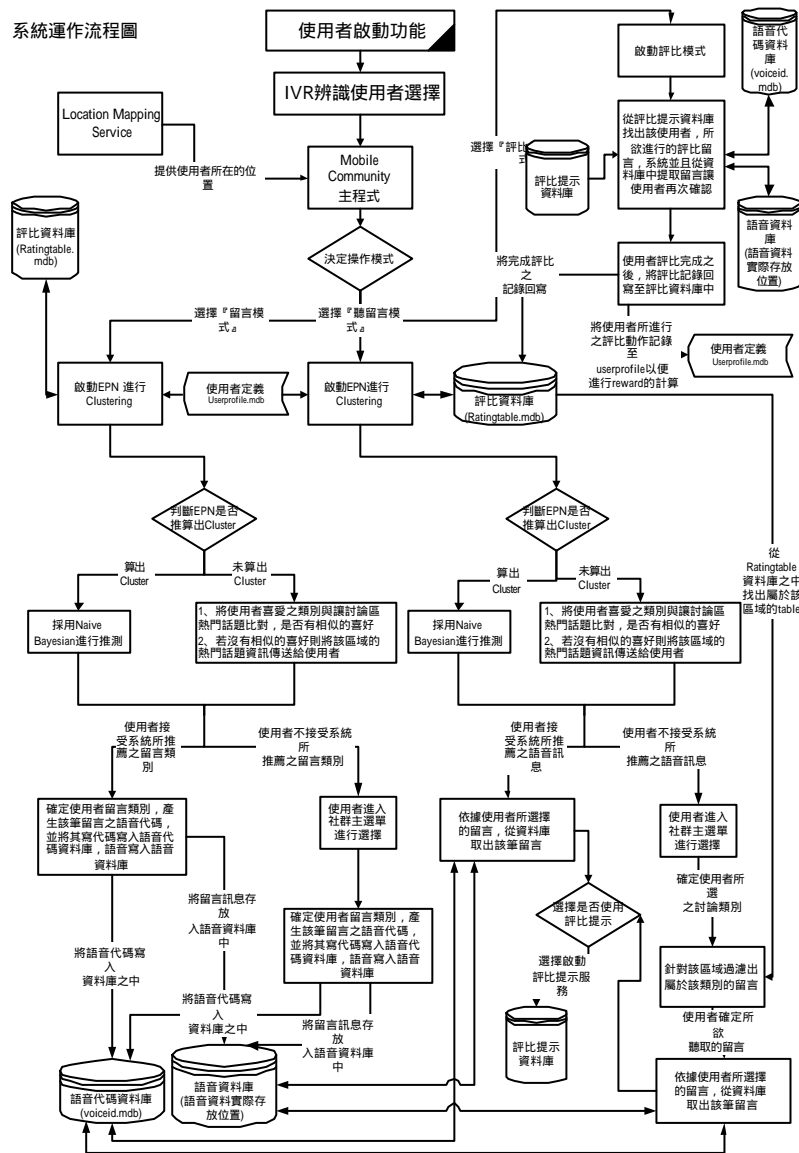


圖 2、系統運作流程圖

由圖 2 中我們可以了解到系統在設計時刻意將社群規劃成為分散式架構，即為每一個 location 的 rating table (評比記錄表) 皆只存放與該地相關的評比，而不會存有其他地區的評比，所以資料是屬於分散式的架構，換句話說可以儘量避免資料量過於龐大時，在進行 Naive Bayesian predication 時所造成的困擾。此外圖 2 中右下角有標示『V』的處理，就表示有與語音系統進行互動的程序。

本研究採用 Java 進行系統開發工具，而採用 Java 的原因便是著眼於其跨平台的特性，現階段系統是發展在 Windows Base 平台之上，使用的設備等級為 CPU P4-1.5 GHz 記憶體 256 MB Ram bus 硬碟為 60GB ATA-100 之個人電腦。雖然現階段是以 windows base 為發展平台，但以 Java 的跨平台特性，將有助於本系統在進行系統移轉時更為容易。

為何本研究以美食餐廳為主題來架構行動社群呢？因為餐廳本身的訊息充份符合 Context-Awareness application 所定義的觀念。

一、實體位置 (physical location)

餐廳本身具有不易移動性，因此可以較容易藉由定位的方式找出其所在地，因此適用於定位服務。

二、過去互動的歷史記錄 (past interaction history)

由於餐廳之中所發生的事件不外乎就是有美味的餐點或對服務態度的感受，而這為模組化，只要新增討論區模組便能夠讓系統立即增加新的討論區，如此便能夠讓本研究所提出的架構更具彈性。

有鑑於行動通訊快速的成長，使得行動商務被譽為未來極具潛力的市場，從上述分析中可以清楚的了解到 Context-Aware Voice-Based Mobile Community 的運作模式，而我們只針對內部運作的模式加以分析，然而業者最關心的莫過於 Business Model，因此本研究亦針對美食資訊的行動社群提出概念性的 Business Model，茲將其

些事件不似其他如特賣活動、簽唱會等高變動性的資料，所以便容易累積更多的歷史記錄，故本研究決定採用美食餐廳為主題。

三、行動設備 (mobile device)

以目前使用的行動通訊設備，以手機的普及率為最高，所以本研究架構便是以手機為平台，並且結合現有成熟的技術，而不採用新的網路傳輸技術，其目的就是要讓使用者不用另外申請便能輕鬆使用此服務，所以有操作模式都是建構在使用者的習慣之上，如此的設計將更能貼近使用者的需求。

四、其他使用者在相同的位置 (other users in his location)

『Context-Aware Voice-Based mobile community』有別於以往的社群就是在於此結合地點的互動性，從前的討論區都只能限制在網路之上，但藉由手機的功能我們可以讓不同的使用者在相同的地方發表其對某一相同事物的意見，如此一來使用者之間的互動就和環境息息相關，所有的訊息都結合上了地域的觀念，而系統也更能夠提供使用者『適地』的資訊。

雖然系統建構之初是採用美食餐廳為基礎，但不表示本研究的應用只侷限於美食餐廳類別，只要符合上述條件的主題皆可以加入此一行動社群，以現階段而言例如：『流行資訊』提供購物相關資訊，『旅遊景點資訊』提供旅遊景點和其週遭的相關資訊等服務，都可本研究可以延伸的應用範圍。

所以為了讓社群具有更多不同領域的討論區，故本研究將社群的討論區設計成

分述如下：

一、採用目標行銷：

現階段國內多家知名電信公司所提供的服務，雖然在使用者初次登入系統時會要求使用者輸入其個人相關資訊和喜好，以便於往後達到所謂的目標行銷，可是使用者的習性真的會如同初次登入系統而不會有任何改變嗎？所以藉由本研究架構可以較易掌握使用者習性的變動，真正達到具有時效性和正確性的目標行銷，也正因此廣告主便能獲得更有效的回饋，而不是所謂的亂槍

打鳥的行銷方式，廣告主也樂於在行動社群中購買廣告時數。

二、使用者毋須付費：

部份系統業者提供網內互打前五分鐘免費，此一觀念就是著眼於網內互打系統業者所負擔的成本相對的較為低廉，也正因為如此行動社群亦是建構在網內的機制。

三、提供廣告主另一類選擇：

行動電話有別於其他的行銷管道就是在於其有極高的移動性，使用者幾乎把手機隨身攜帶在身邊，若能善用此一特性便能夠提高行銷的成功率，畢竟在店面附近的訊息較能夠吸引到使用者注意到該店家所在的位置與販售的物品，並且產生較深刻的印象增加該使用者來店消費的可能性。

四、使用者互動產生內容(Content)：

早期的觀念都告訴我們 content 的來源必需藉由 ICP 來提供，但是否從 ICP 所取得的資訊就是正確且符合成本效益的呢？若能透過使用者自主性的回饋資訊，系統業者不但省去資訊維護的成本，也讓使用者本身有更多的參與機會和意願。其次，資訊的來源若是經由 ICP 業者提供的話，系統業者勢必要付出高額的權利金去取得資訊。若是能夠讓使用者免費使用 Mobile Community 機制，進而建立 Mobile Community 的用戶數，更深一層的目的便是促使行動電話用戶數增加，廣告商也願意在此平台上購買廣告時數。

伍、研究限制

本研究認為理想的 Context-Aware Voice-Based Mobile Community 應該具備有完整的功能，但在本研究之中考量到現階段的技術層面與實際的法令，使的本研究具有以下限制：

一、語音識別的限制

雖然現階段的語音識別技術已經愈趨成熟，但背景噪音值仍舊是一個問題，因為使用者所處的環境可能會產生許多干擾的聲音，此時便會影響到系統辨識的能力，再加上行動通訊設備的使用者本身就具有高

度的移動性，使用者可能身處在許多不同的地點，若在使用此項功能時所處的環境過於吵雜，語音識別的能力便會產生誤差，為此本研究便採用直接儲存使用者語音記錄，而不加以識別，目的就是要確保留言的原始性，避免經過識別所產生的誤差，若是遭遇到使用者留言品質不佳時，便會經由使用者的評比過濾，成為品質不良的留言，之後便自動由系統將其淘汰。

二、追蹤使用者移動軌跡

由於個人隱私權的問題，我們並不能藉由定位系統追尋使用者的移動軌跡，也因此我們結合定位系統提供使用者主動式的推播服務，基於這個考量本研究便提出所謂的 Give & Take 的架構，系統的啟動完全由使用者主導，所以不會有侵犯個人隱私權的問題存在。

陸、結論

本論文提出以語音輸入為基礎結合定位化 (Mobile Positioning) 的 Context-Aware Voice-Based Mobile Community 的嶄新 Business Model，也就是將使用者角色轉換成為 ICP 業者，並且透過社群機制加以區隔不同的使用者，並且採用免費使用的方式來吸引使用者加入社群，廣告商可以選擇其喜好的族群來達到目標行銷，此一架構將有別於以往由系統業者主導一切服務，而本研究中所提出的系統架構將有助於社群自動化調整經營模式和協助社群成長，相信此一觀念將會打破傳統手機使用者是獨立個體的印象，讓團體的動力成為在行動商務中的一股洪流。

參考文獻

- [1]楊堤雅，「網際網路虛擬社群成員之角色與溝通互動之探討」，國立中正大學企業管理研究所，碩士論文，民國八十九年六月，
- [2]吳有順，「網路社群知識分享過程之研究-以企業教學網站為例」，國立政治大學企業管理學系所，碩士論文、民國八十九年六月，
- [3]Amrit Tiwana and Ashley Bush，「Peer-to-Peer Valuation as Mechanism for Reinforcing Active Learning in Virtual

- Communities : Actualizing Social Exchange Theory” , Proceedings of the 33rd Hawaii International Conference on System Sciences ,IEEE 2000
- [4]Berardina De Carolis,Fiorella de Rosis And Sebasitano Pizzutilo , ”Context-Sensitive Information Presentation” , University of Bari 1999
- [5]Chumki Basu,Haym Hirsh, and William Cohen. , “Recommendation as classification : Using social and content-based information in recommendation. In Proceedings of the Fifteenth National Conference on Artificial Intelligence”,P714-720 , July 1998
- [6]David M. Pennock and Eric Horvitz , ”Collaborative Filtering by Personality Diagnosis : A Hybrid Memory - and Model-Based Approach” , David Pennock, Eric Horvitz, Steve Lawrence, C Lee GilesProceedings of the 16th Conference on Uncertainty in Artificial Intelligence, UAI 2000
- [7]Ericsson<http://www.ericsson.com.tw/ericsson/technology/t0605.htm> , 2001
- [8]GSM Association”http://www.gsmworld.com/news/press_2001/press_releases_22.html” , 2001
- [9]ITU(International Telecommunication Union) , <http://www.itu.int/ti/industryoverview/index.htm> , 2001
- [10]IBM , ” Collaborative filtering : Community Values” , Centers for IBM e-business Innovation 2001
- [11]M. Ester, H.-P.Kriegel, J. Sander, And X. Xu. , ”A density-based algorithm for discovering clusters in large spatial databases” , In Proc. 1996 Int. conf. Knowledge Discovery and Data Mining(KDD’96) , P226-231 , Portland,OR.Aug.1996
- [12]NTTDoCoMo”http://www.nttdocomo.com/i/i_m_scr.html” , 2001/7/15
- [13]Paul Resnick , Neophytos Iacovou , Mitesh Suchak , Peter Bergstrom and John Riedl , ”GroupLens : An Open Architecture for Collaborative Filtering of Netnews” , ACM 1994
- [14]Peter J. Brown, John D. Bovey, and Xian Chen. *Context-aware applications: from the laboratory to the marketplace*. IEEE Personal Communications, 4(5):58-64, October 1997
- [15]Rheingold.H. , ”Virtual Community : Homesteading on the Electronic Frontier” , 1st MIT Press ed. Cambridge, Mass. : MIT Press, 2000
- [16]Romm.C. Pliskin.N. and Clarke.R. , ”Virtual Communities and Society : Toward an Integrative three Phase Model” , International Journal of Information Management Vol.17,No.4,P261-270 , 1997
- [17]Resnick, P., et al. GroupLens : An Open Architecture for Collaborative Filtering of Netnews. In Proceedings of ACM CSCW’94 Conference on Computer-Supported Cooperative Work,P.175-186 , 1994
- [18]Yu -Hwan Kim, Shang-Yoon Hahn, Byoung-Tak Zhang , ”Text Filtering by Boosting Naïve Bayes Classifiers” , Seoul National University,ACM,2000
- [19]Yung-Hsin Chen,Edward I. George , ”A BAYESIAN MODEL FOR COOLABORATIVE FILTERING” , 19