

An Effective Algorithm for Interest Aware Opportunistic Advertising by Mining Social and Consuming Data

Chia-Yu Lin^{*a}、Zhi-Feng Jiang^b、Li-Chun Wang^a

國立交通大學電機工程學系^a

工業技術研究院^b

摘要—在擁擠的大賣場或是百貨公司逛街時，常因環境或是外在因素而導致行動網路不穩定，無法順利在有限時間內將特定廣告資訊傳遞至目標消費者，因此本論文在機會網路下結合資料探勘與社群網路技術，提出一個高效率演算法，能在有限時間內將商家或是消費者想要推播的廣告資訊推播給特定目標消費者，並且在推播過程中選擇較有興趣的消費者進而有效提高廣告效益以及商家銷售量。

一、簡介

近年來，由於在現實環境當中存在著許多外在不確定因素導致網路中無法存在一條完整有效路徑，因此機會網路的議題逐漸被重視。而近年同時因為社群網路的蓬勃發展，許多機會網路均與社交網路上的資訊結合，例如 Twitter 或是 Facebook，以進行更有效率地推播[1]-[4]。同時隨著智慧型行動裝置的進步，目前更逐漸在發展只要透過行動裝置就能直接在商家透過手持裝置完成消費的功能，如此，智慧型裝置上記載著相當多的個人行為資料，而我們將可以藉由資料探勘技術分析這些資料，建立個人化的消費行為模型，對於使用者方面，將能進一步提供給使用者一個更加便利而且個人化的生活模式；而對於商家或是軟體開發者而言，也更能夠提供廣告消費券或是優惠方案給有效的消費者，如此將可以提高整體的經濟效益[5]-[9]。

然而，在大賣場或百貨公司這種人潮擁擠的場所下，若想在消費者逛街的有效時間內發送有效地廣告與消費資訊往往會因為行動網路的不穩定而需要花費很長的繞送路徑，又無法在有限時間將廣告送達給消費者，抑或是要發送多次，提高發送成本才能送達消費者。因此[10]提出了 Interest-aware PeopleRank (IPeR) 演算法，在 PeopleRank 的計算上加入使用者的興趣，透過此演算法可以選擇較有興趣的使用者進行廣告推播，如此不僅可以提高推播廣告時的廣告效益，更可以提升找到目標使用者的機率[4]。然而，[10] 所以提出的 IPeR 演算法需要 $O(n^4)$ 的複雜度才能完成推播，若是有大量的使用者要進行廣告的推播將會使得智慧型行動裝置上的負荷太重，且[10]在選擇推播路徑食指使用消費者的社群來判斷消費者的興趣，若是能加入消費歷史資料進行判斷，將能更有效地找到潛在的消費者，進而提升廣告的效益。

因此本論文結合機會網路與社群與消費資料，提出一個消費行為多層模型以及 Interest-aware Opportunistic Advertising 演算法，簡稱 ISC 演算法，在大賣場或是百貨公司的環境下，透過分析社群與消費資

料有效選擇有興趣且具有潛在消費機率的消費者進行廣告的推播，如此可將廣告推播給真正有需求且有興趣的消費者，進而有效提高廣告效益以及商家銷售量，同時 ISC 演算法可以在 $O(n^2)$ 完成廣告的播送運算，在智慧型裝置上將可更有效率。

本論文提出一個消費行為多層模型以及一套機會網路 ISC 演算法，其中消費行為多層模型包含了屬性層、社交層以及商店層。在屬性層當中，我們透過商家之間所擁有的商家屬性來對應至社交層，使得我們可以透過社交層中消費者的社群關係結合屬性層的屬性關係來找出消費者之間的潛在的社群消費關係；而在商店層當中我們透過資料探勘的技術找出店家與店家之間的關係，進一步結合消費資料來找出商店層當中消費者之間潛在的消費關係。最後我們整合所有找出來的這些消費關係去辨別消費者之間的消費相關性程度，進而提供給 ISC 演算法做使用。

而在 ISC 演算法當中，我們使用了消費者興趣資料來判斷消費者對於廣告是否有興趣，若消費者對於廣告有興趣則將廣告推播給他，而若判斷出消費者對於該廣告不感興趣則有可能是資料不足所導致，我們將結合在消費行為多層模型所產生的消費行為相關性程度來找出這些對廣告可能有興趣的隱藏相似消費者。而藉由以上這些中間消費者，最後將有效率的將廣告播送給目標使用者。

本論文除了改善[10]所提出演算法的執行效率問題之外，更進一步能將廣告推播給有需要及有潛在消費機率的消費者，同時大幅提昇廣告的整體效益。

本論文剩下章節的結構依序如下。再接續的第二節將會介紹關於本論文所提出的演算法的背景。第三節將會詳細的介紹本論文提出的演算法流程以及架構。第四節則是介紹實作演算法的模擬環境、資料以及結果。第五節則是結論。

二、背景

機會網路是鑒於在傳統無線隨意網路當中在傳輸資料時要求在來源點以及目標點之間必須存在一條完整路徑的情形之下才適用，但在現實環境之中有非常多外在不能抗拒的因素會導致此條件不成立，於是機會網路的概念才會被提出。機會網路即是一種利用網路中節點的移動所帶來節點與節點之間相遇的機會來實現來源點以及目標點通信稱之[11]。

許多的社群導向轉發式演算法在機會網路研究當中被提出來解決在機會網路當中，將資料傳送至特定目標

或相似節點的問題，舉例如 PeopleRank[4], HiBoP[1], BubbleRap[3]以及 SimBet[2]都是以社群資料為導向來找出特定目標點的轉發式演算法。但是這些演算法週期性的交換資訊內容時，都需要耗費極大的網路效能。因此，在[10]當中提出了 IPeR 演算法，IPeR 是一種以使用者興趣資料為導向的轉發式演算法，IPeR 透過自訂機制來排除在轉發過程當中對廣告不感興趣的使用者節點，以降低轉發過程中所需要轉發的總節點數量，藉此來減少網路整體的花費以及延遲時間。IPeR 演算法是 PeopleRank[4] 演算法加入興趣限制條件判斷而來，PeopleRank 主要是希望透過找到一群人當中最佳的熱門社群點(即在這群人當中最有社群影響力的使用者)來幫忙推播訊息，同時根據[12]的論述，有相同興趣的人很常會聚集在一起，因此除了[4]以外，[10]加入了消費者興趣的判斷，在播送過程中選擇興趣較高的使用者，以提升推播效率以及準確度。

在[10]中用來測試以及實驗的環境是由他們自行所開發的，並非一般其他較為熟識的網路架構模擬環境，其中它們設定賣場中消費者數目為 50 人(其中對廣告有興趣者為 10 人)以及商家數 2 間，在資料方面 IPeR 演算法則是使用 CRAWDDAD[13] 網站中所提供的“IFOCOM06”資料集以及“SIGCOMM 2009”資料集，其中他們使用“SIGCOMM 2009”來產生實驗模擬數據。IPeR 演算法主要是希望只透過對廣告有趣的消費者來幫忙傳送廣告資料以降低在傳送資料的過程當中所需要的花費，其中針對每位消費者都定義了一個 IPeR 變數，此變數值代表該消費者對於正在播送的廣告是否感興趣，該值會隨他們定義的懲罰機制來隨時更新調整。當廣告要往下一位消費者傳送時，即會判斷該消費者 IPeR 值是否有大於閾值以決定是否適合幫忙進行廣告推播，若是，則幫忙進行推播；否則不幫忙進行推播並跳過此消費者。

針對[10]的 IPeR 演算法，我們認為有兩個部分可以加以強化改進。首先，IPeR 演算法後發現他們的演算法內包含了兩層迴圈使得整個演算法的時間複雜度會達到 $O(n^4)$ ，鑒於整個演算法是執行於手持行動裝置上方，若大量地使用 IPeR 演算法進行廣告的推播，將造成行動裝置上的負荷量過大且執行效率差；第二，在過往的研究中[6]-[8]均是使用消費的歷史資料進行消費商品的推薦，然而在[10]中，卻只從消費者的社群資料中判斷消費者的興趣，並推測消費者對於廣告的興趣，如此可能無法有效增加廣告在推播時的效益，因此在機會網路的推播演算法裡除了使用社群資料外，若加入了實際的歷史消費資訊來判斷消費者的興趣以及消費機率將可更有效地選擇推播的消費者同時提升廣告的效益。因此本文使用了社群以及消費資料來選擇有興趣且具有消費潛力的消費者進行廣告的推播，且提出的演算法能在 $O(n^2)$ 內完成廣告的推播，在智慧型裝置執行，相當有效率。

三、方法

本節將探討我們所提出的方法，包含了一個多層模型，如圖 1 以及 ISC 演算法。在多層模型方面，我們一開始會預先做資料蒐集的動作，資料內容包含消費者的消費資訊以及商家資訊，再來我們提出了屬性層，主要用來定義商家的型態，同時此資料將會提供給社交層，

社交層會收集使用者的朋友資訊同時利用曾經消費過的店家進行消費行為的比對，同時商店層也會進行店家間的關聯性比對，此多階層的演算法實行在智慧型手持裝置上將可以為商家或消費者更有效率的播送廣告給附近有共同喜好、共同興趣抑或是有社群關係者，以提升整體廣告效益，以下針對各層做詳細的介紹。

3.1 屬性層

在屬性層當中，此層將消費者的消費資料中所有商家的商家屬性紀錄並提供給社交層使用，其中依照所有商家的屬性將其歸納分成 A 個屬性種類，每間商家的屬性 a_i 可能不只一種，其中 $a_i \in A$ 。舉例來說，一間簡餐店可能會同時具有飲料店以及餐館的屬性標籤存在。

3.2 社交層

在社交層中，依照每位消費者所消費過的商家，將其商家屬性對應紀錄至如表 I 的消費屬性表中，表中，行為目前使用者，列為所有商家屬性值。舉例來說，如有間商家被定義為 a_1 屬性，當消費者 User 的消費資料中有曾經到過該商家的消費紀錄，則於該 User 的 a_1 屬性欄作紀錄；若商家屬性被定義為 a_3 以及 a_5 ，則 a_3 以及 a_5 的屬性欄值各會做一次紀錄，以此類推。

再來將藉由 Cosine Similarity 方法(公式 1)來計算任兩消費者屬性欄位的相似程度，其所求出的值越接近於 1 代表兩者相似程度越高，表示兩者在社群上的關係以及消費行為上的關係相似度很高。而在求算相似程度值之後，我們還將判別兩使用者是否有社群之間的關係(如：兩者是否有朋友關係)，若有社群關係，則將該相似度值乘上一個適當權重值 w 以加重兩者相似程度值；否則以原值表示即可。

表 I
消費屬性表

	a_1	a_2	a_3	a_4	a_5	...	$a_{ A }$
User	1	0	1	0	1	...	0

$$\cos(\text{User}_1, \text{User}_2) = \frac{(\text{User}_1 \cdot \text{User}_2)}{|\text{User}_1| |\text{User}_2|} \quad (1)$$

3.3 商店層

在商店層中，我們先透過資料探勘中的 frequent pattern growth (FP-Growth) 演算法依據所有消費資料找出所有商家與商家之間的關聯性，再將有關聯性的商家歸屬於同一集合 S_i 之內，每間商家有可能被歸屬在不只一個集合之內，並定義集合 Set 用來蒐集所有 S_i ，即 $S_i \in \text{Set}$ 。最後再依據消費者的消費資料將其前往過的商家資料對應至如表 II 所屬的商家集合表中。舉例來說，目前消費者 User 的消費資料中記載他曾經消費過的商家集合內容為 $\{l_1, l_2, l_4, l_7, l_8, l_{10}\}$ ，其中 l_i 皆代表各個不同商家，而透過 FP-Growth 演算法分析找出的商家集合有 $S_1 = \{l_1, l_4, l_7\}$ 、 $S_2 = \{l_2, l_3\}$ 、 $S_3 = \{l_4, l_7\}$ 共三個商家集合，現在依據 User 的消費資料依序將 l_i 所對應到的集合內數值記錄於表 II 中，以此類推。

綜合上述多層模型，最後在每位消費者的智慧型手持裝置上將會存放兩表如表 I 以及表 II，而在下一節當中將會使用這兩表來判斷各種情況應該使用何種廣播機

制較為恰當以及計算使用者之間的社交與消費行為相似程度值。

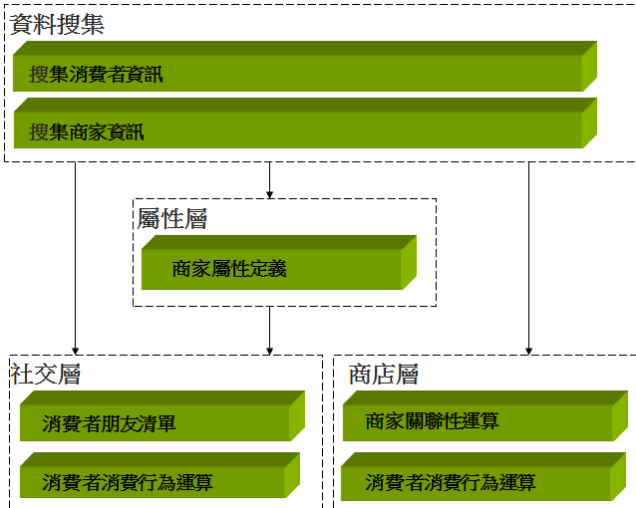


圖 1、多層模型架構

表 II
商家集合表

	S_1	S_2	S_3	...	$S_{ Set }$
User	3	1	2	...	0

3.2 演算法

圖 2 為我們的 ISC 演算法圖形架構，在 ISC 演算法中，我們將智慧型手持裝置中的每則廣告依據演算法中的三個步驟機制來判斷如何廣播至附近的智慧型手持裝置使用者，藉此達到降低花費以及提升廣告效益的目的。

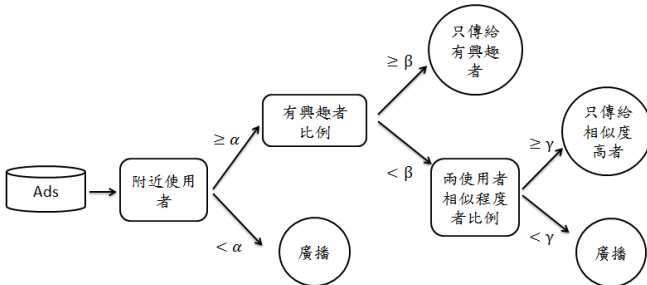


圖 2、ISC 演算法圖形架構

步驟 1：附近使用者人數的多寡將影響整個廣告廣播的花費以及效益，但當附近人數不足時，若還針對廣告加以限制，那整體廣告效益以及將廣告傳達至目標使用者的機率將會很低。所以我們先行一步判斷附近使用者的數量是否有高於定義值 α ，若是，則再進一步透過步驟二判斷；否則將廣告進行廣播給附近所有裝置。

步驟 2：對廣告有興趣者比例的多寡將是影響到整個廣播機制是否能夠提升廣告效益的重要因素，因為有相同興趣者容易聚集在一起[2]。因此，我們將判別附近對廣告有興趣的使用者人數，若是人數比例高達定義值 β 時，則將廣告傳送給附近所有對廣告有興趣的使用者；否則將進入步驟 3 判斷。

步驟 3：透過上一節所記錄的兩個表來開發可能對該廣告有興趣的消費族群。透過 Cosine Similarity(公式 1) 來計算目前使用者與其他使用者的社交與消費行為相似

程度值，當兩使用者所算出的相似程度值越高即代表兩者在社群以及消費相關性都極度相似，如此即有可能也會對該廣告產生興趣而產生消費行為，所以當相似度高於定義值 γ 時，只將廣告傳送給與使用者相似度高的使用者；否則將廣告廣播給附近所有人。

Algorithm ISC

```

Require Neighbors  $\geq 0$ 
1. while(AdsInBuffer)
2.   If (InterestPeopleRate  $\geq \alpha$ ) then
3.     exchangeInterestData()
4.     If(contactWith  $\geq \beta$ ) then
5.       while(InterestPepole)
6.         Send(ad)
7.       End while
8.     Else
9.       while(Neighbors)
10.        If(isSimilarity) then
11.          Send(ad)
12.        Else
13.          Broadcast(ad)
14.        Break
15.      End if
16.    End while
17.  End if
18. Else
19.  Broadcast(ad)
20. End if
21. End while
    
```

演算法一開始需要在裝置上存有附近使用者 Neighbors 的數量以及消費屬性表、商家集合表以及使用者社群朋友清單。AdsInBuffer 記錄目前行動裝置內所包含的所有廣告數量；exchangeInterestData 函式用來判斷正在溝通的對象是否對於目前廣告有興趣；contactWith 代表目前正在溝通的使用者；isSimilarity 則會判斷兩者相似程度高低。

四、模擬結果

4.1 模擬環境

本論文的實驗環境主要是以程式開發軟體 Eclipse[14]支援的 Android 模擬器以及實體智慧手機來模擬演算法在一般生活中手機的實際執行情形，使用的模擬器版本以及智慧型手機型號如表 III 所示。參數部分主要設定(附近消費者人數；廣告數目；朋友個數)，三種參數數目的多寡都會影響到演算法實際執行的效率、決策方式、最後廣告效益、以及裝置所需的記憶體容量。

表 III
模擬裝置表

	模擬器	手機裝置
型號	NexusS	Desire
作業系統	Android 4.1.2	Android 2.2.2
核心版本	armeabi-v7a	Htc-kernel@and18-2 #1

4.2 資料來源

資料方面使用了 CRAWDAD 網站所提供的“SIGCOMM 2009 [13]”會議資料以及英國政府網站釋出

的消費者信用卡消費資料”Wiltshire Council - Credit card spending [15]”。在 SIGCOMM 資料當中，紀錄了共 76 位擁有智慧型手機的參與者資料並安裝了會議提供的軟體來記錄下會議期間的消費者資料，其中包含了消費者興趣以及朋友資料；而在[15]當中我們使用了 2010 年 5 月至 2011 年 8 月共 16 個月的資料來找出消費者之間的相似程度，其中包含 121 個消費者資料以及 1144 間不同店家。

4.3 實作結果

針對於圖 3，此圖是在 Eclipse 上的 Android 模擬器執行 ISC 演算法以及 IPeR 演算法，折線圖代表的意義為兩演算法執行時間，其中橫軸代表在上一節提到的參數設定(附近消費者人數；廣告數目；朋友個數)，總共四組；縱軸所代表的是時間，單位為秒。我們透過折線圖可以清楚發現一般來說 ISC 演算法的執行速度都是快於 IPeR 演算法的，而執行的時間會依廣告數量以及附近人數多寡有明顯的增減，其中在(100;100;70)這組當中，ISC 演算法的效率比 IPeR 演算法足足高出約 4.48 倍。

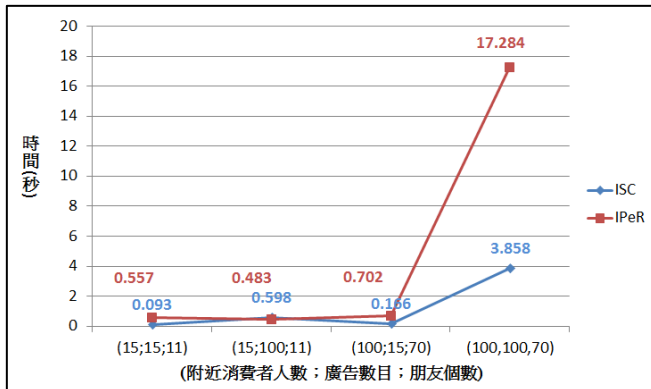


圖 3、Android 模擬器上演算法運算時間

針對於圖 4，此圖是在 Desire 手機上執行 ISC 演算法以及 IPeR 演算法的結果，從結果中可以明顯發現手機的執行效率遠比模擬器的高出很多，而在此結果中更可以明顯觀察出兩演算法的差距，其中各組的倍率差分別是 20.2 倍、1.42 倍、5.01 倍以及 11.18 倍。

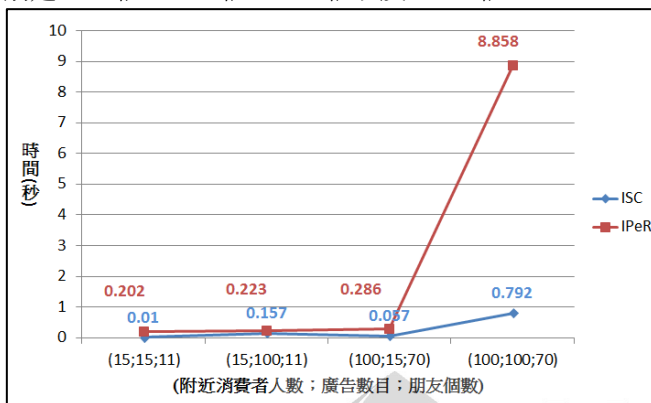


圖 4、Desire 實機上演算法運算時間

最後，針對於圖 3，其中可以觀察出在(15;100;11)這組資料當中 ISC 演算法的處理時間反而比 IPeR 演算法來的高，但在圖 4 的實機上可以在看到 ISC 演算法的處理時間比 IPeR 演算法來的低，所以此狀況應該是模擬器執行上的處理問題，在一般實際手機上我們的 ISC 演算法執行的效率都是比 IPeR 演算法來的好很多。

五、結論

在此篇文章中，我們提出了一個多層模型和ISC演算法，ISC演算法結合了社群、興趣以及消費資訊，能在沒有通訊網路的環境之下透過智慧型手持裝置有效地將來源端的廣告資訊於有限的時間之內推送至目標使用者，並確保在推播的過程中，轉送廣告的中間消費者對轉送的廣告有一定程度的興趣，如此不僅能增加將廣告送到目標消費者的機率更能提升廣告效益，同時更可以從實驗結果中得出本ISC演算法在推播廣告的過程複雜度為 $O(n^2)$ ，縱使需要發送的廣告或是消費者增加，ISC演算法在智慧型裝置上的執行依然相當有效率。透過多層模型與ISC演算法，機會網路成功與社群平台資料結合，廣告的發送更有效率與效益。

參考文獻

- [1] C. Boldrini, M. Conti, and A. Passarella, "Exploiting users social relations to forward data in opportunistic networks: The HiBoP solution," *Pervasive and Mobile Computing*, vol. 4, pp. 633–657, 2008.
- [2] E. M. Daly and M. Haahr, "Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs," in *8th ACM MobiHoc Proceedings*, 2007, pp. 32–40.
- [3] P. Hui et al., "Bubble Rap: Social-based forwarding in Delay Tolerant Networks," in *9th ACM MobiHoc Proceedings*, 2008, pp. 241–250.
- [4] A. Mtibaa, M. May, and M. Ammar, "Social Forwarding in Mobile Opportunistic Networks: A Case of PeopleRank," in *Communication and Social Networks*. Springer, 2012, vol. 58, pp. 387–425.
- [5] Nan Li and Guanling Chen, "Multi-layered Friendship Modeling for Location-Based Mobile Social Networks", *MobiQuitous '09*.
- [6] Vincent S. Tseng, Jeng-Chuan Chang, Kawuu W. Lin, "Mining and Prediction of Temporal Navigation Patterns for Personalized Services in E-Commerce", *SAC'06*, April, 23-27, 2006, Dijon, France.
- [7] Noraswaliza Abdullah, Yue Xu, Shlomo Geva, Jinghong Chen, "Infrequent Purchased Product Recommendation Making Based on User Behaviour and Opinions in E-commerce Sites", *Data Mining Workshops (ICDMW)*, 2010 IEEE International Conference on.
- [8] Yan Zheng Wei, Luc Moreau, and Nicholas R. Jennings, "Learning Users' Interests by Quality Classification in Market-based Recommender Systems", *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, VOL. 17, NO. 12, DECEMBER 2005.
- [9] Xiaoyuan Su and Taghi M. Khoshgoftar, "A Survey of Collaborative Filtering Techniques", *Journal Advances in Artificial Intelligence Volume 2009*, January 2009 Article No. 4.
- [10] Soumaia Al Ayyat, Sherif G. Aly, Khaled A. Harras, "Interest Aware PeopleRank-Towards Effective Social-Based Opportunistic Advertising", *2013 IEEE Wireless Communications and Networking Conference (WCNC): SERVICES & APPLICATIONS*.
- [11] Xiong YP, Sun LM, Niu JW, Liu Y. Opportunistic networks. *Journal of Software*, 2009,20(1):124-137.
- [12] M. McPherson, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, no. 1, 2001.
- [13] "CRAWDAD data set thlab/sigcomm2009 (v.2012-07-15)" Download from <http://crawdad.cs.dartmouth.edu/meta.php?name=thlab/sigcomm2009>
- [14] Eclipse <http://www.eclipse.org/>
- [15] "Wiltshire Council - Credit card spending" data set Download from <http://data.gov.uk/dataset/wiltshire-council-credit-card-spending>
- [16] Gérard Hermet and Jacques Combet, "Mobile Internet Monetization: A Methodology to Monitor in Real Time the Cellular Subscriber Transactional Itinerary, from Mobile Advertising Exposure to Actual Purchase", *Mobile Business (ICMB)*, 2011 Tenth International Conference on.
- [17] Li Liu and Yu Liu, "A Resource Discovery Algorithm Based on Mobile Ad hoc Network and Its Implementation", *Wireless Communications Networking and Mobile Computing (WiCOM)*, 2010 6th International Conference on.
- [18] Li Dongsheng, Xiang Yong, Shi Meilin, Wang Xiaodong, "Busy-zone Avoidable Energy Consume Balance Routing in Mobile Ad Hoc Networks", *Computational Engineering in Systems Applications, IMACS Multiconference on*, 2006, Page(s): 1383-1387.