

雲報專欄：行動社群媒體之影響力最大化

交通大學資訊工程系 彭文志教授/技術專家委員會 委員、 朱文園 博士生

傳統的廣告行銷方法，是在不同的傳播媒體上購買廣告，像是電視、廣播和報紙。然而，在不同傳播媒體上購買廣告花費的金額是非常龐大的，而且觀眾未必對此廣告有興趣，有時甚至造成反感。而在社群網路中病毒式行銷 (Viral Marketing) 利用使用者之間互相傳播廣告訊息，快速地擴大影響力。病毒式行銷在社群網路中成功的原因在於使用者較能接受認識人所傳播的訊息，且社群網路提供良好的傳播媒介。此行銷手法能夠有效減少廣告成本，卻可以帶來明顯的收益。在社群網路中，影響最大化 (Influence Maximization) 是找出固定個數的種子使用者，讓這些種子使用者能夠最大化被影響的使用者個數。解決此問題能夠在有限的經費下得到最大的影響效果，如此除了有效地節省廣告散佈的費用，也能夠有效地增加廣告收益。在行動社群網路中，影響最大化中的目標訊息是指一個需要吸引愈多使用者到此地點，像是商店、旅館和餐廳，這些地點希望有更多使用者到訪以增進商業收益。

在 [1] 中討論行動社群網路中的影響最大化問題，給定一地理範圍的情形下，選取一組最佳的種子使用者使範圍內被影響的人數最多。但在其研究中使用者只有一個代表地點，無法表示使用者在行動社群網路中的移動行為；另外此研究並沒有說明如何選擇被影響的區域。最後在這研究中，使用者間的傳播機率 (Propagation Probability) 僅用和其他使用者的連接數來計算，因此並不能準確的反映出現實行動社群網路中的影響模式。

行動社群網路上的影響最大化和傳統社群網路上的影響最大化是很類似的問題，不過在行動社群網路中有些獨特的特性需要考量。在傳統社群網路中，使用者間的傳播機率僅考慮使用者的社交關係和使用者偏好，並未將位置資訊納入考量。而在行動社群網路中，因為使用者必須實際到達要行銷的地點後才能將資訊傳播出去，因此需考量使用者是否會去行銷的地點。在 [2] 的研究中說明使用者在行動社群網路中的行為是受到自身的喜好為主，使用者之間的影响為最弱，需要從使用者在行動社群網路中的打卡資料推測使用者是否會在行銷地點打卡。此外每位使用者針對不同的地點會去的機率皆不相同，根據行銷地點不同，使用者間的傳播機率也會受影響。所以，在傳播機率的設定上，必須將行銷地點和使用者的行為模式納入考量，才能精確的反映真實情形。

在行動社群網路的影響最大化問題中，為了評估使用者之間訊息的傳播機率，必須從使用者的打卡資料中探勘使用者的移動行為，藉此推估使用者是否會去被行銷的地點。因此，為了評估使用者和被行銷地點的關係，從使用者的打卡行為，推論出使用者的移動模式，進而推斷使用者是否會去行銷地點。

為了從軌跡資料中找出使用者移動的行為，目前已經有許多研究專注於此問題上。由於軌





跡資料是由地點並且依照時間順序所組成的序列，所以許多研究專注於從軌跡中找出序列特徵 (Sequential Patterns)。再者，軌跡不只包含了空間的資料，也包含了時間的資料。因此，時空特徵 (Spatio-temporal Patterns) 也可以從軌跡中被取出，用來代表使用者移動的行為。在 [4] 的研究中，從軌跡中取出不同的段落 (Segments) 用以代表使用者移動的行為。另外在 [5] 中，使用一個以分群 (Clustering-based) 為主要的概念，找出在一段時間內相似的移動行為。而在 [6] 和 [7] 中，使用 T-Pattern 表示在相似時間間隔內頻繁出現的地點序列 (Frequent Sequence of Locations)，並用 T-Pattern 代表使用者移動的行為。最後在 [8] 的研究中，從軌跡資料中推測使用者是使用何種交通工具，這表示不再只是從軌跡中找尋特徵，而可以從軌跡中推測出更高一層的含意。另一方面，有一些研究是找出個別使用者的移動特徵。在 [9] 的研究中，使用一個混和的模型代表個別使用者移動的行為，並且使用此模型預測使用者在特定時間的位置。另外在 [10] 中，延伸之前的 T-pattern，設計出 T-pattern Tree 用以代表個別使用者的移動行為，並使用此資料結構預測使用者下一步的位置。不過上述的研究都是針對採樣率 (Sample Rate) 較高的軌跡中找出使用者移動的特徵，在行動社群網路中，使用者的打卡資料平均一天只有一筆資料，因此，打卡資料無論是在空間上或時間上都很稀疏，以往的研究無法適應此種情形。如何克服打卡資料的空間時間的稀疏性，找出使用者的移動行為模式是一個重要的研究議題。

在行動社群網路中移動和社交之關係

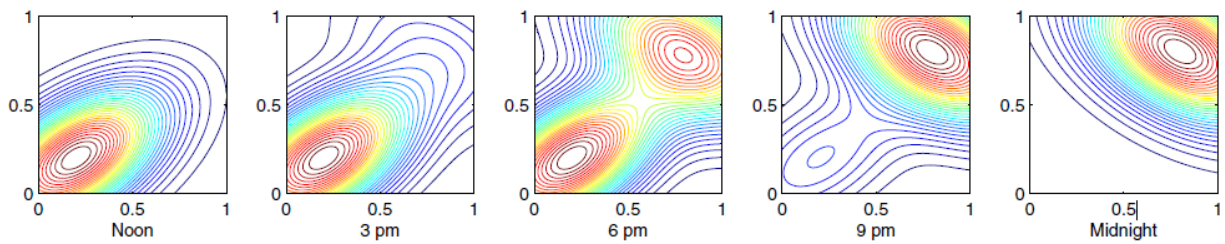
在 [11] 中，認為打卡行為可以分為兩個狀態，家 (Home) 和工作 (Work)，其使用週期性移動模型 PMM (Periodic Mobility Model) 如下所示：

$$P[x(t) = x] = P[x_w(t) = x | c_w(t) = H] \cdot P[c_w(t) = H] + \\ P[x_w(t) = x | c_w(t) = W] \cdot P[c_w(t) = W]$$

使用者在時間 t 時在位置 x 的機率可以分為兩種：一個是在家 (H) 的機率分布而另一個為工作 (W) 的機率。而此兩種機率又可以透過計算 $P[x_w(t) = x | c_w(t)]$ 與 $P[c_w(t)]$ 得到。前者為一個二維高斯分布，描述了一個使用者在特定狀態下其位置在地圖上的分布。而後者則是在時間 t 時，使用者在特定狀態的機率，是使用一維高斯分布描述。另外，作者也提出了社交移動模型 PSMM (Periodic Social Mobility Model)，提到移動行為也會被朋友所影響，使用者會去比平常遠的地方，可以用下列式子表示：

$$P[x(t) = x] = P[x_w(t) = x | z_w(t) = 1] \cdot P[z_w(t) = 1] + \\ P[x_w(t) = x | z_w(t) = 0] \cdot P[z_w(t) = 0]$$





圖一、週期性移動模型 PMM：使用者位置的機率密度隨時間由狀態“W”到狀態“H”

其中 $z_u(t) = 1$ 代表時間 t 時，使用者是受社交影響，而 $z_u(t) = 0$ 則代表使用者是在 PMM 中。其中 $P[x_u(t) = x | z_u(t) = 1]$ 會隨 x 以及 t 距離朋友們打卡的時間差與地點的距離成指數減少，細節如下列式子所示：

$$P[x_u(t) = x | z_u(t) = 1] \sim \sum_{\{t_i, x_i\} \in I_u} |t_i - t|^{-\alpha} \cdot \|x_i - x\|^{-\beta}$$

在圖中，可以看到使用者的位置機率分布會因為時間的關係，高密度的地方由工作的地點漸漸往家的位置移動，這個現象與現實體驗相當符合。但是在這個研究當中，時間是影響移動模型的重要因素，使用打卡時間來判斷該使用者現在的狀態（在家或是工作中）。

在行動社群網路中時間對使用者移動模型的影響

在之前的研究中，通常使用了行動社群網路中的社交關係、地理資訊來產生社交歷史 [11]、社交—地理特性 [12]、地理影響 [13] 和地理—社交關係 [14]，但時間因素並沒有被重視，因此在 [15] 中，提出了在一般人的活動模式中，都會有週期性且規律的移動，因此在預測某時間點，該使用者可能所在的地點時，透過該時間點的時間特性 (Temporal State)，分為每日模式 (Daily Pattern) 和每週模式 (Weekly Pattern)，相同特性的時間點預期使用者會有較高機率在該時間出現在某個地點。將時間特性加進之前的方法中，來提高推測的正確率。

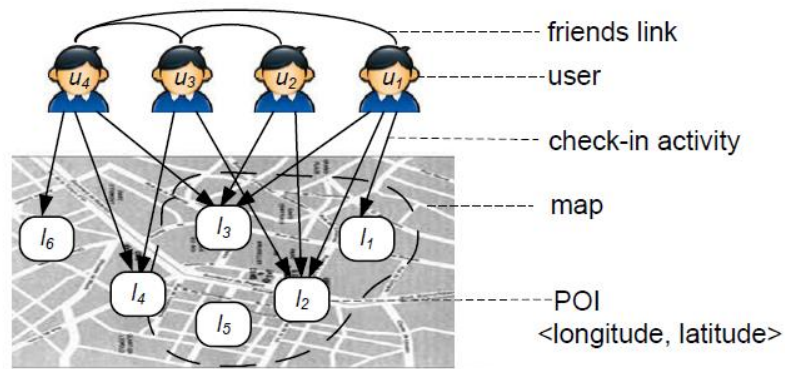
地理資訊對興趣點 (POI) 的影響

在 [2] 這篇研究中，作者將影響某一使用者的 POI 的因素分為三種：使用者偏好 (User preference)、社交影響 (Social Influence) 以及地理資訊影響 (Geographical Influence)。使用者偏好是指由使用者自身對 POI 的喜好程度。社交影響則是朋友對使用者前往 POI 的影響，包含了使用者較可能前往朋友曾經去過的 POI。而地理資訊則是說明了使用者對於熱門地點以及熱門地點附近的 POI 會有更高的意願前往。圖二顯示使用者前往 POI 的機率不僅受社交關係影響，也會和附近熱門的 POI 有關。

在[2]中，發現使預測打卡地點最佳化的混合參數為 0.7:0.1:0.2（使用者偏好：社交影響：



地理資訊影響)。此結果和 [3] 的研究結果類似，因此可以知道，在行動社群網路中，使用者的打卡行為是受到自身影響為主，其次是因為地點很熱門而被影響，最後才是被朋友所影響。若要設定行動社群網路中的傳播機率，因為在行動社群網路上的影響是非常小的，所以更必須考慮使用者的移動行為，才能反映行動社群網路中資訊真實的傳播情形。



圖二、使用者前往 POI 的機率不僅受社交關係影響，也會和附近熱門的 POI 有關

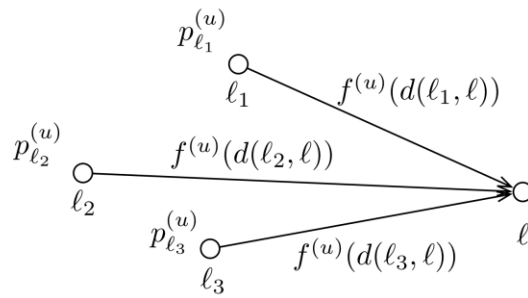
行動社群之位置(商店) 推薦問題定義

在一個行動社群網路中，給一個被行銷地點 l 和一個常數 k ，則地點行銷問題是選擇一個集合 S 包含 k 個種子使用者，使 S 能夠最大化 $\sigma(S)$ ，其中 $\sigma(S)$ 為選擇 S 為種子使用者後被影響到 l 打卡使用者個數的期望值。

建立使用者在行動社群網路中移動模型

在 [2] 中，指出使用者的打卡行為主要是受到使用者移動行為的影響，而受朋友影響只佔了很低的比例。此外，在 [3] 的研究中，分析了在行動社群網路中，社交和移動的關係，並提出了兩種移動模型 PMM 和 PSMM。這兩種模型都是基於混和高斯模型 (Mixture Gaussian Distribution)，並考慮了不同時間有著不同的移動模型，而 PSMM 則多考慮了社交的關係。不過由於此模型難以從機率密度 (Probability Density) 推斷使用者在行銷地點打卡的機率，所以此模型很難直接套用在計算傳播機率上。因此，在我們所發表之論文 [16] 提出以距離為主的 (Distance-based) 移動模型 DMM (Distance-based Mobility Model)，用以推斷使用者之間的傳播機率。





圖三、DMM 的概念示意圖

由於每個使用者的打卡資料大約只有 10-100 筆紀錄，相對於一般軌跡資料動輒上千至上萬筆資料顯得非常稀少。因此很難單純使用資料的位置和密度判斷使用者去某一個地點的機率 [9, 10]。為了從打卡資料中取得每個使用者的移動模型，我們提出 DMM 來解決打卡資料的稀疏性。由於資料的稀疏性，單純從二維空間座標建立機率模型會很不準確。因此我們使用距離的概念，由於距離是一維資料，相較於二維資料，距離較能夠應付資料的稀疏性。DMM 的概念是先取得每個去過地點的平穩態 (Stationary Distribution)，接著預測使用者從去過地點到行銷地點的機率。根據以上的討論，給定一個行銷地點 l ，則使用者 u_i 的 DMM 可以被表示成

$$p_{u_i}(l) = \sum_l p_l^{(i)} f^{(i)}(d(l, l))$$

在此等式中 $p_{u_i}(l)$ 代表著使用者 u_i 到一個地點 l 的機率密度函數， $p_l^{(i)}$ 代表 u_i 對於地點 l 的穩態機率，而 $f^{(i)}(d(l, l))$ 則代表著 u_i 從 l 移動到 l 的機率，其中 $d(l, l)$ 代表著 l 和 l 的距離，其概念如圖三所示。為了計算 $p_l^{(i)}$ ，我們使用重啟型隨機漫步 (random walk with restart) 模擬使用者在行動社群網路中移動的行為。為了計算 $f^{(i)}(d(l, l))$ ，需要選擇一個合適的機率分布成為 $f^{(i)}$ 。然而， $f^{(i)}$ 代表著 u_i 的移動距離的分布，因此，觀察 u_i 到下一個打卡位置的距離。透過觀察 u_i 的移動距離，我們可以選擇出一個合適的分布用於 $f^{(i)}$ 。

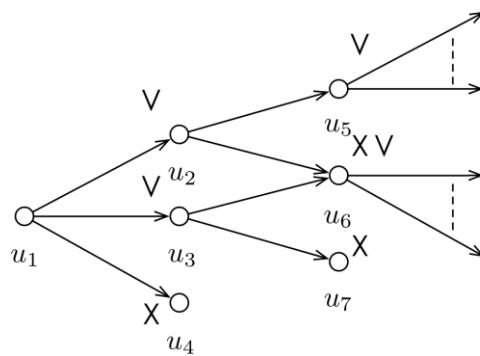
行動社群網路的散播模型

散播 (Diffusion) 代表一個新的文章或是訊息被傳播到其他使用者的情形。其中，活動中的 (Active) 使用者代表他發送了此訊息，反之則稱為非活動中的 (Inactive) 使用者。在病毒式行銷中，目前已經有許多散播模型被提出。針對一個使用者是否被影響 (Influence) 目前主要有兩種模型，LTM (Linear Threshold Model) 和 ICM (Independent Cascade Model)。在 LTM 中，一個使用者是否被影響是根據被影響的程度是否有超過定義的活動門檻值 (Activation Threshold)，若超過此門檻值，則此使用者為活動中的使用者。當一個使用者變成活動中的使





用者時，此使用者會一直持續地影響有連接的非活動中使用者。另一方面，在 ICM 中，當一個使用者成為活動中時，此使用者會有一個機率讓連接的非活動中使用者成為活動中使用者，每個連接的非活動中使用者變成活動中使用者的機率是各自獨立的。LTM 和 ICM 最大不同的地方在於 ICM 影響連接的非活動中使用者只有一次機會，而 LTM 則會一直持續地影響下去直到連接的非活動中使用者變成活動中使用者。以上這兩種模型都是建立在訊息散播的次數上。此外，HDM (Heat Diffusion Model) 是一種新興的散播模型，其概念是從熱力學中熱傳導的概念而來。使用熱的多寡表示影響的程度，若使用者含的熱超過一個門檻值則變成活動中的使用者。另外，HDM 是基於時間模擬，和 ICM、LTM 基於訊息散播的次數上極為不同。並且，HDM 和 LTM 類似，活動中使用者會一直持續地影響下去直到連接的非活動中使用者變成活動中使用者，不過使用者之間的影响是根據每個使用者所攜帶的熱所決定，而不需變成活動中使用者才可以影響其他使用者。在行動社群網路中，可以把打卡行為拆成兩個部分，影響和打卡，所以相較於其餘散播模型，可以輕易地使用 ICM 描述此行為。因此，在所發表之論文 [16] 使用 ICM 描述行動社群網路中資訊散播的行為。



圖四、一個使用 ICM 描述資訊在行動社群網路中資訊散播的例子

圖四顯示一個使用 ICM 描述資訊在行動社群網路中資訊散播的例子。首先，假設使用者 u_1 是一個種子使用者，且 u_2 、 u_3 和 u_4 是 u_1 的朋友。 u_2 、 u_3 和 u_4 都會收到 u_1 在行銷地點打卡的訊息。在 u_1 的朋友 u_2 收到此訊息後， u_2 也在被行銷地點打卡，但是 u_1 的朋友 u_4 卻沒有。 u_2 和 u_3 都因為被共同的朋友 u_1 所影響在行銷地點打卡，則他們的共同朋友 u_6 則會收到兩個關於在行銷地點打卡的訊息。 u_6 有兩次決定是否要在行銷地點打卡，而這決定是基於 u_6 自己的移動行為。在這例子中， u_6 在第二個決定中決定了在行銷地點打卡，並且影響其餘沒有在行銷地點打卡過的朋友。

所發表之論文 [16] 主要的貢獻是透過使用者的過去歷史打卡行為，提出對使用者的出沒地點機率，建立一個模型機率，然後根據這個使用者出沒地點模型機率，估算在行動社群網路上，其擴散機率的估計值，並因此估算機率值，進一步能夠選出 K 個具有影響力的使用者能夠有很大的機率到所欲推薦商家地點進行打卡，影響更多的使用者。在我們所發表之 KDD





2015 論文中 [16]，針對了擴散機率模型，分別提出了靜態機率，GMM-based 的機率模式與 distance-based 機率模式。隨著行動社群媒體的興起，社群媒體的內容也反映使用者在實體世界的行為模式，而透過這些行為模式的探勘與分析，將挖掘出具商業價值之知識，所發表的論文即在展示如何透過行動社群媒體找尋可能的顧客群，相信在未来隨著行動付款方式的便利，行動社群媒體之影響力最大化研究更顯示其研究的重要性與實務之價值。

參考文獻

- [1] G. Li, S. Chen, J. Feng, K.-I. Tan, and W.-S. Li, "Efficient Location-Aware Influence Maximization," in *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2014
- [2] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting Geographical Influence for Collaborative Point-of-Interest Recommendation," in *ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 2011.
- [3] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and Mobility: User Movement in Location-based Social Networks," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2011.
- [4] H. Cao, N. Mamoulis, and D. W. Cheung, "Mining Frequent Spatio-temporal Sequential Patterns," in *IEEE International Conference on Data Mining (ICDM)*, 2005.
- [5] P. Kalnis, N. Mamoulis, and S. Bakiras, "On Discovering Moving Clusters in Spatio-temporal Data," in *International Symposium on Spatial and Temporal Databases (SSTD)*, 2005.
- [6] F. Giannotti, M. Nanni, and D. Pedreschi, "Efficient Mining of Temporally Annotated Sequences," in *SIAM International Conference on Data Mining (SDM)*, 2005.
- [7] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory Pattern Mining," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2007.
- [8] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding Mobility based on GPS Data," in *International Conference on Ubiquitous Computing (UbiComp)*, 2008.
- [9] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, "A Hybrid Prediction Model for Moving Objects," in *IEEE International Conference on Data Engineering (ICDE)*, 2008.
- [10] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "Wherenext: A Location Predictor on Trajectory Pattern Mining," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2009.
- [11] D. Brockmann, L. Hufnagel, and T. Geisel, "The Scaling Laws of Human Travel," *Nature*, vol. 439, pp. 462-465, 2006.
- [12] J. Cranshaw and T. Yano, "Seeing a Home Away from the Home: Distilling





- Proto-neighborhoods from Incidental Data with Latent Topic Modeling," in *CSSWC Workshop at NIPS*, 2010.
- [13] J. Fieberg, "Kernel Density Estimators of Home Range: Smoothing and the Autocorrelation Red Herring," *Ecology*, vol. 88, pp. 1059-1066, 2007.
- [14] J. Chang and E. Sun, "Location3: How Users Share and Respond to Location-based Data on Social Networking Sites," in *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2011.
- [15] H. Gao, J. Tang, X. Hu, and H. Liu, "Modeling Temporal Effects of Human Mobile Behavior on Location-based Social Networks," in *ACM International Conference on Information and Knowledge Management (CIKM)*, 2013.
- [16] W.-Y. Zhu, W.-C. Peng, L.-J. Chen, Kevin Zheng and Xiaofeng Zhou, "Modeling User Mobility for Location Promotion in Location-based Social Networks," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2015.

