ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2017 June: 23(2): 89~105 http://dx.doi.org/10.13088/jiis.2017.23.2.089

Hot spot DBC: Location based information diffusion for marketing strategy in mobile social networks*

Jegwang Ryu
The Department of Computer Science, Yonsei University
(*jg0103@yonsei.ac.kr*)

Sung-Bong Yang
The Department of Computer Science, Yonsei University
(*sbyang@cs.yonsei.ac.kr*)

As the advances of technology in mobile networking and the popularity of online social networks (OSNs), the mobile social networks (MSNs) provide opportunities for marketing strategy. Therefore, understanding the information diffusion in the emerging MSNs is a critical issue. The information diffusion address a problem of how to find the proper initial nodes who can effectively propagate as widely as possible in the minimum amount of time. We propose a new diffusion scheme, called Hotspot DBC, which is to find k influential nodes considering each node's mobility behavior in the hotspot zones. Our experiments were conducted in the Opportunistic Network Environment (ONE) using real GPS trace, to show that the proposed scheme results. In addition, we demonstrate that our proposed scheme outperforms other existing algorithms.

Key Words : mobile social networks, information diffusion, machine learning, NCCU, viral marketing

Received : March 18, 2017 Revised : April 30, 2017 Accepted : May 15, 2017
Publication Type : Regular Paper(Fast-track) Corresponding Author : Sung-Bong Yang

1. Introduction

Online Social Networks (OSNs) such as Facebook and Twitter are an important role for sharing information among users. Many companies exploit social information of OSNs for marketing strategy (Ma et al., 2008). For example, to effectively minimize the marketing cost, companies want to find the people who are influential customers for promotion of the goods by *word-of-mouth* advertising technique, wherein the customers themselves advertise the products and events to other people (Richardson and Domingos, 2002). In addition, with increase of mobile devices, people can go even more easily than before. In addition, the evolution of mobile devices and the technological advancements in wireless network technique, ad-hoc communication has been a new concept of network which can deliver a message by *a store, carry and forward* manner

* This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (2016R1A2B4010142).

without the help of traditional cellular infrastructures (Conti et al., 2010). As a result, the emerging of mobile social networks (MSNs) open opportunities for marketing strategy (Chen et al., 2010). Before we use MSN as a platform for marketing strategy, many challenges have to be addressed. The vital research one of the challenges for marketing applications is the information diffusion from the proper initial users by *word-of-mouth*. In this paper, we address a problem of how to find the proper initial small users who can effectively propagate as widely as possible and minimize the diffusion time, referred to as the *diffusion minimization*. Diffusion minimization is appropriate in MSNs because each node can spread information by a store-carry-forward technique to enable communication unlike different cellular infrastructures information. Several papers solve the problem through information dissemination in various social networks (Myers et al., 2012; Lu et al., 2014; Chen and Xiong, 2015). They have addressed to solve the problem of diffusion minimization in MSNs. However, this scheme is not suitable for dynamically changed and large complex networks because they do not consider node's mobility pattern. To solve this problem, we propose a new information diffusion scheme that minimize the spreading time in MSNs. The proposed scheme is hotspot-based information diffusion using DBSCAN, is called Hotspot DBC; the basic idea exploits the hotspot zones which peoples have the relationship each other frequently in. For example, students frequently visit same locations such as school, home and library. If one of them is selected as influential node, they have more high contact probability because they are same community. Therefore, in our scheme, we consider mobile social behaviors only in hotspot zones instead of entire network area. Each node periodically records its location information and analyzes the regularity through the recorded log and social network analysis methods. After, influential nodes are selected in communities. Note that an influential node is the most active node in each community. We were conducted in ONE simulator (Keränen et al., 2009) using NCCU (Tsai and Chan, 2015), to show that the proposed scheme results in MSNs. In addition, we demonstrate that our proposed scheme outperforms other existing algorithms with various communication range and the number of nodes and ratio of k influential nodes.

The contributions of this paper can be summarized as follows.

- We introduce a new scheme to solve the problem of diffusion minimization in MSNs, by exploiting the social behavior of each node, where Hotspot DBC extracts only essential location information. Our scheme using the mobile patterns is critical to marketing strategy in mobile social networks.
- We refine the contact probabilities using the community concept for large-scale MSNs: We employ the machine learning techniques by DBSCAN and k-means to classify community structures, where the contact probabilities be disseminated between devices in wireless

network environments.

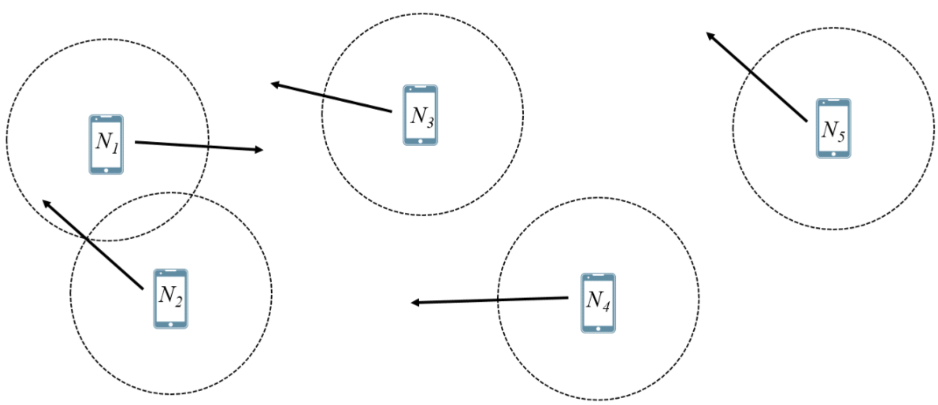
- We were conducted extensive simulations in ONE simulator using GPS trace, and compare the results of our proposed and existing schemes.

The rest of this paper is organized as follows. Section 2 explains the background and related work and we explain to help better understand our aim to solve the diffusion minimization, the assumptions for the system model, and the problem statement. In Section 3, we describe how the proposed scheme works in detail. In Section 4, we present the simulation results and the conclusion is presented in Section 5.

2. Background and Related Work

2.1 Mobile Social Networks

Recently, most smart devices have the technique to communicate other devices in Mobile Social Networks (MSNs) through Bluetooth 4.0 and Wi-Fi Direct. Therefore, MSNs are a new concept of network which can deliver a message by *a store, carry and forward* manner without the help of traditional infrastructures such as 3G and 4G. MSNs are known as Delay Tolerant Network (DTNs) or Opportunistic Networks (OPPNETs) (Conti et al., 2010). But MSNs consider user's social behavior. <Figure 1> illustrates an example of exchange information in MSNs. There are node N_1, N_2, N_3, N_4 and N_5, where they have circles and arrow which indicate the moving direction and communication range, respectively. N_1 and N_2 can exchange information without the help of traditional infrastructures. But N_3, N_4 and N_5 send the message because they do not within communication range. Therefore, nodes can opportunistically deliver/receive messages to/from each other, many studies exploit human behavior for publishing information in the research of wireless networks.



〈Figure 1〉 Information exchange in MSNs

2.2 Influence maximization

As result of emerging of OSNs such as Facebook or Twitter, the influence maximization is study to find a small group of influential individuals for viral marketing. Domingos and Richardson (Domingos and Richardson, 2001) first have addressed the problem. Kemp et al. (Kemp et al., 2003) proposed a greedy algorithm with approximation ratio of $(1-\frac{1}{e})$. Wang et al. (Wang et al., 2010) designed a community-based greedy algorithm. In real world, influence maximization is utilized in various fields of biology (Christakis and Fowler, 2007), marketing (Domingos and Richardson, 2001), and data science (Bakshy et al., 2012), based on user behaviors. The operational models of influence maximization are two fundamental stochastic models, which are the independent cascade (IC) model (Kemp et al. 2003) and the linear threshold (LT) model (López-Pintado, 2008); these two models propagate through individual social relationship. Influence maximization is summarized by the following equation:

$$\underset{I \subset V}{\operatorname{argmax}}\, \sigma(I), \quad |I| \leq k \tag{1}$$

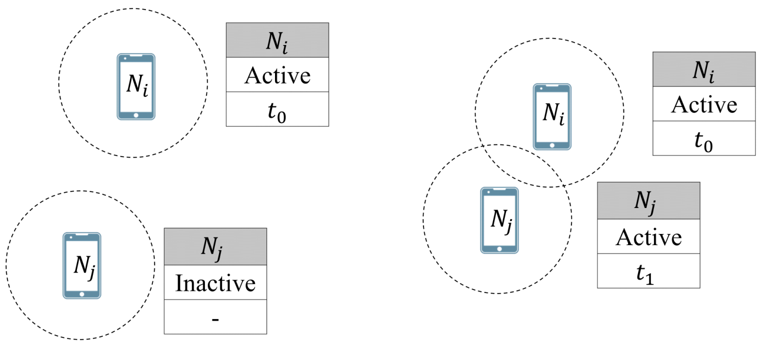
where V is the set of nodes in social network graph, I is the set of initial influential nodes, and $\sigma(I)$ is the expected number of total diffused nodes from I at the end of the influence maximization process in social networks.

2.3 Information diffusion in MSNs

With different from the influence maximization problem, information diffusion is studies of how to find the proper initial small users who can effectively propagate as widely as possible and how to minimize the diffusion time. The diffusion strategy mainly exploits social behaviors such as contact frequency and mobility pattern, to publish information because the network topology changes dynamically with time. Lu et al. (Lu et al., 2014) first proposed a community based algorithm and a distributed set-cover algorithm. Recently, Chen and Xiong (Chen and Xiong, 2015) designed social feature-based diffusion algorithm using machine learning technique. For mobile cellular data offloading scheme, Han et al. (Han et al., 2012) suggested a data offloading scheme formulated from research on information diffusion in MSNs.

2.4 Information diffusion model

Unlike influence maximization model such as IC or LC, Information diffusion in MSNs consider dynamically social behavior and diffusion time. Therefore, we introduce our diffusion model and the assumption for the operational model of information diffusion. Each mobile node has either active or inactive state: active node can diffuse the information. However, inactive node cannot deliver the information and receive the information from only active nodes. They are switched the state from inactive to active and record the first contact time from the active nodes. <Figure 2> (a) and (b) show examples of the diffusion between node N_i



(a) Time t_0 (b) Time t_1

〈Figure 2〉 Diffusion at t_0 and t_1

and N_j at time t_0 and t_1, where each node has a circle which indicates the communication range. Node N_i has active and N_j has inactive at time t_0. When N_i and N_j are in a communication range at time t_1, they have a contact. Then, N_j become active and records first contact time.

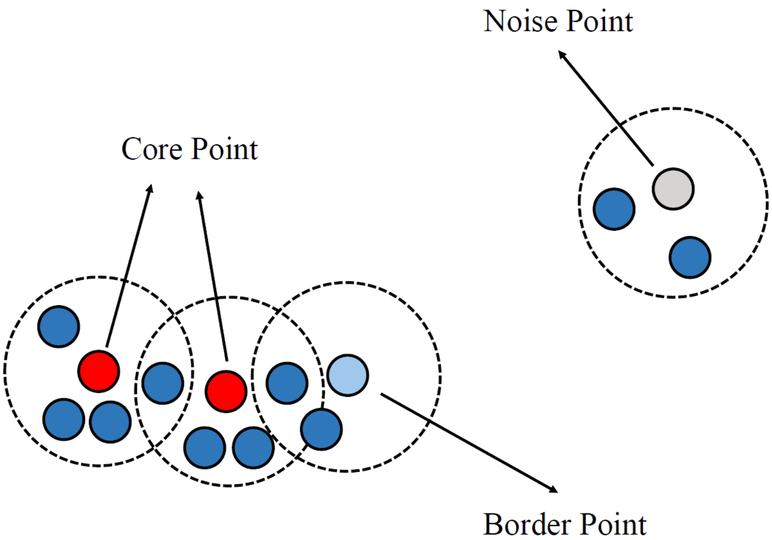
2.5 NCCU Trace Data

Real trace mobility models play an important role in MSNs; understanding the state of real network can predict future dynamically changed contacts for effectively deliver the message by *a store, carry and forward* manner. Recently, NCCU Trace Data (Tsai and Chan, 2015) was designed by a location-aware and behavior-aware Android application. The app installed on the smart devices of National Chengchi University students and collected their daily Global Positioning System (GPS). NCCU real trace model is closely linked with their mobility behavior in the real world. According to the student's location information in each trajectory, we assume that nodes have a contact if they located in certain communication range.

2.6 DBSCAN clustering

Compared to *k*-means, DBSCAN has many advantages because density based clustering technique does not consider number of clusters and can remove outliers (Ester et al., 2012). Basic concept of DBSCAN algorithm in (Ester et al., 2012) can be defined as follows. The points are classified as core points, density reachable points and outliers.

- Eps ε: the radius of each point p_i
- *MinPts*: minimum number of points within ε.
- Core point: A point if at least *MinPts* are within ε.
- Directly Density-reachable: A point p is directly density-reachable from a core point q_i if p_i is within ε of q_i.



〈Figure 3〉 DBSCAN clustering: core, border and noise points with minPt = 5

Algorithm 1 DBSCAN (ε, *MinPts*, *D*)

D: Dataset
ε: the neighborhood distance
Minpts: the minimum number of points in ε

```
for each point x in dataset D
    if x is 'visited'
        continue next point
    end if
    mark x as 'visited'
    if x is not yet classified then
        if FindNeighborPts (x, ε, MinPts) >= MinPts then
            Collect all points density-reachable from x and assign them to a new cluster.
        end if
    end if
    else
    Assign x to noise
    end else
end for each
return Assigned clusters or noises
```

- Border point: A Point $border_i$ is a border point if it is not a core point q_i but density-reachable from another core point q_j.
- Noise point: A noise point $noise_i$ is any point that is not a core point or a border point.
- Density-reachable : A point p_i is density-reachable from a point q_i with respect to ε and *minPts* if there is a chain of points between p_i and q_i is directly density-reachable from p_i.
- Density-connected : two point p_i and p_j are density-connected if there is a point p_k such that p_i and p_j are density-reachable form p_k.

In <Figure 3>, the dataset represents result of a clustering approach that takes distance ε and requires minimum number of points mints. As shown in <Figure 3>, when *minPts* is selected as 5, a group is marked as a cluster because they contain sufficient number of points to form a cluster. But two points are marked noise points because they are not a core point or a border point. In Algorithm 1, the pseudo-code of DBSCAN algorithm is given in detail. In this manner, we use DBSCAN to find the hotspot zones. Dataset *P* is node's location information. Each node shares their location information and then, they build and update DBSCAN themselves during warm-up period. Note that each cluster area is a hotspot zone, where accumulated and shared hotspot zones.

3. Proposed scheme

3.1 Overview

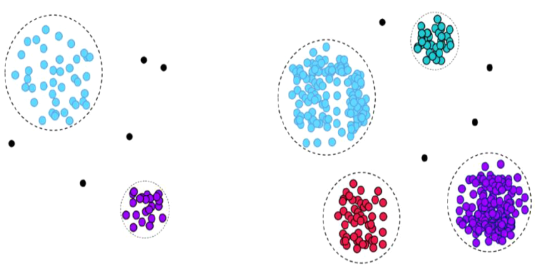
Peoples frequently visit specific locations such as their home, school, a park and a movie theater. Since they have personalities, interests, age, gender, household income, geographic location, and education level. We call an active place a hotspot. If people frequently go to several hotspot zones, they are in same community and the information is likely to be well distributed. Within a community, nodes frequently contact each other. Hence, location-based and community-based diffusion concept can be good strategies for information dissemination in MSNs. We proposed an information diffusion scheme, the basic idea of our proposed scheme exploit community concept. The proposed scheme consists of two steps.

Step 1 [Warm-up period]: In this step, each node records its locations as x, y coordinates and constructs hotspot zones by clustering technique. After, they calculate contacts between their location and hotspot zones, and then classify the mobility patterns. Each node find the influence nodes by sharing the *k* influential node set *I*. Only *k* nodes can be initial active.

Step 2 [Diffusion Stage]: In the beginning of this step, inactive nodes receive information from *k* influential nodes. When all nodes have in active states from *k*-node set *I*, the diffusion process is terminated. Next, we will describe the proposed scheme in details.

3.2 Construction of a hotspot

Hotspot zones are determined at the end of the warm-up period. To construct the hotspot, each node periodically records its current location, which represents the coordinate <x, y> during warm-up period t_τ, where τ is the number of interval. After each node collects recorded coordinate <x, y> in every interval t, they cluster based on DBSCAN in algorithm 1 to find certain hotspot zones. <Figure 4> (a) and (b) show result of DBSCAN data clustering in entire network area at time t_i, t_{i+j}. As shown in <Figure 4>, hotspot zones are different as time goes on because network topology is changed dynamically. As a result, each node exactly find active places while also considering dynamic network environment.



(a) Time t_i (b) Time t_{i+j}

〈Figure 4〉 Network topologies at t_i and t_{i+j}

3.3 Mobility pattern in hotspot zones

Each node exchanges, updates its recorded location information and determines hotspot zones at the end of warm-up period. Then, each node calculates an indicate vector V, consisting of m components, which is presented such as $<v_1, v_2, v_3, \cdots, v_m> = <\frac{F_1}{F_t}, \frac{F_2}{F_t}, \frac{F_3}{F_t}, \ldots, \frac{F_m}{F_t}>$,where m is number of the hotspot, F_i is number of times that node meet with hotspot i and F_t is total contact frequency of hotspot zones. v_i is a contact probability with hotspot zone i during warm-up period. <Table 1> show contact information between virtual hotspot zones and node N_i. As shown <Table 1>, N_i have the contact information of hotspot zones from the coordinate <x, y> in every interval. N_i calculates V_i to extract a pattern in hotspot zones.

〈Table 1〉 Sample contact information between virtual hotspot zones and node N_i

Time	Hotspot ID
t_1	2
t_2	3
t_3	2
t_4	1
…	…
t_τ	4

3.3 Identifying the k influential node set

As construct the communities, similarity between the nodes need to be found. Similarity measure can be written as in the following equation, where $D(N_i, N_j)$ is the distance between a pair of nodes, N_i, N_j

$$D(N_i, N_j) = \frac{\sqrt{\sum_{n=1}^{m}(v_n^i - v_n^j)^2}}{\sqrt{m}} \tag{3}$$

The similarity matrix between nodes is consist of each node' vector V which is a contact probability with hotspot zones. The similarity matrix classified into k-communities by k-means clustering technique. Algorithm 2 describe k-means clustering in detail.

Algorithm 2 k-classification
Input: Similarity matrix
Pick the k points as the initial centroids
while The centroids do not change
Assign all the points to the closet centroid
Recompute the centroid of cluster k
end while
return assigned community ID

After identifying the k-node set, we devise the seed node selection in each community. The k-node set S is selected as follows. First, let the central node in the community i denotes $CNode_i$ and then each $CNode$ is selected through learning the social network. The method of how to learn the social network is the degree of centrality. The degree of centrality calculated as

$$\frac{\sum_{j=1}^{n} L_{ij}}{(N-1)}$$

, where L_{ij} is relationship between the nodes N_i and N_j and they are in same community. In MSNs, the relationship between the nodes presented by the contact information. L_{ij} = 1 if there is a contact between N_i and N_j. On the contrary to this, L_{ij} = 0 if N_i dose not contact N_j. Algorithm 3 is used to determine the k influential nodes. After a warm-up period, the state of the inactive nodes are switched to active from the top-k influential nodes. When all the nodes have been contacted by k-node set I, the diffusion process is terminated.

Algorithm 3 Diffusion node selection
Top-k influential node set: I = { ∅ }
for i = 1 ... k **do**
$CNode_i$ ← the most active node in its community i
$I \leftarrow I \cup \{CNode_i\}$
end for
return k-node set I

4. Performance Evaluation

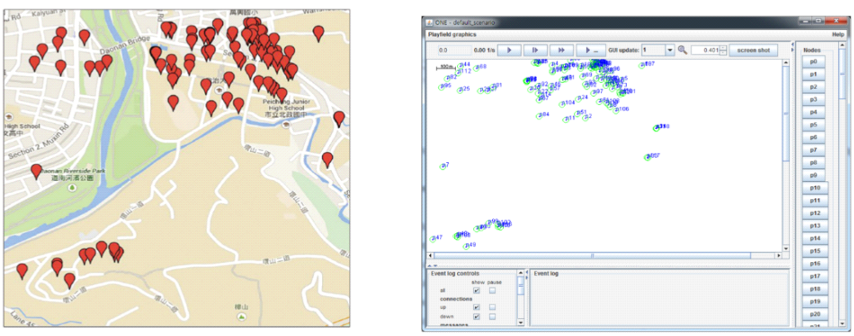
4.1 Simulation environment

In this section, we present the simulation environment. The hotspot-based scheme using DBSCAN (Hotspot DBC) is compared to three schemes—Random, Naïve, and a hotspot-based diffusion scheme using k-means clustering (Hotspot kmeans). The top-k influential nodes in Random are selected randomly. The Naïve algorithm for diffusion minimization is based on Degree centrality (Opsahl et al., 2010) in graph G. In our scheme, the relationship among nodes in G is represented by the contact frequency during warm-up period. A hotspot-based diffusion scheme using k-means clustering is a community-based. Hotspot kmeans is to show performance between clustering techniques, similar to our proposed scheme because we only change clustering technique. In our experiment, we use NCCU Trace

Data (Tsai and Chan, 2015) was designed by a location-aware. NCCU real trace model is closely linked with their mobility behavior in the real world. We use ONE simulator (Keränen et al, 2009) to evaluate our proposed scheme, Hotpot DBC, due to analyze its result. The nodes in a network area move according real dataset, which have GPS location information obtained from NCCU (Tsai and Chan, 2015). <Table 2> shows the parameters of the simulation environment, in detail. The number of nodes is 115. Since k should be a small value in the real world, the number of top-k influential nodes is smaller than 20% of total nodes in each setting. The size of the network is set to 2000 m × 2000 m. The radius of communication range varies from 5 to 20 m. DBSCAN algorithm does not require number of clusters, two initial input parameters, namely ε (the radius of the cluster) and *minPts* (the minimum data objects required inside the cluster). We control *minPts* and ε. The warm-up period is set to 1500 s to collect the movement patterns of the nodes. Due to sparse node contacts in real trace, the information cannot diffuse to every node in simulation time. Therefore, we measured ratio of the contact node set $\sigma(I)$ by k-node set I in $\gamma(I)$. We set to 15000 s as the simulation time. <Figure 5> (a) and (b) show a snapshot of NCCU student's GPS and our simulated network, respectively. As shown <Figure 5>, we show that many nodes frequently move to several spots. In our experiment, we implemented that every node issues GPS and a contact log at current context, and then communicate each other within radius of communication range.

〈Table 2〉 Simulation parameters

Parameter (unit)	Value (default)
number of nodes	115
interval of message creation	student behavior
movement model	NCCU Trace Data
size of the network (m^2)	2000 × 2000
number of k influential nodes	4, 8, 12, 16 (4)
radius of communication range (m)	5, 10, 15, 20 (10)
MinPts	50,70,90,100 (100)
ε	15,35,105,115 (115)
warm-up period (s)	1500
simulation time (s)	15000



(a) (b)

〈Figure 5〉 NCCU trace and ONE simulator

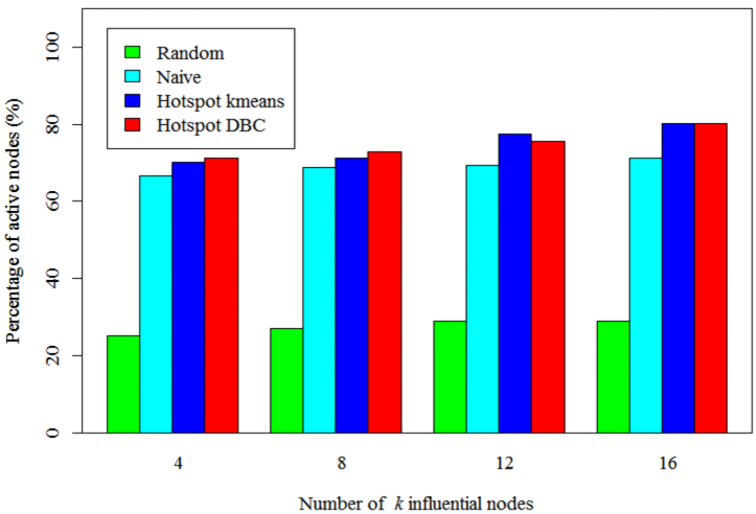
4.2 Simulation results

4.2.1 Number of influential nodes

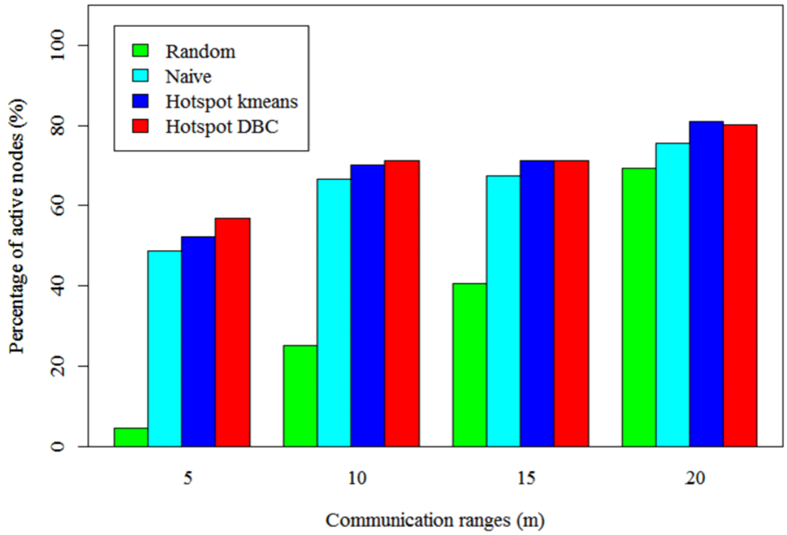
We estimated percentage of active nodes with varying the number of *k* diffusion nodes on each scheme. The number of *k* diffusion nodes are set to 4,8,12 and 16. <Figure 6> shows percentage of active nodes for different *k* influential nodes at the simulation time 15000 s. Random has worst performance while Naive performs better than Random. The hotspot-based schemes (Hotspot *k*-means and Hotspot DBC) always demonstrate better performances than Random and Naive scheme. Hotspot DBC has the best performance. Our schemes can spread the information to more 80 % of nodes with k less than 16. But Naïve and Random cannot spread the information more than Hotspot *k*-means and Hotspot DBC. In addition, percentage of active nodes in Naive does not change much with the increase in the ratio of *k* influential nodes, when compared to the hotspot-based schemes. Since Random and Naive do not consider the social behaviors between the nodes, it difficult to meet the isolated nodes for diffusion process. Meanwhile, in the hotspot-based two schemes, the isolated nodes have high contact probabilities because the one of the nodes who have the similar mobility pattern with them is selected as *k* influential node. Therefore, hotspot-based schemes have better performances than other schemes.

4.2.2 Communication ranges

We examine the effect of the nodes' communication range on each scheme. The communication ranges are set to 5m, 10m, 15m and 20m. In <Figure 7>, as the communication range increases, diffusion time of our schemes and other scheme decreases. Random has still worst



〈Figure 6〉 Percentage of active nodes with different influential nodes

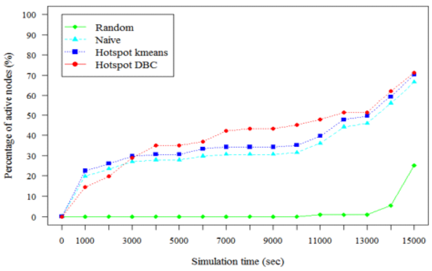


〈Figure 7〉 Percentage of active nodes with different communication ranges

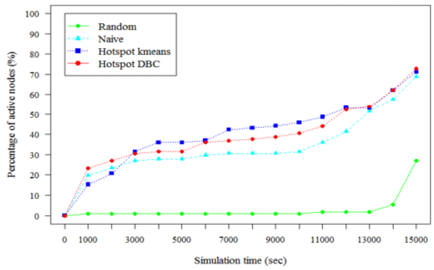
performance and Naive outperforms Random. However, Hotspot kmeans and DBC schemes have better performances than Random and Naive. When the range is 5 or 10 m, the differences in the performances among schemes are much large than 15 m and 20 m because of sparse MSNs. In addition, all scheme have good performance in 20 m. Therefore, hotspot-based schemes can be implemented well in sparse networks.

4.2.3 Communication ranges

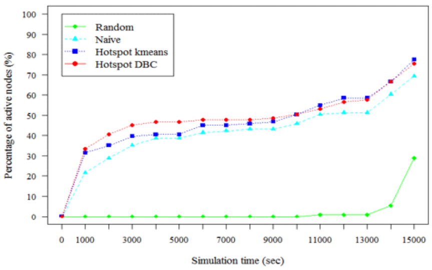
Lastly, we compare the percentage of active nodes on each scheme during the simulation time. <Figure 8> and <Figure 9> show the percentage of active nodes for the simulation time. Random has the worst performance compare to other schemes during the simulation time. Naive has a better performance than Random. The hotspot-based schemes diffuse to more active nodes than Naïve. Our schemes have the best performance for most of the diffusion time. As we mentioned, in <Figure 8> (a), (b), (c) and (d), our schemes propagate the information more than Random and Naive. In <Figure 9> (d), we can see that all the methods become increasingly indifferent because of the simulation in sparse environment. In summary, the hotspot-based diffusion schemes not only reduce the term of the diffusion process in the sparse network for propagation, but also finding isolated nodes, where they encounter mobility patterns, similar to the one of *k*-node set



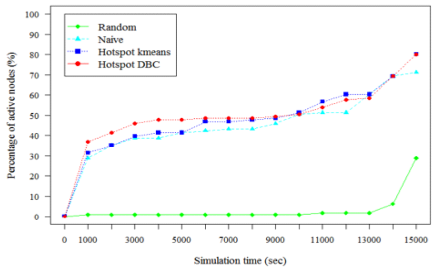
(a) $k = 4$



(b) $k = 8$

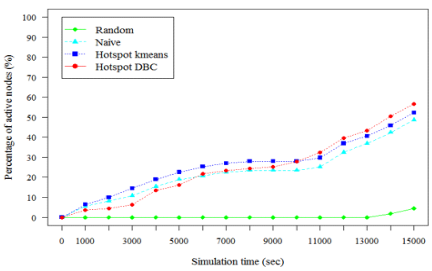


(c) $k = 12$

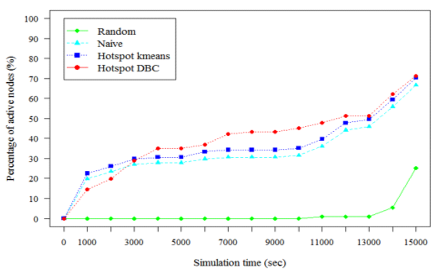


(d) $k = 16$

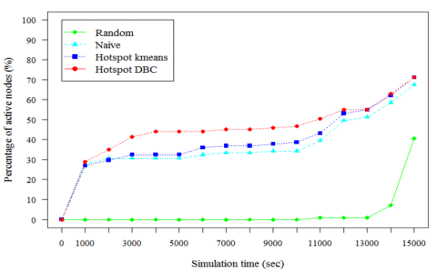
〈Figure 8〉 Percentage of active nodes with different influential nodes



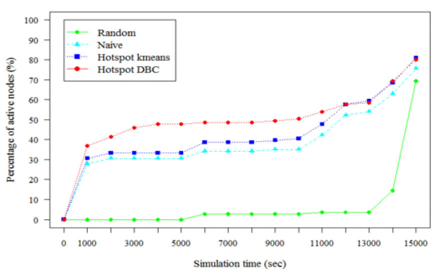
(a) Communication range 5m



(b) Communication range 10m



(c) Communication range 15m



(d) Communication range 20m

〈Figure 9〉 Percentage of active nodes with different communication ranges

5. Conclusion

We addressed a problem of finding the top-*k* influential nodes to propagate information effectively to nodes in a dynamic network as soon as possible, referred to as the diffusion minimization problem. We studied the solution for the diffusion minimization problem in MSNs by proposing the hotspot-based information diffusion, which is a novel diffusion scheme and influential nodes are selected through node behaviors based on social information in the hotspot zones. In our performance evaluation, it was more effective to solve the diffusion minimization, because the influential nodes were selected within the location-based communities instead of the entire network topology. As the platforms for viral marketing, targeted vaccination, and offloading, our diffusion scheme solves the information diffusion problem by node's mobility behavior. Thus, our schemes can be applied to marketing application and disseminated to the communication among mobile nodes in MSNs. We also provide future work relevant to our research which can be implemented as approaches of deep learning, and mobile applications for Internet of Things.

References

Ma, H., H. Yang, M. R. Lyu and I. King, "Mining social networks using heat diffusion processes for marketing candidates selection," *Proceedings of the 17th ACM conference on Information and knowledge management*, (2008), 233~242.

Richardson, M. and P. Domingos, "Mining knowledge-sharing sites for viral marketing," *Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2002), 61~70.

Conti, S., Giordano, M. May, and A. Passarella, "From opportunistic networks to opportunistic computing," *IEEE Communications Magazine*, Vol.48, No. 9(2010), 126~139.

Chen, W., C. Wang and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2010), 1029~1038.

Lu, Z., Y. Wen and G. Cao, "Information diffusion in mobile social networks: The speed perspective," *Proceedings of IEEE INFOCOM*, (2014), 1932~1940.

Chen, X. and K. Xiong, "Dynamic social feature-based diffusion in mobile social networks," *Proceedings of IEEE/CIC International Conference on Communications in China (ICCC)*, (2015), 1~6.

Myers S. A., C. Zhu and J. Leskovec, "Information diffusion and external influence in networks," *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2012), 33~41.

Tsai T. C. and H. H. Chan, "NCCU Trace: social-network-aware mobility trace," *IEEE Communications Magazine*, Vol.53, No. 10(2015), 144~149.

Domingos P. and M. Richardson, “Mining the network value of customers,” *Proceedings of the 7th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2001), 57~66.

Kempe D., J. Kleinberg and É. Tardos, “Maximizing the spread of influence through a social network,” *Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2003), 137~146.

Wang Y., G. Cong, G. Song and K. Xie, “Community-based greedy algorithm for mining top-k influential nodes in mobile social networks,” *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2010), 1039~1048.

Christakis N. A. and J. H. Fowler, “The spread of obesity in a large social network over 32 years,” *New England journal of medicine*, Vol.357, No. 4(2007), 370~379.

Bakshy E., I. Rosenn, C. Marlow and L. Adamic, “The role of social networks in information diffusion,” *Proceedings of the 21st international conference on World Wide Web. ACM*, (2012), 519~528.

López-Pintado D., “Diffusion in complex social networks,” *Games and Economic Behavior*, Vol.62, No. 2(2008), 573~590.

Han, B., P. Hui, V. S. A. Kumar, M. V. Marathe, J. Shao and A. Srinivasan, “Mobile data offloading through opportunistic communications and social participation,” *IEEE Transactions on Mobile Computing*, Vol.11, No. 5(2012), 821~834.

Ester M., H. P. Kriegel, J. Sander and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” *kdd*, Vol.96 No. 34(1996), 226~231.

Opsahl T., F. Agneessens and J. Skvoretz, “Node centrality in weighted networks: Generalizing degree and shortest paths,” *Social networks*, Vol.32 No. 3(2010), 245~251.

Keränen A., J. Ott and T. Kärkkäinen, “The ONE simulator for DTN protocol evaluation,” *Proceedings of the 2nd International Conference on Simulation Tools and Techniques*, (2009), 1~10.

국문요약

Hotspot DBC: 모바일 소셜 네트워크 상에서 마케팅 전략을 위한 위치 기반 정보 유포

류제광* · 양성봉**

모바일 디바이스의 무선 네트워크 통신 기술과 온라인 소셜 네트워크 발전으로 모바일 소셜 네트워크는 모바일 기기 사이에 마케팅 전략의 기회를 제공한다. 이에 따라 모바일 소셜 네트워크 상에서 정보 유포는 중요한 문제가 되었으며 여러 기법을 제안해왔다. 정보 유포 연구 정의는 메시지와 같은 정보를 가진 초기 노드로부터 최소한의 시간에 최대한 많은 유저에게 정보를 전달하는 기법이다. 본 논문에서 우리는 새로운 정보 유포 기법인 기계학습과 소셜 위치정보 기반의 Hotspot DBC를 제안한다. 위치기반 정보 유포 기법으로써 핫스팟 지역을 사용한다. 웜업 기간에 움직임 패턴을 활용하여 초기 영향력 있는 노드를 찾는다. 이후 전체 네트워크 지역을 고려하는 것이 아닌 특정 핫스팟 지역에서만 패턴을 추출하여 찾는다. 웜업 기간 끝나는 시점에서 각 노드는 움직임 패턴을 추출한다. 마지막으로 각 패턴에서 소셜 관계를 분석함으로써 영향력 있는 노드 k개가 선정된다. 우리는 기회적 네트워크 환경에서 GPS 위치 기록의 실제 모바일 노드의 움직임 데이터를 ONE 시뮬레이터 환경에서 실험하였다. 추가적으로 통신범위와 초기 정보 유포 k 노드 수를 다양하게 실험하여 기존 기법보다 더 나은 결과를 확인할 수 있다.

주제어 : mobile social networks, information diffusion, machine learning, NCCU, viral marketing

논문접수일 : 2017년 3월 18일　논문수정일 : 2017년 4월 30일　게재확정일 : 2017년 5월 15일
원고유형 : 일반논문(급행)　교신저자 : 양성봉

* 연세대학교 공과대학 컴퓨터과학과
** 교신저자 : 양성봉
연세대학교 공과대학 컴퓨터과학과
D813, The 4rd Engineering Building, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722
Tel: +82-02-2123-7239, Fax: +82- 02-365-2579, E-mail: sbyang@cs.yonsei.ac.kr

저 자 소 개

Jegwang Ryu
He is currently an M.S. candidate in computer science at Yonsei University in Korea. His research interests include mobile social networks, social network analysis and machine learning.

Sung-Bong Yang
He received his M.S. and Ph.D. from the Dept. of Computer Science at the University of Oklahoma in 1986 and 1992, respectively. He has been a professor at Yonsei University since 1994. His research interests include graph algorithms, mobile computing, machine learning and social network analysis.