PINOCCHIO: Probabilistic Influence-based Location Selection over Moving Objects

(Extended Abstract)

Meng Wang*, Hui Li†, Jiangtao Cui*, Ke Deng‡, Sourav S Bhowmick§, Zhenhua Dong¶

*School of Computer Science and Technology / †School of Cyber Engineering, Xidian University, Xi'an, China

‡School of Computer Science and Information Technology, RMIT University, Melbourne, Australia

§School of Computer Science and Engineering, Nanyang Technological University, Singapore

¶Huawei Noah's Ark Research Lab, Hongkong, China

wamengit@sina.com, {hli, cuijt}@xidian.edu.cn, ke.deng@rmit.edu.au, assourav@ntu.edu.sg, dongzhenhua@huawei.com

***Abstract*—The *location selection* (LS) problem aims to mine the optimal location to place a new facility from a set of candidates such that the benefit or influence on a given set of objects is maximized. State-of-the-art LS techniques assume each object is static and can only be influenced by a single facility. However, in reality, objects (*e.g.,* people, vehicles) are mobile and are influenced by multiple facilities. Consequently, classical LS solutions fail to select locations accurately. In this work, we introduce a *generalized* LS problem called PRIME-LS which takes mobility and probability factors into consideration to address the aforementioned limitations. To solve the problem, we propose an algorithm called PINOCCHIO, which leverages two pruning rules based on a novel distance measure, and further extend it by incorporating two *optimization strategies*. Experimental study over two real-world datasets demonstrates superiority of our framework in comparison to state-of-the-art LS techniques.**

I. Introduction

Location selection (LS) problem has received considerable research attention due to its role in a wide spectrum of applications such as marketing, urban planning, resource allocation, etc. Given a set of objects Ω with their positions and a set of candidate locations C, existing LS techniques [1] aim to mine a candidate $c \in C$ such that c can *influence* (affect) the maximum number of objects. Despite the significant progress made by these techniques, they are often not effective in real scenarios where objects are mobile due to at least one of the following three drawbacks. Firstly, objects are assumed to be static. Hence, for mobile objects, considering only a single instead of a set of positions fails to provide a complete description of these objects' activities. Secondly, the influence of a location on an object is considered binary (either influenced or not) based on a distance metric between them. However, this criterion may not be able to reflect the influence in real world, which is in fact a probabilistic event [2]. Thirdly, the assumption that an object is influenced by only one facility (*e.g.,* the nearest neighbor) is too rigid as it jettisons other facilities which might also exhibit influence. Since an object can be influenced by multiple facilities simultaneously, excluding other facilities in selecting location may reduce the accuracy of the solution. Hence, in this work, we present an overview of a *generalized* LS problem called PRIME-LS (**PR**obabilistic **I**nfluence-based **M**obility-awar**E** **L**ocation **S**election) [3], which is the first effort to take mobility and probability into consideration to address the aforementioned limitations of traditional LS techniques.

II. Problem Definition

We model a moving object O by a set of discrete positions $O = \{p_1, p_2, \dots, p_n\}$, where a position p is a (geographic) coordinate in a 2-D Euclidian space. Candidate locations of a new facility are denoted as $C = \{c_1, c_2, \dots, c_m\}$. As the probability of an object for a facility is inversely proportional to the distance [2], we assume the influence probability only depends on distance, and use a probability function PF, which is monotonically decreasing to distance, to depict the behavior pattern of influence. Moreover, the probability that a moving object O is influenced by a candidate c at any position $p_i \in O$, denoted by $Pr_c(p_i)$, is independent of those at other positions. Hence $Pr_c(p_i) = PF(dist(c, p_i))$, where $dist(c, p_i)$ is the distance between them. Inspired by Influence Model [4], O is *cumulatively* influenced by c if and only if there is at least a position of O that is influenced by c. Hence, we can redefine the concept of *influence* in our problem setting as follows. Given a candidate location c and a moving object O with n positions $\{p_1, p_2, \dots, p_n\}$, the *cumulative influence probability* of O being influenced by c, denoted by $Pr_c(O)$, is defined as: $Pr_c(O) = 1 - \prod_{i=1}^{n}(1 - Pr_c(p_i))$. Then, given an influence threshold τ, c *influence*s O if and only if $Pr_c(O) \geq \tau$. Further, given a set of moving objects Ω, the *influence* value of c, denoted as $inf(c)$, is the number of moving objects in Ω that are influenced by c.

Based on the aforementioned definition of influence, we can now define the PRIME-LS problem as follows.

Definition 1: Given a set of candidate locations C, a set of moving objects Ω, a certain distance-based probability function PF and a user-specified influence threshold τ, the PRIME-LS *problem aims to mine the optimal candidate $c \in C$ such that $\forall c' \in C - \{c\}$, $inf(c) \geq inf(c')$.* ∎

By varying *PF*, PRIME-LS can be applied to various mobility patterns (*e.g.,* check-in pattern [2]), and even degenerate to traditional LS problems.

III. Solution to PRIME-LS

To address PRIME-LS efficiently, we present an overview of a novel algorithm called PINOCCHIO (**P**robabilistic **IN**fluence-based L**OC**ation Sele**C**tion Tec**H**nique over Mov**I**ng **O**bjects) [3], which comprises of two phases, *i.e., pruning* and *validation*. In the *pruning phase*, we propose a novel pruning strategy to filter out inferior candidate locations. In the *validation phase*, the remaining candidates are refined.

A. *minMaxRadius and Pruning Rules*

PRIME-LS focuses on moving objects, whose activity regions are highly overlapping and may enclose some candidates. Hence, the single point- and nearest neighbor-based pruning techniques cannot be adopted here. To this end, we propose a novel pruning measure, *minMaxRadius*, to quantify the cumulative influence probability. Specifically, given a moving object O with n positions and a probability function PF, the ***minMaxRadius*** of O based on a user-specified probability threshold τ, is defined as: $minMaxRadius(\tau, n) = PF^{-1}(1 - (1 - \tau)^{\frac{1}{n}})$. Based on this *minMaxRadius*, we design two pruning rules, *influence arcs* (IA) *rule* and *non-influence boundary* (NIB) *rule*. Figure 1 and 2 illustrate the two rules, where the minimum bounding rectangle (MBR) encloses the activity region (*i.e.,* all positions) of an object. IA rule determines an area bounded by four influence arcs, each of which centers at a MBR corner with *minMaxRadius* as the radius length. For each moving object, any candidate falling into the IA area definitely influences the object. NIB rule finds a relatively larger area for each moving object such that any candidate outside the area definitely fails to influence that object. The delimitation of NIB area is a rounded rectangle such that for any point on it, the minimum distance to the object's MBR is exactly *minMaxRadius*.

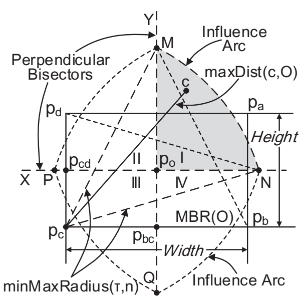


Fig. 1: IA Rule

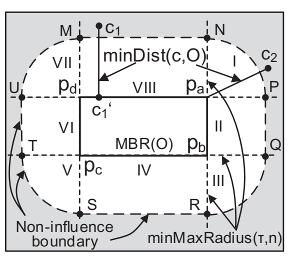


Fig. 2: NIB Rule

B. *Overview of* PINOCCHIO

The key idea of PINOCCHIO is as follows. 1) For each moving object, we calculate its $minMaxRadius$, based on which the corresponding IA and NIB areas can be obtained. 2) With the help of IA rule, we identify the candidates that influence the object. 3) For the remnant candidates, NIB rule is used to exclude the candidates that cannot influence the object. 4) The remnant candidates are verified using the definition of influence in Section II.

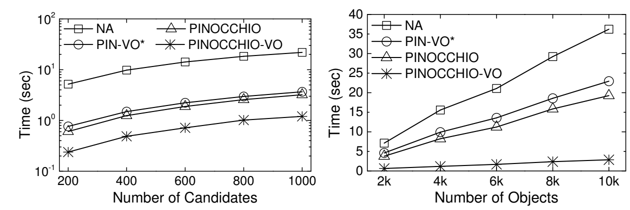
We further enhance the PINOCCHIO algorithm, referred to as PINOCCHIO-VO, by incorporating two optimization strategies, *upper and lower bounds of influence* and *early stopping*. For the former, on condition that $maxInf(c) < maxminInf$, candidate c cannot influence the largest number of moving objects, where $maxInf(c)$ is the maximum of the possible influence of c, $minInf(c)$ is the identified influence of c and $maxminInf = \max_{c \in C} minInf(c)$. Hence, c can be pruned and no further validation is needed. A *Max Heap* of C can efficiently benefit from the strategy. For *early stopping*, the validation of a moving object O is accomplished by computing only partial n' positions of O instead of all the n positions, if $Pr_c^{n-n'}(O) \leq 1 - \tau$, where $Pr_c^{n-n'}(O) = \prod_{i=n'+1}^{n}(1 - Pr_c(p_i))$ $(n' \in N, n' < n)$.

IV. Summary of Experimental Evaluation

We have investigated the performance of our proposed PRIME-LS framework over real-world check-in datasets, where the check-in behaviour model [2] is set as *PF*.

We validate the effectiveness of PRIME-LS in comparison to nearest neighbor-based LS (BRNN*) and range-based LS (RANGE). Extended from MaxBRNN [5], BRNN* selects for each object O a best candidate, which influences the most positions in O, and the result is the candidate selected by the most objects. For RANGE, an object is deemed to be influenced if at least some proportion of its positions lie within a given range of a candidate. With real check-in logs as the ground-truth, we employ *Precision@K* (abbr., *P@K*) and *Average Precision@K* (abbr., *AP@K*) to rank *Top-K* (K=10, . . . , 50) of 200 candidates selected by PRIME-LS, BRNN* and RANGE. On average, PRIME-LS is around 20%(*P@K*)-35%(*AP@K*) and 8%(*P@K*)-12%(*AP@K*) better than BRNN* and RANGE.

Compared to NA (baseline method that exhaustively scans all object-candidate pairs), PINOCCHIO-VO has the best scalability, followed by PINOCCHIO and PIN-VO* (PINOCCHIO-VO without pruning) as shown in Figure 3. PINOCCHIO-VO prunes nearly $2/3$ candidates and significantly reduces the time by orders of magnitude.



(a) *Varying* $|C|$ (b) *Varying* $|O|$

Fig. 3: Performance of different algorithms.

Using $24-48$ positions, we can achieve a tradeoff between accuracy and cost. We also observe that, if we expect a certain number of objects to be influenced, the resulting locations are identical or very close with high accuracy, regardless of how n and τ are set. Finally, our results demonstrate that PINOCCHIO-VO can handle different *PF*s, all of which are variations of the *Log-sigmoid transfer* function.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 61672408 and 61472298), Huawei Innovation Research Program, China 111 Project (No. B16037) and Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2015JM6351).

References

[1] L. Zhan, Y. Zhang, W. Zhang, and X. Lin, "Finding top k most influential spatial facilities over uncertain objects," in *CIKM*, 2012, pp. 922–931.

[2] B. Liu, Y. Fu, Z. Yao, and H. Xiong, "Learning geographical preferences for point-of-interest recommendation," in *KDD*, 2013, pp. 1043–1051.

[3] M. Wang, H. Li, J. Cui, K. Deng, S. S. Bhowmick, and Z. Dong, "Pinocchio: Probabilistic influence-based location selection over moving objects," *TKDE*, vol. 28, no. 11, pp. 3068–3082, 2016.

[4] D. Kempe, J. M. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *KDD*, 2003, pp. 137–146.

[5] R. C.-W. Wong, M. T. Özsu, P. S. Yu, A. W.-C. Fu, and L. Liu, "Efficient method for maximizing bichromatic reverse nearest neighbor," *PVLDB*, vol. 2, no. 1, pp. 1126–1137, 2009.