Optimal citizen-centric sensor placement for citywide environmental monitoring: A submodular approach

Chenxi Sun, Victor O. K. Li, Jacqueline C.K. Lam
The University of Hong Kong
{cxsun, vli, jcklam}@eee.hku.hk

Abstract—The general environmental monitoring problem refers to the task of placing sensors or stations to optimize certain objectives under budget constraints. Application scenarios include monitoring temperature, water contamination, air quality etc. As citizens are increasingly concerned about the surrounding environment, it is important to provide sufficient and accurate information to the public. In this study, we focus on the problem of optimal citizen-centric sensor placement, i.e., given a set of locations within the city, we aim at placing sensors or stations at locations that will benefit as many citizens as possible under budget constraints. We prove that the problem is NP-hard, yet the objective function has the nice monotone and submodular property. Then the efficient greedy algorithm and its variants can be adopted with a guaranteed approximation ratio of \((1-1/e)\) for the unit cost case and \(\frac{1}{2}(1-1/e)\) for the general cost case. Finally we demonstrate the effectiveness of the proposed approach by comparing with two baseline algorithms through a case study.

Keywords—Sensor Placement, Submodular Maximization, Environmental Monitoring

I. INTRODUCTION

The environmental monitoring problem refers to the task of placing sensors or stations to optimize certain objectives such as information obtained under budget constraints, with applications in monitoring temperature, water contamination, air quality etc.

Take air quality as an example. As people are increasingly concerned about the surrounding air quality, it is important to provide sufficient and accurate information to the public, especially when the air is bad. Studies have shown that exercise outdoors when the air pollution level is high will cause adversarial effect on health [1]. Furthermore, the air quality information is pivotal for health studies.

Generally, the environment information can be obtained either through low-cost mobile wireless sensors or official monitoring stations with sophisticated measurement equipments and calibration. In this study, we focus on deployment strategies for static monitoring stations as their measurements are more accurate and reliable. However, deploying static monitoring stations can be very costly, incurring not only construction cost but also operational cost. Hence only a limited number of stations can be placed under a total cost constraint. There have been much work on sensor placement for environment monitoring. For example, [2] studied the problem of selecting a most informative subset of correlated random variables under size constraints. [3] proposed a near optimal sensor placement strategy for temperature monitoring by maximizing mutual information under size constraint. [4] studied the wind monitoring problem.

However, all the studies mentioned above aim to place sensors or stations at the most informative locations that are defined either in terms of entropy or mutual information. Furthermore, by modeling each location as a random variable, the proposed approach would require the covariance between any two random variables, usually through a strong Gaussian Process assumption on the underlying field.

In this paper, we take a different perspective on the sensor placement problem. Specifically, we are interested in citizen-centric sensor placement for citywide monitoring: given a fixed cost constraint \(C\), we aim to find the set of locations within the city that will bring maximum benefit to the citizens.

The main contribution of this study are as follows:

1) we formulate the citizen-centric sensor placement problem whose aim is to maximize human satisfaction on the deployment scheme and prove its NP-hardness. As far as we know, this is the first formulation of the sensor placement problem in the context of maximizing overall citizen satisfaction.

2) We prove the monotonicity and submodularity of the objective function, i.e., our problem is a submodular maximization optimization problem under general cost constraints, which can be solved with greedy algorithm and its variants with approximation guarantee.

3) We provide a case study in Hong Kong to demonstrate the effectiveness of our approach.

The rest of the paper is organized as follows. In Section [I] we introduce the problem formulation. In Section [II] we describe the proposed approach. In Section [IV] we discuss how to incorporate other objectives to this formulation. In Section [V] we provide a case study in Hong Kong. We draw conclusion in Section [VI].

II. PROBLEM FORMULATION

In this paper, we focus on the utility gain of placing sensors with respect to citizen satisfaction. Noticing that people will naturally rely on the observations from the nearest station to obtain environment information when there are a number of
stations available within the city, we model an individual’s satisfaction with the sensor placement scheme as a function of his/her distance to the nearest sensor.

Following common practice in environmental monitoring [3, 5], we divide the city into discrete equal-sized square grids, in which sensors will be placed, with at most one sensor in each grid. Let \( V = \{ v_i : i = 1, 2, \ldots, n \} \) denote the set of grids in the region of interest, where \( n = |V| \) is the total number of grids. Let \( p_i \) denote the percentage population living in the \( i \)-th grid \( v_i \) and \( (x_i, y_i) \) denote its location (on the whole map). Then we can use the Hamming distance to model the distance between any two grids \( i, j \), i.e.,

\[
d(i, j) = |x_i - x_j| + |y_i - y_j|.
\]

We further define the minimum distance between grid \( i \) and a set of grids \( A \) as follows:

\[
d(i, A) = \min_{j \in A} d(i, j). \tag{2}
\]

When \( A = \emptyset \), we set \( d(i, A) = \infty \). Then we can define the average satisfaction ratio of choosing a set of grids \( A \) for placing sensors as

\[
f(A) = \sum_{i \in V} p_i \exp \left( -\frac{d(i, A)\theta}{d(i, A)} \right) \tag{3}
\]

where \( \theta \) is an exponential decay parameter controlling the decay speed. By default we can choose \( \theta = 1 \). A smaller \( \theta \) will make the satisfaction vanish much more rapidly with the increase in distance.

Let \( c(\cdot) \) be the cost associated with subset \( A \). Then our objective is to select a subset of locations \( A \subseteq V \) that can maximize the average satisfaction ratio under the total cost constraint \( c(A) \leq C \).

Formally, our optimal citizen-centric sensor placement problem is as follows:

\[
\begin{array}{ll}
\max & f(A) \\
\text{s.t.} & c(A) \leq C, A \subseteq V \tag{4}
\end{array}
\]

In the simplest case, the cost is uniform for all locations, and the cost constraint is reduced to the cardinality constraint:

\[
\begin{array}{ll}
\max & f(A) \\
\text{s.t.} & |A| = k, A \subseteq V \tag{5}
\end{array}
\]

where \( k \) is the total number of sensors that can be placed subject to the total cost constraint \( C \).

### III. Solution Approach

#### A. Hardness of the problem

We prove the NP-hardness of the problem in the unit-cost case by showing its reduction from the k-median of a network, which is known to be NP-hard [6]. The k-median (of a network) problem is defined as follows. A network \( G = (V, E) \) has a nonnegative number \( w(v) \) associated with each of its \( |V| = n \) vertices and a positive number \( l(e) \) associated with each of its \( |E| \) edges. Let \( X_k = \{ x_1, x_2, \ldots , x_k \} \) be a set of \( k \) points on \( G \). The goal is to find \( X_k^* \) such that the distance sum of the set \( H(X_k) = \sum_{v \in V} w(v) \min_{1 \leq i \leq k} \{ d(v, x_i) \} \) is minimized.

**Reduction:** The corresponding relationship between these two set systems are as follows: the weight \( w(v) \) of \( v \in V \) corresponds to the population \( p_i \) of grid \( i \in V \) and the distance \( d(v, x_i) \) between \( v \) and \( x_i \in X_k \) corresponds to the non-negative satisfaction loss \( 1 - \exp(1 - \frac{d(v, x_i)}{\theta}) \) of location \( i \) by placing sensor at location \( j \). Then selection of a set of points \( X_k \subseteq V \) on \( G \) corresponds to selecting a set of locations \( A \subseteq V \) with size \( |A| = k \).

#### B. Properties of the objective function

Despite the hardness of the problem, We show that the objective function \( f \) has the nice monotone and submodular properties.

**Lemma 1 (Monotonicity).** For any \( A \subseteq V \) and a sensor \( s \in V \), we have

\[
f(A \cup \{ s \}) - f(A) \geq 0. \tag{6}
\]

**Proof.** By definition we have \( f(\emptyset) = 0 \). For all \( i \in V \), we have \( d(i, A) = \min_{j \in A} d(i, j) \geq \min_{j \in A \cup \{ s \}} d(i, j) = d(i, A \cup \{ s \}), \forall A \subseteq V, s \in V \). Hence we have \( f(A \cup \{ s \}) = \sum_{i \in V} w_i \exp(-d(i, A \cup \{ s \})) \geq \sum_{i \in V} w_i \exp(-d(i, A)) = f(A) \) for any \( A \subseteq V, s \in V \).

Furthermore, we can also show that the objective function \( f \) is submodular.

**Lemma 2 (Submodularity).** For all placements \( A \subseteq B \subseteq V \) and a sensor \( s \in V \setminus B \), we have

\[
f(A \cup \{ s \}) - f(A) \geq f(B \cup \{ s \}) - f(B). \tag{7}
\]

**Proof.** By definition of \( f \) we need to prove that \( \sum_{i \in V} w_i \exp(-d(i, A)) - \sum_{i \in V} w_i \exp(-d(i, A \cup \{ s \})) \geq \sum_{i \in V} w_i \exp(-d(i, B)) - \sum_{i \in V} w_i \exp(-d(i, B \cup \{ s \})) \). It suffices to show that for each \( i \in V \), we have \( \exp(-d(i, A)) - \exp(-d(i, A \cup \{ s \})) \geq \exp(-d(i, B)) - \exp(-d(i, B \cup \{ s \})) \). Suppose that \( d(i, A) = d(i, a) \) and \( d(i, B) = d(i, b) \). Since \( A \subseteq B, a \in A, \) and \( b \in B \), we have \( d(i, b) \leq d(i, a) \).

Here we have three scenarios.

1. \( d(i, s) \geq d(i, a) \).
   - Then \( d(i, A \cup \{ s \}) = d(i, A) + d(i, B) \).
   - Hence \( \exp(-d(i, A)) - \exp(-d(i, A \cup \{ s \})) = \exp(-d(i, B)) - \exp(-d(i, B \cup \{ s \})) = 0 \).

2. \( d(i, b) \leq d(i, s) < d(i, a) \).
   - Then \( d(i, A \cup \{ s \}) = d(i, s) < d(i, A) \).
   - Hence \( \exp(-d(i, A)) - \exp(-d(i, A \cup \{ s \})) > \exp(-d(i, B)) - \exp(-d(i, B \cup \{ s \})) = 0 \).

3. \( d(i, s) < d(i, b) \).
   - Then \( d(i, A \cup \{ s \}) = d(i, B) \).
   - Hence \( \exp(-d(i, A)) - \exp(-d(i, A \cup \{ s \})) = \exp(-d(i, B)) - \exp(-d(i, B \cup \{ s \})) = 0 \).

Hence \( f \) satisfies the above property. 

Hence the the optimal citizen-centric sensor placement problem \( \text{(4)} \) is a monotone submodular maximization problem.
C. Placement constraints

We start with the simplest case for the cost constraint when the cost is uniform for all locations. In this case, a simple greedy approach which selects

$$s_i = \arg \max_{s \in V \setminus A_{i-1}} f(A_{i-1} \cup \{s\}) - f(A_{i-1})$$  \hspace{1cm} (8)

at the $i$th iteration until the cardinality constraint is no longer satisfied has been shown to achieve at least $1 - 1/e$ of the optimal solution.

Theorem 1 (\cite{7}). If $f$ is a submodular monotone set function, the greedy algorithm finds a set $A^*$ such that $f(A^*) \geq (1 - 1/e) \max_{|A|=k} f(A)$ with at most $O(nk)$ evaluations of $f$.

The station deployment algorithm with unit cost constraint is summarized in Algorithm 1.

Algorithm 1 Station deployment algorithm with unit cost constraint

Input: Sensor number constraint $k$, a set of grids $V$ with associated population $\{w_i\}_{i=1}^n$, distance function $d$, exponential decay parameter $\theta$

Output: A subset of locations $A \subseteq V$

A = \emptyset

for $i$ = 1 to $k$ do

$$\Delta_s = \sum_{i \in V} w_i \left( \exp \left( -\frac{d(i, A_{i-1} \cup \{s\})}{\theta} \right) - \exp \left( -\frac{d(i, A_{i-1})}{\theta} \right) \right)$$

$s^* = \arg \max_{s \in V \setminus A_{i-1}} \Delta_s$

$A = A \cup \{s^*\}$

return $A$

A more general scenario is when the cost is non-uniform for all locations, which is also known as the knapsack constraint. In this case, the simple greedy algorithm fails to provide a satisfactory result when placing station at a location is much more expensive but only provides a marginally better utility gain. A modified greedy selection rule that takes cost into account is:

$$s_k = \arg \max_{s \in V \setminus A_{k-1}} \frac{f(A_{k-1} \cup \{s\}) - f(A_{k-1})}{c(s)}$$  \hspace{1cm} (9)

However the result for this condition can also be arbitrarily bad. Let $A_{CG}$ denote the greedy selection result using criteria (8) and $A_{CEG}$ denote the greedy selection result using criteria (9). Then the following theorem shows that the better of these two solutions provides $\frac{1}{2}(1 - 1/e)$ approximation guarantee with at most $O(nk)$ function evaluations.

Theorem 2 (\cite{8}). If $f$ is a submodular monotone set function, we have

$$\max\{f(A_G), f(A_{CEG})\} \geq \frac{1}{2}(1 - 1/e) \max_{A, c(A) \leq C} f(A)$$  \hspace{1cm} (10)

The station deployment algorithm with general cost constraint is summarized in Algorithm 2.

Algorithm 2 Station deployment algorithm with general cost constraint

Input: Cost constraint $C$, a set of grids $V$ with associated population $\{w_i\}_{i=1}^n$, distance function $d$, exponential decay parameter $\theta$, a cost function $c$

Output: A subset of locations $A \subseteq V$

$A_1 = A_2 = \emptyset$, $V' = V'' = V$, $C'_1 = C'_2 = 0$

while $V'_1 \neq 0$ do

for all $s \in V'_1$ do

$$\Delta_s = \sum_{i \in V} w_i \left( \exp \left( -\frac{d(i, A_{i-1} \cup \{s\})}{\theta} \right) - \exp \left( -\frac{d(i, A_{i-1})}{\theta} \right) \right)$$

$s^* = \arg \max_{s \in V \setminus A_{i-1}, s \neq \emptyset} \Delta_s$

if $c(s^*) + C'_1 \leq C$ then

$C'_1 = C'_1 + c(s^*)$, $A_1 = A_1 \cup \{s^*\}$

end if

$V'_1 = V'_1 \setminus \{s^*\}$

end for

while $V'_2 \neq 0$ do

for all $s \in V'_2$ do

$$\Delta_s = \sum_{i \in V} w_i \left( \exp \left( -\frac{d(i, A_{i-1} \cup \{s\})}{\theta} \right) - \exp \left( -\frac{d(i, A_{i-1})}{\theta} \right) \right)$$

$s^* = \arg \max_{s \in V \setminus A_{i-1}, s \neq \emptyset} \Delta_s$

if $c(s^*) + C'_2 \leq C$ then

$C'_2 = C'_2 + c(s^*)$, $A_2 = A_2 \cup \{s^*\}$

end if

$V'_2 = V'_2 \setminus \{s^*\}$

end for

$A = \arg \max_{A \in \{A_1, A_2\}} G(A)$

return $A$

IV. DISCUSSION

A. Incorporating other factors

Many objective functions in existing sensor placement formulations are submodular in nature, for example, maximum entropy sampling \cite{2}, mutual information maximization \cite{3}, outbreak detection \cite{8} etc. Hence our formulation can easily incorporate other objectives and be reformulated as a multi-objective optimization problem. Then we can find Pareto-optimal solutions \cite{9} by weighted sum transformation:

$$\max \sum_i \lambda_i f_i(A)$$  \hspace{1cm} (11)

s.t. $c(A) \leq C$, $A \subseteq V$

where $\lambda_i$ is the weight factor of the $i$-th objective function $f_i$.

V. A CASE STUDY IN HONG KONG

A. Dataset and setup

We conduct a case study in Hong Kong with the LandScan 2010TM data\cite{LandScan} to evaluate the proposed algorithm. The dataset contains the finest resolution population distribution information at approximately 1km resolution in 2010 and there are 1612 non-empty grids in total for Hong Kong. Figure 1 shows the gridded population density in Hong Kong in 2010. The population density is associated population with grid resolution in Hong Kong in 2010.

proposed algorithm performs better than randomSelect and maxSelect in all three scenarios with different $\theta$. The reason is that maxSelect does not take into account the satisfaction gain of placing sensors nearby and randomSelect chooses population-dense areas with only small probability. By the definition of average satisfaction ratio, each sensor will benefit a broader region as $\theta$ increases. Hence when placing the same number of sensors, a larger $\theta$ will result in a higher average satisfaction ratio. This also explains the improved performance of random selection as its randomness will help spread out the deployment.

Figure 2 and Figure 3 visualize the placement results by the proposed approach in Hong Kong with $\theta = 1, 5$. The black triangles on the map indicate the 16 candidate locations for deploying stations. As can be seen, the locations are usually the center of population-dense areas and large $\theta$ will cause the stations to spread more evenly over the city. The results are also consistent with most of the existing locations.

VI. CONCLUSION

In this paper we study the citizen-centric sensor placement problem. We show the NP-hardness of the problem by reduction from the k-median on network problem. We prove the submodularity and monotonicity of the objective function, and hence the greedy algorithm and its variants are proposed to solve the problem with a guaranteed approximation ratio of $(1 - 1/e)$ for the unit cost case and $\frac{1}{2}(1 - 1/e)$ for the general cost case. Finally, we perform a case study in Hong Kong to evaluate the effectiveness of the proposed approach.

ACKNOWLEDGEMENTS

This research is supported in part by the Theme-based Research Scheme of the Research Grants Council of Hong Kong, Under Grant No. T41-709/17-N. We thank Ms Yu Yangwen for the helpful discussions.

REFERENCES


