Context-aware Crowd-sensing in Opportunistic Mobile Social Networks

Phuong Nguyen
Department of Computer Science
University of Illinois at Urbana-Champaign
Email: pvnguye2@illinois.edu

Klara Nahrstedt
Department of Computer Science
University of Illinois at Urbana-Champaign
Email: klara@illinois.edu

Abstract—In this paper, we study the physical crowd-sensing problem and draw the connection to the vertex cover problem in graph theory. Since finding the optimal solution for minimum vertex cover problem is NP-complete and the well-known approximation algorithms do not perform well with under crowd-sensing scenario, we propose the notions of node observability and coverage utility score and design a new context-aware approximation algorithm to find vertex cover that is tailored for crowd-sensing task. In addition, we design human-centric bootstrapping strategies to make initial assignment of sensing devices in the physical crowd based on social information about the users (e.g., interests, friendship). Our experiments on real-world mobile traces show that the proposed approach significantly outperforms the baseline approximation algorithms in terms of sensing coverage.

Keywords—Crowd-sensing; Opportunistic mobile ad hoc network; Approximation algorithm.

I. INTRODUCTION

There have been previous experiments (e.g., SIGCOMM’09 [6], [1] or UIM [7]) on collecting opportunistic mobile traces that can be used to answer a variety of questions about the crowd, such as how many people are in the crowd, or how does the current opportunistic contact graph between devices look like (i.e., to understand the ability of data dissemination within the crowd). However, in those experiments, the settings are usually configured in a certain manner. For example, the number of sensing devices and the sensing interval are fixed when devices are given to participants on a voluntary basis.

In this paper, we treat the above experimental settings as the given constraints, and optimize the assignment of sensing tasks to appropriate devices to maximize the sensing coverage. Under this scheme, we show the connection between the crowd-sensing problem and the vertex cover problem in graph theory [2]. While finding the optimal solution of minimum vertex cover is NP-complete [5], we show that the constrained versions of existing approximation algorithms can be used to derive approximate solutions. However, since those algorithms are designed for generic graph, they do not take into account the spatio-temporal and human-centric characteristics of the crowd-sensing tasks, and they optimize only individual coverage of covering vertex instead of combined coverage.

To overcome this problem, we propose a new context-aware approximation algorithm that is tailored for crowd-sensing task. Particularly, we propose the notion of node observability and coverage utility score to optimize the combined sensing coverage objective. In addition, we incorporate the out-of-band, social-related information about the participants, such as personal interests, and social relationship, to improve the bootstrapping of the crowd-sensing tasks. The experimental results on real-world mobile traces show a significant improvement in sensing coverage while satisfying the optimization constraints.

II. CROWD-SENSING MODEL AND PROBLEM DEFINITION

Let us denote $V$ as the set of devices in a physical crowd event, and $V_{in} \subset V$, i.e., internal devices, as the devices that participate in the physical crowd-sensing experiment to collect opportunistic Bluetooth contacts in the mobile ad hoc networks. The remaining devices, i.e., $V_{ex}$ or external devices, do not participate in the experiment. A sensing application is installed on each internal device and periodically connects to a centralized server to i) send collected sensing data, and ii) receive sensing instructions.

Since sensing tasks are energy-consuming and data from multiple sensing devices could be overlapped, we only require a subset of registered devices doing the sensing tasks at a time. At any point of time, an internal device can be either in sensing mode, i.e., belongs to $V_{in}^n$, or non-sensing mode, i.e., belongs to $V_{in}^p$. Each internal device operates under two different time windows. During sensing time window $t_s$, devices in $V_{in}$ periodically sense the neighboring environment and store the data locally before sending them to the server at the end of sensing interval. The value of $t_s$ is chosen as a multiplication of inquiry interval of wireless sensor on each device $\tau$: $t_s = T \times \tau$ (with $T$ is a fixed integer). During decision time window $t_d$, all devices listen to the instructions from the centralized server, which specify the set of devices doing the sensing task in the next round.

Let us denote $E_{ts}$ as the set of wireless contacts obtained from devices in $V_{in}^n$. This data can be used to construct a contact graph $G_{ts} = (V_{ts}, E_{ts})$. Since $V_{in}^n \subseteq V_{in}$, the set of observed pairwise contacts might not be complete: $E_{ts} \subseteq E_{ts}^o$, where $E_{ts}^o$ is the set of contacts observed if $V_{in}^p = V_{in}$. The objective is thus to find a $V_{in}^n$ that maximizes the number of observed contacts.

Problem Definition: Given a set of internal devices $V_{in} = V_{in}^n \cup V_{in}^p$, for each sensing time interval $t_s$, find an optimal set of sensing devices $V_{in}^n$, whose size equals a predefined $n$ ($n \leq |V_{in}|$), that maximizes the number of wireless contacts $|E_{ts}|$ observed during $t_s$:

$$V_{in}^n = \arg\max_{V_{in}^n \subseteq V_{in}} |V_{in}^n| |E_{ts}|$$

In this definition, $|E_{ts}|$ represents the coverage capability and is used as the maximization objective. Equivalently, we can also use the ratio $|E_{ts}|/|E_{ts}^o|$, i.e., sensing coverage ratio, to measure the coverage capability.

III. CONTEXT-AWARE APPROXIMATION AND HUMAN-CENTRIC BOOTSTRAPPING FOR CROWD-SENSING

Since every edge in a graph is incident to at least one vertex in vertex cover [2], a vertex cover could essentially...
“observe” all contacts (i.e., edges) of the graph. In addition, since the number of sensing devices is limited to \( n \), which is preferred to be small, it is desirable to find a vertex cover of minimum size in crowd-sensing. However, finding minimum vertex cover is a NP-complete problem \[5\], and there have been efforts to come up with approximate solutions \[4\], \[3\]. We consider the constrained approximation algorithms (i.e., constrained by \( n \) ): Top-\( n \) Random-based Approximation (based on randomization), and Top-\( n \) Greedy Approximation (based on greedy selection).

In this paper, we design a new context-aware approximation algorithm and human-centric bootstrapping methods for the physical crowd-sensing task that take into account the spatio-temporal interactions between devices and the social context of the crowd.

A. Human-centric bootstrapping strategies

Beside random bootstrapping, we propose to use “out-of-band” social information of the users who carry the internal devices to initialize the selection of sensing nodes. Our approach is motivated from the fact that human mobility and interaction are influenced by how users are connected to each other socially. In the scope of this paper, we verify our approach by considering two types of social relationship between people: friendship and common interests. For friendship-based bootstrapping, we build a friendship social network of participants in \( V_n \), and select top-\( n \) people with highest node degree as the sensing devices. For interest-based bootstrapping, we group people into groups of similar interests, and for each group in the top-\( n \) most popular groups, select a member with the most number of groups he/she belongs to.

B. Context-aware approximation algorithm

Our proposed approximation algorithm is based on the notion of node observability and coverage utility score.

**Definition 1 (Node observability):** During a sensing interval \( t_s \), node observability \( \sigma_{t_s}(v) \) of a non-sensing node \( v \notin V_{in} \) is defined as:

\[
\sigma_{t_s}(v) = \{ u \mid u \in V_{in}, (u,v) \in E_{t_s} \}
\]

**Definition 2 (Coverage utility score):** Coverage utility score \( \Delta_{t_s}(u) \) of a sensing node \( u \in V_{in} \) during a sensing time interval \( t_s \) is defined as:

\[
\Delta_{t_s}(u) = \sum_{v \in (u,v) \notin E_{t_s}, v \notin V_{in}} \sigma_{t_s}^{-1}(v)
\]

Node observability is used to account for different levels of visibility of non-sensing nodes. As it does not help improve sensing coverage when highly observable nodes are observed, the coverage utility is designed so that it takes into account different “importance” levels of observed contacts (i.e., measured by each contact’s inverse observability).

In our approximation algorithm, we first calculate the observability scores of all non-sensing nodes and coverage utility scores of the sensing nodes from the collected sensing data. Top \( k \) (\( k < n \)) nodes with highest coverage utility scores are doing sensing tasks in the next interval, while \( (n-k) \) nodes with lowest coverage utility scores are replaced by the ones with highest observability.

IV. Evaluation

In our experiments, we use a simulation-based approach with real-world datasets \[1\] collected during SIGCOMM’09 \[6\] (with the social profile of participants) and UIM experiment \[7\]. In terms of sensing coverage, our proposed approach, i.e., \( HCONTEXT \), outperforms RANDOM and GREEDY approximation in different scenarios (Figure 1a and 1b). In terms of bootstrapping methods, during the poster/demo session, the proposed social relationship-based strategies outperform the randomization method (Figure 1c). In terms of sensing parameters, our results suggest that the length of sensing interval should be chosen carefully based on the context of the environment. For the number of sensing devices \( n \), the result (Figure 1d) shows that \( HCONTEXT \) is able to withstand the limited number of sensing devices as it quickly reaches a higher sensing coverage after just a few sensing rounds. This result is highly desirable as lower percentage of sensing devices means saving energy and cost.

**Fig. 1:** Performance evaluation

V. CONCLUSION AND FUTURE WORK

In conclusion, in this paper, we have proposed the notions of node observability, coverage utility score, and design a new context-aware approximation algorithm with human-centric bootstrapping strategies for physical crowd-sensing. We have also verified the effectiveness of our proposed approach via comprehensive experiments on real-world datasets. For the future work, we would like to investigate new algorithms that can automatically adjust sensing interval and number of sensing nodes based on the context of the crowd, while maintaining a desirable sensing coverage.

**REFERENCES**