Quality Evaluation of Crowdsensed Fingerprints for Indoor Localization

***Abstract*—**

I. Introduction

The past decade has witnessed a flourishing of indoor localization systems based on wireless techniques [1], where the fingerprinting based methodology has been widely adopted due to its convenient deployability [?], [?]. The fingerprinting based indoor localization system has two phases: In the offline phase, the site surveyor observes the received signal strength (RSS) of Wi-Fi access points (APs) termed as RSS fingerprints at each reference point, and submit the fingerprints and the location information of the reference point to the localization database; in the online phase, a user needs localization service could submit the observed fingerprints to the database, which then returns the location of the reference point that matches the fingerprints best as the estimated location of the user.

The fingerprinting based method utilizes Wi-Fi APs widely existing in buildings and has no need for other dedicated infrastructure; however, the site survey in the offline phase requires substantial efforts, which is hardly accomplished by any single entity. The recent advances of fingerprinting localization systems utilize mobile crowdsensing approach to collect fingerprints [?], [?], [?], [?], [?], [?]. Mobile crowdsensing is a cost-effective approach to collect large scale data for mobile applications, where individuals with hand-held mobile devices collectively contribute sensing data so that information of certain events could be retrieved [?], [?]. Although sensing participants could receive certain rewards for the efforts and resources spent on the sensing activity, the cost of mobile crowdsensing is still much lower than deploying the dedicated sensing networks [?].

As the crowdsensing data are collected by unprofessional participants with non-dedicated equipment, the sensing data obtained are usually with considerable noise. The quality of the sensing data is the crux for evaluating contribution of the participants, which is the vitally important for effective utilizing rewards to incentivize participants to accomplish sensing tasks satisfactorily. However, how to evaluate the quality of the crowdsensing data is a challenging issue, because there is no ground truth for the collected data to be compared with. Efforts have been made to evaluate the crowdsensing data quality [?], [?], and the task allocation scheme [?], [?] and incentive mechanisms considering the data quality are proposed [?], [?], [?], [?].

While the efforts have been made to study quality evaluation of crowdsensing data for mobile applications in general framework [], how to evaluate the quality of fingerprints with crowdsensing approach in the indoor localization case is still not fully investigated, which is the focus of this work. Our motivation is two-fold. On one hand, the existing work for data quality evaluation in the literature [] can not be applied to the indoor localization system in practice; on the other hand, we want to find out how the very nature of the indoor localization system could facilitate evaluating quality of crowdsensed fingerprints. Our contributions are as following.

First,

Second, .

Third,

II. Related Work

Many related work has been done. About the basic theory of online learning, [?] provided an introduction of online convex programming and an effective algorithm: Generalized Infinitesimal Gradient Ascent for this problem, making for the fundamental construction of online learning. However, it only derives a general model aiming at online convex optimization, with a number of details to be modified and renovated in specific situations. [?] comprehensively displayed online convex optimizing theory, on which our online learning researches are predicated.

With regard to our particular problem setting: minimizing the overall regret with a fixed upper limit of budget,[?] presented an all-round dissection of it. [?] firstly utilized the classical online learning algorithm: Follow the Regularized Leader(FTRL) to update steadily the hypothesis in every round the mechanism should give, in light of information the mechanism has acquired in previous rounds about offered hypotheses and suffered loss. Then [?] embedded FTRL in their main algorithm: Mechanism for no-regret data-purchasing problem, which resolved the problem with an upper regret bound $O(T/\sqrt{B})$.

In terms of concrete techniques applied in [?], importance weighting is an important one, playing a crucial role in estimating unbiasedly the cumulative loss suffered until the current round. This technique is also a research point in our work, and we focus on providing an alternative form of the unbiased estimator to reach better regret bound. One specific application of importance weighting methods–binary classification is shown and analyzed in [?].

Moreover, a bottleneck in solving this problem is the complex form of our objective function and budget constraint, with unknown distributions of costs in each round and an inequality containing integral. [?] made use of inequality zooming methods to clear the integral part in budget constraint, rendering a more computable convex optimizing problem. However, this method can only derive an approximated solution and the corresponding bound may not be tight enough. In our work we try to get access to the accurate solution of this problem, by means of calculus variation. [**?**] illustrates the applicability and methods of calculus variation in detail. A concrete example taking advantage of it lies in [**?**]. [**?**] dealt with the problem that minimizing the variance of estimator with a given price distribution and a fixed budget by calculus variation. Nevertheless, it concentrated on the offline situation, deviating from our online background.

III. Localization Model

IV. Low-cost data purchasing problem

In many situations, we could not get access to all the data for both the reason that the data has a cost and our budget is limited. In this section, we will define the problem of thedesigning of the effective mechanism to acquire the RSS information collected by the crowds. However, the mechanism have no means to know either the data is good enough for our localization or there will be a better one coming after. We implement the online machine learning algorithm in our mechanism.

A. Preliminaries and basic assumption

We first define the loss function according to the localization model above.

$$f_t(h_t) = \tag{1}$$

After we acquire the loss function, we give the defininition of the regret function.

$$R(T) = \sum_{t=1}^{T} f_t(h_t) - \min_{h^* \in H} \sum_{t=1}^{T} f_t(h_t^*)$$

where h^* is the optimal choice, causing the least loss in our solution space H. We also make some assumptions for this problem

1)
2)
3)

B. Online Learning Algorithms

We will here use the classical Follow the Regularized Leader(FoRL) algorithm to work as the Online Algorithms. The FoRL has a upper bound of regret of $O(\sqrt{T})$, which ensures that the average regret tends to zero when.There are many kinds of other Online Algorithms which can be found in **??**, etc. The FoRL is described in **??**.

C. Problem formulation

The problem can be described as follows.

1) a sequence of data $d_1, , , d_T$ coming in time $1, , , , , T$ with each data possessing a posted price c_t, $c_t \in [0, M]$.
2) The mechanism post a price p_t according to a probablity $g_t(p_t)$.
3) If the $p > c_t$ agent accepted the price, the mechanism get the loss function and send it back to the OLA and the mechanism will pay for the posted price c_t. If the agent rejected the price, the mechanism would send a null data to the OLA .

D. Importance Weighting technique

In tradational online learning problem, all the data will be used, and we can consider . In out low-cost purchasing problem, not all the loss function are used, and the estimation of loss is $E(\sum_{t=0}^{T} \delta_t f_t) = \sum_{t=0}^{T} q_t f_t$, where δ_t is the function showing whether the data is procured. Noticing that the definition of regret still includes all the loss, in order to get an unbiased estimator, we define

$$lf_t(h) = \begin{cases} \dfrac{f_t(h_t)}{q_t} & \text{if the is} \quad (2) \\ 0 & \text{else} \quad (3) \end{cases}$$

E. online batch to conversion

We give our final results by averaging every hypothesis h_t acquired in each Details will be added later.

V. The regret minimization senario

In this senario, the mechanism has a fixed budget. The main purpose of the mechanism is to get a high accuracy of localization informaiton, which is consistent with our definition of loss function and regret.

A. Estimate the upper bound of regret

We will first find the upper bound of the regret. [] gives a quite well estimation as shown in the following lemma

Lemma 1. *The regret bound of problem* **??** *using the OLA of FoRL is bounded by*

$$R(T) = \frac{\beta}{\eta} + E(\sum_{t=1}^{T} \frac{\Delta^2_{h_t, f_t}}{q_t})$$

B. Derivation of the Regeret Minimization Problem

First consider the unconstrained problem. If $\overline{y}$ is the local minimum of the functional $J(y)$ if y is the local niminum of $J(y)$, then it holds that $\forall \hat{y}$ in a function space V

$$\delta J|_y(\hat{y} - \overline{y}) \geq 0$$

where ,$\delta J|_y(\hat{y} - \overline{y})$is the Gateaux deravative of J in the direction of $\hat{y} - \overline{y}$.

We then consider the constraint problem. For the constraint optimization problem, we have that if y is the extremal of the constraint problem, then it is also the extremal of the augmented cost functional(Lagarangian) $J(y) + \lambda C(y)$,where the λ is the Lagarange Multiplier, and $C(y)$ is the constraint.

We come back to our problem. We first give our function space $V = \{y | y(0) = 0, y(M) = 1\}$. And we denote our cost function as

$$M(F_1, , , F_n) = \sum_{i=1}^{n} \frac{\alpha_i}{1 - F_i(c_i)}.$$

Then the augmented cost function is derived as

$$J(F_1,,,F_n,\lambda) = M(F_1,,,F_n) + \lambda(\sum_{i=1}^{n}\int_{c_i}^{M} xdF_i(x) - B)$$

According to the calculation, we obtain that for $\forall \hat{F} \in V$

$$\delta J|_{F_t}(\hat{F}_t - F_t) = \int_{c_t}^{M}(-\frac{\alpha_t}{(1-F_t(c_t))^2} + \lambda x)(\hat{f}(x) - f(x))dx$$

if $\overline{F}$ is the local minimum, then we have

$$\delta J(\hat{F}_t - \overline{F_t}) \geq 0$$

holds for every $\hat{F} \in V$.Noticing that

$$\int_{0}^{M} f_t(x) - f(x)dx = 0$$

We must have

$$-\frac{\alpha_t}{(1-F_t(c_t))^2} + \lambda x \geq 0$$

hold on every where on $[c_t, M]$ thus we obtain that

$$F_t(c) = 1 - \sqrt{\frac{\alpha_t}{\lambda c}} \quad c \in (0, M)$$

Noticing that $F(x)$ is not continuous, according to Stieltjes Integral, we rewrite the constraint as following

$$\sum_{i=1}^{n}(\int_{c_i}^{M} xf_i(x)dx + (1 - F_i(M)M) \leq B$$

According to the property of Lagarangian, we have

$$\frac{\partial J}{\partial \lambda} = \sum_{i=1}^{n}(\int_{c_i}^{M} xf_i(x)dx + (1 - F_i(M)M) - B = 0$$

we can solve the λ when we have the form of F_t

C. Interpreting of the result

VI. The budget minimization senario

In this senario, the mechanism do not have a certain amount of budget, instead, an upper bound of regret R_{min} is required as a constraint and the optimization target changes to the minimum of budget.

$$\min E(B)$$

$$s.t. R(T) \leq R_{min}$$

VII. Experiments and Simulations

VIII. Conclusion and Future Work

References

[1] Z. Yang, Z. Zhou and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp.1-32, 2013.