

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 [?], [?], [?], [?], [?], [3]. 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

A. Industrial context

The industrial context our work pertains to is fingerprinting localization based on mobile crowdsensing. In regard to localization via fingerprints, Kaemarungsi and Krishnamurthy draft out the analytical model of it [2], and Wen *et al.* further dig into this model combined with received signal strength, making out theoretical analysis of its limits on indoor localization [3]. Concerning mobile crowdsensing, Ganti *et al.* presents a brief overview of concrete applications about this technique, dissects its distinctive useful features and furthermore blueprints its future utilization and challenges[4].Meanwhile, mobile crowdsensing approach has been applied to various domains such as transportation [5], environment surveillance [6], [7] and location based service [8], [9],evincing the research value of this method in reality.

B. Basic theory

The basic theory of our work is online learning, which is applied to the fingerprinting data procurement phase of indoor localization. Zinkevich 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[10]. 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. Online convex optimizing theory has been comprehensively displayed in [11], on which our online learning researches are predicated.

One of the kernel issues in online learning is how to evaluate the quality of data acquired. In online learning, an acknowledged way to assessing procured data is the regret, which can be viewed informally as the difference of overall loss between the strategy one chooses and the theoretical best strategy reaching the least loss [11].Abernethy *et al.*

crystallize a particular online problem setting with a regret analysis attached[12].

C. Our particular problem

Based on the theoretical tool mentioned above, Our particular problem can be interpreted as: minimizing the overall regret in online learning with a fixed upper limit of budget. With regard to this problem setting, Abernethy *et al.* presented an all-round dissection of it[12]. They 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 They embedded FTRL in their main algorithm: Mechanism for no-regret data-purchasing problem, which resolved the problem with an upper regret bound $O(T/\sqrt{B})$. Compared with what Abernethy *et al.* have done[12], what is special in our work is to further lower the regret bound of online learning in the context of indoor localization—a brand new research background in online learning realm.

In terms of concrete techniques applied in this particular problem, importance weighting is an important one, playing a crucial role in estimating unbiasedly the cumulative loss suffered until the current round[12]. 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 [13].

Moreover, a bottleneck in solving this particular 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. Abernethy *et al.* made use of inequality zooming methods to clear the integral part in budget constraint, rendering a more computable convex optimizing problem[12]. 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. The applicability and methods of calculus variation are illustrated in detail in [14]. A concrete example taking advantage of it lies in [15], which 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 the designing 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.

$$f_t(h_t) =$$

After we acquire the loss function, we give the definition 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, \dots, d_T coming in time $1, \dots, T$ 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 probability $g_t(p_t)$.
- 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 traditional online learning problem, all the data will be used, and we can consider . In our 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

$$lf_t(h) = \begin{cases} \frac{f_t(h_t)}{q_t} & \text{if the is} \\ 0 & \text{else} \end{cases} \quad (1)$$

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. 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*

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

B. Randomized posted price setting

There still many details be determined here.

C. The optimization problem

Now we can change the problem into a more single form.

$$\begin{aligned} & \min \sum_{t=1}^T \frac{\Delta_{h_t, f_t}^2}{q_{c_t}} \\ & s.t. \sum_{t=0}^T \int_{c_t}^M x dq(x) \leq B \end{aligned}$$

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.

$$\begin{aligned} & \min E(B) \\ & s.t. R(T) \leq R_{min} \end{aligned}$$

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.
- [2] K.Kaemarungsi, and P.Krishnamurthy, "Modeling of indoor positioning systems based on location fingerprinting," in *Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, 2004, vol. 2, pp. 1012-1022.
- [3] Y. Wen, X. Tian, X. Wang and S. Lu, "Fundamental limits of RSS fingerprinting based indoor localization," in *Proc. IEEE INFOCOM*, 2015, pp.2479-2487.
- [4] R.K.Ganti, F.Ye, H.Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol.49, no.11, pp.32-39, 2011.
- [5] S.Hu, L.Su, H.Liu, H.Wang, and T.F.Abdelzaher, "Smartroad: Smartphone-based crowd sensing for traffic regulator detection and identification," *ACM Transactions on Sensor Networks*, vol.11, no.4, pp.55, 2015.
- [6] M.Mun, S.Reddy, K.Shilton, N.Yau, J.Burke, D.Estrin, M.Hansen, E.Howard, R.West, and P.Boda, "PEIR, the personal environmental impact report, as a platform for participatory sensing systems research," in *Proc.ACM MobiSys*, 2009, pp.55-68.
- [7] R.Rana, C.Chou, S.Kanhere, N.Bulusu, and W.Hu, "Earphone:An end-to-end participatory urban noise mapping," in *Proc.ACM/IEEE IPSN*, 2010, pp.105-116.
- [8] R.Gao, M.Zhao, T.Ye, F.Ye, Y.Wang, K.Bian, T.Wang, and X.Li, "Jigsaw: Indoor floor plan reconstruction via mobile crowdsensing," in *Proc.ACM MobiCom*, 2014, pp.249-260.
- [9] Y.Wen, J.Shi, Q.Zhang, X.Tian, Z.Huang, H.Yu, Y.Cheng, and X.Shen, "Quality-driven auction-based incentive mechanism for mobile crowd sensing," *IEEE Transactions on Vehicular Technology*, vol.64, no.9, pp.4203-4214, 2015.
- [10] M.Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," School of Computer Science, Carnegie Mellon University, 2003.
- [11] S.Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol.4, no.2, pp.107-194, 2011.
- [12] J.Abernethy, Y.Chen, C.J.Ho, and B.Waggoner, "Low-cost learning via active data procurement," in *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, 2015, pp.619-636.
- [13] A.Beygelzimer, S.Dasgupta, and J.Langford, "Importance weighted active learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp.49-56.
- [14] D.Liberzon, "Calculus of variations and optimal control theory: a concise introduction," Princeton University Press, 2012.
- [15] A.Roth, and G.Schoenebeck, "Conducting truthful surveys, cheaply," in *Proceedings of the 13th ACM Conference on Electronic Commerce*, 2012, pp.826-843.