

Efficient Calibration for RSSI-based Indoor Localization by Bayesian Experimental Design on Multi-task Classification

Masamichi Shimosaka
Tokyo Institute of Technology
Tokyo, Japan
simosaka@miubiq.cs.titech.ac.jp

Osamu Saisho
The University of Tokyo
Tokyo, Japan
saisho@miubiq.cs.titech.ac.jp

ABSTRACT

RSSI-based indoor localization is getting much attention. Thanks to a number of researchers, the localization accuracy has already reached a sufficient level. However, it is still not easy-to-use technology because of its heavy installation cost. When an indoor localization system is installed, it needs to collect RSSI data for training classifiers. Existing techniques need to collect enough data at each location. This is why the installation cost is very heavy. We propose a technique to gather data efficiently by using machine learning techniques. Our proposed algorithm is based on multi-task learning and Bayesian optimization. This algorithm can remove the need to collect data of all location labels and select location labels to acquire new data efficiently. We verify this algorithm by using a Wi-Fi RSSI dataset collected in a building. The empirical results suggest that the algorithm is superior to an existing algorithm applying single-task learning and Active Class Selection.

Author Keywords

indoor localization; RSSI; information gathering; multi-task learning; Bayesian optimization

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Received signal strength indicator (RSSI)-based indoor localization systems are getting a lot of attention thanks to the development of wearable devices and IoT technology. It is expected that they can be used for recognizing an elderly person's activities of daily living (ADLs) and quality of life (QOLs) recognition or for predicting and analyzing supermarket traffic lines [15]. A number of researchers have already attempted to establish variable RSSI-based indoor localization techniques over the past 15 years. As a pioneering work in this field, in 2000, Bahl and Padmanabhan proposed RADAR [2], which applies k-nearest neighbor method.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UbiComp '16, September 12 - 16, 2016, Heidelberg, Germany

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4461-6/16/09...\$15.00

DOI: <http://dx.doi.org/10.1145/2971648.2971710>

Schwaighofer et al. [14] and Ferris et al. [5] leveraged Gaussian process to model RSSI distribution from training data collected beforehand. In 2014, McGibney et al. [1] developed a Wi-Fi RSSI and particle-filter-based tracker that achieved a 1.6-m accuracy as the second ranked performance among the state of the art techniques in the Microsoft Indoor Localization Competition¹. This means that localization systems have already reached sufficient accuracy as a way to recognize human indoor activities. However, the technologies they feature are still not easy-to-use. One of the biggest factors is the deployment cost needed for system installation.

In recent years, easy calibration methods for indoor localization systems have also been studied actively. Kawajiri et al. [8] proposed a semi-supervised approach and Rai et al. [13] proposed automatic calibration approach utilizing other sensors included in smartphones. These can reduce the calibration cost thanks to machine learning technology, however, they remain a problem that they don't consider which point to gather training data at for higher efficiency.

Kawajiri et al. [9] proposed a technique combining a calibration efficiency method based on Active Class Selection(ACS) [10] and a technique controlling crowd incentives. The main purpose of their research was to develop a novel technique for steering crowds. However, applying ACS also created critical issues in terms of information gathering. Although information gathering is a research field aiming to collect data efficiently, the technique proposed by Kawajiri et al. [8] created two major problems due to the simplicity of ACS. First, it needs data associated with all locations because learning classifiers at each label is independent. Single-task learning can't share any data with the other tasks. Second, it doesn't consider the potential in system improvement because of the framework of ACS. ACS can consider only the average of estimated location error distance.

Our research objective is to build an efficient calibration algorithm for an RSSI-based indoor localization system. Specifically, the main purpose is removing the need to collect data corresponding to all labels and selecting location to acquire data in turn for efficient calibration. We first enhance learning efficiency by enabling all classifiers to learn simultaneously. We also enhance the information gathering efficiency by considering the potential of each location in minimizing the total loss of the indoor localization system.

¹<http://research.microsoft.com/en-us/events/ipsn2014indoorlocalizationcompetition/>

The two main contributions of this paper are as follows. First, we eliminate the need for collecting data associated with all location labels by applying of multi-task learning. Multi-task learning is a machine learning method that learns problems simultaneously and efficiently considering the relationships to other tasks [22]. Second, we construct an information gathering technique that takes into account both of the expected loss and its uncertainty at each location based on the framework of Bayesian optimization. Bayesian optimization [11] is an approach to select an optimal location to maximize cumulative rewards sequentially. The algorithm we propose will also be helpful for other applications that require constructing spatio-temporal models.

RELATED WORKS

Indoor Localization Calibration

Indoor localization systems require training data for machine learning. Thanks to a number of researchers, the localization accuracy has already reached a sufficient level. However, the technologies the systems use are still not easy to use because of their heavy installation cost. Thus, cost reduction techniques have been studied actively in recent years.

Semi-supervised machine learning technologies have been studied to reduce the annotation cost. The technique proposed by Kashima et al. [7] applies a semi-supervised approach and defines similarity measures between data. Thus it needs to collect only a small amount of labeled data if it can collect a large amount of unlabeled data. However, it can't remove the need to acquire one or more data elements for each location.

Transfer learning and domain adaptation have also been applied for reducing the cost of indoor localization calibration [12, 3]. These methods collect training data at a source domain and test data at a target domain. However, they also need to acquire at least one data element at each location. Zheng et al. [21] solved the problem of individual differences by utilizing robust feature representation. It doesn't need any target domain data for calibration. However, Zheng et al. needs enough source domain data.

The method Kawajiri et al. [8] proposed saves the labor involved in stopping and gathering data at each location by semi-supervised learning. Rai et al. [13] proposed an automatic calibration method utilizing an acceleration sensor, a compass sensor and a gyro sensor, which are installed in smartphones and floor maps. In addition to Rai et al. [13], not a few researchers such as Shin et al. [17] and Kawajiri et al. [9] have developed methods that utilize crowds' cooperation. However, except for Kawajiri et al. [9] their methods don't consider which label to get new data and Kawajiri et al. is not enough.

The method developed by Kawajiri et al. [9] applies an ACS framework [10] for score calculation. This score indicates the need for gathering information at each location and steers crowds to acquire the data of high score locations. However, this indicator can't express the improvement expectation value because ACS takes into account only the average of estimated location error distance. Moreover, it can't learn classifiers efficiently because the learning method is based

on single-task learning for multi-class classifiers. Single-task learning can't share any data with the other tasks, so it needs data corresponding to all location labels. Our approach applies multi-task learning and Bayesian optimization to solve these problems.

Multi-task Learning

Multi-task learning is a learning method that learns some problems simultaneously and efficiently considering the relationships to other tasks [22]. An indoor localization system needs to construct classifiers at each location in the scope. The classifiers at two locations close to each other should be learned simultaneously because they will be similar and the relationships between them will be clear. Multi-task learning can express the relationships between tasks by learning the parameters of classifiers while taking spatial relationships and the characteristics into account. Moreover, learning infinite or continuous tasks with high accuracy is possible thanks to Takeuchi et al. [20]. They developed a method to mixture learning models of discrete and finite tasks.

Bayesian Optimization

Bayesian optimization [11] is an approach to select an optimal location to maximize cumulative rewards sequentially. Specifically, this approach repeats the following information gathering process. First, the expected value and the standard deviation of rewards at each location in the scope are obtained by using Gaussian process that is trained by a reward map based on already obtained data. The need to gather information at each location is indicated by acquisition function A based on the expected value and the standard deviation. Gaussian Process Upper Confidence Bound (GP-UCB) [18], an indicator of common acquisition function, is defined as

$$A(\mathbf{x}) = \mu(\mathbf{x}) + C\sigma(\mathbf{x}) \quad (1)$$

where $\mu(\mathbf{x})$, and $\sigma(\mathbf{x})$ respectively denote the expected value and the standard deviation at \mathbf{x} and C is a positive constant for adjusting the balance between $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$. Then, a new data element is acquired at the location whose acquisition function is the maximum in the scope, and Gaussian process is renewed.

ACS is another approach equivalent to (1) when $C = 0$. Thus Bayesian optimization based on GP-UCB is different from ACS; it can express not only the expected value but also its uncertainty of loss.

RSSI-BASED INDOOR LOCALIZATION AS MULTI-CLASS CLASSIFICATION

In wave propagation theory, received signal strength is in inverse proportion to the propagation distance. An RSSI based indoor localization system applies this property. It measures RSSI from access points and localizes on the basis of the data obtained in a way such as triangulation (Figure 1). However, the effects of reflection and absorption from indoor structures and people make the measurement unstable, and it is hard to predict them without real data. Instead, using machine learning technology is a popular way to estimate locations stochastically. In previous methods, enough training data was collected at each location and indoor localization problems were formulated as multi-class classification problems.

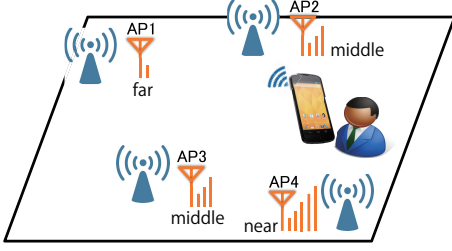


Figure 1: how to localize by using RSSI

The number of Wi-Fi access points is S , and $\mathbf{x} \in \mathbb{R}^S$ is the vector of RSSI obtained from the access points. The output of trained classification model \mathbf{y} is the likelihood vector of locations. Namely, at each location \mathbf{r} , the likelihood of \mathbf{r} is $y_{\mathbf{r}}$, the classification model can be represented as (2), and the localization result can be represented as (3):

$$y_{\mathbf{r}} = f(\mathbf{x}|\boldsymbol{\theta}_{\mathbf{r}}) \quad (2)$$

$$\hat{\mathbf{r}} = \operatorname{argmax}_{\mathbf{r} \in \mathcal{R}} y_{\mathbf{r}} \quad (3)$$

where $\boldsymbol{\theta}_{\mathbf{r}}$ is a parameter vector. In this research, we formulate an indoor localization problem as a multi-class classification problem, and this problem can be represented as

$$f(\mathbf{x}|\boldsymbol{\theta}_{\mathbf{r}}) = \boldsymbol{\theta}_{\mathbf{r}}^T \boldsymbol{\phi}(\mathbf{x}) \quad (4)$$

where $\boldsymbol{\phi}(\mathbf{x})$ is a feature vector of RSSI. The feature design of $\boldsymbol{\phi}(\mathbf{x})$ follows Shimosaka et al. [16]. The m -th element ($m \in \{1, \dots, M\}$) of a feature vector $\boldsymbol{\rho}(x_s) \in \mathbb{R}^M$ ($s \in \{1, \dots, S\}$) coming from the s -th sensor can be represented as $\boldsymbol{\rho}(x_s)_m = \exp(-(x_s - \mu_m)^2/\sigma^2)$ where the all μ_m are in arithmetic progression in the range of the measured RSSI. The $\boldsymbol{\phi}(\mathbf{x}) \in \mathbb{R}^{MS}$ is a feature vector independent of location labels, where $\boldsymbol{\phi}(\mathbf{x}) = (\boldsymbol{\rho}(x_1)^T, \dots, \boldsymbol{\rho}(x_S)^T)^T$.

PROPOSED ALGORITHM

In this research, we construct a novel efficient information gathering method for calibrating of indoor localization systems. The proposed algorithm enables function f whose argument is RSSI (not place) to learn considering the proximity of the place. It is based on multi-task learning and Bayesian optimization.

Multi-task Training Setting

Even though classifiers constructed at different locations are different from each other, classifiers constructed at two closely located locations should be similar if there is no barrier and environmental change between them. Thus we introduce multi-task learning into an RSSI-based indoor localization system. Thanks to multi-task learning, the system can learn classifiers at all locations in the scope simultaneously.

In this research, we assume that the scope can be represented as a graph structure for which each node is a location label. For multi-task learning with a graph structure, we apply regularization learning. Regularization is a machine learning technique to introduce additional information for improving the generalization of models. The regularization learning can be represented as

$$\hat{\Theta} = \operatorname{argmin}_{\Theta} \sum_{(\mathbf{x}^{(r)}, y^{(r)}) \in \mathcal{D}} l(\mathbf{x}^{(r)}, y^{(r)}, \boldsymbol{\theta}_{\mathbf{r}}) + R(\Theta) \quad (5)$$

where Θ denotes concatenation of all $\boldsymbol{\theta}_{\mathbf{r}}$ and $\mathbf{r}^{(y)}$ denotes \mathbf{r} corresponding to the location of answer label y . The localization result is defined as $\hat{\mathbf{r}}$. Loss term $l(\mathbf{x}^{(r)}, y^{(r)}, \boldsymbol{\theta}_{\mathbf{r}})$ and regularization term $R(\Theta)$ can be respectively represented as (6) and (7).

$$l(\mathbf{x}^{(r)}, y^{(r)}, \boldsymbol{\theta}_{\mathbf{r}}) = d(\hat{\mathbf{r}}, \mathbf{r}^{(y)}) [1 - \boldsymbol{\theta}_{\mathbf{r}^{(y)}}^T \boldsymbol{\phi}(\mathbf{x}) + \boldsymbol{\theta}_{\hat{\mathbf{r}}}^T \boldsymbol{\phi}(\mathbf{x})]_+ \quad (6)$$

$$R(\Theta) = \lambda_1 \sum_{\mathbf{r} \in \mathcal{R}} \|\boldsymbol{\theta}_{\mathbf{r}}\|_2^2 + \lambda_2 \sum_{(\mathbf{r}_i, \mathbf{r}_j) \in \mathcal{E}} \|\boldsymbol{\theta}_{\mathbf{r}_i} - \boldsymbol{\theta}_{\mathbf{r}_j}\|_2^2 \quad (7)$$

where $d(\cdot, \cdot)$ denotes the distance between two locations defined as

$$d(\mathbf{r}_i, \mathbf{r}_j) = \|\mathbf{r}_i - \mathbf{r}_j\|_2 \quad (8)$$

and $[z]_+ = \max(0, z)$, \mathcal{E} denotes the set of edges in the graph structure and λ_1 and λ_2 are positive constants related to the strength of regularization. The loss term definition follows Shimosaka et al. [16], i.e., hinge loss taking the estimation error distance into account.

It is usually difficult to optimize loss terms and regularization terms simultaneously. Moreover, parameters are related to each other because of the structure regularization term in (5). Therefore, we repeat their updates alternately by using online proximal gradient descent type of alternating direction multiplier method (OPG-ADMM) [19], which is one type of stochastic ADMM approach to calculate them effectively. OPG-ADMM is an approach combining forward backward splitting (FOBOS) [4], which is often used for achieving calculation efficiency in online learning while taking regularization terms into account, and alternating direction multiplier method (ADMM) [6], which is often used in batch learning.

Information Gathering by Bayesian Experimental Design

The ACS-based method proposed by Kawajiri et al. [9] takes into account only the expected loss at each location when it determines the next location to acquire new data. Practically, $l_{\mathbf{r}}$ is an expected loss value based on \mathcal{D} and includes uncertainty that expresses the potential for improvement of the loss function. Efficient information gathering method should not be determined by the expected loss function values (as the method proposed by Kawajiri et al. does), but should also take uncertainty into account. Thus we introduce the frameworks of mixing Bayesian optimization and experimental design.

In this research, we define $l_{\mathbf{r}}(\mathcal{D}, \Theta, l)$ as the loss function of each task \mathbf{r} and express $l_{\mathbf{r}}$ as $\mathcal{N}(\mu_{\mathbf{r}}, \sigma_{\mathbf{r}}^2)$ by using Gaussian process for taking the uncertainty into account. The expected value of the decrease in the loss function thanks to the new data acquired at \mathbf{r} is represented by acquisition function A_Q defined as GP-UCB (9).

$$A_Q(\Theta^{(n)}, \mathcal{D}^{(n)}, \mathbf{r}) = \mu_{\mathbf{r}} + C\sigma_{\mathbf{r}} \quad (9)$$

Here, $\Theta^{(n)}$ and $\mathcal{D}^{(n)}$ are the values for the n -th information gathering and C denotes a positive constant for adjusting the balance between $\mu_{\mathbf{r}}$ and $\sigma_{\mathbf{r}}$. A_Q takes both of the expected value and the uncertainty of the loss function into account thanks to the definition of GP-UCB. Thus we introduce (10)

as a determining formula instead of ACS.

$$\mathbf{r}^{(n+1)} = \operatorname{argmax}_{\mathbf{r} \in \mathcal{R}} A_Q(\Theta^{(n)}, \mathcal{D}^{(n)}, \mathbf{r}) \quad (10)$$

In order to calculate l_r , we utilize the distance of the ideal output and the output from the current optimized prediction model. The ideal output can be calculated from \mathbf{y} for each instance in \mathcal{D} and measurement error. $l'(\mathbf{x}^{(r)}, \mathbf{y}^{(r)}, \theta_r^{(n)})$ denotes the loss for each instance belonging to \mathcal{D} calculated from the two outputs. For example, Kullback-Leibler divergence can be applied to l' if the output is probability distribution, and mean square error can be applied to l' if the output is a real number. If information gathering is conducted at all \mathbf{r} belonging to \mathcal{R} , the problem can be solved by calculating the average of empirical error at each \mathbf{r} . In this case, we can replace l_r with \tilde{l}_r defined as (11).

$$\tilde{l}_r = \frac{1}{M_r^{(n)}} \sum_{(\mathbf{x}^{(r)}, \mathbf{y}^{(r)}) \in \mathcal{D}} l'(\mathbf{x}^{(r)}, \mathbf{y}^{(r)}, \theta_r^{(n)}) \quad (11)$$

where $M_r^{(n)}$ denotes the count of data acquisition at \mathbf{r} at n -th information gathering. However, the definition by (11) can't calculate the loss function at \mathbf{r} with no data. In order to solve the problem, we apply the data similarity of two locations close to each other. At such locations, not only the parameters but also the error function should be similar. In order to take this similarity into account, we apply Gaussian process to the error function and complement missing data. Gaussian process is a method often utilized for spatio-temporal data analysis. Specifically, we create a dataset $\{(\mathbf{r}, l'(\mathbf{x}^{(r)}, \mathbf{y}^{(r)}, \theta_r^{(n)}))\}$ whose size is equal to that of $\mathcal{D}^{(n)}$ and optimize Gaussian process $\mathcal{GP}(\mathbf{r}|\mathcal{D}^{(n)})$. Thanks to Gaussian process, $l_r(\mathcal{D}^{(n)}, \Theta^{(n)})$ at each \mathbf{r} belonging to \mathcal{R} can be expressed as Gaussian distribution $\mathcal{N}(\mu_r^{(n)}, \sigma_r^{(n)2})$ by the equation $l_r(\mathcal{D}^{(n)}, \Theta^{(n)}) = \mathcal{GP}(\mathbf{r}|\mathcal{D}^{(n)})$. Thus, we can calculate A_Q without the necessity of acquiring data at all \mathbf{r} belonging to \mathcal{R} . The algorithm for calculating acquisition function is summarized in Algorithm 1. In this research, \mathcal{D} is constructed by experimental design with acquisition function. The framework of the proposed algorithm is summarized in Algorithm 2.

Algorithm 1 update acquisition functions by GP

```

given:
 $\mathcal{D}^{(n)} = \{(\mathbf{x}^{(r)}, \mathbf{y}^{(r)})\}$ 
for  $\mathbf{r}$  in  $\mathcal{R}$  do
  for  $m = 1$  to  $M_r^{(n)}$  do
    calculating error function  $l'(\mathbf{x}^{(r)}, \mathbf{y}^{(r)}, \theta_r^{(n)})$ 
  end for
end for
learning Gaussian process  $\mathcal{GP}(\mathbf{r}|\mathcal{D}^{(n)})$  by
 $\{(\mathbf{r}, l'(\mathbf{x}^{(r)}, \mathbf{y}^{(r)}, \theta_r^{(n)}))\}$ 
calculating  $(\mu_r, \sigma_r)$  ( $\forall \mathbf{r} \in \mathcal{R}$ ) by  $\mathcal{GP}(\mathbf{r}|\mathcal{D}^{(n)})$ 
for  $\mathbf{r}$  in  $\mathcal{R}$  do
 $A_Q(\Theta^{(n)}, \mathcal{D}^{(n)}, \mathbf{r}) = \mu_r + C\sigma_r$  ( $\forall \mathbf{r} \in \mathcal{R}$ ) (9)
end for

```

Table 1: dataset information

space size	about 65 meters \times 70 meters
# of location labels	47
# of scans	1862
# of found access point	1478
# of participants	18
the experimental period	5 weeks

EXPERIMENT

This section describes an empirical method and evaluation to verify the effectiveness of the proposed algorithm.

Experimental Environment and Comparative Method

For evaluation, we utilized a part of the Wi-Fi RSSI dataset made available on the Internet by Kawajiri et al. [9]. We utilize data acquired at the second floor of a building, even though this dataset includes data acquired at multiple floors. Since the second floor is connected to a facade, almost all users use it when they walk into and out of the building. The information for the dataset we used is showed in Table 1. Regarding RSSI-based indoor localization systems, differences in the devices used can produce gaps in the output that cannot be ignored. Thus, taking the devices used into account, we divide data uniformly into a training dataset and a test dataset considering used devices. We do not, however, use all data in the training dataset because our research objective is to construct an efficient information gathering algorithm for indoor localization systems. That is to say, what we want to do is to select a location that will enable the proposed algorithm to acquire new data, extracting labeled RSSI data from the training dataset, and training classifiers with only the extracted data.

In this experiment, we introduce multi-task learning by using the way the map expressed the second floor as a graph structure. The map and the graph structure are shown in Figure 2. This graph structure consists of four undirected graphs comprising 47 nodes and 48 edges in total. All nodes correspond to location labels for acquiring data. The blue area is semi-external space deviding the yellow, purple, and red areas.

Algorithm 2 multi-task experimental design

```

given:
 $\mathcal{R} = \{\mathbf{r}\}$ 
 $\mathcal{D}^{(0)} = \emptyset$ 
 $n = 0$ 
while do
   $n = n + 1$ 
  if  $n = 1$  then
    select  $\mathbf{r}^{(1)}$  randomly
  else
     $\mathbf{r}^{(n)} = \operatorname{argmax}_{\mathbf{r} \in \mathcal{R}} A_Q(\Theta^{(n-1)}, \mathcal{D}^{(n-1)}, \mathbf{r})$ 
  end if
  sampling:  $\mathcal{D}^{(n)} = \mathcal{D}^{(n-1)} \cup \{(x_n^{(r)}, y_n^{(r)})\}$ 
  optimizing  $\Theta$  by multitask learning using  $\mathcal{D}^{(n)}$ 
  calculating  $A_Q(\Theta^{(n)}, \mathcal{D}^{(n)}, \mathbf{r})$  ( $\forall \mathbf{r} \in \mathcal{R}$ ) by Gaussian
  process (Algorithm 1)
end while

```

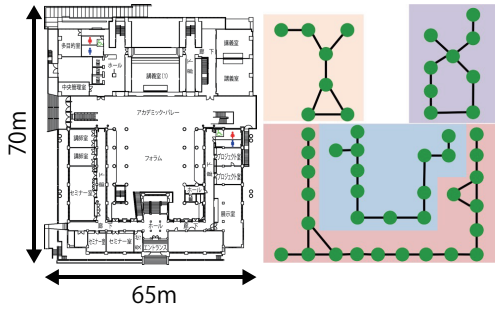


Figure 2: experiment details. left:map, right:graph structure

This is why the graph in Figure 2 is divided into four connected graphs. RSSI and classifiers of edge nodes of different areas may not be similar because of the difference of environment even though they are located closely. Moreover, if there are barriers such as walls and pillars between two close nodes, we didn't connect them. All RSSI data elements belonging to the dataset consisting of acquired data are labeled; we call them the "acquired dataset". When information gathering is conducted, labeled data elements are extracted from the training dataset and added to the acquired dataset. To accelerate the experiment, we fix the first location for acquiring data and acquire two labeled data elements when gathering information. The first location is the upper right hand corner (colored in purple) of Figure 2.

In order to verify the proposed algorithm's effectiveness, we conduct an experiment that compares the localization accuracy to the accuracy of a conventional algorithm combining single-task classifier learning and ACS-based information gathering. The performance criteria we assumed are the average, median, and 95-percentile value of the localization error distance for each number of acquired data elements. All of these criteria indicate that smaller values mean better accuracy.

Experimental Results

Figure 3 shows the three graphs of the average, median and 95-percentile value of the error distance, respectively for our proposed algorithm (proposed) and the conventional algorithm (Kawajiri+) proposed by Kawajiri et al. When there are four or less data elements, there is little difference between the results these algorithms produce. However, when there are six or more data elements, in all three graphs the blue line (proposed) is almost always under the yellow line (Kawajiri+). This means that our proposed algorithm is better than the conventional algorithm. Moreover, for the average and median graphs, the convergence speeds are very high and the results are stable when there are 6 to 20 data elements. This is a further indication of our proposed algorithm's superiority. Namely, the proposed algorithm enables indoor localization system to calibrate with only a few training data.

Figure 4 shows the changes in acquisition function. the location of red circle is the location where the acquisition function is maximal. The size of the other blue circles expresses the relative size of acquisition function at the location. When we apply the conventional algorithm, the corner with no data tends to be selected, but when we apply our proposed algorithm, the location slightly to the center from the corner with

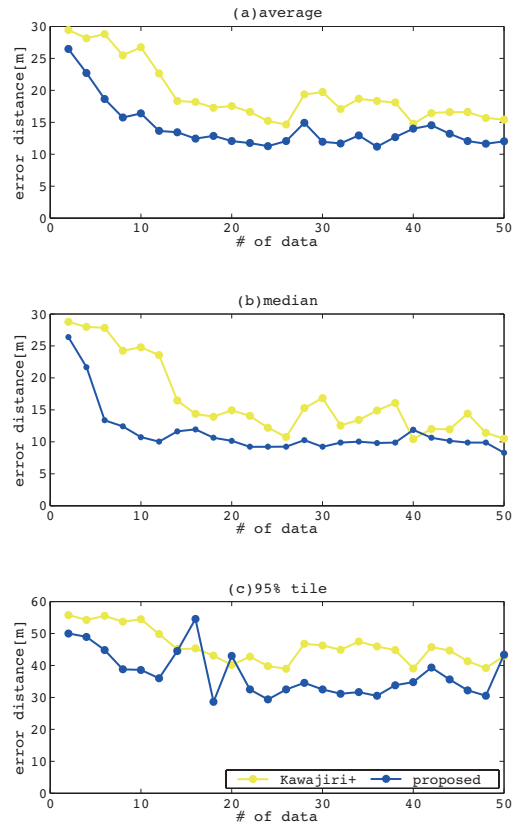


Figure 3: experimental results

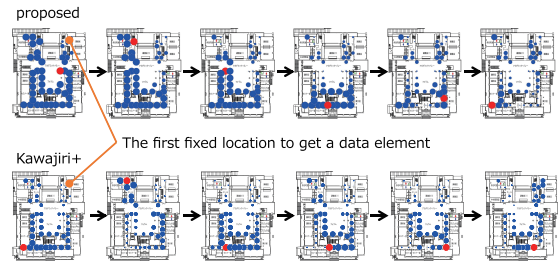


Figure 4: change of acquisition function

no data tends to be selected. Data acquired at these locations is more useful for multi-task learning because its complementary range is wider. This also indicates our proposed algorithm's effectiveness.

CONCLUSION

In our research, we constructed an effective information gathering method combining multi-task learning and Bayesian optimization for efficiently calibrating RSSI-based indoor localization systems. Obtained experiment results verified our proposed algorithm's efficacy and showed it is more efficient than conventional algorithm. Our proposed algorithm removes the need to acquire data for all locations, thus reduced calibration cost. For future work, we will generalize the algorithm for variable types of information gathering and to it add a mathematical guarantee.

ACKNOWLEDGEMENT

This work was partly supported by CREST, JST and by JSPS KAKENHI Grant Number 25700026.

REFERENCES

1. A. Mcgibney, C. B., and Klepal, M. MapUme: Smartphone localisation as a service a cloud based architecture for providing indoor localisation services. In *Proc. of IPIN* (2012).
2. Bahl, P., and Padmanabhan, V. RADAR: An in-building rf-based user location and tracking system. In *Proc. of INFOCOM* (2000), 775–784.
3. Baktashmotlagh, M., Harandi, M., Lovell, B., and Salzmann, M. Unsupervised domain adaptation by domain invariant projection. In *Proc. of ICCV* (2013), 769–776.
4. Duchi, J., and Singer, Y. Efficient online and batch learning using forward backward splitting. *J. Mach. Learn. Res.* 10 (2009), 2899–2934.
5. Ferris, B., Fox, D., and Lawrence, N. WiFi-SLAM Using Gaussian Process Latent Variable Models. In *Proc. of IJCAI* (2007), 2480–2485.
6. Gabay, D., and Mercier, B. A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. *Comput. Math. Appl.* 2 (1976), 17–40.
7. Kashima, H., Suzuki, S., Hido, S., Tsuboi, Y., Takahashi, T., Ide, T., Takahashi, R., and Tajima, A. A semi-supervised approach to indoor location estimation (2007).
8. Kawajiri, R., Shimosaka, M., Fukui, R., and Sato, T. Frustratingly simplified deployment in wlan localization by learning from route annotation. In *Proc. of ACML* (2012), 191–204.
9. Kawajiri, R., Shimosaka, M., and Kashima, H. Steered crowdsensing: Incentive design towards quality-oriented place-centric crowdsensing. In *Proc. of UbiComp* (2014), 691–701.
10. Lomasky, R., Brodley, C., Aernecke, M., Walt, D., and Friedl, M. Active class selection. In *Proc. of ECML* (2007), 640–647.
11. Mockus, J. Application of bayesian approach to numerical methods of global and stochastic optimization. *J. of Global Optimization* 4 (1994), 347–365.
12. Pan, S., Zheng, V., Yang, Q., and Hu, D. Transfer learning for WiFi-based indoor localization. In *Adj. Proc. of AAAI* (2008), 43–48.
13. Rai, A., Chintalapudi, K., Padmanabhan, V., and Sen, R. Zee: Zero-effort crowdsourcing for indoor localization. In *Proc. of MobiCom* (2012), 293–304.
14. Schwaighofer, A., Grigoras, M., Tresp, V., and Hoffmann, C. GPPS: A Gaussian Process positioning system for cellular networks. In *Proc. of NIPS* (2003), 579–586.
15. Shimosaka, M., Masuda, S., Takeichi, K., Fukui, R., and Sato, T. Health score prediction using low-invasive sensors. In *Adj. Proc. of UbiComp* (2012), 1044–1048.
16. Shimosaka, M., Saisho, O., Sunakawa, T., Koyasu, H., Maeda, K., and Kawajiri, R. Zigbee based wireless indoor localization with sensor placement optimization towards practical home sensing. *Advanced Robotics* 30, 5 (2016), 315–325.
17. Shin, H., Chon, Y., Kim, Y., and Cha, H. A participatory service platform for indoor location-based services. *Pervasive Computing* 14, 1 (2015), 62–69.
18. Srinivas, N., Krause, A., Kakade, S., and Seeger, M. Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE Trans. Information Theory* 58, 5 (2012), 3250–3265.
19. Suzuki, T. Dual averaging and proximal gradient descent for online alternating direction multiplier method. In *Proc. of ICML* (2013), 392–400.
20. Takeuchi, I., Hongo, T., Sugiyama, M., and Nakajima, S. Parametric task learning. In *Advances in NIPS* (2013), 1358–1366.
21. Zheng, V., Cao, H., Gao, S., Adhikari, A., Lin, M., and Chang, K. Cold-start heterogeneous-device wireless localization. In *AAAI* (2016), 1429–1435.
22. Zhou, J., Chen, J., and Ye, J. Multi-task learning: Theory, algorithms, and applications. *SDM tutorials*, 2012.