Adaptive Indoor Localization with Wi-Fi Based on Transfer Learning

Anmin Hu and Lijun Zhang
School of Electronic and Information Engineering
Beijing Jiaotong University
Beijing 100044, China

Abstract--**Many approaches used in The Wi-Fi based indoor location system (WILS) typically assume the distribution of the signal strength data is time invariant. However, the assumption does not hold in real world, which degrades the location accuracy. We propose an algorithm that can adjust the distribution of training data by mixing a fraction of new data. Experimental results show that our algorithm can greatly improve the localization accuracy and reduce a great amount of the calibration effort**
Keywords: Wi-Fi; indoor localization; machine learning; transfer learning

I. INTRODUCTION

The distribution of signal strength may change a lot when the indoor environment alternates [1], which serves against the performance of localization. Simultaneously, it is costly, even infeasible to collect new data in a large-scale environment.

In some previous works, many approaches have been proposed to handle the problem [2, 3]. Among others, transfer learning in indoor localization has attracted much attention in machine learning research due to its feasibility [3]. Transfer learning [4] is proposed to deal with the problem introduced by the training data from source domain and test data from target domain in different feature spaces. This technique has been applied in two major contexts. One is instance-based approach where new weights of instance in a source domain can be learnt. The other is feature-based approach which finds the common feature in a low dimension. Several feature-based transfer learning algorithms have been proposed to handle the problem such as in [2] and [3]. However, none of them considers instance transfer learning.

In this essay, we propose an algorithm based on instance transfer learning, which adapts an out-of-date localization model to a disparate signal distribution by only collecting a fraction of new data.

II. ALGORITHM

A. Problem description

Suppose totally m APs are deployed in a large-scale environment. The Received Signal Strength (RSS) is defined as a signal vector $X=(x^1,x^2,...,x^j,...,x^n)$, where x^j stands for the fingerprint of the j_{th} position . For $x^j=(x_1^j,x_2^j,...,x_i^j,...,x_m^j)\in R^m$, x_j^i stands for RSS value received from AP_i at the j_{th} position. The position label vector $Y=(y^1,y^2,\square\ y^j,\square\ ,y^n)$, where y^j stands for the location of the j_{th} position. Then, we use these data to train a model which can map fingerprint to location. Consider the Wi-Fi fingerprint data collected in two different time periods, say T_0 and T_1. We have enough labeled data (n instances) collected in period T_0 as X_{T_0} and a small amount of labeled data (k instances) collected in period as X_{T_1}. We want to locate in period T_1. However because the signal distribution of two data set is different, we cannot directly use data collected in T_0 to train model and locate in period T_1. Generally, this situation in WILS can be formulated as a uniform learning problem [1]. Although distributions are distinct, they are based on a common physical space, which makes the transfer learning feasible [4]. Our goal is to obtain an adaptive model by using these data.

B. Our algorithm

When environment changes, some instances of X_{T_0} become out of date, thus causing the old localization inaccurate. We can select part of data from X_{T_0} that have similar distribution to X_{T_1} by adjusting the weights of instances.

Suppose all instances in X_{T_0} have the same weight $1/n$ at the beginning, instances in X_{T_1} also have the same weight $1/k$ .we use X_{T_0} and a fraction of X_{T_1} to train N classifiers $h_t(x_i)$ in N times iterations to adjust the weights. Through N iterations, we can vote on each useful data instance based on the value of weights.

Define the weight adjustment as follows：

$$W_i^{t+1}==\begin{cases}W_i^t\beta^{|f(x_i)|} & x_i\in X_{T_o}\\ W_i^t\beta_T^{-|f(x_i)|} & x_i\in X_{T_1}\end{cases}\quad(1)$$

$$f_t(x_i)=\begin{cases}1 & if\ \ h_t(x_i)\neq y_{x_i}\\ 0 & if\ \ h_t(x_i)=y_{x_i}\end{cases}\quad(2)$$

$$\varepsilon_{T_1}=\sum_{i=1}^{k}\frac{w_i^t.f(x_i)}{\sum_{i=1}^{k}w_i^t}\quad(3)$$

$$\beta_{T_1}=\frac{\varepsilon_{T_1}}{1-\varepsilon_{T_1}}\quad \beta=\frac{\varepsilon_{T_1}}{1+\sqrt{\ln\frac{n}{N}}}\quad(4)$$

C. Classifier combination

Define β_{T_1} as the importance of classifier in different iteration, and combine the classifiers to get the final model.

$$F(x)=\sum_{i=1}^{t}\beta_{T_1}^{t}.h_t(x_i) \tag{5}$$

Algorithm: Adjusting the distribution

input: labeled data X_{T_0} , new labeled data X_{T_1} , a based algorithm learning and the maximum number of iteration *N*.
output: Location model
begin
1. Initialize the initial weight vectors
2.**repeat (** $t \le N$ **)**
   a) Set $P_t=\frac{W}{\sum_{i=1}^{n+m} w_i^t}$
   b) Call learner, providing it the X_{T_0} and X_{T_1} . Then, get back a hypothesis $h_t: X \to Y$
   c) Calculate the error of h_t on the. X_{T_0} and X_{T_1}
   d) Update the β_{T_1} (4)
   e) Update the weight vector as (1)
3. Rank the weights of instance in X_{T_0} and select instances with large weights.
4. Output the hypothesis

$$F(x)=\sum_{i=1}^{t}\beta_{T_1}^{t}.h_t(x_i)$$

end

D.Analysis of Algorithms

When indoor environment changes, we only need a small fraction of new data to adjust the distribution of the previous training data through our algorithm. Compared with the original training data, the distribution of the new training data is more similar to the distribution of the test data collected in the new environment. And the positioning accuracy is improved.

III. EXPERIMENTS

A. Dataset

The UJIIndoorLocn dataset [5, 6] from UCI Machine Learning Repository is used, which consists of 21,048 Wi-Fi fingerprints. This data set is divided into a training set with 19,937 fingerprints and a testing set with 1,111 fingerprints. The testing data are performed 4 months later than when the training data gathered. Therefore, the distributions of the training data and testing data are different, and to locate these data is challenging.

B. Setup and Results

In our experiment, we split the original testing data into two-part, one part as new training data to train classifier and the other as validation for our algorithm.

To prove the validity of our algorithm, we do other two experiments. One uses original data to train model and use original testing data to test the model, as experiment 1. The other uses partial testing data to train model and uses another part to test model, as experiment 2.

As shown in Table 1, the accuracy of building and floor detection in experiment 1 only achieves 72%, and the accuracy in experiment 2 is 80%. Our algorithm can achieve 90%. We believe that the 18% difference between experiment 1 and our algorithm is caused by the different distributions between training data and testing data, and the 10% difference between experiment 2 and our algorithm is caused by insufficient data.

Table 1. Accuracy comparison among different methods

	Experiment1	Experiment 2	Proposed
Accuracy	74%	78%	90%

IV. CONCLUSION

An algorithm based on instance transfer learning for indoor localization is proposed, which only requires a fraction of new signal data to modify the existed model. In comparison with other machine learning approaches, our algorithm is promising in terms of accuracy and implementation complexity.

ACKNOWLEDGEMENT

This work was supported in part by National Natural Science Foundation of China under grant no. 61872031 and 61572071.

REFERENCE

[1] V. Zheng, E. Xiang , Q. Yang, *et al*., "Transferring localization models over time," *National Conference on Artificial Intelligence*. AAAI Press, 2008.

[2] Z. Sun, Y. Chen, J. Qi , *et al.*, "Adaptive localization through transfer learning in indoor Wi-Fi environment," *Seventh International Conference on Machine Learning and Applications, ICMLA 2008*, San Diego, California, USA, Dec. 11—13, 2008.

[3] K. Weiss, T. Hoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, pp. 9, 2016.

[4] W. Dai, G. Yang, R. Xue, *et al.,* "Boosting for transfer learning," *International Conference on Machine Learning*, 2007.

[5] E. Lohan, J. Torres-Sospedra, H. Leppäkoski, *et al.*, "Wi-Fi crowdsourced fingerprinting dataset for indoor positioning," *Data*, vol. 2, no. 4, pp. 32, 2017.

[6] F. Zafari, A. Gkelias, and K. Leung, "A survey of indoor localization systems and technologies," arXiv preprint arXiv:1709.01015, 2017.