Learning Methods for RSSI-based Geolocation: A Comparative Study

Kevin Elgui, Pascal Bianchi, François Portier, Olivier Isson

Learning Methods for RSSI-based Geolocation: A Comparative Study

Kevin Elgui*†, Pascal Bianchi*, François Portier*, Olivier Isson†,
*Télécom, ParisTech, Paris, France
†Sigfox, Toulouse, France
Email: * firstname.lastname@telecom-paristech.fr
† firstname.lastname@sigfox.com

***Abstract*—In this paper, we investigate machine learning approaches addressing the problem of geolocation. First, we review some classical learning methods to build a radio map. In particular, these methods are splitted in two categories, which we refer to as likelihood-based methods and fingerprinting methods. Then, we provide a novel geolocation approach in each of these two categories. The first proposed technique relies on a semi-parametric Nadaraya-Watson estimator of the likelihood, followed by a maximum a posteriori (MAP) estimator of the object's position. The second technique consists in learning a proper metric on the dataset, constructed by means of a Gradient boosting regressor: a k-nearest neighbor algorithm is then used to estimate the position. Finally, all the proposed methods are compared on a data set originated from Sigfox network. The experiments show the interest of the proposed methods, both in terms of location estimation performance, and of ability to build radio maps.**

Keywords: LPWA Network, localization, maximum likelihood, metric learning

I. Introduction

Approaches based on the measurement of the received signal strength indicator (RSSI) to geolocate connected objects have witnessed tremendous success since Internet of Things (IoT) is on the rise. In the last few years, IoT has raised a great deal of attention in very diverse fields such as agriculture or health care. Experts agree (in [1]) that 30 billions objects will be part of the IoT by 2023 and 40% of these objects will need to be geolocated (e.g for freight transport). To guarantee reliable connectivity between a multitude of connected devices, researchers have been developing various Low Power Wide Area Network (LPWAN) standards [2]. The IoT requires LPWAN standards to support long-range communications. Moreover, ultra-low power consumption is a crucial aspect for the lifetime devices.

Several standard methods such as channel-fingerprinting provide satisfying results in the situation where the propagation channel exhibits enough frequency diversity as shown in [3]. Nevertheless, in many network of interest, every message transmitted occupies an Ultra Narrow Band (UNB). For instance, for the Sigfox network, the message occupies a band of 100 Hz within the Industrial, Scientific and Medical Band which corresponds to the frequency between 868 MHz and 868.2 MHz in Europe. As a consequence, the geolocation by means of channel-fingerprinting becomes irrelevant because of the absence of frequency diversity.

When the BS's of a network are time-synchronized, time based approaches as Time Difference Of Arrival (TDOA) provide accurate methods for geolocation [4] and [5]. However, when the BS's are *not* time-synchronized, the collection of the Received Signal Strength Indicator (RSSI) observed at all the BS is, by default, the main source of information allowing to geolocate the source.

In the present paper, we focus on a baseline probabilistic RSSI-only localization algorithm. The main challenge comes from the large range of fluctuations of the observed RSSI values, for a given source location. In such data, the observed signals can be very noisy, especially in urban environment (RSSI based methods are often assisted with accelerometers, gyroscopes or Bluetooth beacons to improve their accuracy [6]). It may also happen that, due to range limitation or network sensitivity, some messages are not detected by some BS's. Experience has shown that the performance is increased when the information of non reception is taken into account. Very few models in the literature chose to regard the information given by the non reception, though. Most of the time, the RSSI in the case of non reception, is replaced by the lowest RSSI amongst all observed RSSI's, as *e.g* in [7].

In this paper, we reviewed important off-the-shelf methods for RSSI based geolocation. Based on this review , we have observed that two methods arose amongst the most performant methods for the task of location estimation. First, ensemble methods (used *e.g* in [8]) as XGBoost Regressor, and k-NN regressor (*e.g* [6]).

Contributions.

- We provided a method exploiting the advantages of both ensemble methods and $k-$NN regressor. The idea, borrowed from [9], is to learn the metric used by the k-NN explicitly for the location estimation task. That is, build a metric to compare two RSSI's vectors, such that the k-NN regressor can chose the most appropriate neighbours for the location estimation task. The main idea here is to learn this metric d such that for a couple of RSSI's vector $(\mathbf{r}, \mathbf{r}')$: $\mathsf{d}(\mathbf{r}, \mathbf{r}')$ is a good predictor of the euclidean distance between the two emitters locations $\|z - z'\|_2$. We thus expect that through this metric, the k-NN regressor will be able to chose (within the training set) the k nearest

points of z and then compute their mean. We propose to learn d as a sum of T regression trees. Those trees are obtained through a XGBoost algorithm. The benefits w.r.t. a classic k-NN regressor are twofold:
  - it takes into account the information of reception/non reception of the signal at a BS;
  - it improves the model by optimizing the metric explicitly for the task of geolocation. This drives to better performances of the model (see Section V).
- We proposed a semi-parametric model relying on a relevant likelihood of the RSSI's given the object's position. The shape of the likelihood, is based on a model assumption, of Naive Bayes type: given the emiter position, the coordinates of the RSSI vector are independent. Throughout this paper, this assumption will be accepted. The main benefit of this assumption is to allow a low complexity of the model and to make it numerically tractable. The distribution of a RSSI at a given BS, given the location of the emitter will be model by a Gaussian distribution. The mean, and the standard deviation of this distribution are obtained by a non-parametric estimator of type Nadaraya-Watson. Finally, the location estimation will be obtained using a Maximum-A-Posteriori (MAP). This proposed estimator enables us to take into account the Boolean variable modeling the reception/non reception of the signal at BS's. The advantages of the provided method are manyfold:
  - it provides good results, even on small training data sets. Moreover, its performances are relatively stable when the number of training points decreases;
  - it offers a statistical framework through which density level sets and confident regions on the location estimate can be easily computed, when classical machine learning methods (as k-NN), are not able to do so.
- We provide detailed experiments results to compare these methods using real data originated from Sigfox network.

The rest of the paper is organized as follows. In Section II, we introduce the problem setting. Section III investigates several popular geolocation techniques of the literature. Section IV introduces the proposed predictors. Finally, Section V is devoted to the numerical experiments and discussions.

II. Problem setting

The network under consideration is dedicated to long-range and low-power consumption IoT communications. The range of transmission is up to 100 km, and the battery life-time is about 20 years. The network is composed of K fixed BS, say $(\mathrm{BS}_1, \ldots, \mathrm{BS}_K)$, whose respective coordinates $(z_1, \ldots, z_K)$ in the complex plane are known.

Consider a connected device whose position Z is a random variable in some given subset $\mathcal{Z}$, typically an open subset of $\mathbb{R}^2$. The device sends packets/messages which are collected by the neighboring BS. For a given message, each BS k ($k = 1, \ldots, K$) computes a RSSI R_k as the temporal mean of the

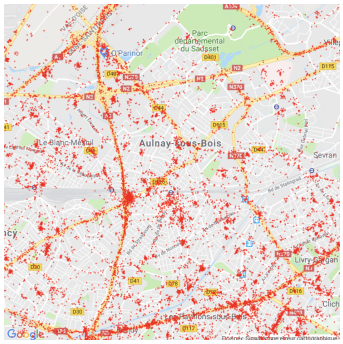

Fig. 1. Image of a sample of the locations emitters.

BS 1	BS 2	...	BS K	Lat	Long
-102	NaN	...	-83	49.15434	2.24928
NaN	-98	...	NaN	48.865584	2.44567

Fig. 2. Sample from the Sigfox dataset

received signal strength. The RSSI R_k is typically real-valued in a certain subset $\mathcal{R} \subset \mathbb{R}$. However, due to range limitation and network sensitivity, some messages may not be detected by some BS, in which case we just set $R_k = \texttt{NaN}$, where NaN stands for an unobserved value. We thus assume that for every $k = 1, \ldots, K$, R_k is a random variable in the set $\tilde{\mathcal{R}} := \mathcal{R} \bigcup \{\texttt{NaN}\}$.

The aim of this paper is to predict the unknown position Z from the observation of the RSSI-vector

$$\boldsymbol{R} := (R_1, \ldots, R_K)\,.$$

A predictor is a function $\hat{Z} : \tilde{\mathcal{R}}^K \to \mathcal{Z}$. We evaluate the performance of a predictor w.r.t. to the risk $\mathbb{E}(\ell(Z, \hat{Z}(\boldsymbol{R})))$ where $\mathbb{E}$ is the expectation, $\ell : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ is a loss. In typical settings, $\ell(z, \hat{z}) = \|z - \hat{z}\|^2$.

To achieve this task, we assume that the network operator has collected a dataset of fully supervised examples. The dataset is built by gathering observed RSSI's of devices equipped with GPS. As represented in Fig. II, every row of the dataset corresponds to a message. The features are the RSSI's at the receiving BS's and the label is the GPS coordinates of the transmitting device at the instant when the packet is sent. Formally, the dataset is represented by a collection of n random samples $\mathcal{X}_n := \{(Z^i, \boldsymbol{R}^i) : i = 1, \ldots, n\}$, assumed to be iid copies of $(Z, \boldsymbol{R})$.

III. Review of Geolocation Approaches

In this section, we discuss different off-the-shelf predictors which can be used to solve the geolocation task introduced above.

A. Likelihood-based methods

We refer to as Likelihood-based methods the methods which learn from the dataset a likelihood model $p(\boldsymbol{r}|z)$ for the conditional probability of the RSSI vector $\boldsymbol{R}$ given the position Z.

One first learns from the observed data $\mathcal{X}_n$ a mapping $p(\boldsymbol{r}|z)$ which represents the conditional probability density function (pdf) of $\boldsymbol{R}|Z$ namely, the likelihood. To this end, a way is to introduce a parametric likelihood model, such as the path-loss model discussed at the end of this paragraph, and to learn the parameters of this model from the dataset. Non-parametric methods can be used as well (see Section IV). One of the main advantages is that some prior hypotheses on the form of the likelihood $p(\boldsymbol{r}|z)$ can be easily introduced, based on physical considerations. One such hypothesis is the following:

Assumption 1. *The components $R_1, \dots, R_K$ of the random vector $\boldsymbol{R}$ are independent conditionally to Z.*

Assumption 1 is often used in the literature [10], [11], [12]. Discussing its validity is out of the scope of this paper, but we refer the interested reader to [13] where a independence kernel based test is proposed. Under this hypothesis, the likelihood admits the following decomposition:

$$p(\boldsymbol{r}|z) = \prod_{k=1}^{K} p_k(r_k|z)\,,$$

where $\boldsymbol{r} = (r_1, \dots, r_K)$ and where $p_1, \dots, p_K$ are conditional marginals to be learned.

Once the likelihood model $p(\boldsymbol{r}|z)$ has been obtained, the predictor $\hat{Z}(\boldsymbol{R})$ can be easily defined from standard statistical methods. Assume now that a new message arises from the unknown position Z with a RSSI vector $\boldsymbol{R}$. A legitimate (but often computationally intractable) choice is to define the predictor $\hat{Z}(\boldsymbol{R})$ as a minimizer w.r.t. $\hat{z}$ of the estimated risk:

$$\int \ell(z, \hat{z}) p(z|\boldsymbol{R}) dz \tag{1}$$

where, according to the Bayes formula, $p(z|\boldsymbol{R}) \propto p(\boldsymbol{R}|z)p_Z(z)$ and where $p_Z(z)$ is the prior distribution of the r.v. Z supposed to be known (typically uniform on $\mathcal{Z}$ as in [14], or inferred from the dataset $\mathcal{X}_n$). As the computation and the minimization of (1) can be difficult, an alternative is to consider the Maximum-a-Posteriori (MAP) estimator given by:

$$\begin{aligned}\hat{Z}_{MAP}(\boldsymbol{R}) &:= \arg\max_{z \in \mathcal{Z}} p(z|\boldsymbol{R}) \\ &= \arg\max_{z \in \mathcal{Z}} \sum_{k=1}^{K} \log p_k(R_k|z) + \log p_Z(z). \end{aligned} \tag{2}$$

To conclude this paragraph, we briefly discuss the broadly used *log-loss* (or *path-loss*) parametric model [6], [15]. The model is widely used to model the coupling between the received power at the receiver antenna and the distance between the received and emitter. The conditional distribution $p_k(r|z)$ of $R_k|Z$ is supposed to have the form $p_{\theta_k}(r|z)$ where $\theta_k = (P_{0,k}, \nu_k, \sigma_k^2)$ is a triplet of parameters $p_{\theta_k}(.|z)$ is a Gaussian distribution of variance σ_k^2 and mean $P_{0,k} - 10\nu_k \log_{10} d_v(z, z_k)/d_0$. Here, d_0 is some reference distance and d_v stands for the Vincenty distance, the parameters $P_{0,k}, \nu_k$ respectively represent the power in dBm at distance d_0 and ν_k is the so-called path-loss exponent. The parameter vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$ is estimated from the dataset $\mathcal{X}_n$ using a standard maximum likelihood approach.

B. Fingerprinting Methods

Fingerprinting methods directly map the vector $\boldsymbol{R}$ into a position Z, typically by means of a supervised learning algorithm. In the following, we present several popular learning algorithms to perform the task of geolocation.

1) k-Nearest Neighbors (k-NN): The method is used in [7] in the context of outdoor geolocation. We endow the space of RSSI's vectors with the Euclidean distance. For this purpose, [7] suggests to replace all the `NaN` values either by the lowest RSSI amongst all observed RSSI, or by an arbitrary value (the value -200 is used in [7]). For every K-dimensional RSSI vector $\boldsymbol{R}$, we let $(\boldsymbol{R}^{(1)}, Z^{(1)}), \dots, (\boldsymbol{R}^{(n)}, Z^{(n)})$ be a reordering of the dataset $\mathcal{X}_n$ such that $\|\boldsymbol{R} - \boldsymbol{R}^{(1)}\| \leq \dots \leq \|\boldsymbol{R} - \boldsymbol{R}^{(n)}\|$. The unknown position Z is finally estimated by $\hat{Z}(\boldsymbol{R}) := k^{-1} \sum_{i=1}^{k} Z^{(i)}$, where the integer k is an hyperparameter (see [16] for a discussion on the choice of k).

2) Ensemble Trees Methods: A Random Forest model has been applied as a classifier for a indoor-context geolocation in (see [8]). In this paper, this method gets better accuracy than a k-NN based method. The goal of such ensemble methods is to combine the predictions of several base estimators built with a given learning algorithm in order to improve the robustness and the ability to generalize over a single estimator. Two important families are bagging methods such as random forests [17], and boosting methods such as Gradient Tree boosting. The final estimate has the form $\hat{Z}(\boldsymbol{R}) = \sum_{t=1}^{T} f_t(\boldsymbol{R})$ where T is an integer and $f_1, \dots, f_T$ are regression trees learned on the dataset $\mathcal{X}_n$ by one of the above methods.

IV. Proposed Geolocation Methods

We propose two localization methods, one for each category.

A. Semi-Parametric Likelihood-Based Method

We propose the following semi-parametric likelihood model for $p(\boldsymbol{r}|z)$, the conditional density of $\boldsymbol{R}$ given Z. As often in geolocation [10], [11], [12], [6], we strongly rely on the conditional independence Assumption 1. Using the later hypothesis, it is sufficient to provide a model for the marginal conditional distributions $p_k(r_k|z)$ of R_k given Z, for every $k = 1, \dots, K$. Here, we recall that R_k is a random variable over the set $\mathbb{R} \cup \{\texttt{NaN}\}$. Densities are thus considered w.r.t. the reference measure $\lambda + \delta_{\texttt{NaN}}$ where λ is the Lebesgue measure and $\delta_{\texttt{NaN}}$ is the Dirac measure at the `NaN`-value. We define $\pi_k : \mathcal{Z} \to [0, 1]$ as

$$\pi_k(z) := \mathbb{P}(R_k = \texttt{NaN}|Z = z)$$

and we constrain the model by assuming that, given Z and given that $R_k \neq \texttt{NaN}$, R_k follows a Gaussian distribution

whose mean and variance are respectively denoted by $m_k(z)$ and $\sigma_k^2(z)$:

$$m_k(z) := \mathbb{E}(R_k | Z = z, R_k \neq \texttt{NaN})$$
$$\sigma_k^2(z) := \mathrm{Var}(R_k | Z = z, R_k \neq \texttt{NaN}) .$$

We denote by $r \mapsto \Phi(r; m, \sigma^2)$ the normal density of mean m and variance σ^2. We summarize our model is as follows:

1) $R_1, \ldots, R_K$ are independent given Z;
2) For every k,

$$\mathbb{P}(R_k \in dr | Z) = \pi_k(Z)\delta_{\texttt{NaN}}(dr) + (1 - \pi_k(Z))\Phi(r; m_k(Z), \sigma_k^2(Z))dr .$$

Based on this model, the likelihood $p(\boldsymbol{r}|z)$ is fully determined by the mappings π_k, m_k and σ_k^2 for all $k = 1, \ldots, K$. The remaining task is to estimates these quantities using our dataset $\mathcal{X}_n$. To this end, we propose to use a non-parametric approach, and to replace these mappings with their Nadaraya-Watson estimates [18]. Let $K : \mathcal{Z} \to \mathbb{R}_+$ be a kernel, i.e., nonnegative, symmetric function integrating to one, and let $h > 0$ be a scalar (the so-called *bandwidth*). Define $K_h(z) = h^{-1}K(h^{-1}z)$ for all $z \in \mathcal{Z}$. The Nadaraya-Watson estimates are respectively given for every k by

$$\hat{\pi}_k(z) := n^{-1} \sum_{i=1}^{n} \mathbb{1}_{\texttt{NaN}}(R_k^i) K_h\left(Z^i - z\right)$$
$$\hat{m}_k(z) := D_k(z)^{-1} \sum_{i=1}^{n} \mathbb{1}_{\mathbb{R}}(R_k^i) R_k^i K_h\left(Z^i - z\right)$$
$$\hat{\sigma}_k^2(z) := D_k(z)^{-1} \sum_{i=1}^{n} \mathbb{1}_{\mathbb{R}}(R_k^i)(R_k^i - m_k(z))^2 K_h\left(Z^i - z\right)$$

where $D_k(z) := \sum_{i=1}^{n} \mathbb{1}_{\mathbb{R}}(R_k^i) K_h\left(Z^i - z\right)$. Under standard technical conditions, $\hat{\pi}_k$, $\hat{m}_k$ and $\hat{\sigma}_k^2$ converge uniformly towards π_k, m_k and σ_k^2 as $n \to \infty$ and $nh \to \infty$ [18]. Finally, the MAP location estimator can be written as:

$$\hat{Z}(\boldsymbol{R}) = \arg\max_{z \in \mathcal{Z}} \sum_{k \in \mathcal{I}_{\boldsymbol{R}}} (1 - \hat{\pi}_k(z)) \log \Phi(R_k; \hat{m}_k(z), \hat{\sigma}_k^2(z)) + \sum_{k \in \mathcal{I}_{\boldsymbol{R}}^c} (1 - \hat{\pi}_k(z)) + \log \hat{p}_Z(z) ,$$

where $\mathcal{I}_{\boldsymbol{R}} := \{k = 1, \ldots, K : R_k \neq \texttt{NaN}\}$ stands for the set of the receiving BS's, and where $\hat{p}_Z(z)$ stands for an estimation of the prior on Z, which we suggest to estimate from the dataset $\mathcal{X}_n$ through the kernel density estimator:

$$\hat{p}_Z(z) = n^{-1} \sum_{i=1}^{n} K_h\left(Z^i - z\right) .$$

B. Metric-Learning Fingerprinting Method

In this paragraph, we tackle the problem of learning an adapted metric (see [19]) on $\mathcal{R}^K$ to improve basic k-NN using the standard Euclidean distance. We recall that NaN values are here replaced by a fixed real value as discussed in Section III-B. The idea is to build a mapping $\mathsf{d} : \mathcal{R}^K \times \mathcal{R}^K \to [0, +\infty)$ such that *close RSSI* (w.r.t. to the metric d) *correspond to close object positions* (w.r.t. to the Vincenty distance d_v on $\mathcal{Z}$). In that sense, a "good" metric (see Fig. 3) is a mapping d for which the empirical risk

$$ER_n(\mathsf{d}) := \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\mathsf{d}(R^i, R^j) - d_v(Z^i, Z^j)\right)^2$$

is small. The main trick, borrowed from [20], [9] is to search for a mapping d minimizing $\mathbb{E}R_n(\mathsf{d})$ within a relevant hypothesis class. More precisely, we search for d under the form

$$\mathsf{d}(\boldsymbol{r}, \boldsymbol{r}') := \sum_{t=1}^{T} f_t(\varphi(\boldsymbol{r}, \boldsymbol{r}')) ,$$

where $f_1, \ldots, f_T$ is a collection of T regression trees, and where $\varphi : \mathcal{R}^K \times \mathcal{R}^K \to \mathbb{R}^K \times \mathbb{R}^K$ is given by:

$$\varphi(\boldsymbol{r}, \boldsymbol{r}') := \begin{pmatrix} |\boldsymbol{r} - \boldsymbol{r}'| \\ \frac{1}{2}(\boldsymbol{r} + \boldsymbol{r}') \end{pmatrix} .$$

In practice, the minimization of $ER_n(\mathsf{d})$ w.r.t. $f_1, \ldots, f_T$ is untractable. An alternative is to use a Random Forest or an XGboost regressor, which separately optimizes the T regression trees. In practice, the learning stage is thuse as follows:

- Compute the pairwise features $\varphi(\boldsymbol{R}^i, \boldsymbol{R}^j)$ for all couples (i, j) in the dataset;
- Use a regression tree ensemble method to predict the labels $d_v(Z^i, Z^j)$ based on the features $\varphi(\boldsymbol{R}^i, \boldsymbol{R}^j)$.

Note that the obtained mapping d, though symmetric, is not mathematically speaking a metric. This point is however irrelevant regarding the application of interest. Given the obtained metric and given an observed RSSI vector $\boldsymbol{R}$, the k-NN estimate of Z is computed as in Section III-B.

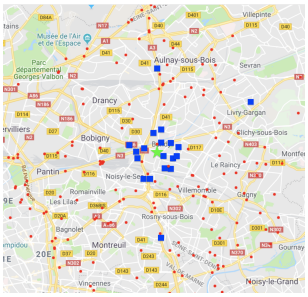

Fig. 3. Scatter plots of the k nearest neighbors. In red, a sample of the 200 neighbors according to the euclidean distance. In blue, the 25 neighbors according to the learned metric. In green stars, the true position of the emitter.

V. Numerical Experiments

A. Performance Analysis

To compare the performances of the different methods, we use the Sigfox dataset $\mathcal{X}_n$ composed of $n = 1.5 \cdot 10^6$ observations. The dataset was randomly split in a training subset (90%) and a test subset (10%). The training subset was used to perform cross-validation (each fold containing 10%

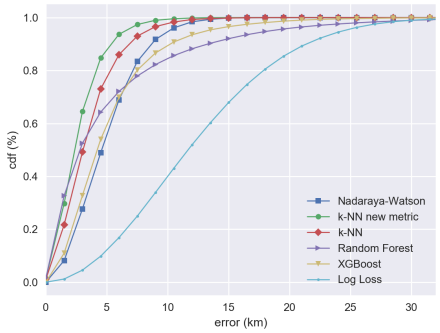

Fig. 4. Comparisons of the presented methods in terms of their performances.

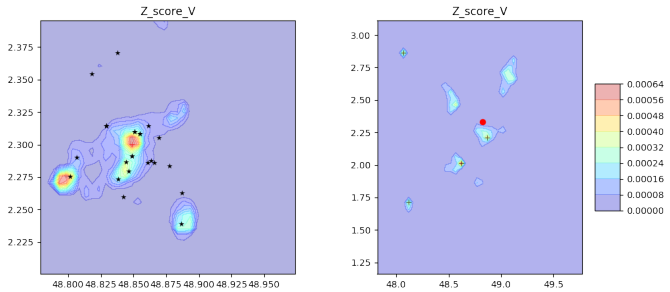

Fig. 5. Heat map of the position $Z|\boldsymbol{R}$ for two different observations of $\boldsymbol{R}$. The red dots show the true positions. The black dots are the observed positions in the test dataset corresponding for the same observations of $\boldsymbol{R}$.

of the training set) in order to find the optimal parameters of our algorithms. The test subset was employed to evaluate the accuracy of the methods in competition.

To compute the errors we employ the Vincenty distance between estimated and actual location. Fig. 4 shows the cumulative distribution function of the estimation error for all the presented methods. The k-NN with the learned metric turns out to outperform the other methods of the paper. By contrast, the Log Loss model is not relevent for this noisy urban dataset.

B. Heat Map estimation

A major benefit of the Semi-Parametric Likelihood-Based method is that density level sets can be computed easily. Thanks to the statistical framework, this method is able to evaluate the probability density of $Z|\boldsymbol{R}$ at all $z \in \mathcal{Z}$. This density level sets are regions in which Z is most likely to lie given the observation of $\boldsymbol{R}$. This is shown in Fig. 5 where further information is provided on the uncertainty of the estimation.

Conclusion

In this paper, we investigated machine learning approaches addressing the problem of geolocation. We presented most popular methods that can be found in the literature. Then, we proposed two new techniques: one based on a likelihood and the other on a learned metric for a k-NN. To compare these methods, 1,5M observations were collected from the Sigfox network. Results have shown that the metric learning method has obtained the highest accuracy on this dataset. As for the semi-parametric method, it goes beyond the simple estimation by providing heat maps and level sets, making it, a suitable methods for industrial applications.

References

[1] C.-L. Hsu and J. C.-C. Lin, "An empirical examination of consumer adoption of internet of things services: Network externalities and concern for information privacy perspectives," *Computers in Human Behavior*, vol. 62, pp. 516–527, 2016.

[2] U. Raza, P. Kulkarni, and M. Sooriyabandara, "Low power wide area networks: An overview," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 855–873, 2017.

[3] G. Sun, J. Chen, W. Guo, and K. R. Liu, "Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 12–23, 2005.

[4] B. Denis, J. Keignart, and N. Daniele, "Impact of nlos propagation upon ranging precision in uwb systems." Citeseer.

[5] R. J. Fontana, "Experimental results from an ultra wideband precision geolocation system," in *Ultra-Wideband, Short-Pulse Electromagnetics 5*. Springer, 2002, pp. 215–223.

[6] S. Yiu, M. Dashti, H. Claussen, and F. Perez-Cruz, "Wireless rssi fingerprinting localization," *Signal Processing*, vol. 131, pp. 235–244, 2017.

[7] N. Patwari, "Location estimation in sensor networks," Ph.D. dissertation, University of Michigan, 2005.

[8] E. Jedari, Z. Wu, R. Rashidzadeh, and M. Saif, "Wi-fi based indoor location positioning employing random forest classifier," in *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*. IEEE, 2015, pp. 1–5.

[9] A. Bellet, A. Habrard, and M. Sebban, "A survey on metric learning for feature vectors and structured data," *arXiv preprint arXiv:1306.6709*, 2013.

[10] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for wlan location fingerprinting," in *Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004. The First Annual International Conference on*. IEEE, 2004, pp. 14–23.

[11] S. Mazuelas, A. Bahillo, R. M. Lorenzo, P. Fernandez, F. A. Lago, E. Garcia, J. Blas, and E. J. Abril, "Robust indoor positioning provided by real-time rssi values in unmodified wlan networks," *IEEE Journal of selected topics in signal processing*, vol. 3, no. 5, pp. 821–831, 2009.

[12] X. Li, "Rss-based location estimation with unknown pathloss model," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, 2006.

[13] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf, "Measuring statistical dependence with hilbert-schmidt norms," in *International conference on algorithmic learning theory*. Springer, 2005, pp. 63–77.

[14] M. Ibrahim and M. Youssef, "Cellsense: A probabilistic rssi-based gsm positioning system," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*. IEEE, 2010, pp. 1–5.

[15] M. Bshara, U. Orguner, F. Gustafsson, and L. Van Biesen, "Fingerprinting localization in wireless networks based on received-signal-strength measurements: A case study on wimax networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 1, pp. 283–294, 2010.

[16] P. Hall, B. U. Park, and R. J. Samworth, "Choice of neighbor order in nearest-neighbor classification," *The Annals of Statistics*, pp. 2135–2152, 2008.

[17] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[18] A. Tsybakov, "Apprentissage statistique et estimation non-paramétrique," 2013, course.

[19] Y. Xie, Y. Wang, A. Nallanathan, and L. Wang, "An improved k-nearest-neighbor indoor localization method based on spearman distance." *IEEE Signal Process. Lett.*, vol. 23, no. 3, pp. 351–355, 2016.

[20] C. Xiong, D. Johnson, R. Xu, and J. J. Corso, "Random forests for metric learning with implicit pairwise position dependence," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 958–966.