GEOLOCATION AND NAVIGATION
IN SPACE AND TIME

A semi parametric model for RSSI-based localization
Modèle semi paramétrique pour la localisation par RSSI

***Pascal Bianchi*[2], *François Portier*[3], *and Kevin Elgui*[1]**

[2] *Télécom ParisTech, pascal.bianchi@telecom-paristech.fr*
[3] *Télécom ParisTech, francois.portier@telecom-paristech.fr*
[1] *Télécom ParisTech, kevin.elgui@telecom-paristech.fr, kevin.elgui@sigfox.com*

Keywords: statistics, Internet of Things (IoT), semi-parametric, geolocation, RSSI
Mots-clefs: statistiques, Internet des objets, semi paramétrique, géolocalisation, RSSI

Abstract:
Accurate and reliable geolocation of an object in the context of an Internet of Things (IoT) network must deal with a lack of available information. The approach followed in this paper is based on Received Signal Strength Indicators (RSSI) measured at all base stations (BS) as a useful information to infer the object's position. The proposed technique is based on a maximum a posteriori (MAP) estimator derived within a flexible semi-parametric model. The non-parametric components of the model are estimated with the help of Nadaraya-Watson type estimators. Our proposal has been tested and compared with an other method found in the literature: a simple model consisting of taking a weighted barycenter of the BS as the estimator of the real position. Our approach, although computationally less efficient, has revealed to outperform other methods and to come up with some very interesting information on the behaviour of the studied variables.

Résumé:
La problématique de la géolocalisation par le réseau se heurte rapidement au manque de pertinence des informations disponibles. L'approche proposée dans ce papier est basée sur l'estimation de la position d'un objet du réseau à partir des puissances de signal reçues (RSSI) aux antennes de base (BS). Pour cela, nous proposons un estimateur maximum à posteriori (MAP) dans le cadre d'un modèle semi-paramétrique. Les composantes non paramétriques de ce modèle ont été estimés à l'aide d'estimateurs de Nadaraya-Watson. Nous testerons et nous comparerons notre approches à une méthode trouvées dans l'état de l'art, consistant à estimer la position par une barycentre des BS récepetrices.

1 Introduction

In the last few years, the Internet of Things raised a great deal of attention in very diverse fields such as agriculture or health care. Experts agree that 30 billions objects will be part of the IoT by 2020 [1] and 40% of these objects might need to be geolocated. This paper is devoted to the geolocalization of connected objects in Sigfox wireless network, called LPWAN (Low Power Wide Area Network). The network has been specifically deployed in order to offer an international connectivity for objects in IoT in more than 40 countries. It provides a low energy, and economic solution for transmissions of messages. Beyond the traditional wireless access to the internet cloud, a targetted application is to use the network to geolocalize the objects. However, the singularity of the Sigfox network makes the task especially challenging.
Indeed, every message transmitted by an object in the Sigfox network occupies an Ultra Narrow Band (100 Hz - 600 Hz). This makes it difficult to address the geolocalization by means of *e.g.* channel fingerprinting, because of the absence of frequency diversity of the propagation channel. In addition, the base stations (BS) of the network are not time-synchronized, and geometric approaches such as Time Difference of Arrival ([2], [3]) are not relevant. The approach followed in this paper consists in the use of the set *Received Signal Strength Indicators* (RSSI) measured at all base station (BS) as a useful information to infer the object's position. RSSI based localization is known to be a difficult problem. On the top of that, a quick inspection of the available RSSI observations reveals that the data are very noisy: RSSI measurements are subject to a significant variablility. This is due to several factors (urban environment, indoor or outdoor transmission, etc.) The latter variability of the data typically rules traditional parametric approaches based on path-loss model, which aim at relating the RSSI with the source-destination distance. [4], [5].
Our contribution are as follows:

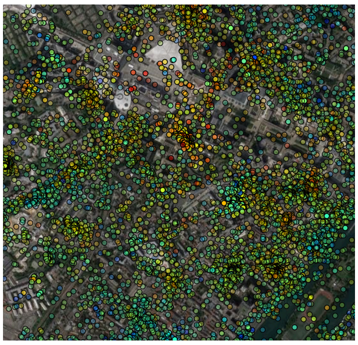

Figure 1 – RSSI received at a particular base station

- We propose to formulate the localization as a Maximum a Posteriori estimation problem of the position given the vector of RSSI values gathered at all BS of the network.
- We use a large amount of data in order to model the likelihood of the RSSI measurements given the position. To that end, we introduce a semiparametric model well suited to the problem at stake.
- We provide detailed numerical experiments showing that the proposed method outperform off-the-shelf solutions.

2 Semiparametric framework

We consider that an object is located at an unknown random position noted Z and belonging to a subset $\mathcal{Z} \subset \mathbb{R}^2$. The network is formed by K base stations numbered from 1 to K. The message leads to K values of RSSI, at each BS, denoted by $\mathbf{R} = (R_1, R_2, \ldots, R_K)$. We allow the values of the RSSI to take the value `NaN` which will encode that the message is actually not received by a BS. Our first assumption deals with the parametric part of the model : the distribution we consider for $\mathbf{R}$ given Z is Gaussian. This Gaussian assumption is illustrated in Fig. 2. We have tested this hypothesis for several z, all of them give similar histograms. Besides, this hypothesis is widely accepted in the literature (e.g. [6]).

Assumption 1 *The conditional distribution of* $\mathbf{R}$ *given* Z *is Gaussian with mean* $\mathbf{m}(Z)$ *and variance* $\mathbf{\Sigma}$*.*

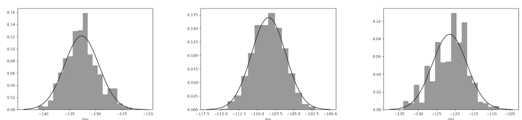

Figure 2 – Histograms of received RSSI for three emitting position.

Our second assumptions is supported by the intuition that the vector $(\mathbf{R}_1, \ldots, \mathbf{R}_K)$ should be conditionally independent given the position Z (See Fig. 3).

Assumption 2 *For all* $k \neq \ell$*,* $\mathbf{\Sigma}_{k,\ell}(Z) = 0$ *a.s.*

Hence, the likelihood function of the observations $\mathbf{R}$ given Z can be decomposed as $p(\mathbf{R}|Z) = \prod_k p(R_k|Z)$. Under assumption 1 we have for all k:

$$(R_k = r|Z = z) = \mathcal{N}(r\,;\, m_k(z), \sigma_k^2)\,,$$

where $\mathcal{N}(\,.\,; m, \sigma^2)$ stands for the Gaussian p.d.f. of mean m and variance σ^2. Of course, m_k needs to be evaluated. To that end, we use the knowledge of a dataset composed of indepent copies $(Z^{(i)}, \mathbf{R}^{(i)})_{i=1\ldots n}$ of

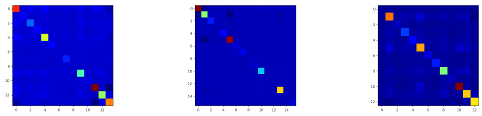

Figure 3 – HeatMap of the covariance matrix of $\mathbf{R}$ *for three different value of* z.

$(Z, \mathbf{R})$. To estimate the non-pararametric part of the model, namely the unknown function m, we employ Nadaraya-Watson type estimates as presented for instance in [7] and defined as:

$$\hat{m}_k(z) = \frac{\sum_{i=1}^{n} R_k^{(i)} K_h\left(z - Z^{(i)}\right)}{\sum_{i=1}^{n} K_h\left(z - Z^{(i)}\right)}.$$

Where $K_h : \mathbb{R}^2 \to \mathbb{R}_+$ is a kernel function such that $\int K_h = 1$, and $h > 0$ is called the bandwidth of the kernel. In the above equation and due to the lack of space, we overlooked the fact that RSSI may take `NaN` values, but this can be easily addressed in practice by using a slight modification of our model.

2.1 MAP Estimator of Z

The MAP estimator of Z when $\mathbf{R}$ has been observed is as follows:

$$\begin{aligned} Z^{MAP} &= \arg\max_{z \in \mathcal{Z}} \mathbb{P}\left(Z \in dz, \mathbf{R} = \underline{\mathbf{r}}\right) \\ &= \arg\max_{z \in \mathcal{Z}} \prod_k \mathbb{P}\left(R_k = \underline{r}_k \mid Z \in dz\right) \end{aligned}$$

Consequently, we define the MAP estimator $\hat{Z}^{MAP}$ replacing the unknown quantities in Z^{MAP} by their estimates, that is:

$$\hat{Z}^{MAP} = \arg\max_{z \in \mathcal{Z}} \prod_k \mathcal{N}\left(\underline{r}_k; \hat{m}_k(z), \sigma_k^2\right)$$

An exhaustive grid search is a possible way to solve this equation.

3 Experiments

3.1 Barycenter model

We propose a very simple model to estimate the location of a device when $\mathbf{R}$ has been observed. The location estimation is as follow:

$$\hat{Z}_{\text{bary}}(\tilde{\mathbf{r}}) = \sum_{k=1}^{K} \omega_k(\underline{r}_k) z_{BS_k}.$$

It is simply weighted barycenter of the receiving base stations, and where the weights are increasing functions of the signal power. Despite its simplicity, this model has proven to perform well in certain contexts (e.g.high density of base stations).

3.2 Comparisions

We are giving three metrics to compare the two models. First, the expectation of distance between estimates and real position.

$$\begin{aligned} \mathbb{E}_{test}(d(Z, \hat{Z}_{\text{bary}})) &= 3.5673km \\ \mathbb{E}_{test}(d(Z, \hat{Z}_{\text{MAP}})) &= 2.6141km \end{aligned}$$

One may want to bound the error, thus we give the following information:

$$\begin{aligned} \max_{test}(d(Z, \hat{Z}_{\text{bary}})) &= 8.0573km \\ \max_{test}(d(Z, \hat{Z}_{\text{MAP}})) &= 6.1792km \end{aligned}$$

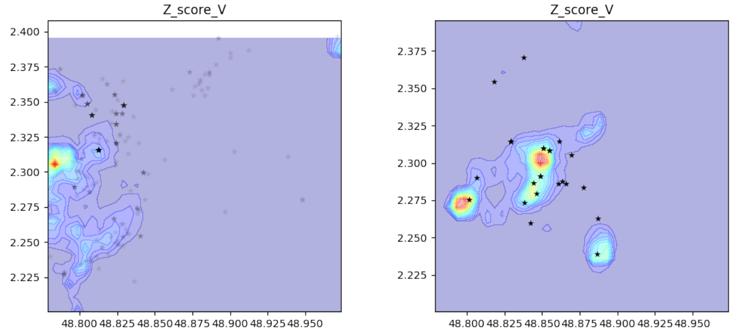

Figure 4 – Colormaps of the contour line of our learned density for two observed R

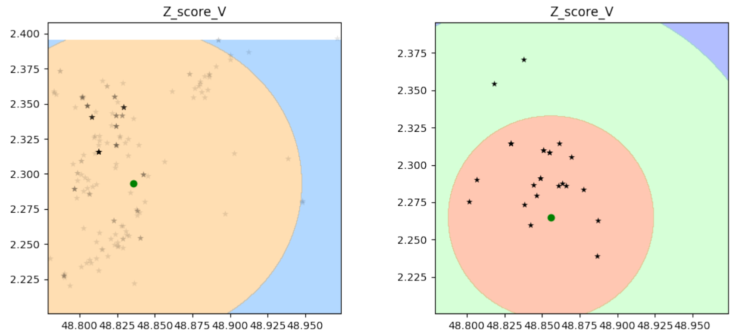

Figure 5 – Colormaps of the contour line of the barycenter approach density for two observed R.

4 Conclusion

In this paper, we propose a model for geolocatlisation, with the consideration, that subjected to a disturbed environment, a message can not be received by a base station. Model performance improvements are possible in many places. At first, the relaxation of certain independence hypotheses, and thus search for "patterns" of receiving antennas, characteristic of certain positions. Finally, the reception of a message seems to carry more information about Z than the variable R, this consideration can lead us to search for $\hat{Z}$ with a selection of candidates in several time.

5 References

[1] C.-L. . al., “An empirical examination of consumer adoption of internet of things services: Network externalities and concern for information privacy perspectives,” in *Computers in Human Behavior*, 2016.

[2] R. K. . al., “Accuracy analysis for tdoa localization in sensor networks,” 14th International Conference on Information Fusion Chicago, 2011.

[3] L. Mailaender, “On the geolocation bounds for round-trip time-of-arrival and all non-line-of-sight channels,” EURASIP Journal on Advances in Signal Processing, 2008.

[4] S. M. . al., “Robust indoor positioning provided by real-time rssi values in unmodified wlan networks,” IEEE Journal of Selected topics in signal processing, 2009.

[5] S. C. . D. K. Dhaka, “Path loss prediction models for wireless communication channels and its comparative analysis,” in *International Journal of Engineering, Management Sciences*, 2015.

[6] E. Elnahrawy, X. Li, and R. P. Martin, “The limits of localization using signal strength: a comparative study,” in *2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004.*, pp. 406–414, Oct 2004.

[7] A. Tsybakov, “Apprentissage statistique et estimation non-paramétrique.” Course, 2013.