Decentralized detection in sensor networks - sensor localization with kernel methods

Pan XU, Xin WANG
(Key Laboratory of the Ministry of Education for Advanced Control in Light Industry Process, Jiangnan University, Wuxi, 214122, China)

Abstract: This paper demonstrates the use of kernel methods in a challenging problem of localization in sensor networks. We show that the coarse-grained localization problems for ad hoc sensor networks can be posed and solved as a pattern recognition problem using kernel methods from statistical learning theory. We provide an evaluation of our algorithm on simulated sensor networks.

Key words: decentralized detection, sensor networks, localization, kernel methods

1 Introduction

A sensor network can be viewed as a distributed pattern recognition device. In the pattern recognition approach, rather than transforming sensor locations and sensor readings into Euclidean, world-centric coordinates, we work directly with the (non-Euclidean) co-ordinate system given by the physical sensor readings themselves. Using the methodology of “kernel functions,” the topology implicit in sets of sensor readings can be exploited in the construction of signal-based function spaces that are useful for the prediction of various extrinsic quantities of interest, using any of a variety of statistical algorithms for regression and classification. In the current chapter we illustrate this approach in the setting of a localization problem [1-3].

The localization problem that we study is that of determining the location of a large number of sensors of unknown location, based on the known location of a small number of base sensors. Let $X_1, \ldots, X_m$ denote a set of $m$ sensors, and let $x_i$ denote the position in $\mathbb{R}^2$ of sensor $X_i$. Suppose that the locations of the first $n$ sensors are known, i.e., $X_1 = x_1, \ldots, X_n = x_n$, where $n \ll m$. We want to recover the positions of $X_{n+1}, \ldots, X_m$ solely on the basis of the receive/transmit signals $s(x_i, x_j)$ between pairs of sensors.

An important characteristic of radio or light signal strength is the relationship of the signal attenuation as a function of distance [4]. For instance, for radio signals in an idealized environment, given that the sending and receiving antennas are focused on the same radio frequency, we have:

$$s \propto P d^{-\eta},$$

(1)

Where $\eta > 2$ is a constant, and $P$ is the sending signal voltage. Such relationships provide the basis for a variety of localization algorithms in the literature, which consist of two main steps: (1) a ranging procedure which involves estimating the distance from a sensor to another sensor based on the signal strength of the signals transmitted/received between the two, and (2) a procedure that recovers the locations of the sensors based on their pairwise distance estimates either by triangulation or by least-squares methods [5-7]. Unfortunately, however, the idealized model in Eq. (1) can be highly inaccurate due to variability caused by multipath effects and ambient noise interference as well as device-specific factors such as the frequencies of node radios, physical antenna orientation, and fluctuations in the power source [8-9]. Methods based on ranging inherit these inaccuracies and improvements are possible only if difficult problems in signal modeling are addressed.

The paper is organized as follows. We begin with a brief background of classification using kernel methods, and motivate our application of kernel methods to the localization problem based on sensor signal strength. Next, the localization algorithm and its error analysis are described. We then present details of the implementation of the algorithm and its computational cost, followed by an evaluation of our algorithm with simulated sensor networks. Finally, we present our discussions in the final section.
2 Classification using kernel methods

In a classification algorithm, we are given as training data n samples \((x_i, y_i)\)\(\text{for } i=1, \ldots, n\) in \(X \times \{+1, -1\}\), where \(X\) denotes the input space. Each \(y_i\) specifies whether the data point \(x_i \in X\) lies in a class \(C \subseteq X\) (\(y_i = +1\)) or not (\(y_i = -1\)). A classification algorithm involves finding a discriminant function \(y = \text{sign}(f(x))\) that minimizes the classification error \(P(Y \neq \text{sign}(f(X)))\).

Central to a kernel-based classification algorithm (e.g., the SVM) is the notion of a kernel function \(K(x, x')\) that provides a measure of similarity between two data points \(x\) and \(x'\) in \(X\). Technically, \(K\) is required to be a symmetric positive semidefinite function.

For such a function, Mercer’s theorem implies that there must exist a feature space \(\mathcal{H}\) in which \(K\) acts as an inner product, i.e., \(K(x, x') = \langle \Phi(x), \Phi(x') \rangle\) for some mapping \(\Phi(x)\). The SVM and related kernel-based algorithms choose a linear function \(f(x) = \langle w, \Phi(x) \rangle\) in this feature space. That is, they find a vector \(w\) which minimizes the loss
\[
\sum_{i=1}^{n} \theta(y_i f(x_i)) 
\]
subject to \(\|w\| \leq B\) for some constant \(B\). Here \(\theta\) denotes a convex function that is an upper bound on the 0-1 loss \(P(Y \neq \text{sign}(f(X)))\). In particular, the SVM algorithm is based on the hinge loss \(\phi(y f(x)) = (1 - y f(x))^+\), By the Representer Theorem \([10]\), it turns out that the minimizing \(f\) can be expressed directly in terms of the kernel function \(K\):
\[
f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x) 
\]
for an optimizing choice of coefficients \(\alpha_i\).

There are a large number of kernel functions that satisfy the positive semidefinite proper-erty required by the SVM algorithm. Examples include the Gaussian kernel:
\[
K(x, x') = \exp(-\|x - x'\|^2 / \sigma) 
\]
as well as the polynomial kernel:
\[
K(x, x') = (\gamma + \|x - x'\|)^{-\sigma} 
\]
for parameters \(\sigma\) and \(\gamma\). Both of these kernel functions decay with respect to the distance \(\|x - x'\|\), a property that is shared by most idealized signal strength models. In particular, the radio signal model (1) has a form similar to that of a polynomial kernel. In [11], the authors justify the use of an acoustic energy model for localization that has the form of the Gaussian kernel above. These relationships suggest a basic connection between kernel methods and sensor networks. In particular, a naive usage of kernel methods could be envisaged in which signal strength is used directly to define a kernel function. In general, however, signal strength in real sensor networks need not define a positive semidefinite function. Nonetheless, it is the premise of this chapter that signal strength matrices provide a useful starting point for defining kernel-based discriminant functions. We show how to define derived kernels which are stacked on top of signal strength measurements in the following section.

3 Localization in ad hoc sensor networks

We assume that a large number of sensors are deployed in a geographical area. The input to our algorithm is a set of \(m\) sensors, denoted by \(X_1, \ldots, X_m\). For each \(i\) we denote by \(x_i\) the position in \(\mathbb{R}^2\) of sensor \(X_i\). Suppose that the first \(n\) sensor locations are known, i.e., \(X_1, \ldots, X_n = x_n\), where \(n \ll m\). For every pair of sensors \(X_i\) and \(X_j\), we are given the signal \(s(x_i, x_j)\) that sensor \(X_i\) receives from \(X_j\). We want to recover the positions of \(X_{n+1}, \ldots, X_m\).

We first aim to obtain a coarse location estimate for \(X_{n+1}, \ldots, X_m\). Given an arbitrarily constructed region \(C \subseteq \mathbb{R}^2\), we ask whether \(X_i \in C\) or not, for \(i = n+1, \ldots, m\). This can be readily formulated as a classification problem. Indeed, since the location of the base sensors \(X_1, \ldots, X_n\) are known, we know whether or not each of these base sensors are in \(C\). Hence we have as our training data \(n\) pairs \((x_i, y_i) = \text{sign}(x_i \in C)\)\(\text{for } i = 1, \ldots, n\). For any sensor \(X_p\), \(j = n+1, \ldots, m\), we can predict whether \(X_j \in C\) or not based on the sign of the discriminant function \(f(x_j)\):
\[
f(x_j) = \sum_{i=1}^{n} \alpha_i K(x_i, x_j). 
\]
We emphasize that the value of \(f(x_j)\) is known because the values of the kernels, \(K(x_i, x_j)\), are known, despite the fact that we do not know the position \(x_j\) per se.

Next, we turn to the definition of the kernel matrix \(K = (K(x_i, x_j))_{1\leq i,j \leq m}\). In general we envision
a hierarchy of kernels based on the signal matrix. An example of such a hierarchy is as follows:

1. We might simply define \( K(x_i, x_j) = s(x_i, x_j) \). We call this naive choice a first-tier kernel. If the signal matrix \( S = \left( s(x_i, x_j) \right)_{i,j=1}^{n,m} \) is a symmetric positive semidefinite Gram matrix then this approach is mathematically correct although it may not yield optimal performance. If \( S \) is not symmetric positive semidefinite, then a possible approximation is \( (S + ST)/2 + \delta I \). This matrix is symmetric, and is positive semidefinite for sufficiently large \( \delta > 0 \) (in particular, for \( \delta \) larger in absolute value than the most negative eigenvalue of \( (S + ST)/2 \)).

2. Alternatively, define \( K = S^2 S \) to be referred to as a second-tier linear kernel. \( K \) is always symmetric positive semidefinite. This kernel can be interpreted as an inner product for a feature space \( \mathcal{H} \) which is spanned by vectors of the form:

\[
\Phi(x) = (s(x, x_1), (x, x_2), \ldots, (x, x_m))
\]

Specifically, we define:

\[
K(x_i, x_j) = \sum_{l=1}^{\rho} s(x_i, x_l) s(x_j, x_l)
\]

Intuitively, the idea is that sensors that are associated with similar vectors of sensor readings are likely to be nearby in space.

3. Finally, it is also possible to evaluate any kernel function (e.g., Gaussian) on the feature space \( \mathcal{H} \) induced by the second-tier kernel. This yields a symmetric positive semidefinite matrix, to be referred to as a third-tier kernel. Specifically, a third-tier Gaussian kernel has the following form, for a parameter \( \sigma \):

\[
K(x_i, x_j) = \exp\left\{-\frac{\|\Phi(x_i) - \Phi(x_j)\|^2}{\sigma}\right\}
\]

\[
= \exp\left\{-\frac{\sum_{l=1}^{\rho} s(x_i, x_l) - s(x_j, x_l))^2}{\sigma}\right\}.
\]

Given training data \( (x_i, y_i)_{i=1}^{n} \) a kernel function \( K \), we apply the SVM algorithm to learn a discriminant function \( f(x) \) as in Eq. (3).

Our classification formulation has several noteworthy characteristics. First, the training points correspond to the base sensors, and thus may be limited in number, making the learning problem nominally a difficult one. However, because we are free to choose the target region \( C \), the problem can in fact be made easy. This ability to design the geometry of the boundary to fit the geometry of the classifier distinguishes this problem from a traditional pattern recognition problem.

The second characteristic is that we require that the network be relatively dense. As seen in Eq. (6), the prediction of position is based on a sum over sensors, and an accurate prediction can be achieved in general only if there are enough non-zero terms in the sum for it to be statistically stable.

A related point is that it is not necessary that the network be completely connected. If the sensor reading \( s(x_i, x_j) \) is generally small or zero for a pair of sensors, then that term does not perturb the kernel calculation or the discriminant calculation. If readings fluctuate between small values and large non-zero values, then the prediction will generally be degraded. Given that the approach is a statistical approach, however, with predictions based on an aggregation over neighboring sensors, it should be expected to exhibit a certain degree of robustness to fluctuations.

4. Experimental results

We consider a network of size 10×10 square units. The base sensors are distributed uniformly in a grid-like structure. There are a total of \( n \) such sensors. We are concerned with recognizing whether a sensor position \( x \), characterized by the signal reading \( s(x_i, x) \) for \( i = 1, \ldots, n \), lies in a region \( C \) or not.

We first define a signal model: Each sensor location \( x \) is assumed to receive from a sensor located at \( x' \) a signal value following a fading channel model:

\[
s(x, x') = \exp\left(-\frac{\|x-x'|^2}{\sigma}\right) + N(0, \tau)
\]

where \( N(0, \tau) \) denotes an independently generated normal random variable with standard deviation \( \tau \). This signal model is a randomized version of a Gaussian kernel. We have also experimented with a signal strength model that is a randomized version of the polynomial kernel:

\[
s(x, x') = (\|x-x'|\) + N(0, \tau)
\]

The results for the polynomial kernels are similar to the Gaussian kernels,
and are not presented here. It is emphasized that although the use of these models have been motivated elsewhere as signal models[9], in our case they are used merely to generate the signal matrix S.

Next, we define a region C to be recognized. In particular, C consists of all locations x that satisfy the following equations: \((x-v)^T H (x-v) \leq R\), where \(v=[5 5]^T\), \(H_1=[1 0; 0 1]\), and \(R=2\). The radius R is used to describe the size of C. For each simulation set-up with different box constraint parameter of the SVM, we learn a discriminant function \(f\) for the region C using the training data given by the base sensor positions. Once the discriminant function \(f\) is learned, we can draw the boundary of the region C.

In the experiments, we found that the box constraint parameter is related to the smoothness of the obtained boundary, as shown in the Fig. 1.

Fig. 1. The plot of the boundary obtained by our method, with the box constraint parameter 5000, 5 and 0.005.

We can see that the less the box constraint parameter, the smoother the obtained boundary.

5 Conclusions

In this paper, we introduce the use of kernel methods in decentralized detection and estimation by considering a challenging problem of localization in ad hoc sensor network. We have presented a nonparametric learning algorithm for coarse-grained localization for ad hoc wireless sensor networks. Our approach treats the signal strength as measured by sensor motes as a natural coordinate system in which to deploy statistical classification and regression methods. For the localization problem, this approach avoids the ranging computation, a computation which requires accurate signal models that are difficult to calibrate.

References


