Engineering Applications of Artificial Intelligence 50 (2016) 125–134

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

A machine learning approach for user localization exploiting connectivity data

Pietro Cottone, Salvatore Gaglio, Giuseppe Lo Re, Marco Ortolani *

DICGIM, University of Palermo, Viale delle Scienze, ed. 6 – 90128 Palermo, Italy

ARTICLE INFO

Article history:
Received 26 June 2015
Received in revised form
5 November 2015
Accepted 30 December 2015

Keywords:
Wireless sensor networks
Range-free localization
Support vector machines

ABSTRACT

The growing popularity of Location-Based Services (LBSs) has boosted research on cheaper and more pervasive localization systems, typically relying on such monitoring equipment as Wireless Sensor Networks (WSNs), which allow to re-use the same instrumentation both for monitoring and for localization without requiring lengthy off-line training.

This work addresses the localization problem, exploiting knowledge acquired in sample environments, and extensible to areas not considered in advance. Localization is turned into a learning problem, solved by a statistical algorithm. Additionally, parameter tuning is fully automated thanks to its formulation as an optimization problem based only on connectivity information.

Performance of our approach has been thoroughly assessed based on data collected in simulation as well as in actual deployment.

1. Introduction

Currently available technology for environmental monitoring makes it practically feasible to design and deploy a controlled network of sensors and actuators throughout the surroundings of an intelligent system. Besides the obvious intended purpose, the interconnected sensing devices indirectly enable advanced functionalities: for instance, they are suitable for making the system aware of the presence of users so that it can capture their preferences and eventually act on the environment to adapt it to their needs (Lo Re et al., 2013). Such vision falls within the scope of Ambient Intelligence (AmI), whose aim is to support the user in carrying out everyday activities in an unobtrusive way, and in this context LBSs play a central role as location is a very valuable piece of information in order to deliver the correct service to the user. A very common choice for the sensing equipment consists in WSNs, whose characteristics of pervasiveness and inconspicuity meet two of the most specific requirements of the AmI philosophy (Marsh et al., 2004).

In this work we propose an approach to the design of a WSN-based localization module, so as to allow the user to assess the accuracy obtainable when taking into account the practical limitations arising from the actual deployment. The aim is clearly to minimize the amount of nodes scattered through the environment, without excessively losing precision. Our proposal is characterized by a fully automated parameter selection, so that it may be easily adapted to different scenarios without user intervention. The corresponding module is fed with online data elaborated during the localization process. As regards localization, specifically, our proposal was inspired by a state-of-art algorithm, which was deeply modified in order to operate on different kinds of connectivity data, and to take advantage of our automated parameter selection process. Moreover, new kernels were used for the learning algorithm, in order to make it more resilient to a change in the external conditions, and, finally, we introduced a new mechanism to make the system aware of changes in settings and adapt to them.

The underlying idea is that information about connectivity, such as Received Signal Strength Indicator (RSSI) or Link Quality Indicator (LQI), can be used to get an approximated location of a sensor node whose position varies slowly within the borders of the covered area, as widely reported in the literature (Langendoen and Reijers, 2003; Mao et al., 2007; Wang et al., 2010). A common assumption is that the node is carried by a user moving through the monitored area. For the sake of generality, however, we adopt a novel approach with respect to the majority of works in this field, and we avoid making use of any a priori information, such as network topology or the layout of the monitored area; rather, we just require knowledge about the actual position of a limited subset of nodes in the network.

It is well-known that link quality estimators are characterized by very poor reliability (Song et al., 2012), so that the performance of any system based on them must be carefully assessed; a typical approach consists in arranging ad-hoc testbeds, which is often impractical as it depends on expensive, and hard to maintain

* Corresponding author.
E-mail addresses: pietro.cottone@unipa.it (P. Cottone), salvatore.gaglio@unipa.it (S. Gaglio), giuseppe.lore@unipa.it (G. Lo Re), marco.ortolani@unipa.it (M. Ortolani).

http://dx.doi.org/10.1016/j.engappai.2015.12.015
0952-1976/

Nomenclature

Acronyms

AoA	Angle of Arrival
AmI	Ambient Intelligence
BN	Beacon Node
BS	Base Station
CI	Connectivity Information
HS	Harmony Search
LBS	Location-Based Service
LQI	Link Quality Indicator
k-NN	k-Nearest Neighbor
ML	Maximum Likelihood
Pr	Received Power
RF	Radio Frequency
RSSI	Received Signal Strength Indicator
SVDD	Support Vector Data Description
SVM	Support Vector Machine
TDoA	Time Difference of Arrival
WSN	Wireless Sensor Network

deployment of nodes in complex scenarios (Aitsaadi et al., 2011). Moreover, this does not allow to test for application scalability, nor to evaluate the system behavior across different configurations, so an alternative strategy involves simulating the environment, the sensing equipment, and the interaction with the system in order to get an estimate of the overall performance (Chowdhury and Boutaba, 2010).

The concern regards in this case is the reliability of the simulation, so we chose to rely for our analysis on models generated by a *hybrid* simulator for WSNs, where virtual nodes coexist with real ones (Lalomia et al., 2009). The use of simulated nodes allows to limit the deployment to just a minimal set of real nodes, which may serve as realistic data model generators to steer the behavior of their virtual counterpart. This is useful to allow the simulation of environmental scenarios wider than those actually at researchers' disposal, as long as "tunable" models may be created for data, as discussed in De Paola et al. (2011). Models can be instantiated according to real past sensory readings, and tuned with respect to on-line incoming ones. We tested our approach on simulated as well as on empirical data, demonstrating that localization can be abstracted from the specific information source, so that the parameters of the localization system can be directly inferred from the statistical model describing the monitored area. Thus, the problem of localization can be easily adapted to the available information, independently of a specific signal model, and can take advantage of data from multiple information sources (De Paola et al., 2011).

This aspect sets our approach apart from a great number of localization systems that are designed and tuned for some specific signal process only. Those systems often outperform more general ones, but they are typically overadapted to the specific environment and WSN. Even a little change in the monitoring system or the environment can significantly degrade the localization accuracy. On the other hand, accuracy of our system is not heavily affected by the particular setting or scenario, as our experiments demonstrate.

The remainder of the paper is organized as follows. Section 2 provides some background on documented methods for localization, and Section 3 describes our proposal focusing on obtaining a generalizable approach. Section 4 presents our experimental results on simulated and actual data, and finally Section 5 states our conclusions.

2. Related work

Localization has been a widely studied topic during the last decades.

Range-based schemes can be considered the earliest approaches to localization in WSN, according to the taxonomy developed in Hightower and Borriello (2001). They compute location by estimating the absolute pairwise distance between two nodes, using RSSI, Time Difference of Arrival (TDoA), or Angle of Arrival (AoA). For example, the system proposed in Bahl and Padmanabhan (2000), uses RSSI registered at multiple base stations combined with signal propagation models in order to estimate node location. The main drawback of these approaches is the difficulty in obtaining a reliable model of the signal transmission for the ranging process. As a consequence, an accurate estimate requires very expensive hardware. Often, the localization accuracy needed by most WSN applications does not justify the high costs due to range-based methods. Thus, research has shifted toward alternatives providing a coarser yet sufficient accuracy, while requiring fewer hardware and software resources.

As a result, *range-free* algorithms have been conceived. They do not assume that a direct distance or angle estimation is available, thus avoiding the creation of a model of the signal propagation. However, they use signal strength information to narrow the area the sensor might be in (the so-called *area-based* approach).

The range-free methods can be parted into two major classes: optimization approaches and machine-learning based ones. Systems belonging to the former class transform localization into an optimization problem, whose constraints arise from physical characteristics (Radio Frequency (RF) transmission system, topology, etc). On the contrary, machine-learning based approaches correlate positions to the input, without using any a-priori model. The first class of algorithms can provide more accurate results, but is less adaptable to different scenarios, unless user intervention is done to accurately adapt the model to the new environment. Machine-learning localization schemes are more flexible and, if carefully designed, can adapt to new scenarios without any user intervention, implicitly capturing the new main physical features.

One of the first examples of range-free optimization systems is presented in He et al. (2003), whose authors propose *APIT*, a localization scheme where few nodes (beacons) are aware of their own position (via GPS or other mechanism) and are equipped with high-powered transmitters. The environment is parted into overlapping triangular regions, whose vertices are the beacons. Since the attenuation of the signal strength is monotonically decreasing in a certain direction, in the absence of obstacles, then each node can determine whether it resides within a triangle by using the signal strength received from beacon nodes, and comparing it to the one coming from its neighbors. Location is estimated as the center of gravity of the most probable areas the sensor belongs to (i.e., intersections of triangles the node is assigned to).

In Doherty et al. (2001), the authors solve the localization problem modeling the underlying network as a set of geometric constraints on the node position. The global solution, yielded by convex optimization, provides reliable estimate of the unknown positions. Constraints derive from connections, RF transmitter model and receiver model.

A similar approach was used in Li and Qin (2013), where a range-free localization problem is solved by finding a feasible solution to a class of non-linear inequalities defined on a graph