Computer Networks 000 (2016) 1–9

Contents lists available at ScienceDirect

Computer Networks

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

Sample Size Determination Algorithm for fingerprint-based indoor localization systems

Loizos Kanaris[a,*], Akis Kokkinis[a], Giancarlo Fortino[b], Antonio Liotta[a], Stavros Stavrou[c]

[a] Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[b] Department of Electronics, Informatics and Systems, University of Calabria, Italy
[c] Faculty of Pure and Applied Sciences, Open University of Cyprus, Nicosia, Cyprus

A R T I C L E I N F O

Article history:
Received 2 August 2015
Revised 12 November 2015
Accepted 14 December 2015
Available online xxx

Keywords:
Internet of Things
Sample size
Evaluation
Indoor positioning systems
Localization error
Positioning accuracy

A B S T R A C T

Provision of accurate location information is an important task in the Internet of Things (IoT) applications and scenarios. This need has boosted the research and development of fingerprint based, indoor localization systems, since GPS information is not available in indoor environments. Performance evaluation of such systems and their related localization algorithms, is usually based on sampling collection in predetermined test environments. The sample size determination and sampling methodology can significantly affect the reliability of the outcome. This work proposes an algorithm that calculates the minimum sample size of positioning data required for objective performance evaluation of fingerprint based localization systems. The use of a correct, independent, unbiased and representative sample size can speed up the training, evaluation and calibration procedures of a fingerprint based localization system, while ensuring that the system's true accuracy is achieved. The proposed Sample Size Determination Algorithm (SSDA) takes into consideration the desired confidence level, the resulting standard deviation of a small size preliminary sample as well as the error approximation with respect to the actual error of the system and proposes the final sample size for the evaluation and/or calibration and/or training of the utilized radio-maps. Additionally, the SSDA, assumes random sample allocation in the area of interest in order to avoid biased results. Risks arising from the selection of a sample of convenience are also investigated. Finally, the performance of the proposed algorithm is tested in both measured and simulated radio-maps.

1. Introduction

In the Internet of Things (IoT), several applications and scenarios envision the integration of a great variety of wireless technologies that will provide services based on the user behavior [15]. Such services often require the localization and tracking of the user in indoor environments of smart cites (such as malls, hospitals, underground stations) and smart houses [1,9,14,28]. Fingerprint-based positioning is one of the mostpopular indoor localization techniques implemented by Real Time Localization Systems (RTLS). This technique typically utilizes the Received Signal Strength (RSS) to perform positioning. Other radio parameters can be also used or combined, such as Power Delay Profile (PDP), Angle of Arrival (AOA) etc. RTLS may also utilize non-radio parameters, such as inertial measurements or prior knowledge of environment constraints, in an aim to improve accuracy [11]. Fingerprint-based positioning requires the generation of a dataset of measurements,

* Corresponding author.
E-mail addresses: l.kanaris@tue.nl (L. Kanaris), a.kokkinis@tue.nl (A. Kokkinis), g.fortino@unical.it (G. Fortino), a.liotta@tue.nl (A. Liotta), stavros.stavrou@ouc.ac.cy (S. Stavrou).

http://dx.doi.org/10.1016/j.comnet.2015.12.015
1389-1286/

usually the RSS, during an off-line phase. This dataset, called radio-map, requires calibration before being utilized for the estimation of the user location during the on-line phase. Calibration is important in order to train and configure the positioning algorithms to perform better for the specific radio-map [17]. The calibration techniques use a sample of measurements, which is taken in the area of the indoor environment. The sample size, as well as the allocation of the samples in the area of interest, influence the overall accuracy of the localization system. A similar sample of measurements is also used for the performance evaluation of the system. The calibration and evaluation procedures are two important steps that are influenced by the quality of the sample measurements. Selecting a small sample size or a biased sample can result in misleading calibration parameters and a degraded accuracy during location estimation. On the other hand, large sample sizes are more time consuming and more expensive to carry out. To the authors best knowledge, no previous work exists for selecting the aforementioned sample sizes in fingerprint-based indoor localization systems. The main goal of this paper is to develop and suggest an algorithm that will define the minimum sample size which will ensure correct calibration or training of the RTLS, and objective evaluation of the system's performance.

The rest of the paper is organized as follows: Section 2 presents related work on fingerprint-based methodologies and performance evaluation techniques. Section 3 introduces and analyses the proposed Sample Size Determination Algorithm (SSDA). Sections 4 and 5 describe the evaluation of SSDA based on an experimental indoor localization platform. Finally, Section 6 summarizes the conclusions.

2. Related work

Fingerprint-based positioning requires the construction of a fingerprint database during an off-line phase, in which a number of radio parameters are stored in the form of vectors. The aforesaid database is generated either by performing a measurement campaign or through simulation procedures. In the latter case, statistical, semi-deterministic or fully deterministic models are utilized [3,4,6,10,17,18,26]. A calibration/training procedure is then implemented, using a measurement sample, in order to identify the optimum configuration parameters for the positioning algorithms and minimize localization errors.

The location estimation is performed by the user during the on-line phase, by implementing various deterministic or probabilistic localization algorithms. Examples include the K-Nearest Neighbor (KNN) and the Weighted K-Nearest Neighbor (WKNN) as analyzed in [13], the Minimum Mean Square Error (MMSE) presented in [19] and the Maximum A Posteriori (MAP) [27].

The evaluation of the Fingerprint-based RTLS is also performed by retrieving a measurement sample. The importance of sampling is highlighted in [23], where the authors proposed the development of a benchmark standard. They specifically state that samples constitute the core of any benchmark for location systems, since they are used to compute the position estimates. They proposed that the benchmark specification should state the number of samples recorded per second and the duration of the measurements per location.

When trying to review common practices, literature suggests that researchers tend to utilize a diverse number of sample sizes for the purpose of evaluating their research work, without necessarily clarifying the rationale behind the sample selection. In [19], 40 observations were recorded for a set of 155 calibration points, that were used as training data to eliminate the randomness of human behavior. In [27], authors measured one sample per second, for a period of five minutes (300 samples total), while trying to investigate wireless channel changes over time.

Authors in [5] proposed a dynamic hybrid projection (DHP) technique for improved 802.11 localization. During their experiments they collected 802.11 RSS data at 27 different reference locations in the area of interest, on different days and at four different user orientations. Out of this sample they selected 15 locations with a step of 1.5–2 m, which they then used as training data.

A different sample size was used in [25], where different filtering strategies for real life indoor 802.11 positioning systems were analyzed and compared. The authors measured the radio distribution at 250 uniformly distributed grid points in an area of 15 m x 35 m.

In [7], the differences among the received signal strengths from a number of 802.11 adapters were investigated. The authors of this work conducted system validation using a total of 3120 positioning requests.

In another aspect of indoor positioning, authors of [12] introduced several fault models to capture the effect of failures in the wireless infrastructure. During the investigation of fault tolerance of positioning methods and evaluation in terms of their performance degradation, they contacted experiments based on a radio-map consisting of 107 distinct reference locations having a step of 2–3 m. A total of 3210 reference fingerprints, corresponding to 30 fingerprints per reference location, were collected at the rate of 1 sample/s. For testing purposes they collected fingerprints along a path, consisting of 192 locations. Authors of [24] proposed a novel indoor localization scheme based on sub-area fingerprint determination and surface fitting. During the performance evaluation of the proposed technique, they performed experiments in an area 16.2 m x 28.5 m by setting 25 reference points per room and randomly selecting 200 test points in the environment.

Finally, authors of [20,21], worked towards the challenge of deployment load reduction in RSS based indoor localization systems. In their work, they proposed an interesting scheme that combines the data retrieved from a ray tracing simulator with a limited number of measurements (15%–30% of the complete fingerprint dataset) and performs localization using manifold alignment. The aforementioned methodology leads to a significant load reduction but assumes that the utilized fingerprint datasets have stronger correlation among neighboring data points compared to other points. This assumption is not always valid in indoor environments with strong multipath effects in Rayleigh channels.

Summarizing, our literature review suggests that authors calibrate, train, test and evaluate indoor positioning systems by utilizing a variety of sample sizes and