Research Article

Full-Band GSM Fingerprints for Indoor Localization Using a Machine Learning Approach

Iness Ahriz,¹ Yacine Oussar,¹ Bruce Denby,^{2,1} and Gérard Dreyfus¹

¹ Signal Processing and Machine Learning (SIGMA) Laboratory, ESPCI—ParisTech, 10 rue Vauquelin, 75005 Paris, France

² Université Pierre et Marie Curie—Paris VI, 4 place Jussieu, 75005 Paris, France

Correspondence should be addressed to Bruce Denby, denby@ieee.org

Received 1 October 2009; Accepted 25 March 2010

Academic Editor: Simon Plass

Copyright © 2010 Iness Ahriz et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Indoor handset localization in an urban apartment setting is studied using GSM trace mobile measurements. Nearest-neighbor, Support Vector Machine, Multilayer Perceptron, and Gaussian Process classifiers are compared. The linear Support Vector Machine provides mean room classification accuracy of almost 98% when all GSM carriers are used. To our knowledge, ours is the first study to use fingerprints containing all GSM carriers, as well as the first to suggest that GSM can be useful for localization of very high performance.

1. Introduction

Location-based services for cellular telephone networks are today very much in the public eye [1]. Global Positioning System, or GPS, receivers integrated into cellular handsets can provide very accurate positioning information; however, few mobiles are so equipped at present, and GPS furthermore performs poorly in the indoor and urban canyon environments which are prevalent in wireless networks. For these reasons, the study of localization techniques based upon the radio networks themselves is also a very active area. Most commercially installed systems still rely on cell-ID, in which the mobile station's position is reported as that of the serving base station. Although improvement is possible using triangulation, time of arrival, and the like, the accuracy of such methods is in practice compromised by the path loss and multipath characteristics inherent in the radio channel [2].

The database correlation method [3] allows to overcome channel effects to a certain extent. In this method, a mobile is localized by comparing one of the regularly emitted Received Signal Strength (RSS) measurements to a position-labelled database of such measurements, which are often called fingerprints. Existing localization services implemented in some GSM networks rely on Network Measurement Reports (NMR), which are a part of the GSM norm and contain the RSS and Base Station Identity Code (BSIC) of the serving

cell and six strongest neighboring cells. The resulting 7-component vector allows a localization precision of some tens of meters in outdoor environments (see, e.g., [4, 5]).

As for indoor radio-based localization, most studies which have appeared in the literature have involved WiFi networks, describing “corridor waveguide” scenarios in the workplace, and obtaining performance which, though interesting, can still be improved [6–8]. Another approach, using the household power lines as an antenna, appears in [9]. The notion of using GSM or CDMA networks for localization in indoor environments, particularly in domestic settings, is still somewhat new (see, e.g., [10, 11]). The basic idea is that inside a building, the RSS of the external base stations will be strongly correlated with a mobile's exact position, due to for example the varying absorption of electromagnetic energy by different building materials, and the exact placement of doors and windows. There has also been evidence that including more than the standard 7-carriers of the NMR fingerprint is advantageous in indoor GSM localization [10, 12].

In this article, we present tests of indoor GSM localization using scans containing large numbers of carriers—up to the full GSM band. In order to keep working with such large numbers of carriers tractable, we propose to create a mathematical model mapping fingerprints to position using machine-learning techniques, in this case Support Vector

Machines (SVM), and Multilayer Perceptrons (MLP), often also referred to as neural networks. We demonstrate the superiority of the machine learning approach, for problems with such high input dimensionality, over more traditional classifiers based on Euclidean (K-Nearest Neighbor) and Mahalanobis (Gaussian process) distances. Our results show that in an urban apartment setting, the room in which a handset is located can be identified with nearly 98% accuracy when the full set of GSM carriers is included. To our knowledge, this study, which is an extension of that described in [12], is the first to use fingerprints of all carriers in the GSM band, and the first to demonstrate very good performance on indoor localization using GSM.

The structure of the article is as follows. The data sets used in our study are presented in Section 2, while a discussion of preprocessing and the classifiers tested are given in Section 3. Our results are discussed in Section 4, while our conclusion, as well as some perspectives for the future is outlined in the final section.

2. Data Sets

The TEMS [14] trace mobile system was used to take twice-daily scans of the entire set of 498 GSM carriers in 5 rooms of a 5th floor apartment (top floor) in Paris, France. Both the RSS and the BSIC, where readable, were requested for each carrier in the scans. The layout of the apartment is shown in Figure 1. Acquisitions could be made anywhere within a room; however, in practice, the scans were recorded in those areas where the necessary laptop and cellphone could be conveniently set down and accessed. An exhaustive coverage of all rooms was thus not assured.

3. Data Analysis

3.1. Preprocessing. Ten of the carriers were found to contain no energy and were removed from the study. As the BSICs of the remaining 488 proved unreadable in many instances, a decision was made to exclude the BSICs entirely from the subsequent analysis, despite the possibility this engenders of confusing carriers at the same frequency in separate cellular motifs. The data set contained a total of 241 scans—approximately 48 scans per class, where a class is defined here simply as the index of the room within the apartment, indicated in Figure 1. To obtain a measure of the statistical significance of our classification results, cross-validation was performed with ten independent randomly selected splits of our data, each one containing 169 training examples and 72 validation examples. In a given split, the training and validation examples were uniformly distributed over time during the one-month acquisition period.

3.2. Dimensionality Reduction and Fingerprint Types. The relatively small size of our dataset is a reflection of the difficult, time-consuming nature of obtaining labeled scan data—a point to which we will return later. Its high dimensionality (488 carriers) also limits the complexity of the classifiers which may be applied. To deal with these

issues, signal strength-based carrier selection was initially carried out so as to define the four fingerprint types defined below. Further dimensionality reduction of any fingerprint can be obtained by a subsequent application of Principal Component Analysis (PCA).

Three vectors are used in defining the fingerprints:

$$\mathbf{g}_j^7 = \left\{ i = 1 \cdots 488, \sum_k \mathbf{1}_{\text{RSS}(i,j) < \text{RSS}(k,j)} \leq 6 \right\},$$

$$\mathbf{G}^7 = \bigcup_j \mathbf{g}_j^7, \quad (1)$$

$$\mathbf{G}^{35} = \left\{ i = 1 \cdots 488, \sum_k \mathbf{1}_{\langle \text{RSS}(i,j) \rangle_j < \langle \text{RSS}(k,j) \rangle_j} \leq 34 \right\},$$

where $\mathbf{1}$ is the so-called indicator function, and $\langle \rangle_j$ represents the mean over the index j . The first, $\mathbf{g}_j^7$, contains the indices of the 7 strongest carriers, i , in example j . The vector $\mathbf{G}^7$, composed of the indices of the carriers which were among the strongest 7 in at least one scan of the training set, contains between 36 and 40 of such “good” carriers, depending upon the random split used. The third vector, $\mathbf{G}^{35}$, consists of the indices of the 35 carriers which were the strongest on average, over the whole training set. The fingerprints may then be defined as follows.

(1) *Current Top 7.* These seven carrier fingerprints, $\text{RSS}(\mathbf{g}_j^7)$, are meant to mimic standard “top 7” NMRs, which were not present in our scans. Indeed, NMRs are only logged during a communication, while our scans were obtained in idle mode. Validation set fingerprints can in fact contain less than 7 elements if certain carriers were not represented in the training set. For classifiers requiring fixed labeling of input vectors, such as KNN, SVM, and MLP, the seven $\text{RSS}(\mathbf{g}_j^7)$, values are entered at the corresponding positions in a vector of length $\|\mathbf{G}^7\|$, and the rest of the elements are set to zero.

(2) *Top 7 with Memory.* This fingerprint, defined as $\text{RSS}(\mathbf{G}^7)$, includes the values of all of the 36–40 “good” carriers; they are thus “wider” than the *Current Top 7* fingerprint defined above.

(3) *35 Best Overall.* The *35 Best Overall* fingerprint, of length 35, is defined as $\text{RSS}(\mathbf{G}^{35})$. It thus gives another way of assessing the “goodness” of a carrier, by the size of its average RSS value over the whole training set.

(4) *All 488.* All of the active carriers’ RSS values are included in the fingerprint, that is, no selection is in fact made.

3.3. Classifiers. Four types of classifier were tested:

(1) *Support Vector Machines (SVM).* A 2-class SVM classifier [15] finds the separating surface which maximizes the distance (or “margin”) between that surface and the data points on either side of it. The SVM can be linear and operate

FIGURE 1: Schematic of apartment layout.

directly upon the data, or first map the data onto a higher-dimensional space using a non-linear transformation, before finding the maximum margin surface. The SVM decision rule takes the sign of

$$f(\mathbf{x}) = \sum_{i=1}^{N_s} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + b \quad (2)$$

with $\mathbf{x}$ the RSS vector to be localized, N_s the number of support vectors $\mathbf{s}_i$ (training vectors which are on the boundary of the optimal margin), and $y_i = \pm 1$ the class label of the vector $\mathbf{s}_i$. $K(\cdot)$ here is the selected kernel, and b as well as the α_i are parameters determined in the search for the optimal separating surface. For large, well-behaved data sets, the SVM rule approximates the Bayes decision rule [15].

In the case of a linear SVM, the kernel function is just the scalar product $K(\mathbf{s}_i, \mathbf{x}) = \mathbf{s}_i \cdot \mathbf{x}$. The standard Gaussian kernel was adopted in our tests of non-linear SVMs,

$$K(\mathbf{s}_i, \mathbf{x}) = e^{-|\mathbf{s}_i - \mathbf{x}|^2 / \sigma^2}, \quad (3)$$

where the variance, σ^2 , is optimized in the cross-validation stage. Since a “soft margin” approach was used (i.e., some training examples were allowed to lie within the margin), a regularization parameter controlling the complexity of the separating surface [15] was also estimated by cross-validation. For m classes, it is traditional, using the “conventional recipe” [16], to construct m binary, one-versus-rest classifiers, and take as the output class that of the classifier having the largest output value before thresholding. This procedure is illustrated for the case of $m = 5$ in Figure 2. The Spider SVM package [17] was used in all of our analyses.

(2) *Multilayer Perceptron (MLP)*. A multilayer Perceptron is a multivariate, nonlinear, scalar or vector function, which is a combination of parameterized elementary nonlinear functions called neurons [18]. A neuron is usually a function of the form $f = \tanh(\boldsymbol{\theta} \cdot \mathbf{x})$ where $\boldsymbol{\theta}$ is the vector of parameters of the neuron and $\mathbf{x}$ is the vector of variables. A single-output “multilayer Perceptron” $g(\mathbf{x})$ is a combination of N_h “hidden” neurons $f_i (i = 1 \text{ to } N_h)$ and of a constant equal to 1. Denoting by Θ_1 the vector of parameters of the linear combination (of size $N_h + 1$), by Θ_2 the $(N + 1, N_h)$ matrix whose elements are the parameters of the “hidden” neurons, and by $\mathbf{f}$ the vector (of size $N_h + 1$) of functions computed by the N_h hidden neurons with an additional component equal to 1, the multilayer Perceptron function is of the form

$$g(\mathbf{x}) = h(\Theta_1 \cdot \mathbf{f}(\Theta_2 \mathbf{x})). \quad (4)$$

Multilayer Perceptrons are frequently described pictorially as shown in Figure 3.

The parameters of the multilayer Perceptron are estimated from the available training data by minimizing the least squared cost function

$$J(\Theta_1, \Theta_2) = \sum_{k=1}^n (y_k - g(\mathbf{x}_k))^2 \quad (5)$$

with respect to all parameters, where $\mathbf{x}_k$ is the vector of variables pertaining to example k and y_k is the measured value of the quantity of interest for example k . In the present study, the gradient of the cost function was computed by a computationally efficient algorithm known as “backpropagation”, and the optimization of the cost function was

FIGURE 2: Architecture combining five one-versus-rest SVM classifiers to predict the class of an RSS vector from one of the carrier sets.

FIGURE 3: A multilayer Perceptron with a single output.

performed by the conjugate gradient algorithm with Powell-Beale restarts [19].

In a two-class (A, B) classification problem, $y_k = +1$ for all examples of class A and $y_k = -1$ for all examples of the other class. After training, an unknown example described by vector $\mathbf{x}$ is assigned to class A if $\text{sgn}(g(\mathbf{x})) = +1$, and to class B otherwise. In the present study, function h was taken identical to function f . For a c -class problem, example k , belonging to class i ($1 \leq i \leq c$), is assigned a vector $\mathbf{y}_k$, of dimension c , that encodes the class in a 1-out-of- c code: all components are equal to -1 , except component i , which is equal to $+1$. The number of output neurons is equal to the number of classes, so that the output of the multilayer Perceptron is a vector $\mathbf{g}(\mathbf{x})$ of dimension c . The cost function that is optimized during training is

$$J(\Theta_1, \Theta_2) = \sum_{k=1}^n \|\mathbf{y}_k - \mathbf{g}(\mathbf{x}_k)\|^2. \quad (6)$$

In the present study, two strategies were compared for multiclass classification with multilayer Perceptrons.

- (i) All functions h were taken identical to f (sigmoid functions), so that the output vector of the multilayer Perceptron was

$$\mathbf{g}(\mathbf{x}) = \mathbf{h}(\Theta_1 \mathbf{f}(\Theta_2 \mathbf{x})), \quad (7)$$

where Θ_1 is the $(c, N_h + 1)$ matrix of the parameters of the output neurons.

- (ii) Output i ($1 \leq i \leq c$) of the multilayer Perceptron was computed as

$$g_i(\mathbf{x}) = \frac{\exp[(\Theta_1 \mathbf{f}(\Theta_2 \mathbf{x}))_i]}{\sum_{j=1}^c \exp[(\Theta_1 \mathbf{f}(\Theta_2 \mathbf{x}))_j]} \quad (\text{softmax function}). \quad (8)$$

In either case, and example described by $\mathbf{x}$ was assigned to the class j such that

$$j = \text{argmax}_{1 \leq i \leq c} (g_i(\mathbf{x})). \quad (9)$$

In the second case, the components of vector $\mathbf{g}$ belong to $[0, 1]$ and sum to 1, so that they can be interpreted safely as estimates of the posterior probability of class c given the observed vector $\mathbf{x}$.

(3) *K-Nearest Neighbor (K-NN)*. As a first step, K -NN ranks the training vectors according to their RSS-space Euclidean distances from a test vector to be localized. The predicted class of this test vector is then the class most often represented in the K “nearest” vectors according to the defined metric. The K parameter is chosen empirically, to optimize performance. When a single best neighbor is used, $K = 1$, and the classifier is called *1-NN*.

TABLE 1: Percentage of correct radio fingerprint classifications on the 4 carrier sets described in the text. Figures quoted are averages and standard deviations over 10 randomly selected validation sets. All classifiers achieve their best performance when all 488 carriers are included. The most effective classifier for this case is the linear SVM.

Classifier			Fingerprint Type						
			Current Top 7 (≤ 7 carriers) ¹	Top 7/Memory (36–40 carriers)	35 Best Overall (35 carriers)	All 488 (488 carriers)			
Linear SVM			71.3 $\pm$ 7.2	84.6 $\pm$ 3.6	90.4 $\pm$ 3.5	97.8 $\pm$ 1.5			
Gauss. SVM	w/o PCA		72.2 $\pm$ 3.6	89.2 $\pm$ 2.9	93.2 $\pm$ 3.4	— ²			
	w/PCA ³		71.8 $\pm$ 3.2	85.6 $\pm$ 5.3	92.0 $\pm$ 3.0	96.4 $\pm$ 1.5			
Linear Perceptron			66.9 $\pm$ 4.1	73.2 $\pm$ 5.1	79.7 $\pm$ 5.1	94.4 $\pm$ 2.6			
MLP (one versus all)	w/o PCA		66.9 $\pm$ 7.1	87.2 $\pm$ 3.3	91.8 $\pm$ 3.4	— ²			
	w/PCA ³		68.1 $\pm$ 3.4	87.5 $\pm$ 4.5	89.6 $\pm$ 2.5	95.7 $\pm$ 2.1			
MLP (multiclass) sigmoids	w/o PCA		56.8 $\pm$ 7.1	80.4 $\pm$ 12.9	92.6 $\pm$ 3.2	— ²			
	w/PCA ³		66.4 $\pm$ 5.7	85.1 $\pm$ 9.5	89.4 $\pm$ 3.6	96.1 $\pm$ 1.1			
MLP (multiclass) softmax	w/o PCA		64.3 $\pm$ 7.5	85.7 $\pm$ 15.8	91.2 $\pm$ 4.2	— ²			
	w/PCA ³		67.7 $\pm$ 5.7	88.2 $\pm$ 3.9	90.4 $\pm$ 3.1	96.6 $\pm$ 2.4			
K_{best}	K-NN	5	59.3 $\pm$ 3.5	26	85.1 $\pm$ 3.0	20	93.3 $\pm$ 2.1	20	94.9 $\pm$ 1.9
1-NN			58.1 $\pm$ 5.2	74.7 $\pm$ 3.7	86.0 $\pm$ 2.9	87.2 $\pm$ 2.8			
GP ($\sigma = 5$ dB)			78.8 $\pm$ 3.7		— ⁴				

¹SVM and K-NN can have < 7 carriers if some did not show up in the training set.

²Small training set size precludes training a nonlinear classifier due to Cover's theorem [13].

³Best result obtained using the first 4 principal components.

⁴Gaussian process is equivalent to 1-NN for fixed input vector length.

(4) *Gaussian Process (GP)*. As in the case of K -NN, GP starts by comparing the test RSS vector to be localized to every vector in the training set. The probability, P_1 , that the compared vectors correspond to measurements at the same geographical position is assumed to be Gaussian in the Euclidean RSS distance between the two vectors, using a fixed variance σ^2 which is determined empirically. If a carrier appears in one of the compared vectors, but not in the other, GP presumes that the missing value was below the reception threshold in the vector lacking it. A penalty term probability, P_p , is introduced, in which the missing RSS value is replaced by an estimate of the reception threshold, taken to be the smallest RSS in the vector which is missing the carrier. The overall GP probability, P , is the product of P_1 and P_p .

To be more precise, let A and B be sets of indices of carriers contained in a training set vector, and a test set vector, respectively. We define the set of common carriers as $C = A \cap B$, and the noncommon carrier sets as $D = A - C$ and $E = B - C$, for the train and test sets, respectively. We then have

$$\begin{aligned}
 P_1 &= \sqrt{|C|} \prod_{i \in C} e^{-|\text{RSS}_i^A - \text{RSS}_i^B|^2 / \sigma^2}, \\
 P_p &= \sqrt{|D|} \prod_{j \in D} e^{-|\text{RSS}_j^D - \min_B(\text{RSS}^B)|^2 / \sigma^2} \\
 &\quad \times \sqrt{|E|} \prod_{k \in E} e^{-|\text{RSS}_k^E - \min_A(\text{RSS}^A)|^2 / \sigma^2}, \\
 P &= P_1 \cdot P_p,
 \end{aligned} \tag{10}$$

where RSS_i^A is the signal strength of the i th carrier of set A , and the order of each root normalizes the probability to the number of carriers in the corresponding term. GP is actually the only classifier tested which is able to handle missing carriers in a natural way. When input vectors are of fixed length—a requirement for SVM, MLP, and KNN—and all variables must be represented, GP is equivalent to a 1-NN classifier. As a caveat, however, as we do not use the BSIC information, in some cases, carriers with the same index can belong to different cellular motifs, which would penalize the GP method.

4. Results

We define the localization performance of a given classifier as the average of the validation scores obtained over our ten random splits, expressed as a percentage of correctly identified locations. The standard deviation over the ten splits is also calculated. The results are shown in Table 1.

A few preliminary remarks about the table are in order. First, when the *All 488* fingerprint is used, it is not meaningful to apply a non-linear classifier to the data. This is because of Cover's theorem [13], which states that the examples of a training set are always linearly separable when the number of input variables exceeds the number of examples. The corresponding table entries are thus left blank (footnote 2 in the table). Secondly, on the other hand, dimensionality reduction by principal component analysis is known to often make examples nonlinearly separable, giving poor performance (nonlinear separability of the training examples was verified using the Ho-Kashyap algorithm [20]).

TABLE 2: Confusion Matrices for 35 Best Overall and All 488 carrier sets, using a Linear SVM classifier. Figures quoted are in percent. Using the full number of carriers tightens up the diagonal to give individual room classification efficiencies near 100%.

(a)					
Confusion Matrix	True class 35 Best overall				
Pred. Class	1	2	3	4	5
1	95	5.3			3.3
2	1.4	93.3	3.6		
3	0.7	1.3	77.9	11.4	
4			16.4	87.9	0.7
5	2.9		2.1	0.7	96

(b)					
Confusion Matrix	True class All 488				
Pred. Class	1	2	3	4	5
1	100		0.7		
2		100			
3			91.4	1.4	1.3
4			5.7	98.6	
5			2.2		98.7

For this reason, linear classifiers are not applied in those cases where PCA is used. A few further details are explained in the remaining footnotes of Table 1.

The table shows that the performance of all classifiers tested improves as more carriers are added to the fingerprint, but that very good performance—for example, our best result of 97.8% in the case of the linear SVM—is only obtained on the *All 488* carrier fingerprint. The implication is that indoor position can indeed be deduced from the RSS of GSM cell towers, but that commonly used 7-carrier NMRs and even “wide” fingerprints are insufficient: high performance requires fingerprints of very high dimensionality. It is reassuring to see that this conclusion is supported by all the classifiers tested, including a simple K -NN, even if the best results are obtained with SVM and MLP machine learning techniques. MLP performance appears slightly worse than that of linear SVMs, within the statistics of our sample, with the best MLP performance, 96.6%, obtained on a multiclass MLP with the softmax output function applied to *All 488* carriers, after an input dimensionality reduction.

A more detailed look at our conclusion is given in Table 2, where the confusion matrices for the linear SVM classifier on the *35 Best Overall* and *All 488* fingerprints appear. The table shows once again that the ability to sharply discriminate between rooms comes only with the inclusion of the full GSM carrier set. The deviation of our global result from 100% is in fact dominated by the confusion between class 3 and class 4, which appears to be the most difficult case.

5. Conclusions and Perspectives

We believe this study, which is an extension of that presented in [12], to be the first to include the full set of GSM carriers in RSS fingerprints for localization. Although confirmation with more extensive databases will be required, our results strongly suggest that high-performance room-level localization is possible through the use of such fingerprints. The fact that good performance is obtained irrespective of the machine learning technique used (MLPs or SVMs,) is a further confirmation that the useful information for localization is obtained by taking into account many GSM carriers, including those which may be rather weak. Finally, it is interesting to note that our result is robust against time-dependent effects—network modifications, propagation channel changes, meteorological effects, and so forth, as our dataset was acquired over a period of one month.

Acquiring datasets and labeling scans is a tedious and time-consuming activity. To address this issue, two independent solutions are currently being investigated. First, experiments with semisupervised classification techniques using kernel methods (see, e.g., [21]) are being carried out, which will permit to take advantage of the unlabeled scans during the training procedure. The second approach entails the design and construction, in our laboratory, of a set of ten autonomous scanning devices which will allow the acquisition of large datasets simultaneously in different rooms, labeled with very little human intervention. These devices will also enable to test the efficiency of our approach when implemented using mixed datasets of scans acquired both indoors and in nearby outdoor areas. For larger outdoor areas, preliminary results indicate that a regression approach using x - y coordinates seems more suitable than the room-by-room classification used here for indoor localization.

References

- [1] A. Küpper, *Location-Based Services: Fundamentals and Operation*, John Wiley & Sons, New York, NY, USA, 2005.
- [2] H. Laitinen, et al., “CELLO: cellular network optimization based on mobile location,” Internal Report IST-2000-25382-CELLO, EU IST Project, October 2001, <http://www.telecom.ntua.gr/cello/documents/CELLO-WP2-VTT-D03-007-Int.pdf>.
- [3] D. Zimmermann, J. Baumann, M. Layh, F. Landstorfer, R. Hoppe, and G. Wölflé, “Database correlation for positioning of mobile terminals in cellular networks using wave propagation models,” *Proceedings of the 60th IEEE Vehicular Technology Conference*, vol. 7, pp. 4682–4686, September 2004.
- [4] M. Chen, T. Sohn, D. Chmelev, et al., “Practical metropolitan-scale positioning for GSM phones,” in *Proceedings of the 8th International Conference on Ubiquitous Computing*, P. Dourish and A. Friday, Eds., vol. 4206 of *Lecture Notes in Computer Science*, pp. 225–242, Springer, Orange County, Calif, USA, September 2006.
- [5] B. Denby, Y. Oussar, and I. Ahriz, “Geolocalisation in cellular telephone networks,” in *Proceedings of the NATO Advanced Study Institute on Mining Massive Data Sets for Security*, F. Fogelman-Soulié, D. Perrotta, J. Piskorski, and R. Steinberger, Eds., IOS Press, Amsterdam, The Netherlands, 2007.

- [6] M. Brunato and R. Battiti, "Statistical learning theory for location fingerprinting in wireless LANs," *Computer Networks and ISDN Systems*, vol. 47, no. 6, pp. 825–845, 2005.
- [7] A. M. Ladd, K. E. Bekris, A. P. Rudys, D. S. Wallach, and L. E. Kavraki, "On the feasibility of using wireless ethernet for indoor localization," *IEEE Transactions on Robotics and Automation*, vol. 20, no. 3, pp. 555–559, 2004.
- [8] Q. Yang, S. J. Pan, and V. W. Zheng, "Estimating location using Wi-Fi," *IEEE Intelligent Systems*, vol. 23, no. 1, pp. 8–9, 2008.
- [9] S. N. Patel, K. N. Truong, and G. D. Abowd, "Powerline positioning: a practical sub-room-level indoor location system for domestic use," in *Proceedings of the 8th International Conference on Ubiquitous Computing (UbiComp '06)*, Orange County, Calif, USA, September 2006.
- [10] V. Otsason, A. Varshavsky, A. LaMarca, and E. de Lara, "Accurate GSM indoor localization," in *Proceedings of the International Conference on Ubiquitous Computing (UbiComp '05)*, M. Beigl, et al., Ed., pp. 141–158, Springer, Berlin, Germany, 2005.
- [11] W. ur Rehman, E. De Lara, and S. Saroiu, "CILoS: a CDMA indoor localization system," in *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp '08)*, pp. 104–113, Seoul, Korea, 2008.
- [12] B. Denby, Y. Oussar, I. Ahriz, and G. Dreyfus, "High-performance indoor localization with full-band GSM fingerprints," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC '09)*, Dresden, Germany, June 2009.
- [13] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Transactions on Electronic Computers*, vol. 14, pp. 326–334, 1965.
- [14] Test Mobile System, <http://www1.ericsson.com/solutions/tems>.
- [15] N. Cristianini and J. Shawe-Taylor, *Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge, UK, 2000.
- [16] Y. Lee, Y. Lin, and G. Wahba, "Multicategory support vector machines: theory and application to the classification of microarray data and satellite radiance data," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 67–81, 2004.
- [17] The Spider, <http://www.kyb.tuebingen.mpg.de/bs/people/spider>.
- [18] G. Dreyfus, *Neural Networks: Methodology and Applications*, Springer, New York, NY, USA, 2005.
- [19] M. J. D. Powell, "Restart procedures for the conjugate gradient method," *Mathematical Programming*, vol. 12, no. 1, pp. 241–254, 1977.
- [20] Y.-C. Ho and R. L. Kashyap, "An algorithm for linear inequalities and its applications," *IEEE Transactions on Electronic Computers*, vol. 14, no. 5, pp. 683–688, 1965.
- [21] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*, MIT Press, Cambridge, Mass, USA, 2006.