A HYBRID CLASSIFICATION APPROACH TO IMPROVING LOCATION ACCURACY IN A BLUETOOTH-BASED ROOM LOCALISATION SYSTEM

YI-MIN GUO[1], SHENG-YI PAN[2], HAI-YING WANG[1], HUI-RU ZHENG[1]

[1] School of Computing of Mathematics, University of Ulster, Co. Antrim, UK
[2] University of Sheffield, UK
E-MAIL: h.zheng@ulster.ac.uk

Abstract:

It has been well recognised that the use of localisation techniques in home environment are beneficial to the development of health monitoring and activity recognition systems. The Bluetooth devices, as a kind of effective sensor with remarkable characteristics such as low cost, have been widely used in our daily life. Research has been carried out to integrate cellular network signal measurements and Bluetooth link measurements in developing home localisation systems. This paper presents a hybrid classification approach, based on the combination of Bayesian statistics and supported vector machines, to supporting the development of the Bluetooth-based room localisation system. The proposed approach considers the dependency between features and non-linear overlapping of features between rooms. The results show that the prediction accuracy has been improved in comparison to the traditional Naive Bayes classifier and the hidden Markov model used in previous studies.

Keywords:

Bluetooth; Bayesian statistics; Support vector machine; Room localisation systems

1. Introduction

The pervasive sensing technologies and machine learning techniques have offered great opportunities in the development of smart environments for home based telehealth and telemedicine. The first smart room set up in the Media Laboratory at the Massachusetts Institute of Technology in 1991 led the trend of the research in smart environment technology [1]. The growing of aging population, the cost of formal health care and the desires of remaining independent in own homes have driven the research forward. The focus of the research has been monitoring activities of daily living and life styles. However, before a smart environment can begin to figure out what people are doing, it needs to identify their location at the first place [1]. Thus, localisation systems installed in smart room are becoming a promising area for the researchers to investigate.

A simple model of smart rooms consists of controllers and several sensors. A computer acts as a controller to communicate with various kinds of sensors which can be used to provide location information for people in the room.

A number of researches on determining and tracking indoor location of people or sensors are carried out. For example, in the study published by Pentland [1], the problem of location detection was addressed by installing cameras to record the motions of people. However, this method required expensive equipments and it may increase the system complexity. Furthermore, such a approach needs to address the issue of privacy. Other approaches include using the Wireless Local Area Network (WLAN) infrastructure, which is practicable in an office environment where IEEE 802.11 wireless access points (APs) and devices (e.g. PDAs, laptop computers) are often available [2]. However, this may not be the case for home environment in which the infrastructure is not sufficient enough to support this kind of development in most of cases and the WLAN devices are not usually available.

The introduction of Bluetooth technology gives us the opportunity to address some major problems raised in the in-home localisation systems. Its attractiveness lies not only in the low power property, but also in the advantage that it has already been integrated into most of portable electronic devices such as mobile phones. Therefore, the Bluetooth devices are easily for people to carry at home, which may result in the reduction in the cost while arranging an in-home localisation system. Furthermore, within the maximum permitted power, the Bluetooth signals can transmit in the range of 100 meters, which is sufficient for the in-home environment test for most of cases. Kelly *et al* [3, 4] proposed an affordable localization system using a Bluetooth enable mobile device, Blue Radios BR-SC30N class 1, as a single AP connected to a base station computer. The signals received by the base station were used to predict the user's location. A hidden Markov Model (HMM)

978-1-4244-6527-9/10/$26.00 ©2010 IEEE

localization method was applied to improve the prediction accuracy over a Naive Bayes (NB) classifier. The maximum prediction accuracy was 72% for HMM classifier and 62% for the NB approach.

In this paper, we propose a hybrid approach to supporting the development of Bluetooth-based room localisation system. It aims to improve the prediction accuracy by incorporating the dependency between features and non-linear overlapping of features between rooms into the prediction process.

The rest of the paper is organised as follows. Section 2 describes the dataset under study, followed by a description of methodology in Section 3, which introduces the related classification methods, and details the proposed hybrid classification approach. Results and discussion are presented in Section 4. The paper concludes by a summary and future work in Section 5.

2. Dataset under study

The dataset under study was provided by [3]. It consists of three signal measurements collected by a Bluetooth-based localisation system used in [3, 4] from a four-room, denoted as *Room1*, *Room2*, *Room3* and *Room4*, setting environment (see Figure 1). The three measurements are link quality (LQ), received signal strength indicators (RSSI) and cellular signal quality (CSQ). The dataset contains a total of 880 cases, among which there are 253, 284, 203 and 140 cases measured from Room1, Room2, Room3 and Room4 respectively. The problem imposed in this dataset is, given the value of LQ, RSSI and CSQ, we want to differentiate whether the subject is in the given room (positive case) or not (negative case).

2.1. LQ

LQ is a measurement available during the course of communications between a BlueRadios BR-SC30N chip and another Bluetooth device [3]. It measures how much information lost in transmission. When transmitting a command, it has a linear relationship with Bit-Error-Rate (BER), which means that the quality of signal is influenced by the transmitter-receiver separation distance and obstructions in the environment [3]. The value of LQ ranges from 172 to 255.

2.2. RSSI

The RSSI gives a measurement of the intensity of the electromagnetic power incident on the receiver. Due to the usage of power control in the Bluetooth device, the receiver

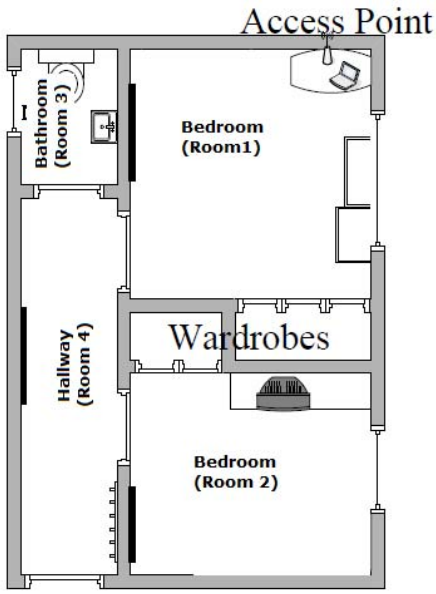

Figure 1 Layout of rooms where data were collected (adapted from [4]).

is able to notify the transmitter to adjust its power level automatically in order to maintain the transmitting signal strength in a specified "golden range" [3]. According to the specific value of RSSI, the receiver can tell whether or not any other Bluetooth devices are within the range of the golden range and may calculate the strength of power above or below a "golden value". Then most important factor that affects the RSSI is the distance between the receiver and transmitter. RSSI is intended to be used as a "relative value" in the chipset. In this case, the lowest value obtained is -10. This is the evidence of the transmitter being unable to increase its transmitting power level any further. Nevertheless, a value of 0 indicates the device is moving within the golden range. From the results of the experiment, it was noted that the RSSI contains no less information than LQ [3, 4]. In addition, it shows that there is different information encoded between them. It has been also found that a device which is outside the golden range may result in the weak correlation between RSSI and the actual signal strength, resulting in a value less than zero.

2.3. CSQ

CSQ represents the strength of a cellular basement. LQ varies as a function of distance and obstructions between their access point and the Bluetooth device. Nevertheless, the use of LQ on its own is not sufficient to locate a stationary subject. The CSQ is complementary to LQ signal. When a connection is established between a base station and a mobile phone, the mobile phone will reply with CSQ containing the signal strength corresponding to the current associated cellular base station in response to the command sent from the base station. The data range of CSQ would be around -102 to -93.

In the experiment conducted by [3, 4], a mobile phone was carried by a human moving between the four rooms within the home environment. All the measurement data training were obtained from dynamic object. In comparison to the data obtained by the mobile phone placing in a stationary location, this approach is more practical.

3. Methodology

Based on a combination of likelihood ratio (LR) -based Bayesian network with support vector machine (SVM), a LR-SVM hybrid classification approach is proposed in this study.

3.1 LB-based Bayesian networks

Given a two-classes (positive and negative) problem, the posterior probability of a positive case given its value of n features, $f_1, f_2, \ldots, f_n$, can be defined as [6]

$$P(pos \mid f_1, f_2, \ldots, f_n) = \frac{L(f_1, f_2, \ldots, f_n)}{L(f_1, f_2, \ldots, f_n) + P(neg)/P(pos)} \quad (1)$$

where $P(pos)$ and $P(neg)$ give the prior probability of positive and negative and

$$L(f_1, f_2, \ldots, f_n) = \frac{P(f_1, f_2, \ldots, f_n \mid pos)}{P(f_1, f_2, \ldots, f_n \mid neg)} \quad (2)$$

is the likelihood ratio. If features, $f_1, f_2, \ldots, f_n$, are conditional independent, then calculation of the likelihood would be greatly simplified which can now be expressed as,

$$L(f_1, f_2, \ldots, f_n) = \prod_{i=1}^{n} \frac{P(f_i \mid pos)}{P(f_i \mid neg)} = \prod_{i=1}^{n} L(f_i) \quad (3)$$

In this study, the dependency between features was assessed on basis of the computation of Pearson's Correlation Co-efficient (PCC) and Mutual Information (MI) [7]. Given two N dimensional vectors $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_N)$, where X and Y can represent the feature values, the PCC can be calculated as:

$$PCC(X,Y) = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}} \quad (4)$$

The MI between X and Y denoted as $MI(X;Y)$ is computed using the following equation.

$$MI(X;Y) = \sum_{i=1}^{N} x_i \sum_{j=1}^{N} y_j \, p(x_{i,} y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j} \quad (5)$$

Where $p(x)$ and $p(y)$ represent the probability distribution and $p(x, y)$ stands for the joint distribution of X and Y. The values of PCC drop the interval of [-1, 1]. For both indicators, the larger the value is, the stronger correlation two features have.

In order to predict whether an input sample case denoted as $S = (f_1, f_2, \ldots, f_n)$ is positive, we wish to determine whether the posterior probability as expressed in Equation (1) is greater than 0.5. This condition will be met whenever

$$L(f_1, f_2, \ldots, f_N) > \frac{P(neg)}{P(pos)}. \quad (7)$$

3.2 SVM

A SVM performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories [8].

Given a training set of instance-label pairs (x_i, y_i), $i = \{0, 1, 2, \ldots, k\}$ where $x_i \in R^n$ and $y_i \in \{1, -1\}$ being the class label indicating the class to which x_i belongs, we can find a separating hyperphane (for linearly separable cases), which takes the form as:

$$w^T x + b = 0 \quad (8)$$

where the vector w is normal to the heperplane and the value of $|b|/\|w\|$ decides the perpendicular distance from the hyperplane to the origin. For a given separating hyperplane, both positive and negative samples in the training data set need to satisfy the following constraints:

$$w^T x_i + b \geq 1, \text{for positive cases with } y_i = +1 \quad (9)$$

$$w^T x_i + b \le -1, \text{for negative cases with } y_i = -1 \quad (10)$$

For non linear cases we need to map the original nonlinear problem into a higher-dimensional with the kernel trick. Let $K(x_i, x_j) = \varphi(x_i^T)\varphi(x_j)$ be a kernel function satisfying the Mercer's theorem, the classification with SVM can be represented as

$$\begin{cases} \omega^T \phi(x_i) + b \le -1, \text{if } \ f_i = -1 \\ \omega^T \phi(x_i) + b \ge +1, \text{if } \ f_i = +1 \end{cases} \quad (11)$$

There are four basic kernel functions, namely linear, polynomial, radial basis function (RBF) and sigmoid. In this research the SVM model with RBF as the kernel function was used.

3.3 A combination of LR-SVM approach

The LR-based Bayesian network has ability to consider the dependency between features while SVM has the advantages of differentiating classes with non-linear features. To integrate the advantages demonstrated by both LR-Bayesian networks and SVM classifiers, a LR-SVM hybrid classification approach is proposed to localise room positions from Bluetooth signals as illustrated in the Algorithm 1 shown below.

Algorithm 1 LR-SVM based hybrid classification approach

1: Input the feature f_i values f_{ij}(*i=1, 2, …, n; j=1,2,.., m*), where *n* is the number of features and *m* is the number of samples.
2: Discretisation f_{ij} for each feature f_i
3: Calculate likelihood ratio tables
4: if $P(pos \mid f_1, f_2, \ldots, f_n) \ge 0.5$
5: apply LB classifier
6: else apply SVM classifier
7: Output classification results

4. Results and discussion

In an attempt to assess the dependency between features, the pair-wise PCC and MI between features were computed (shown in Table 1). As can be seen from the table, all three features are highly correlated with all the PCC value greater than 0.5, while the information overlapping of LQ with RSSI, LQ with CSQ are greater than the overlapping of RSSI and CSQ. This is consistent with the correlation degree of the three features. Since all three features are not independent as indicated by their PCC and MI information, the NB-based technique may not be suitable for combining the evidence encoded in these three features. This is consistent with the analysis provided by Kelly et al. [3].

TABLE 1 CORRELATION AND MI BETWEEN THREE FEATURES. THE FIRST VALUE IS PCC AND THE SECOND ONE IS MI MEASURE

PCC/MI	LQ	RSSI	CSQ
LQ	-	0.82/0.79	0.67/0.56
RSSI	0.82/0.79	-	0.69/0.35
CSQ	0.67/0.56	0.69/0.35	-

The data discretisation was achieved using alternating decision tree (ADTree) provided by the open source Weka software package [5] followed by merging ranges having low likelihood ratios. The example of the calculation of the associated LR with each individual feature is illustrated in Tables 2 to 4.

TABLE 2. LIKELIHOOD RATIOS OF LQ FOR ROOM1. FOR EACH INTERVAL, THE NUMBER OF POSITIVE AND NEGATIVE CASES FOUND IN THE DATASET, AND THEIR CORRESPONDING PROBABILITIES WERE LISTED ALONG WITH THE RELATED LIKELIHOOD RATIO VALUE.

Discretisation Interval	Neg	Pos	P(value|neg)	P(value|pos)	LR
[172, 210]	448	33	0.71	0.13	0.18
[211, 255]	179	220	0.29	0.86	**2.97**

TABLE 3. DATA DISCRETISATION AND LIKELIHOOD RATIOS OF CSQ FOR ROOM1.

Discretisation Interval	Neg	Pos	P(value|neg)	P(value|pos)	LR
[-102, -100]	505	104	0.81	0.41	0.51
[-99, -93]	122	149	0.19	0.59	**3.11**

TABLE 4. DATA DISCRETISATION AND LIKELIHOOD RATIOS OF RSSI FOR ROOM1

Discretisation Interval	Neg	Pos	P(value|neg)	P(value|pos)	L(S_1)
[-6, 0]	54	188	0.086	0.74	**8.64**
[-9, -7]	164	53	0.26	0.21	0.80
[-10, -10]	409	12	0.65	0.047	0.073

In this study we binned the LQ into 2 bins, CSQ values into 2 intervals and RSSI into 3 intervals, as described in the first column in Tables 2 to 4. After each feature values were discretised, three features were integrated and the intervals were combined. As shown in Algorithm I, if the maximum probability of locating a subject in a room given the values of LQ, CSQ, and RSSI calculated using Equation (1), is less than 0.5, the SVM with the RBF as a kernel function is used. Table 5 summarises the prediction results with the proposed LR-SVM approach. The performance was assessed using a 10-fold cross-validation, i.e. the dataset dataset is divided into 10 subsets with roughly equal size. Each subset in turn is used as the test set while the other subsets are used as a training set. The following three statistical measures, accuracy (*Acc*), sensitivity (*Se*) and Specificity (*Sp*), were employed to assess the performance of the prediction.

$$Acc(\%) = TP/ (TP+FP+TN+FN) \times 100 \quad (12)$$
$$Se\ (\%) = TP/ (TP+FN) \times 100 \quad (13)$$
$$Sp\ (\%) = TN/ (FP+TN) \times 100 \quad (14)$$

where *TP* is true positive (samples in Room *i* correctly classified as Room *i*, for *i*=1, 2, 3, 4), *FN* is false negative (samples incorrectly classified as not belonging to Rooms *i*), *FP* is false positive (samples incorrectly classified as Room i), and *TN* is true negative (samples correctly classified as not belonging to Room *i*).

As shown in Table 5, the better classification performance was achieved using the proposed LR-SVM hybrid approach in comparison to NB and HMM approaches presented in [3]. The classification accuracy for Room1, Room2, Room3 and Room4 reached 0.74, 0.96, 0.75 and 0.86 respectively. The average accuracy of four rooms is 0.83, which is much higher than that using the NB classifier (0.63) and the HMM classifier (0.73) [3]. This indicates that taking into account of the dependency between features and potential non-linear overlapping of features between different rooms can improve the prediction performance. A closer examination of the output results reveals that the classification of Room1 and Room2 were generated with the LR-based Bayesian network while Room3 and Room4 were generated from SVM, This indicated that the Bluetooth signals from Room3 and Room4 may have a non-linear overlapping, which may not be able to be differentiated by using the LR-based Bayesian network alone. From Figure 1, we observed that the Room3 is closely related to the Room4 and the Room3 is the smallest room among the four rooms. Whether such a layout contributes to the overlapping between the Room3 and Room4 deserves further investigation.

TABLE 5 THE PREDICTION PERFORMANCE FOR EACH ROOM WITH THE PROPOSED LR-SVM APPROACH

Measurements	*Acc*	*Se*	*Sp*
Room1	74%	78%	83%
Room2	96%	64%	97%
Room3	75%	78%	62%
Room4	86%	86%	83%
Average	83%	77%	81%

5. Summary and future work

The Bluetooth based room localisation system has the advantage of easy setting, low cost and few requirements on hardware. However, the appropriate use of classification models is critical to achieve a high performance of location prediction. Based on the integration of the advantages demonstrated by the LR-based Bayesian model and the SVM, this research proposed a LR-SVM based hybrid approach to improving the location accuracy from the work presented in [3]. Results obtained show that the prediction of location accuracy has been improved when considering the dependency between features and non-linear overlapping between rooms. The high classification accuracy for Room2 (96%) was achieved while the average accuracy was 83%, which represents a big improvement in comparison with the results obtained using the NB and HMM [3].

It is worth noting that the HMM classifier adopted in [3] has a merit of considering the room access pathway, which has not been considered in the LB-SVM approach. The combination of the access pathway knowledge with the hybrid classification model would be an important part of our future work.

Currently, the proposed method has only been tested on the dataset collected from one domestic environment. Moreover, the size of the dataset under study is relatively small. The behaviour of the proposed model in large datasets collected from different room layouts will be further investigated.

Acknowledgements

We would like to thank Dr. D. Kelly in National University of Ireland for the useful discussion and kindly providing the dataset used in this study.

References

[1] A.P. Pentland, “Smart Rooms Smart Clothes”, Scientific American, Vol. 274, No. 4, pp. 68-76, Apr. 1996
[2] P. Bahl and V. Padmanabham, “RADAR: An In-building RF-based User Location and Tracking System,” in INFOCOM 2000. Proceeding of Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, pp. 775-784, 2000.
[3] D. Kelly, S. McLoone, and T. Dishongh, “A Bluetooth-based Minimum Infrastructure Home Localisation System”. Proceedings of 5th IEEE International Symposium on Wireless Communication Systems (ISWCS 2008), 2008.
[4] D. Kelly, S. McLoone, T. Dishongh, M. McGrath, and J. Behan, “Single access point location tracking for in-home health monitoring”, Proceedings of 5th Workshop on Positioning, Navigation and Communication, pp. 23-29, Mar. 2008.
[5] I.H. Witten and E. Frank, “Data Mining Practical Machine”, ELSEVIER, San Francisco, 2005.
[6] H. Zheng, H. Wang, D. Glass, “Integration of genomic data for inferring protein complexes from global protien-protein interation networks”, Proceedings of IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 38, pp. 5-16, 2008.
[7] L. J. Lu, Y. Xia, A. Paccanaro, H. Yu and M. Gerstein, “Assessing The Limits of Genomic Data Integration for Prediction Protein Networks”, Genome Res., vol. 15, pp. 945, 2005.
[8] C. Cortes and V. Vapnik, “Support-vector Network”, Machine Learning, vol. 20, pp. 273–297, 1995.