Pre-processing of Data in RSS Signature-Based Localization

C. M. Takenga, Student Member IEEE[1], K. Kyamakya[2]

[1]IKT, University of Hannover, Hannover, Germany, takenga@ant.uni-hannover.de
[2]Alpen Adria University Klagenfurt, Chair of Transportation Informatics, Klagenfurt Austria, kyamakya@isys.uni-klu.ac.at

Abstract –Most applications for location based services rely on a precise determination of the mobile terminal (MS) location. This paper proposes a method of locating a MS in a GSM cellular system based on the radio signal strength (RSS) and use of neural network (NN). The proposed algorithm is best suited for urban and suburban environments. Due to the big effort during RSS collection from the streets to realize a fingerprint database, predicted RSS-data of the area of concern are rather used for this experiment. An exact modelling of all effects in prediction formula in a complex environment is not appropriate in a practical implementation. Therefore, the predicted data differ from the real one. A system is developed which calibrates the radio propagation prediction data on the basis of sample real collected measurements. During this correction procedure, the instability or stochastic behavior of the RSS is taken into account. An algorithm which reduces the noise in the collected RSS is applied. Having the corrected predicted data, there is no need of collecting the RSS from all the streets to train the NN. The positioning accuracy is compared for different pre-processing methods and calibration shows to reduce the positioning error.

1 Introduction.

Intelligent systems require training or expert knowledge. A large amount of training data is required to build the database for training the network. This training data is often hard to obtain, time consuming and may not be a good representation of the total dataset. The case in this paper is a fingerprint-based positioning, the signatures are the RSS. Due to the big effort during RSS collection from the streets to realize a fingerprint database, predicted RSS-data of the area of concern are rather used in our experiment. The wave propagation modelling in urban areas is a very complex task. Various propagation effects have to be considered. An exact modelling of all these effects in such a complex environment is not appropriate in a practical implementation due to computing time constraint. Only the most important factors are taken into account. Due to this above factor, the predicted RSS provided differ to the real data. For this reason a calibration system will be developed to correct these predicted data on the basis of sample RSS measurements. The high sensitivity of the RSS to the environmental changes makes it difficult to get the best estimated value. A system which remedies to this issue is applied in this work.

The remainder of this paper is organized as follows. In section 2, we survey related works in RSS-fingerprint position determination technologies. In the third section, we describe our system and the experimental environment. In section 4, we discuss our research methodology for calibration. The noise cancellation algorithm appears in section 5. The positioning algorithm and the positioning results are in section 6. Finally, we conclude in section 7.

2 Related works

A number of papers have paid an attention on this issue of location of a mobile system using the received RSS[1-5]. Some works have used the real collected RSS to train the NN, which requires RSS collection from many streets for a better accuracy[2, 4, 5]. Others have used the predicted data, which was less accurate than the first due to the difference between the prediction and the real RSS. This paper finds a compromise between the demanding effort and accuracy. The predicted data are corrected by only a sample of real collected data.

In[1] an algorithm that is based on the RSS measurements and considers the fluctuation in received signal strength due to shadowing. This algorithm is about twice more accurate than the cell-based algorithm. The RSS were used to train the NN used for positioning in[2] but the issue due to the random behavior of the RSS was not taken into account.

In [3, 5], positioning is tackled as a multi-class classification problem. The area of interest is divided into small square sections and location estimation of the MS section using received power levels from different base stations was the classification task.

3 System description

Many researchers have worked on the RSS prediction models[6-8]. A database of the predicted RSS for our interested area is from the E-Plus Mobile Network (Germany). This database is from a prediction model based on Outdoor and Outdoor-to-indoor coverage in urban areas at 1.8 GHz[6].

The wave propagation modeling in urban areas is a very complex task. Various propagation effects, like diffraction over the roof tops and around building corners, scattering on buildings, multipath and penetration through vegetation have to be considered.

Due to difficulties in implementation, only the most important factors were taken into account. As a result, the predicted RSS provided from this algorithm differ to the real data.

The predicted RSS used are from 10 Base stations, whereby four of them are indoor antennas. Each of the 6 outdoor Base stations has three sectors. The experimental area is an urban area with some high buildings. Data set contains 365 148 data records from each of the 22 cell antennas generated from 365 148 locations (5m x 5m resolution) in an area 3 km x 3 km.

Fig. 1 shows the area used for the experiment and the RSS map from one cell antenna.

The GSM-RSS used to correct the predicted ones are collected along the streets of our experimental area by using a GPS-GSM receiver with a FALCOM A2D-3 Modem.

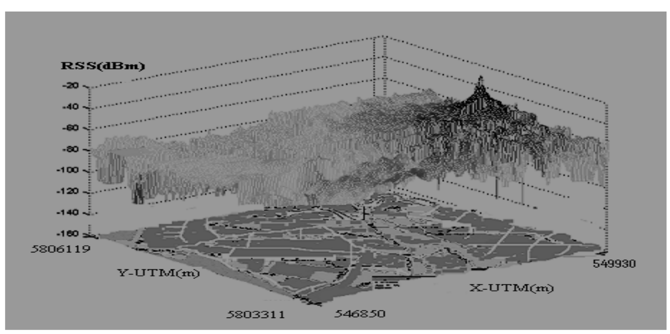

Fig.1 Area of concern (3x3km) and 3D-RSS map from one Cell antenna (UTM=Universal Transversal Mercator Projection)

4 Calibration of the predicted RSS

In this work, calibration is referred to as correction of predicted RSS by a sample of measured data from the streets. Neural network is used to make the relationship between the predicted and measured RSS as shown in Fig.2. The flexibility and ability of the NN to deal with uncertain data is the main reason of its use in this calibration issue[9].

The NN has 24 inputs, i.e. 22 predicted RSS from 22 antennas and 2 inputs which are the coordinates of the points used. The two position coordinates are involved to the input in order to make a better generalization for the entire experimental region. 22 outputs of the NN are compared to a string of 22 RSS collected data from the street.

Due to the limitation of standard GSM receiver, only a maximum of 7 strongest RSS values appear in measured record. Therefore, missing values are assumed to be equal to the limited sensitivity level. This is the cause of a big difference between the predicted values and the real ones when taking the mean value at each point, see Fig.3. The collected RSS from the GSM-GPS receiver are filtered from noise in the Noise cancellation unit.

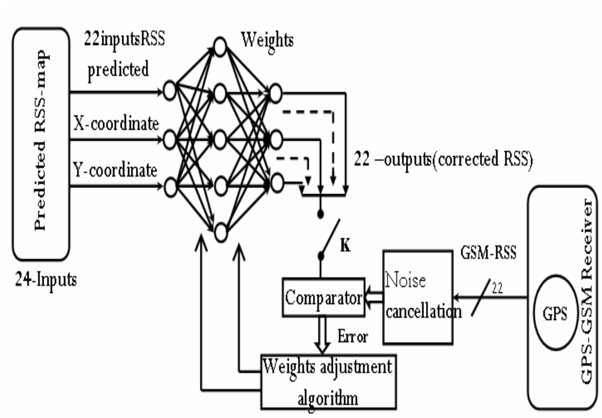

Fig. 2 Block diagram used for calibration

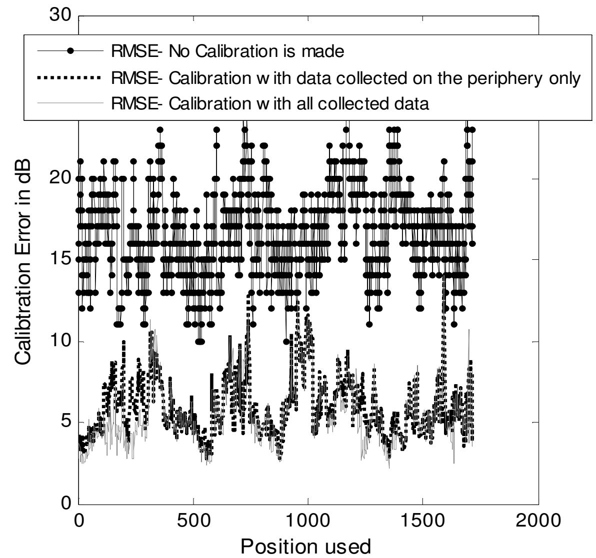

Fig.3 Dependence of calibration error on different scenarios

Most standard RF receivers have a limited receiver sensitivity level. Therefore, all the predicted RSS which power values are smaller than -113dBm or RxLev = 0 are fixed equal to this limited value during training phase.
The error coming from the comparator, which is the difference between the desired output and the actual one, is then utilized in weight-adjusting algorithm to determine the amount of adjustment to be made in the weights in both layers.
In order to start the training process, the connector K is closed and weights are randomly adjusted to small random values. When the input vector with predicted RSS and the corresponding coordinates is applied to the NN, it produces an output vector which is compared to the target vector (filtered GSM-RSS) to produce the error. The weight-adjusting algorithm then modifies the weights in the direction that reduces this error. When the input vector is again applied, it produces a new output and this process is repeated over and over until the error is minimized to some specified value or to an irreducible small quantity. At this time, the NN is said to have been trained to map input vector into desired output vector. The connector K can then be opened and any predicted RSS at the input will give us a corrected value of that predicted RSS at the output.

NN are very sensitive to absolute magnitudes, if some of the input range from 0 to -113 and other are very big values (x-UTM, y-UTM), fluctuations in the big range input will tend to swamp any importance given to the first, even if the first input is much more important in predicting the desired output. To minimize the influence of absolute scale, all inputs to a NN are scaled and normalized so that they correspond to roughly the same range of values 0 to 1.

During the RSS samples collection, it does appear that some data are missing. In spite of the robustness property of the NN that it can work with incomplete data sets, missing data can create serious problems. If a data cannot be found, the common sense and technically correct thing to do is to replace every missing value with the best estimate of what it would have been. In our case, the limited sensitive value appears to be the best estimate.

Fig.3 shows the calibration impact on the data. Without calibration the calibration RMSE is around 17dB. With calibration, the error is reduced to a value around 5dB-difference. We should note that, this big difference is also due to the fact that we are considering the mean value from all the antennas. Missing values in measured data are taken to be equal to the minimum sensitive value of the receiver. It is also shown on Fig.3 that, only a sample of collected data (data collected from the periphery only) is able to calibrate the whole experimental area. This reduces the demanding effort.

$RMSE = \sqrt{\left[RSSmean_k - RSS'mean_k\right]^2}$, where $RSSmean_k = \frac{1}{22} * \sum_{i=1}^{22}\left[RSS_{ki}\right]$, is the mean value of the predicted RSS from the 22 BS antennas at k position, i is the index of the cell antenna, $RMSE$ is the root mean square error, i.e. the difference between predicted and measured values at a k point. $RSS'mean_k$ is the mean value of the measured RSS from the all the cell antennas at k position.

5 Noise cancellation algorithm

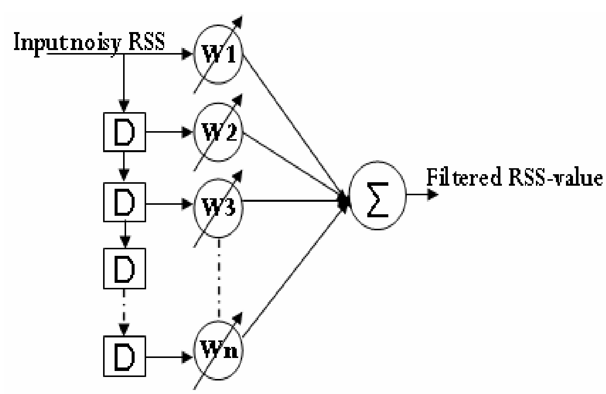

Fig. 4 Noise cancellation with TDNN (Time delay NN)

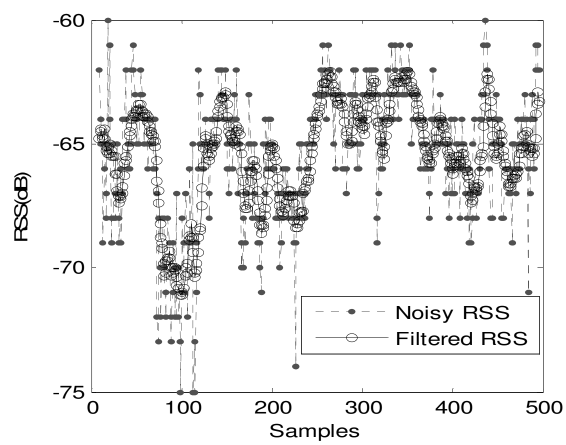

Fig. 5 Noisy and filtered RSS

The noise cancellation block on Fig.2 has a big role. This block solves the problem related to the stochastic behaviour of the received power of the radio signal. The high sensitivity of the RSS to the environmental changes makes it difficult to get the best estimated value. The accuracy of the method based on RSS-positioning depends on how these RSS values are pre-processed before their input to the localization unit. A time delayed NN (a single input adaptive transverse filter) is able to remove a large portion of the noise in the received RSS, see Fig.4 It actually performs this operation by calculating a linear weighted average over a window. In this paper, the window is six time steps long, that means we used 5 delay units (D). This procedure can be performed either online or off line. A string of RSS from a BS is connected to the input of the TDNN, and at the output we have the filtered values. Fig.5 shows us the RSS before and after the noise cancellation unit. The result shows that a large portion of the noise in the received RSS has been removed. This TDNN uses not only the actual RSS to reduce the noise but also the previous ones. The RSS history helps a TDNN to make a best estimation of the values.

6 Positioning with NN

NN training data should be selected to cover the entire region where the network is expected to operate. We use a supervised learning. Our input patterns are the corrected predicted RSS and the correct outputs are their corresponding positions. Selecting a training set is somewhat of an art and somewhat of a trial. We want to keep a training set small so that training is fast, but we also want to exercise the input space well which may require a large training set. Identification of the appropriate NN architecture is of great importance. A large number of hidden layers increases the processing power of the NN but requires

significantly more time for training and a larger number of training examples to train the NN properly. The number of neurons in each layer is determined by the nature of the problem. A systematic analysis of a series of candidate NN architectures was conducted for this RSS-based fingerprint localization.
A multilayer perceptron architecture (MLP) with 22 inputs, 2 hidden layers and 2 outputs was the best candidate. 36 and 32 neurons were applied to the hidden layers respectively, see Fig.6
The training principle is similar to the one on Fig.2 In both cases the function approximation property of the NN is explored.

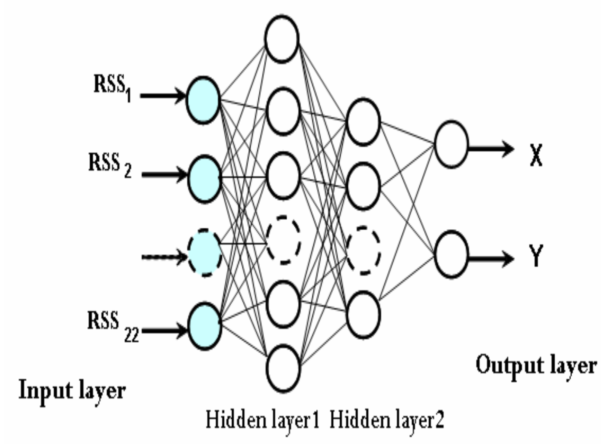

Fig.6 NN architecture used for this positioning issue

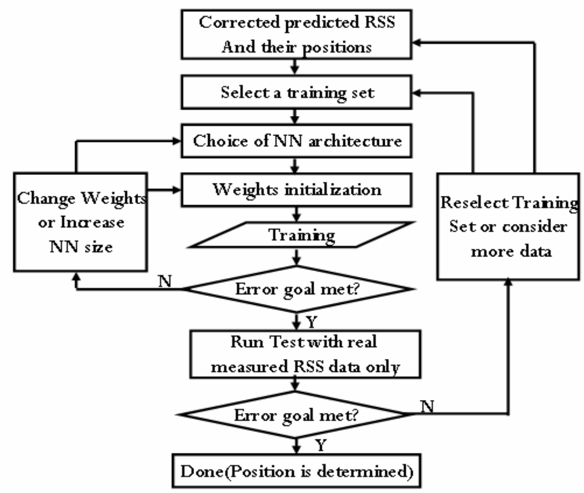

Fig.7 NN training flowchart

A gradient descent algorithm is chosen to adapt the NN weights for its simplicity. The error goal may not be achieved for three reasons:

- A risk of stacking in local minima is present, and in this case is to start over by reinitializing the weights to some new set of small random values.
- The network does not have enough degrees of freedom to fit the desired input/output model. In this case, hidden nodes or layers are added and training is restarted.
- There is no enough information in the training data to perform the desired mapping.

The flowchart on Fig.7 shows the training and test process.

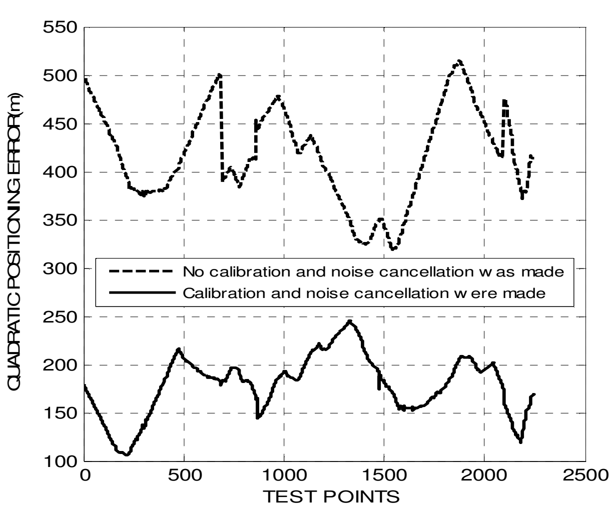

Fig.8 Quadratic positioning error. Comparison of the performances with and without pre-processing.

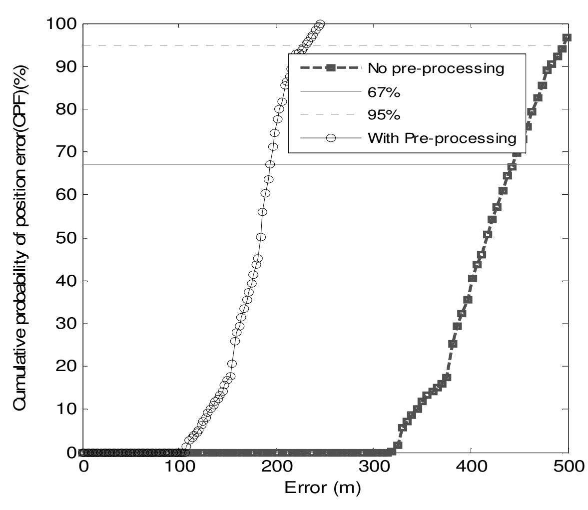

Fig.9 CPF of the positioning error. Comparison of the performances with and without pre-processing

7 Conclusion

This paper presents the results related to the positioning of a mobile station using the signal strength information. It has been shown that, due to the demanding effort during RSS collection from the streets to realize a fingerprint database, predicted RSS of the concerned area can be used for training the NN which will be later on used for positioning purpose. The results on Fig.8,9 show that the difference in prediction and real data causes a big positioning error. The calibration of predicted data by a sample of collected RSS is performed to correct the predicted data. A sample of collected data used to correct the predicted one is first filtered in order to reduce the noise. Results show that with 67% of probability, positioning error is less than 175m instead of 420m when no calibration is performed. And with 95%, the positioning error is less than 220m instead of 490m when pre-processing is not performed. The new results satisfy the FCC requirements.

8 References

[1] R. Yamamoto, H. Matsutani, H. Matsuki, T. Oono, and H. Ohtsuka, "Position Location Technologies using signal strengths in cellular system," presented at VTC, 2001 Spring, Rhodes Island.

[2] C. Nerguizian, C. Despins, and S. Affes, "Indoor Geolocation with received Signal Stregth Fingerprinting Technique and Neural Networks," presented at ICT, Putten, 2004.

[3] Z. Salcic and E. Chan, "Mobile station positioning using GSM cellular phone and artificial Neural networks," *Wireless personal Communications*, vol. 14, pp. 235-254, 2000.

[4] K. R. Anne, K. Kyamakya, F. Erbas, C. Takenga, and J. C. Chedjou, "GSM RSSI-based positioning using extended Kalman filter for training ANN," presented at VTC2004-Fall, Los Angeles, 2004.

[5] C. M. Takenga, K. R. Anne, K. Kyamakya, and J. C. Chedjou, "Comparison of Gradient descent method, Kalman filtering and Decoupled Kalman in training Neural networks used for fingerprint-based positioning," presented at IEEE VTC2004-Fall, Los Angeles, 2004.

[6] T. Kurner and A. Meier, "Prediction of outdoor and outdoor-to-indoor coverage in urban areas at 1.8 GHz," *ieee on selected areas in communications*, vol. 20, 2002.

[7] J. E. Berg, "Building Penetration Loss Along Urban Street Microcells," presented at Cost 231 Final Report, Brussels, 1999.

[8] G. Liang and H. L. Bertoni, "A New Approach to 3-D Ray Tracing for Propagation Prediction in Cities," *IEEE*, vol. 46, pp. 853-863, 1998.

[9] F. Despagne and D. L. Massart, "Neural network in multivariate calibration," in *Analyst*, vol. 123, pp. 157R-178R, 1998.

ACKNOWLEDGMENT

We would like to thank Prof. T.Kurner and his research group from the Institute of communication Technology from Braunschweig Technical University for providing the RSS prediction data of our concerned area. We thank as well the E-Plus mobile network.