Novel Approach for RSS Calibration in DCM -based Mobile Positioning using Propagation Models

Pushpika Wijesinghe∗, Dileeka Dias†
∗Dialog-University of Moratuwa Mobile Communications Research Laboratory
†Department of Electronic & Telecommunication Engineering
University of Moratuwa, Moratuwa, 10400, Sri Lanka
Email:{∗pushpika, †dileeka}@ent.mrt.ac.lk

Abstract—The database correlation method (DCM) is a network based positioning technology which has shown superior in terms of accuracy. DCM is based on a pre-measured database of a location dependent variable such as received signal strength (RSS). Even though the technique has good potential, the practical difficulty in forming the database (fingerprints) using field measurements has become the major challenge in implementing this in large, dynamic networks. A remedy for this is to make use of propagation model predictions instead of field measurements to create the fingerprints. However, due to the considerable deviation between the predictions and the actual measurements, the positioning accuracy diminishes with this approach. In order to overcome this issue, tuning of the predictions using a small number of field measurements can be applied. The work presented in this paper proposes a technique for the correction of such deviations which would improve the performance of DCM. The proposed tuning process, Cell-wise Calibration, is based on artificial neural networks (ANN). Two different training algorithms, particle swarm optimization Algorithm (PSO) and BFGS Algorithm are applied for ANN training. The results of the trials carried out in urban, suburban and rural environments are presented. With the PSO algorithm, the level of accuracy is comparable to that obtained with a measured fingerprint database in urban and suburban environments, and is better in rural environment.

Keywords—Positioning technology, database correlation method, fingerprints, received signal strength, propagation model, calibration, neural networks

I. INTRODUCTION

The estimation of the user’s location from data that is inherently available in the cellular network, which is known as cellular positioning, has become a key technology in mobile communications industry. The focus of cellular positioning covers a wide span, involving a wide range of civilian, commercial, emergency and military applications. Location-based services (LBS) are such applications that utilize knowledge of the physical location of a device to enable a variety of services. From answering the simple question “Where am I?” to providing assistance to a lost hiker, a wide array of applications are possible with precise positioning techniques.

In the United States, people facing a critical emergency situation can request assistance through dialing 911 emergency assistance services. Most of these calls are made from mobile phones with the user being unaware of his or her whereabouts. Hence, the Federal Communication Commission (FCC) has imposed, through E-911 mandate, the mobile operators to precisely locate the callers requesting emergency assistance via 911 [1], [2]. Phase II of this regulation imposes accuracy levels for different location technologies based on their implementation [1]. This was the wakeup call for high enthusiasm in cellular positioning.

A range of cellular positioning technologies have been researched all over the world, yet, none of them have proven superior in terms of accuracy for all environments worldwide [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]. On the other hand, a positioning technology, first proposed in [5] known as database correlation method (DCM), has shown the potential for very good accuracies in different environments [5], [9], [11], [12], [13]. Still, the burden involved in database formation has become a challenging issue preventing the technique being implemented in dynamic networks. Even though a solution of using propagation model predictions to create the database exists, the fact that this approach lowers the accuracy of DCM is another challenge. This paper addresses this issue and proposes a methodology to improve the accuracy of DCM using a predicted database. The work includes investigating the deviation of the propagation model predictions and actual measurements and proposes a technique for correction of such deviations in order to improve the performance of DCM.

Section II is devoted to an overview of the fingerprinting method while Section III presents the authors’ approach for location estimation. Section IV gives the details of the propagation model used in this work. In Section V, a deviation analysis between the predictions and actual measurements is presented. Then, the novel approach for calibration, Cell-wise calibration, is presented in Section VI followed by the description of the environment used for positioning trials in Section VII. Section VIII analyses the positioning results using the proposed calibration method. Finally, the concluding remarks are given in Section IX with possible future directions for further research.

II. OVERVIEW OF FINGERPRINTING METHOD FOR MOBILE POSITIONING

Fingerprinting method, first proposed in [5], is also known as Database Correlation Method as it relies on a pre-measured database of location sensitive parameters. In general, fingerprinting technique involves two phases.
• Database Preparation (Off-line Phase)
• Location Estimation (On-line Phase)

A. Database Preparation

The key idea of DCM is to store the location sensitive parameters seen by the mobile station (MS), within the coverage area, in a database as signal information samples called Fingerprints [5]. A fingerprint consists of signal information seen at a location together with the location coordinate.

Received signal strength (RSS) being selected as the location sensitive parameter, the database consists of RSS fingerprints within the area. A single fingerprint is made up of:

- Cell ID of all hearable cells
- RSS of all hearable cells
- GPS coordinate of the measured location

The original work in [5] has used real measurements for fingerprint formation. The proposed fingerprint collecting process in [5] takes measurements at a very low speed along the selected routes and two measured fingerprints per second are taken on average.

A more feasible approach for collecting database fingerprints is proposed in [13]. The novel method averages the signal strengths of ten consecutive measurements along the route while the median value of ten GPS coordinates is taken as the location coordinate. The averaging is done to compensate variation in signal strength at a given location over time. Further, it proposes a sliding window approach, such that the last five measurements of one fingerprint contribute to the first five measurements of the next fingerprint. This helps to increase the fingerprint resolution.

The work presented in this paper uses signal strength predictions from a propagation model instead of field measurements to create the database. The predictions along the selected roads within the area are used to create fingerprints to develop a predicted database with fingerprints along roads. A predicted fingerprint consists of signal strengths from several hearable cells at a particular location together with the location coordinates. In addition, the cell with the largest predicted strength at the location is selected to be the serving cell and is also stored in the fingerprint.

B. Location Estimation

The location estimation algorithm in the fingerprinting method is a correlation algorithm, which matches the measured signal sample at the location to be estimated, with the fingerprints stored in the database. The output of the algorithm is the location corresponding to the best matching fingerprint. The number of comparisons can be reduced by filtering the fingerprints using other parameters such as serving cell and timing advance [14].

In [5], the fingerprint matching was based on the least mean square approach, which uses (1) to calculate a cost for each fingerprint.

\[ d(k) = \sum_i \left( f_i - g_i(k) \right)^2 + p(k) \]  

(1)

\( f_i \) is the RSS from the \( i^{th} \) cell in the measured sample, \( g_i(k) \) is the RSS from the same cell in the \( k^{th} \) fingerprint. The summation is taken over all the hearable cells that are found in both of the fingerprints. A penalty term \( p(k) \) is introduced for each cell that is found in only one of the samples. The database fingerprint with the lowest value for \( d(k) \) is set to be the best match for the measured sample.

The trial results in urban and suburban areas in Finland show that the above algorithm can provide a positioning error (using a measured database) less than 90m (90%) in urban whereas the error for suburban is 190m (90%) [5].

In addition, different approaches for correlation algorithms have been proposed in [11] and [14] together with their accuracy results. More recent approaches involve the application of neural networks (NN) for fingerprint matching [7], [12], [15], [16].

In this work, the authors propose a novel algorithm for location estimation. It consists of two steps, namely, Fingerprint Filtering and Fingerprint Matching.

III. Fingerprint Filtering and Matching

A. Fingerprint Filtering

A number of fingerprints to be correlated during location estimation can be limited by filtering the database fingerprints using a proper filtering criterion. The idea of fingerprint filtering is to extract a set of potential solutions for the location estimation problem. In addition, this process reduces the number of database comparisons for particular location estimation attempt.

Specifically, the use of serving cell information is much applicable since the serving cell defines a most probable area to locate the MS. However, it was observed that the database with high resolution contains a considerable number of fingerprints having same serving cell. This leads to the selection of a far away fingerprint as the estimated location by removing the closer ones in the matching process. In order to overcome this issue, a further filtering step is proposed to eliminate far-way fingerprints having same serving cell and select only the closest fingerprints to the actual location.

In the novel filtering approach, first the fingerprints are filtered based on the serving cell. Then a score value is calculated for each filtered fingerprint using (2). After that, \( K \)-number of fingerprints having the highest score values are selected as the fingerprints nearest to the location to be estimated.

\[ \text{Score}(i) = \frac{\text{Number of matching cells in measurement \& Fingerprint}}{\text{Max(# of cells in measurement,\# of cells in Fingerprint)}} \]  

(2)

In (2), the number of cells contributing to both fingerprint and the measurement is referred as matching cells. Localization tests were performed for a range of \( K \) values in each environment and the value that minimizes 80% error was taken as the optimum value for \( K \) in each environment. That value for urban environment was found to be equal to 5 while the
values for suburban and rural environments were 3 and 2 respectively.

The proposed filtering method is based on the assumption that the higher the number of matching cells, the closer the fingerprint to the location to be estimated.

B. Fingerprint Matching

The fingerprint matching process finds the best solution out of potential fingerprints extracted in the filtering process. This involves a correlation function, known as the cost function, which calculates a correlation coefficient known as cost for fingerprint matching. The authors propose a novel cost function, which is a modification to the original one proposed in [5], for this purpose. The proposed cost function takes the cell count of the matching cells in both fingerprints and the measured sample and that of the residual cells available in both, in to account as shown in (3).

\[
d(k) = \frac{\sum_{i=1}^{N_1} (f_i - g_i(k))^2}{N_1} + \frac{\sum_{j=1}^{N_2} (f_j - l_{\text{max}})^2}{N_2} + \frac{\sum_{m=1}^{N_3} (l_{\text{max}} - g_m(k))^2}{N_3}
\]

(3)

\(f_i\) is the RSS of the measured sample from the \(i^{th}\) cell, \(g_i(k)\) is the RSS of the \(k^{th}\) fingerprint from the same cell, \(f_j\) is the RSS form the \(j^{th}\) cell present in measurement which is not present in the \(k^{th}\) database fingerprint, \(g_m(k)\) is the RSS from the \(m^{th}\) cell present in the database fingerprint which is not present in the measurement, \(l_{\text{max}}\) is the sensitivity level of the receiver, \(N_1\) is the number of cells present in both measurement and the fingerprint, \(N_2\) is the number of cells present in measurement but not in fingerprint and \(N_3\) is the number of cells present in fingerprint but not in measurement.

IV. WAVE PROPAGATION MODEL

A Wave Propagation Model is a mathematical formulation for characterization of radio wave propagation within the environment, as a function of frequency, distance and other environmental parameters such as, terrain profile, clutter, etc. This work makes use of a propagation model to predict the RSS to be used in predicted database.

The CRC-Predict propagation model, which is a VHF/UHF Propagation Prediction Model used for estimating radio signal strengths on terrestrial paths at VHF and UHF, is used in this work. The model can operate on a topographic database consisting of terrain data to count the effect of the obstruction. The calculation includes diffraction losses due to terrain obstacles (e.g. Hills, trees, buildings, etc.). The diffraction calculation is done by starting at the transmitting antenna and finding the radio field at progressively greater distances. At each step, the field at a point is found by a numerical integration over the field values found in the previous step [17].

The CRC-Predict propagation model tuned to the selected environment by a mobile service provider has been used to obtain the RSS predictions to create the database for fingerprinting method.

V. DEVIATION ANALYSIS

Even though the selected propagation model has been tuned to the selected environment, the predictions still differ from the actual measurements obtained in the same environment. This section analyses the deviations between the predictions and actual measurements in three different environments- urban, suburban and rural.

Here, the deviation is analyzed separately for each cell using the measurements obtained at different locations within the coverage area of the cell and the predictions of the same locations. The root mean square error (RMSE) for each cell is computed using (4).

\[
RMSE_k = \sqrt{\frac{1}{N_k} \sum_{i=1}^{N_k} (RSS_{m}^{i} - RSS_{p}^{i})^2}
\]

(4)

\(RMSE_k\) is the mean square error of \(k^{th}\) cell in dB, \(RSS_{m}^{i}\) is the measured received signal strength of \(k^{th}\) cell at \(i^{th}\) location in dBm, \(RSS_{p}^{i}\) is the predicted received signal strength of \(k^{th}\) cell at \(i^{th}\) position in dBm and \(N_k\) is the total number of test points for the \(k^{th}\) cell.

The results of this analysis are given by Table I, II and III for urban, suburban and rural environments respectively. Accordingly, the deviation between the predictions and actual measurements in urban environment lies in the range 5dB-30dB and the average deviation is 16dB. The deviation in suburban environment is in the range 5dB-20dB with an average of 10dB. Comparatively lower deviation, which lies in the range 3dB-15dB with an average of 7.8dB, was observed in rural environment.

VI. CALIBRATION TECHNIQUE

The deviation analysis showed that there exists a considerable deviation between the propagation model predictions and the actual measurements, which diminishes the accuracy of DCM using a predicted database. Hence, it is worthwhile to investigate techniques to minimize such deviations in order to
Calibration referred to as a small number of measured data is applied and this is an issue. This section describes the authors’ approach to address this issue.

The approach of correcting predicted signal strengths using a small number of measured data is applied and this is referred to as Calibration. Furthermore, the proposed calibration technique corrects the predicted signal strength of each cell separately. Hence we call it Cell-wise Calibration. This requires separately designed calibration techniques for each cell. Apart from that, it is possible to design a global calibration technique for a particular environment such that all the predicted fingerprints could be corrected together (This can be defined as Fingerprint-wise Calibration). However, authors have identified several drawbacks of this approach in practical implementation. The major drawback is the burden involved in upgrading the networks frequently. In large, dynamic networks, new cells are added to the network regularly and the calibration technique should be re-designed as a new cell emerges. In addition, the entire database should be re-calibrated as a new cell is established. Furthermore, complex calibration techniques are required to optimally calibrate the fingerprints for the whole environment. On the other hand, cell-wise calibration requires a lower maintenance work in adding new cells to the network. It is just a matter of calibrating the predictions of new cells separately and inserting the calibrated strengths to the existing database in appropriate locations. In addition, it has the potential for higher accuracies as the predictions are optimized per cell. Due to these facts, cell-wise calibration approach is selected. The designed cell-wise calibration technique is based on artificial neural networks (ANN). ANN is an intellectual abstraction which would enable a computer to work in a similar way to that in which the human brain works [18]. Neural Networks are universal approximators, which can be used to fit any continuous function defined on bounded inputs to a pre-defined arbitrary degree of accuracy. It has the flexibility and the ability to deal with uncertain data. Due to these facts, neural networks are applicable in calibration, which tries to model a relationship between corrupted data and real data.

A. Proposed Neural Network Architecture

The proposed technique is based on the multi-layer perceptron neural network architecture. The two inputs to the neural network are the latitude and longitude values of the location, where the calibration measurements are performed. The output is the error in signal strength loss corresponding to the predicted strength and the measured strength at the same location. The signal strength loss is defined in (5)-(7).

\[
\text{Loss1} = \text{Transmitted Power} - \text{Predicted Strength (dB)} \quad (5)
\]

\[
\text{Loss2} = \text{Transmitted Power} - \text{Measured Strength (dB)} \quad (6)
\]

\[
\text{Error(NN output)} = \text{Loss1} - \text{Loss2} = \text{Measured strength} - \text{Predicted strength (dB)} \quad (7)
\]

Then, the calibrated strength can be derived from (8).

\[
\text{Calibrated Strength} = \text{Predicted Strength} + \text{Error} \quad (8)
\]

The neural network consists of one hidden layer with five neurons and a sigmoid transfer function. The transfer function for the output layer is linear. The performance function was selected to be regularized mean square error, in order to reduce the over fitting during training the network [19]. The complete network consists of 15 parameters that should be trained before applying in calibration process.

In training the neural network, the authors ensured that the number of training examples should be at least five times greater than the number of weights and biases to be trained [19]. Among the numerous training algorithms found in literature, the authors have made use of two sophisticated algorithms called particle swarm optimization (PSO) algorithm and BFGS algorithm for training the neural networks in this work [19], [20].

PSO is a population based optimization algorithm that is motivated from the simulation of the social behavior [20]. If the optimization problem is regarded as a bird swarm looking for food in the sky, then one bird is a particle of PSO algorithm which conducts the search in the solution space. In this regard, the PSO algorithm consists of particles which are flown through the solution space towards the global optimum value. Every particle of the PSO algorithm is one of the solutions, and it adjusts its flying according to its own experience and others. The best position that every particle has experienced during flying (pBest) is the best solution found by itself and the best position that group has experienced (gBest) is the best solution found by the swarm. The fitness value decided by optimization is used to evaluate that the particle is good or bad. Every particle can adjust itself according to pBest and gBest, which makes the particle swarm move to good area. This algorithm has the capability of providing a global best solution. The application of this algorithm for neural network training is described in detail in [21].

On the other hand, the BFGS algorithm is a Quasi Newton algorithm for numerical optimization. Quasi Newton algorithms compute an approximation for Hessian matrix as a function of the gradient. Hence, the computation complexity is lesser than the direct Newton’s method for numerical optimization. BFGS algorithm is the most successful quasi Newton algorithm found so far. Even though it requires more computation in each iteration than simple gradient descent method, it generally converges in less iteration [19].

The neural networks trained for each cell by both algorithms are tested for better accuracies in each environment and the

<table>
<thead>
<tr>
<th>Cell ID</th>
<th>RMSE-(dB)</th>
<th>Cell ID</th>
<th>RMSE-(dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&amp; &amp; &amp; 3</td>
<td>9.29</td>
<td>&amp; &amp; &amp; 5</td>
<td>13.82</td>
</tr>
<tr>
<td>&amp; &amp; &amp; 2</td>
<td>8.04</td>
<td>&amp; &amp; &amp; 6</td>
<td>8.24</td>
</tr>
<tr>
<td>&amp; &amp; &amp; 7</td>
<td>8.14</td>
<td>&amp; &amp; &amp; 9</td>
<td>5.89</td>
</tr>
<tr>
<td>&amp; &amp; &amp; 4</td>
<td>8.51</td>
<td>&amp; &amp; &amp; 1</td>
<td>6.49</td>
</tr>
</tbody>
</table>
TABLE IV
SUMMARY OF FINGERPRINTS AND TEST POINTS

<table>
<thead>
<tr>
<th>Urban</th>
<th>Measured Fingerprints</th>
<th>Predicted/calibrated Fingerprints</th>
<th>Test Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>urban</td>
<td>295</td>
<td>71</td>
<td>63</td>
</tr>
<tr>
<td>Suburban</td>
<td>518</td>
<td>321</td>
<td>312</td>
</tr>
<tr>
<td>Rural</td>
<td>281</td>
<td>170</td>
<td>154</td>
</tr>
</tbody>
</table>

one which gives higher performance is used to calibrate the predictions. Once the predictions are calibrated, those are used to create the calibrated database through a process similar to that discussed in section II-A. Finally, the calibrated database is used for positioning.

VII. THE TEST ENVIRONMENT

The proposed RSS calibration process was tested in three different environments, urban, suburban and rural in Sri Lanka.

A segment of a main highway through the city of Colombo was selected as the urban environment which spans across an area of 2.5 \( km^2 \) and is covered by 70 cells. The selected suburban area spans across 6 \( km^2 \) around the University of Moratuwa. The total number of hearable cells within this area is 45. The rural environment was selected to be around 4 \( km^2 \), with a total of 20 cells hearable within this area.

Table IV summarizes the number of measured fingerprints, predicted/calibrated fingerprints and test points obtained under each environment.

VIII. RESULTS ANALYSIS

This section presents the analysis of the positioning results using the proposed calibration technique in comparison to the results with measured database and predicted database. Furthermore, the results are compared with that obtained from basic Cell_ID method. In all three environments, a superior performance is seen with the proposed algorithms, which is a significant improvement from the Cell_ID method which is commonly used for location-based services. The effect of calibration can also be clearly seen when comparing to the performance with an uncalibrated database.

![Fig. 1. Results Analysis in Urban Environment](image1)

![Fig. 2. Results Analysis in Suburban Environment](image2)

![Fig. 3. Results Analysis in Rural Environment](image3)

Fig. 1. Results Analysis in Urban Environment

Fig. 2. Results Analysis in Suburban Environment

Fig. 3. Results Analysis in Rural Environment

trained by PSO algorithm shows better performance than the one trained by BFGS algorithm in urban environment. The results after calibration are comparable with the results using a measured database and are better than the results using a predicted database without calibration in this environment. The positioning accuracy obtained is 125m (80%) and 180m (90%).

The results shown in Fig. 2 show the success of the PSO algorithm as the training algorithm for the proposed calibration method in the suburban environment. The error curve for the calibrated database with the PSO algorithm takes over the curve for the measured database at higher percentages. The best positioning accuracy is 550m (80%) and 700m in (90%) of the estimates.

A remarkable improvement in the performance can be seen in the rural environment (Fig. 3) after calibration with the neural network trained by PSO algorithm. The neural network trained by BFGS algorithm has a poor performance in rural environment as well. The performance after calibration is higher than that before calibration and is even better compared to the performance using a measured database in rural environment. The best positioning accuracy is 400m (80%) and 500m (90%) in this environment.

IX. CONCLUSION

The paper presents the results of the work carried out to improve the performance of DCM using predicted data instead of measured data. The paper proposes a novel technique, defined as cell-wise calibration, to correct the deviations between
the predicted signal strength and measured signal strength. The proposed technique is based on Neural Networks. Two algorithms, namely, the PSO and the BFGS, were tested for training the neural networks for calibration in three different environments, urban, suburban and rural. The results show that PSO algorithm is the best choice as the training algorithm in designing calibration techniques for all three environments. This is due to the fact that PSO algorithm has the ability of avoiding local maxima/minima and providing a global best solution. Furthermore, the calibration has been able to improve the performance of DCM over that obtained using predicted databases (without calibration) in all three environments. A remarkable performance improvement is seen in rural environment after calibration. Further, the performance of the proposed positioning technique is far superior to the simple cell-id based location technique which is commonly employed to provide location-based services. Though the proposed positioning technique does not have sufficient accuracy for E-911 service, it is applicable for several types of location-based services.

Possible directions for future work include, further testing of the technique in several parts of the country and refining the neural network topology and applying different training algorithms for improved accuracy.

ACKNOWLEDGMENT

The authors are thankful to Dialog Telekom PLC, Sri Lanka, for the support of this work.

REFERENCES