

Classificação *a priori* da Acurácia na Localização de Terminais Móveis Celulares

Rafael Saraiva Campos e Lisandro Lovisolo

Resumo— Neste trabalho apresenta-se um procedimento baseado em aprendizado de máquina para classificação *a priori* da acurácia de estimativas de localização fornecidas por um método de correlação de assinaturas de rádio-frequência em redes de telefonia móvel celular, o RF-FING+RTD-PRED, anteriormente formulado pelos autores. O procedimento proposto utiliza os mesmos parâmetros de rádio-frequência empregados pelo método RF-FING+RTD-PRED, acrescidos de algumas informações, como as distâncias entre a estação servidora e as estações vizinhas listadas no relatório de medidas enviado pelo terminal móvel a ser localizado. A abordagem proposta tem duas fases: redução de dimensão do vetor de entrada utilizando análise de componentes principais, e classificação do erro de localização em três segmentos de acurácia (alta, média e baixa) utilizando uma rede neural artificial *backpropagation*. A identificação *a priori* de estimativas de posição de alta acurácia (erro inferior a 100 metros) é de particular relevância para aplicações de localização críticas, como o rastreamento de veículos e a localização de chamadas de emergência. Em testes conduzidos em redes GSM 850 MHz em áreas urbanas e suburbanas, onde foram coletados 6600 relatórios de medida, o procedimento de classificação proposto atingiu precisão de 89% na identificação *a priori* de estimativas de posição de alta acurácia.

Palavras-Chave— Assinatura de Rádio-Frequência, Acurácia da Localização, Análise de Componentes Principais, Redes Neurais Artificiais

Abstract— This work presents a machine learning based procedure to evaluate the accuracy of position estimates produced by a fingerprinting location method in cellular networks, RF-FING+RTD-PRED, previously formulated by the authors. The proposed procedure employs the same radiofrequency parameters used by RF-FING+RTD-PRED plus some additional data, like the distances between the serving and neighbor base stations. The proposed approach has two phases: principal components analysis to reduce the input vector dimension, and classification of the positioning estimate accuracy (high, medium and low accuracy) by a backpropagation artificial neural network. The *a priori* identification of high accuracy position estimates, i.e., with error below 100 meters, is particularly relevant for critical location based applications, like vehicle tracking and emergency call positioning. In field tests carried out in GSM 850 MHz networks in urban and suburban areas, where 6600 measurement reports have been collected, the proposed classification procedure achieved 89% precision in the identification of high accuracy position estimates.

Keywords— Radio-Frequency Fingerprint, Positioning Accuracy, Principal Components Analysis, Artificial Neural Networks.

I. INTRODUÇÃO

A localização de estações móveis (MSs - *Mobile Stations*) em redes de telefonia móvel celular tem sido alvo de crescente interesse das operadoras, dos usuários e de órgãos reguladores de telecomunicações. Há uma grande variedade de serviços de localização (LCS - *Location Services*) disponíveis atualmente, que de modo geral podem ser agrupados nas seguintes categorias [1]: localização de chamadas de emergência, tarifação baseada em posição, rastreamento de pessoas e veículos, serviços de informação baseados em posição, entretenimento. Dentre estes serviços, a localização de chamadas para números de emergência - polícia, bombeiros, defesa civil - constitui a aplicação mais crítica e, em vários países, órgãos reguladores especificaram requisitos para a localização de MSs [2][3] originando chamadas para números de emergência.

A universalidade desejável no serviço de localização de chamadas de emergência, i.e., a possibilidade de localizar qualquer MS dentro da área de cobertura da rede de acesso rádio (RAN - *Radio Access Network*), torna importante aprimorar métodos que permitam a localização de MSs legados, ou seja, sem qualquer suporte específico para LCS. Isto é particularmente relevante ao verificar-se que, dos cerca de 4 bilhões de celulares em uso no mundo atualmente [4], apenas 163 milhões têm receptor GPS (*Global Positioning System*) [5], ou seja, aproximadamente 4% do total de aparelhos. Este percentual médio inclui os países desenvolvidos, logo, no Brasil, a penetração de terminais com GPS provavelmente é muito inferior a este valor. Além disso, os métodos compatíveis com terminais legados podem atuar como *fallback*, quando métodos mais precisos estiverem indisponíveis devido a limitações sistêmicas - por exemplo, em ambientes *indoor* ou em áreas urbanas densas, onde a recepção de sinais GPS é muitas vezes inviável. Métodos de localização com suporte a terminais legados são também denominados métodos baseados na rede, pois utilizam apenas informações já disponíveis na rede ou nos relatórios de medidas (NMR - *Network Measurement Report*), periodicamente enviados pelos MSs em modo ativo.

Para serviços LCS críticos, estimativas de posição de alta acurácia, i.e., com erro inferior a 100 metros, são desejáveis [6]. Contudo, como o erro de uma estimativa de posição não é conhecido *a priori*, deve-se dispor de um procedimento para que, dado um vetor de parâmetros de rádio-frequência (RF) medidos pelo MS, seja possível classificar a acurácia da localização. Com tal procedimento seria possível identificar e selecionar apenas as estimativas de alta acurácia, resultando em maior confiabilidade para os serviços LCS críticos.

Rafael Saraiva Campos, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Rio de Janeiro, Brasil. E-mail: rafael.saraiva@ig.com.br

Lisandro Lovisolo, Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil. E-mail: lisandro@uerj.br

Para atingir os requisitos de universalidade e de seleção *a priori* de estimativas de posição de alta acurácia, propõe-se um procedimento dividido em duas etapas e que recebe como dados de entrada os parâmetros de RF contidos em um NMR, além de algumas informações adicionais descritas na Seção III. Na primeira fase, realiza-se análise de componentes principais (PCA - *Principal Components Analysis*) para redução de dimensão do vetor de entrada. Na segunda fase, utiliza-se uma rede neural artificial (ANN - *Artificial Neural Network*) *backpropagation* para classificação *a priori* da acurácia da estimativa de posição que seria produzida por um método de localização específico, ao receber como entrada o mesmo NMR. Com esta implementação, a resposta a uma solicitação de cálculo de posição é composta pela 3-upla $(\hat{x}, \hat{y}, \hat{e})$, onde (\hat{x}, \hat{y}) são as coordenadas estimadas do MS e \hat{e} é a classe de acurácia estimada (alta, média ou baixa acurácia). O método de localização selecionado é o **RF-FING+RTD-PRED** (*Predicted Radiofrequency Fingerprint with Round Trip Delay*), originalmente proposto pelos autores em [7].

O algoritmo 1 sintetiza os passos do procedimento proposto. O **RF-FING+RTD-PRED** e o método de classificação *a priori* da acurácia das estimativas de posição por ele retornadas são executados no servidor de localização (SMLC - *Serving Mobile Location Center*). O SMLC recebe da RAN o NMR e um ou mais valores de RTD (*Round Trip Delay*) e consulta o BSA (*Base Station Almanac*), que é uma tabela com parâmetros das estações rádio-base (ERBs). Utilizando as informações recebidas da RAN e obtidas no BSA, o SMLC monta a assinatura de RF (\mathbf{F}_i) e o vetor de entrada do classificador (\mathbf{S}_i). O objetivo é selecionar estimativas de posição de alta acurácia que serão fornecidas a aplicações LCS críticas. Para tal, há um laço, que é repetido até que uma estimativa classificada como sendo de alta acurácia seja obtida, ou até que um número máximo de repetições (N_{max}) seja atingido. Neste exemplo, a classe de alta acurácia é identificada pelo vetor coluna $[1 \ 1]^T$.

Algoritmo 1 Classificação *a priori* da acurácia de localização

```

i ← 0
f ← 1
while f = 1 do
    SMLC recebe NMRi e valor(es) de RTD
    SMLC consulta BSA
    SMLC produz  $\mathbf{F}_i$  e  $\mathbf{S}_i$ 
    RF-FING+RTD-PRED produz  $(\hat{x}, \hat{y})$ 
    Classificador produz  $\hat{e}$ 
    if  $\hat{e} = [1 \ 1]^T$  or  $i = N_{max}$  then
        SMLC retorna  $(\hat{x}, \hat{y}, \hat{e})$ 
        f ← 0
    else
        i ← i + 1
    end if
end while

```

II. O MÉTODO RF-FING+RTD-PRED

O método de localização **RF-FING+RTD-PRED** [7] é um método de correlação de assinaturas de RF (DCM - *Database Correlation Method*) [8]. Uma assinatura de RF é um conjunto de parâmetros de sinal dependentes da posição do MS. Estes

parâmetros são medidos pelo MS - e informados à RAN através do NMR - ou pelos setores âncora. A estrutura da assinatura de RF utilizada pelo método **RF-FING+RTD-PRED** é dada pela matriz $N_a \times 3$:

$$\mathbf{F} = \begin{bmatrix} CI_1 & RSS_1 & RTD_1 \\ \vdots & \vdots & \vdots \\ CI_{N_a} & RSS_{N_a} & RTD_{N_a} \end{bmatrix} \quad (1)$$

onde N_a é o número de setores âncora, CI_i e RSS_i são a identidade (CI - *Cell Identity*) e o nível de sinal recebido (RSS - *Received Signal Strength*) do i -ésimo setor âncora. RTD_i é o tempo de propagação de ida e volta entre o MS e o i -ésimo setor âncora. As linhas são classificadas em ordem decrescente de RSS. Os passos de quantização e as faixas dinâmicas de RSS e RTD, o número máximo de setores âncora e de valores de RTD disponíveis são específicos de cada tipo de rede sem fio. Em redes GSM, por exemplo, o valor de RTD está disponível apenas para o setor melhor servidor [9], que usualmente é o setor âncora com o maior valor de RSS. Os setores âncora são aqueles cujas CIs estão no NMR.

Por ser um método DCM, o **RF-FING+RTD-PRED** estima a posição do MS através da correlação de uma assinatura de RF medida pelo MS (assinatura alvo) com assinaturas de RF armazenadas em um banco de dados de correlação (assinaturas de referência). Cada assinatura de referência está associada a um par de coordenadas geográficas. O banco de dados de correlação (CDB - *Correlation Database*) pode ser gerado a partir de medidas ou de modelagem de rádio-propagação [10].

A comparação da assinatura alvo com todas as assinaturas de referência tornaria o tempo para produzir uma estimativa de posição excessivamente longo. Assim, o método **RF-FING+RTD-PRED** avalia apenas um subconjunto do CDB, denominado espaço de correlação. O espaço de correlação é obtido através de um processo de filtragem em quatro etapas que são sucessivamente aplicadas sobre as assinaturas de referência [7].

Após a filtragem, calcula-se a distância euclidiana no espaço N dimensional de RSS entre a assinatura alvo e cada assinatura de referência. A posição estimada do MS é dada pela média aritmética das coordenadas geográficas das K assinaturas de referência mais próximas da assinatura alvo. Os parâmetros N e K podem ter forte impacto na precisão do método **RF-FING+RTD-PRED** e devem ser cuidadosamente selecionados [11].

III. VISÃO GERAL DO ALGORITMO DE CLASSIFICAÇÃO DE ACURÁCIA

O algoritmo proposto divide-se nas etapas: **i)** construção do vetor de entrada a partir do NMR; **ii)** redução da dimensão do vetor de entrada através de PCA; **iii)** classificação da acurácia pela ANN. ANNs têm sido utilizadas para aproximar a função que mapeia uma assinatura de RF em uma localização geográfica em redes de telefonia móvel celular [12][13][14]. Em [15], uma ANN é utilizada para classificar a morfologia do ambiente de teste (urbana, suburbana ou rural), com o objetivo de aumentar a acurácia da localização por multi-lateração circular. Contudo, na literatura especializada pesquisada, não foi encontrada utilização da ANN igual à proposta neste

trabalho, como classificador *a priori* da acurácia de estimativas de posição.

A. Vetor de Entrada do Classificador

O vetor de entrada do classificador é definido pela matriz coluna $\mathbf{V} = [v_i]_{i=1,\dots,13}$, onde:

- v_1 é o valor de RTD do setor melhor servidor;
- v_2 é o valor de RSS do setor melhor servidor;
- v_3 a v_8 são os valores de RSS de 6 setores vizinhos, dispostos em ordem decrescente;
- v_9 a v_{11} são as distâncias (km) entre o setor melhor servidor e os 3 setores vizinhos com maiores valores de RSS;
- v_{12} é o número de ERBs vizinhas;
- v_{13} é o número de setores vizinhos, i.e., a quantidade total de setores nas ERBs vizinhas ao setor melhor servidor.

Os setores vizinhos são os setores âncora (e portanto listados no NMR), excluindo-se o setor melhor servidor. A ERB que contém o setor melhor servidor é denominada ERB servidora. As ERBs que contém um ou mais setores vizinhos são ditas ERBs vizinhas.

O parâmetro v_1 é calculado pela ERB servidora para cada MS em modo ativo. Os parâmetros v_2 a v_8 são informados no NMR. Os parâmetros v_9 a v_{13} são calculados cruzando as informações contidas no NMR com a configuração da RAN (coordenadas das antenas, identificação dos setores, etc.). Os parâmetros v_9 a v_{13} permitem estimar a densidade superficial de setores σ na vizinhança do MS. Logo, são essenciais para estimar a acurácia das estimativas de posição fornecidas pelo método **RF-FING+RTD-PRED**, uma vez que a acurácia do mesmo depende fortemente do valor de σ [11].

B. Redução de Dimensão do Vetor de Entrada

O PCA é um método para redução de dimensionalidade de conjuntos de dados com múltiplas variáveis. Este método gera um novo conjunto de variáveis mutuamente ortogonais chamadas componentes principais (PCs - *Principal Components*). Cada PC é uma combinação linear das variáveis originais [16].

O conjunto de dados coletados nos testes de campo foi dividido em três subconjuntos: treinamento, teste e validação. Inicialmente, o subconjunto de treinamento, com n vetores de entrada, sofre uma translação através da extração da média amostral em cada dimensão, obtendo-se então a matriz:

$$\mathbf{B} = [v_{i,j} - \bar{V}_i]_{i=1,\dots,13;j=1,\dots,n} \quad (2)$$

onde $v_{i,j}$ é o valor da i -ésima variável no j -ésimo vetor de entrada, e \bar{V}_i é a média amostral da i -ésima variável. Seja \mathbf{U}_i o i -ésimo autovetor da matriz de covariância de \mathbf{B}^T . Note que, se $i < j$, então $\lambda_i > \lambda_j$, onde λ_i é o autovalor associado a \mathbf{U}_i . A matriz de PCs do subconjunto de treinamento é dada por:

$$\mathbf{P}_{13 \times n} = \left(\mathbf{B}^T \cdot [\mathbf{U}_i]_{i=1,\dots,13} \right)^T \quad (3)$$

As linhas de \mathbf{P} são dispostas em ordem decrescente de variância, e cada linha de \mathbf{P} corresponde a uma PC. Para reduzir a dimensão dos padrões de treinamento que serão fornecidos à ANN, apenas as p primeiras PCs são mantidas. Neste

trabalho, o valor de p foi selecionado de forma a preservar, no mínimo, 99% da variância total do conjunto de treinamento. A utilização de padrões de entrada de menor dimensão permite simplificar a topologia da rede (o que também auxilia na prevenção do *overtraining*), acelera o treinamento e aumenta a precisão das classificações fornecidas pela ANN [17].

As médias amostrais e a matriz $[\mathbf{U}_i]_{i=1,\dots,13}$, obtidas durante a aplicação do PCA ao conjunto de treinamento, são utilizadas para projetar cada vetor dos conjuntos de teste e validação no subespaço das componentes principais. Assim como no conjunto de treinamento, apenas as p primeiras PCs são mantidas.

C. Classificação da Acurácia pela ANN

Os valores dos vetores de entrada, após a aplicação do PCA e subsequente redução de dimensão, são linearmente normalizados para o intervalo $[-1, +1]$ antes de serem fornecidos à ANN. Seja a 4-upla $(\mathbf{S}_i, \mathbf{F}_i, x_i, y_i)$, onde \mathbf{S}_i é o i -ésimo padrão de entrada, \mathbf{F}_i é a i -ésima assinatura de RF e a partir da qual \mathbf{S}_i foi obtido, x_i e y_i são as coordenadas de referência do MS fornecidas pelo GPS. O alvo do i -ésimo padrão de entrada, durante o treinamento supervisionado da ANN, é a classe de acurácia (alta, média ou baixa) da estimativa de posição fornecida pelo método **RF-FING+RTD-PRED** para \mathbf{F}_i . A classe de acurácia é obtida após discretização do erro de localização, que é a distância euclidiana entre as coordenadas (\hat{x}_i, \hat{y}_i) , estimadas pelo método **RF-FING+RTD-PRED**, e as coordenadas de referência (x_i, y_i) . Para utilizar a geometria euclidiana, é necessário que as coordenadas geográficas sejam representadas através uma projeção cartográfica retangular, ou seja, utilizando um sistema de coordenadas cartesianas. Assim, neste trabalho, todas as coordenadas geográficas são representadas no sistema UTM (*Universal Transverse Mercator*) [18], que utiliza a projeção cartográfica transversa de Mercator.

A ANN tem p entradas, onde $0 < p < 13$, q neurônios na camada escondida, e 2 neurônios na camada de saída. A função de ativação dos neurônios da camada escondida é a tangente hiperbólica. A função de ativação dos neurônios da camada de saída é a função sigmóide, dada por $f(n) = 1/(1 + e^{-n})$. O método de treinamento é o de *Levenberg-Marquardt* [19], tendo como função de desempenho o erro quadrático médio (MSE - *Mean Square Error*). A saída da ANN expressa a classe de acurácia segundo um código binário: $[1 \ 1]^T$ (alta acurácia), $[1 \ 0]^T$ (média acurácia) e $[0 \ 0]^T$ (baixa acurácia). A especificação dos limites numéricos de cada classe é feita na Seção IV-A.

IV. AVALIAÇÃO EXPERIMENTAL

A. Campanha de Coleta de Medidas

Os dados experimentais foram obtidos em testes realizados em redes GSM na banda de 850 MHz nos bairros Centro (Região I) e Santa Cruz (Região II) do município do Rio de Janeiro. A Região I tem morfologia urbana densa, com elevada densidade de setores e predominância de condições de propagação sem linha de visada. A Região II tem morfologia suburbana, com baixa densidade de setores e predominância de condições de propagação em rádio-visibilidade.

O conjunto de teste foi constituído de um telefone GSM legado modelo SAGEM OT298 e um receptor GPS modelo GARMIN, ambos conectados a um *laptop* no interior de um veículo, em cujo topo foi instalada a antena receptora GPS. Em cada região de teste, foi originada uma chamada e o MS foi mantido em modo dedicado durante a coleta. Para cada NMR transmitido o valor de RTD foi identificado e a posição de referência do MS foi calculada pelo GPS. Assim, para cada NMR, o erro de posicionamento é a distância Euclidiana entre a posição fornecida pelo GPS e a posição estimada pelo método **RF-FING+TA-PRED**. Todas as mensagens trocadas entre o MS e a ERB servidora foram capturadas pelo *software* TEMS *Investigation* da *Ericsson*. As características das rotas de coleta estão resumidas na Tabela I.

TABELA I
ROTAS DE COLETA EM REDES GSM.

Região	I	II
Quantidade de NMRs	4501	7864
Área (km ²)	4,84	100
Qtde de Setores Detectados	114	23
Densidade de Setores (set./km ²)	23,5	0,23

Os padrões de entrada foram construídos a partir de assinaturas de RF que, quando submetidas ao método **RF-FING+TA-PRED**, resultaram em estimativas de posição com erro inferior a 100 metros (alta acurácia), entre 300 e 800 metros (média acurácia) e superior a 1000 metros (baixa acurácia). As classes foram definidas de modo que as faixas de erro de localização não fossem adjacentes. Esta definição foi adotada pois os limites entre classes justapostas são nebulosos, ou seja, uma estimativa de posição cuja acurácia esteja na fronteira entre duas classes adjacentes pertence simultaneamente, com diferentes graus de pertinência, a ambas as classes. Estes graus de pertinência, contudo, podem ser bastante semelhantes, tornando difícil a classificação. Além disso, o erro na identificação da classe de acurácia de uma estimativa de posição, quando são utilizadas classes não justapostas, é mais sério, sob o ponto de vista da aplicação LCS crítica que requer estimativas de alta acurácia. Por exemplo, é menos grave identificar uma estimativa de posição com 150 metros de erro de localização como sendo de alta acurácia, do que cometer este mesmo erro de classificação com uma estimativa de posição com 300 metros de erro de localização.

Foram aleatoriamente selecionados 2200 padrões por classe de acurácia, totalizando 6600 padrões, dos quais 73% foram utilizados para o treinamento supervisionado da ANN, 4% para validação e 23% para o teste. Ao aplicar o PCA sobre o conjunto de treinamento, verificou-se que as 10 primeiras PCs acumulavam 99,77% da variância total. Assim, cada padrão de entrada nos conjuntos de treinamento, validação e teste é uma matriz coluna com 10 elementos.

B. Configuração do Método RF-FING+TA-PRED

As áreas de melhor servidor de cada setor - que são aplicadas no primeiro passo de filtragem para redução do espaço de correlação - e os CDBs utilizados pelo método **RF-FING+TA-PRED** foram construídos a partir de predição

de propagação. Para tal, o modelo de Okumura-Hata [20] foi aplicado sobre mapas digitalizados do relevo (DEM - *Digital Elevation Models*) das regiões de teste, utilizando os parâmetros de RF da RAN (coordenadas, azimute e diagrama de radiação das antenas, potência de transmissão, etc.). Os DEMs são matrizes, onde cada elemento contém a altura média do relevo sobre uma superfície de $r \times r$ m². O parâmetro r é a resolução planar do DEM. As resoluções planares dos DEMs representando as Regiões I e II são 5 e 25 metros, respectivamente. Os valores dos coeficientes do modelo de Okumura-Hata utilizados nas predições de propagação nas Regiões I (área urbana) e II (área suburbana) foram ajustados a partir de rotas de calibração e são especificados em [7]

C. Treinamento e Topologia da ANN

O treinamento da ANN foi realizado no MATLAB. Na topologia final da ANN, a camada escondida tem $q = 8$ neurônios. O tamanho ótimo da camada escondida foi encontrado aumentando o valor de q durante o treinamento, até que nenhuma redução perceptível fosse verificada no MSE. Para prevenir o *overtraining* [21], os *biases* das funções de ativação dos neurônios e os pesos das sinapses que minimizam o MSE do conjunto de validação foram mantidos ao final do treinamento. O valor de MSE mínimo para o conjunto de validação foi alcançado após 17 épocas. A partir daí, o MSE do conjunto de validação cresce continuamente por 6 épocas consecutivas, provocando a interrupção do treinamento.

D. Análise de Resultados

A matriz de confusão é uma ferramenta de visualização tipicamente usada para avaliar o desempenho de classificadores treinados com aprendizado supervisionado [22]. A matriz de confusão obtida com o classificador proposto é representada na Tabela II, onde cada coluna representa as instâncias de uma classe predita e cada linha representa as instâncias de uma classe real.

TABELA II
MATRIZ DE CONFUSÃO DO CLASSIFICADOR.

Classe Real	Classe Predita		
	Alta	Média	Baixa
Alta	376	11	0
Média	48	343	124
Baixa	0	80	518

Seja a matriz de confusão dada por $\mathbf{M} = [m_{i,j}]_{i,j=1,\dots,N_c}$, onde N_c é o número de classes. A partir desta matriz, os seguintes contadores podem ser definidos para cada classe: **i)** Verdadeiros Positivos (VP), que é a quantidade de instâncias pertencentes a uma determinada classe e que foram corretamente identificadas como tal; para a k -ésima classe, tem-se $VP_k = m_{k,k}$; **ii)** Verdadeiros Negativos (VN), que é a quantidade de instâncias não pertencentes a uma determinada classe e que foram corretamente identificadas como tal; para a k -ésima classe, tem-se $VN_k = \sum m_{i,j}$, para $i, j = 1, \dots, N_c$ e $i, j \neq k$; **iii)** Falsos Positivos (FP), que é a quantidade de instâncias não pertencentes a uma determinada classe e que foram incorretamente identificadas como pertencentes a ela; para

a k -ésima classe, tem-se $FP_k = \sum m_{k,j}$, para $j = 1, \dots, N_c$ e $j \neq k$; **iv**) Falsos Negativos (FN), que é a quantidade de instâncias pertencentes a uma determinada classe e que foram incorretamente identificadas como não pertencentes a ela; para a k -ésima classe, tem-se $FN_k = \sum m_{i,k}$, para $i = 1, \dots, N_c$ e $i \neq k$.

Três métricas, construídas a partir destes contadores, foram selecionadas para avaliação do desempenho do classificador: a *acurácia*, que expressa o percentual de classificações corretas para uma dada classe; a *precisão*, que expressa o percentual de classificações positivas corretas para uma dada classe; e a *especificidade*, que expressa o percentual de classificações negativas corretas para uma dada classe. A acurácia, precisão e especificidade da k -ésima classe são dadas pelas equações (4), (5) e (6), respectivamente. Os valores de VP, VN, FP, FN, acurácia, precisão e especificidade obtidos pelo classificador são mostrados na Tabela III.

$$\text{Acurácia}_k = \frac{VP_k + VN_k}{VP_k + VN_k + FP_k + FN_k} \quad (4)$$

$$\text{Precisão}_k = \frac{VP_k}{VP_k + FP_k} \quad (5)$$

$$\text{Especificidade}_k = \frac{VN_k}{VN_k + FN_k} \quad (6)$$

TABELA III
DESEMPENHO DO CLASSIFICADOR.

	Classe Predita		
	Alta	Média	Baixa
VP	376	343	518
FP	48	91	124
FN	11	172	80
VN	1065	894	778
Acurácia	96%	82%	86%
Precisão	89%	79%	81%
Especificidade	99%	84%	91%

A Tabela III indica que o classificador proposto atinge valores elevados de acurácia, precisão e especificidade para as três classes consideradas (alta, média e baixa acurácia de localização). Contudo, no contexto deste trabalho, cujo objetivo primordial é a identificação e seleção de estimativas de posição de alta acurácia que serão fornecidas a aplicações LCS críticas, a métrica mais relevante é a precisão da classe de alta acurácia. Esta métrica indica que 89% das estimativas de posição classificadas como sendo de alta acurácia, de fato o são. Além disso, a primeira coluna da matriz de confusão na Tabela II mostra que nenhuma estimativa de baixa acurácia foi erroneamente classificada como sendo de alta acurácia. Tal erro, se verificado, seria o mais grave, do ponto de vista de aplicações como o rastreamento de veículos e, principalmente, a localização de chamadas de emergência, onde a acurácia das estimativas de posição deve ser a máxima possível [2] [3] [6].

V. CONCLUSÃO

Este trabalho apresentou um procedimento para classificação *a priori* da acurácia de estimativas de posição de MSs produzidas pelo método **RF-FING+RTD-PRED**, que

é um método de correlação de assinaturas de RF baseado na rede. Quando testado em cenários reais, o procedimento apresentado atingiu alto grau de confiabilidade, em especial na identificação das estimativas de posição pertencentes à classe de alta acurácia. Este bom desempenho sugere que o procedimento proposto pode ser empregado para selecionar estimativas de posição que serão utilizadas por aplicações LCS críticas, como a localização de chamadas de emergência.

REFERÊNCIAS

- [1] 3GPP TS 22071, v9.0.0 (2009-12), *Location Services (LCS); Service Description; Stage 1 (Release 5)*. 3GPP, 2009.
- [2] FCC, *FCC Wireless Enhanced E911 Requirements*. USA, 1999.
- [3] European Commission, "Commission Recommendation on the processing of caller location information in electronic communication networks for the purpose of location-enhanced emergency call services 2003/558/EC", *Official Journal of the European Union*, Julho 2003.
- [4] INFO Online, <http://info.abril.com.br/aberto/infonews/012009/06012009-23.shl>. Acesso em 20 Jan. 2010.
- [5] TI Inside Online, <http://www.tiinside.com.br/21/01/2010/nokia-da-passo-para-popularizar-gps-no-celular/ti/163446/news.aspx>. Acesso em 20 Jan. 2010.
- [6] M. I. Silventoinen and T. Rantalainen, "Mobile Station Emergency Locating in GSM", *Proc. of IEEE International Conference on Personal Wireless Communications*, p.232-238, New Delhi, India, 1996.
- [7] R. S. Campos e L. Lovisolo, "Análise Empírica de Métodos de Localização de Terminais GSM Legados utilizando Predição de Cobertura e Alinhamento Temporal", *Anais do XXVI Simpósio Brasileiro de Telecomunicações*, Rio de Janeiro, Brasil, 2008.
- [8] H. Laitinen and J. Lahteenmaki and T. Nordstrom, "Database Correlation Method for GSM Location", *Proc. of IEEE 53rd Vehicular Technology Conference*, p. 2504-2508, Rhodes, Greece, 2001.
- [9] ETSI TS 100 911 V6.2.0 (1998-07), *Digital cellular telecommunications system (Phase 2+); Radio subsystem link control (GSM 05.08 version 6.2.0 Release 1997)*. ETSI, 1998.
- [10] D. Zimmermann et al., "Database Correlation for Positioning of Mobile Terminals in Cellular Networks using Wave Propagation Models", *Proc. of IEEE 60th Vehicular Technology Conference*, p.4682-4686, Los Angeles, USA, 2004.
- [11] R. S. Campos e L. Lovisolo, "Location Methods for Legacy GSM Handsets using Coverage Prediction", *Proc. of 9th IEEE Workshop on Signal Processing Advances in Wireless Communications*, p.21-25, Recife, Brazil, 2008.
- [12] Z. Salcic, "GSM Mobile Station Location Using Reference Stations and Artificial Neural Networks", *Wireless Personal Communications*, no. 19, 2001.
- [13] C. Takenga et al., "A Hybrid Neural Network-Data Base Correlation Positioning in GSM Network", *Proc. of IEEE 10th International Conference on Communication Systems*, p.1-5, Singapore, 2006.
- [14] K. Rao Anne et al., "GSM RSSI-based positioning using Extended Kalman Filter for training Artificial Neural Networks", *Proc. of IEEE 60th Vehicular Technology Conference*, Los Angeles, USA, 2004.
- [15] O. Turkyilmaz, F. Alagoz, G. Gur and T. Tugcu, "Environment Aware Location Estimation in Cellular Networks", *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 276456, 9 pages, 2008.
- [16] I. T. Jolliffe, "Principal Component Analysis", 2nd Edition, Springer, 2002.
- [17] C. Lee and C. Lin, "Automatic Search of the Power Ascension Path for a Boiling Water Reactor Using Genetic Algorithm and Neural Network", *Nuclear Technology*, vol. 159, no. 3, p.256-266, September 2007.
- [18] A. Libault, "Geocartografia", Ed. da Universidade de São Paulo, 1975.
- [19] J. J. Moré,, "The Levenberg-Marquardt Algorithm: Implementation and Theory", *Lecture Notes in Mathematics*, no. 630, p.105-116, 1977.
- [20] M. Hata, "Empirical Formula for Propagation Loss in Land Mobile Radio Services", *IEEE Transactions on Vehicular Technology*, vol. 29, no. 3, September 1980.
- [21] S. Haykin, "Neural Networks: A Comprehensive Foundation", Prentice Hall PTR, 1994.
- [22] F. Provost et. al., "The Case Against Accuracy Estimation for Comparing Classifiers", *Proc. of 15th International Conference on Machine Learning (ICML-98)*, Madison, USA, 1998.