

2

Fundamentos

Neste capítulo apresentaremos alguns conceitos básicos referentes a inferência de localização. Começaremos apresentando o processo de inferência de localização em redes IEEE 802.11 (Seção 2.1). Na Seção 2.2 descreveremos como é realizada a representação do espaço e sinal. Por fim explicaremos dois algoritmos de inferência, o *Multiple Nearest Neighbor* (Seção 2.3) e o *Histogram Method* (Seção 2.4).

2.1

Inferência de Localização

O processo de inferência de localização a partir de informações sobre o sinal de pontos de acesso de redes IEEE 802.11 adotado em muitos sistemas encontrados na literatura [8, 9, 10, 11, 5, 2] é bastante semelhante. Habitualmente esse processo funciona em duas fases:

(i) **Coleta de Dados:** Nesta fase são coletados dados de cada ponto de acesso em pontos bem definidos do ambiente e construído um modelo desses dados para cada ponto. A partir de agora chamaremos tais pontos de “Pontos de Referência”. Os dados coletados podem ser o ponto de acesso corrente, a intensidade do sinal de cada ponto de acesso, a taxa sinal-ruído, etc.

A coleta é realizada através de uma facilidade do padrão IEEE 802.11 conhecido como varredura (*scanning*). Nessa função o dispositivo móvel difunde uma consulta na rede e recebe uma resposta de cada ponto de acesso alcançável na rede. Dessa forma, é possível ao dispositivo móvel determinar a intensidade do sinal de cada ponto de acesso a partir da resposta recebida.

(ii) **Inferência:** Nesta fase, dependendo do modelo criado, é empregada uma estratégia de busca sobre o modelo de forma a encontrar o ponto de referência que mais se assemelha ao dado atual, determinando assim, a localização do objeto.

No caso da intensidade do sinal, por exemplo, a inferência da localização é realizada comparando os sinais coletados para cada ponto de referência e os sinais de rádio-freqüência de cada ponto de acesso medidos no dispositivo, para o qual se deseja determinar sua posição. O ponto de referência com padrão de sinal mais “similar” ao do dispositivo em questão é determinado como a localização do dispositivo. Para isso, dispositivos equipados com interfaces de rede sem fio 802.11 coletam, periodicamente, informações sobre a intensidade do sinal dos pontos de acesso na vizinhança.

Os sistemas de localização diferem nas técnicas utilizadas nas duas fases. A idéia por trás dessas técnicas é utilizar o modelo criado para inferir a localização do usuário com respeito a alguma regra bem definida, a partir dos dados coletados.

Na fase de coleta de dados existem, basicamente, dois métodos para amostragem dos sinais no ambiente. O primeiro, método empírico, consiste na coleta de informações sobre o sinal em diversos pontos na região na qual se deseja inferir a localização de dispositivos móveis.

No segundo, método analítico, são utilizados modelos de propagação de ondas de rádio-freqüência para construir amostras dos sinais em diversos pontos de referência. A principal motivação, a despeito dos problemas de propagação de sinais em ambientes internos, é tentar reduzir a dependência de dados empíricos. Diversos modelos para propagação são encontrados na literatura. Entre eles podemos citar o *Rayleigh Fading Model*, *Rician Distribution Model* e o *Floor Attenuation Factor* [2].

Diversas técnicas têm sido propostas para a inferência da localização, desde a simples identificação do ponto de acesso ao qual o dispositivo está conectado até o uso de redes *Bayesianas*, uma das técnicas probabilísticas comumente utilizadas para estimar a localização de dispositivos.

Neste trabalho experimentamos com duas técnicas distintas para inferência de localização, uma determinística e outra probabilística.

As técnicas determinísticas representam a intensidade do sinal em uma localização por um valor escalar, por exemplo, o valor médio. Diversas estratégias

podem ser utilizadas para determinar a semelhança entre as amostras de sinais, por exemplo, a diferença entre as intensidades dos sinais das amostras coletadas e a amostra atual. A localização do dispositivo móvel é então inferida a partir da localização dos pontos de referência selecionados, ou seja, os pontos de referência com as amostras mais parecidas com a medição atual. Entre os algoritmos utilizados em técnicas determinísticas estão o *Nearest Neighbor (NN)* [8] e o *Multiple Nearest Neighbor (MNN)* [8, 12].

Técnicas probabilísticas usam modelos de distribuição de probabilidades para representar a intensidade dos sinais recebidos de um ponto de acesso e utilizam técnicas probabilísticas para estimar a localização do dispositivo. Entre esses algoritmos estão o *Histogram* [5, 10, 13], *Kernel* [5], e o *Difference* [9].

2.2

A Representação do Espaço e do Sinal

A região de interesse, ambiente no qual se deseja inferir a localização, é representada como sendo um conjunto S de localizações. A cada localização, s_i , está associada uma variável de observação, o_i , de sinais.

Uma localização no espaço é representada por uma tupla que contém as coordenadas da posição e a orientação da antena do dispositivo, por exemplo, no caso bidimensional teríamos:

$$s_i = \{x_i, y_i, \theta_i\}$$

Portanto, a área de interesse é representada por um conjunto, S , dessas tuplas.

$$S = \{s_1, s_2, \dots, s_n\}$$

Outras informações que podem estar presentes em uma variável de localização são uma terceira dimensão, z_i , ou a identificação de um andar.

Uma variável de observação, amostra, o_i é representada por um conjunto de pares, ponto de acesso e valor do sinal.

$$o_i = \{(ap_1, \lambda_1), (ap_2, \lambda_2), \dots, (ap_k, \lambda_k)\}$$

Para cada variável de localização, s_i , são coletadas diversas variáveis de observação, amostras, para minimizar distorções causadas pelo ambiente. Desta forma, a cada localização, s_i , está associado um conjunto de variáveis de observação.

$$s_i \rightarrow \{o_{i1}, \dots, o_{im}\}$$

2.3

Multiple Nearest Neighbor

Este método utiliza uma “função distância” para determinar qual a amostra “mais próxima” da variável de observação para a qual se deseja inferir a posição. O termo “distância” aqui se refere à diferença entre intensidades de sinal e não a uma distância física, como de costume.

Para cada posição s_i , no conjunto de localizações S , é calculada a distância d_i da sua variável de observação, o_i , para a variável de observação, o' , para a qual se deseja inferir a localização.

Uma função distância comumente utilizada é a “Distância Euclidiana”. A diferença é calculada para cada componente $\lambda_{ij}, j \in [1, k]$, da variável de observação o_i , de cada um dos k pontos de acesso, coletados para a posição s_i e a correspondente componente λ'_j do padrão de sinal da localização s' que está se tentando inferir.

$$d_i = \sqrt{(\lambda_{i1} - \lambda'_1)^2 + (\lambda_{i2} - \lambda'_2)^2 + \dots + (\lambda_{ik} - \lambda'_k)^2}$$

De posse das distâncias d_i , o algoritmo seleciona as n localizações s_i mais próximas. Daí o nome *Multiple Nearest Neighbor*. Então, é calculada a média ponderada das coordenadas das n localizações s_i e inferida como a coordenada da variável de observação o' . O peso, utilizado na média ponderada, é calculado com sendo o inverso da distância, $\frac{1}{d_i}$. A Figura 2.1 ilustra esse procedimento.

Neste exemplo podemos perceber que a escolha das ($n = 3$) localizações mais próximas melhora consideravelmente a inferência. Caso não fosse realizada a média ponderada teríamos simplesmente escolhido um dos vizinhos selecionados (o vizinho com menor valor para a função distância) como sendo a localização inferida.

O parâmetro n é de fundamental importância para a precisão do algoritmo. Grandes valores para n podem selecionar vizinhos demasiadamente

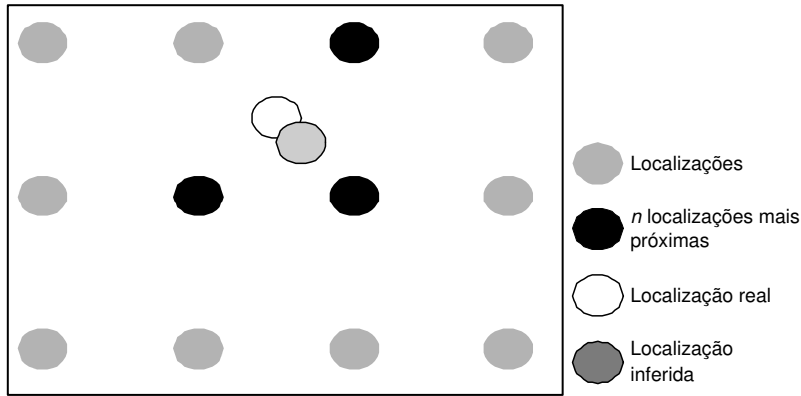


Figura 2.1: Estratégia utilizada para inferência da localização.

distantes da localização real, diminuindo a precisão da inferência. Enquanto que pequenos valores para n tendem a selecionar vizinhos muito próximos, o qual não produz melhorias significativas na inferência.

2.4

Método do Histograma

Geralmente, métodos probabilísticos utilizam a *Regra de Bayes* para inferir a localização. A regra de Bayes nos permite calcular a “Distribuição Posterior” de uma variável de localização.

$$p(s|o) = \frac{p(o|s)p(s)}{p(o)} = \frac{p(o|s)p(s)}{\sum_{s' \in S} p(o|s')p(s')}$$

A distribuição posterior, $p(s|o)$, nos fornece a estimativa de uma posição, s , dada uma variável de observação, o . A “Distribuição Anterior”, $p(s)$, é a probabilidade de um objeto estar na posição s antes de conhecer o valor da variável de observação. A distribuição anterior permite incorporar informações adicionais como perfis de usuários.

Como a probabilidade $p(o)$ não depende da variável de localização, s , tal fator pode ser ignorado nos cálculos sendo útil somente quando é necessário um fator de normalização.

Dado que a distribuição anterior, $p(s)$, seja considerada uniforme, a função de probabilidade determina completamente a distribuição posterior da localização.

A “Função de Probabilidade”, $p(o|s)$, fornece a probabilidade da variável

de observação o para uma variável de localização s . Então, é de extrema importância obter uma função de probabilidade que descreve a distribuição para as variáveis de observação sobre as variáveis de localização da melhor maneira possível. Existem diversas possibilidades para estimar a função de probabilidade a partir das variáveis de observação, sendo nessa função de probabilidade que a maioria das técnicas probabilísticas diferem entre si.

No método *Histogram* [5, 10, 13] é construído um histograma para cada ponto de acesso em cada localização com as probabilidades $p(\lambda_i|ap_i, s_i)$. Essa é a probabilidade que uma medida aconteça para o ponto de acesso ap_i com intensidade de sinal λ_i na variável de localização s_i .

A Figura 2.2 mostra o exemplo de um histograma para um determinado ponto de acesso em uma determinada localização.

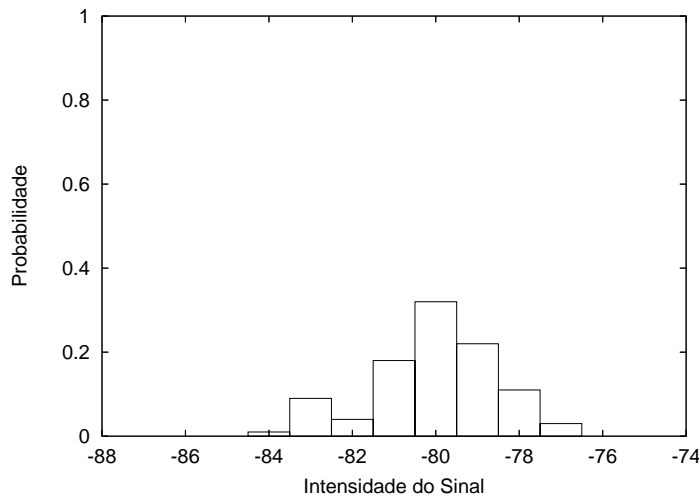


Figura 2.2: Exemplo de histograma.

As probabilidades para cada ponto de acesso em uma mesma localização são multiplicadas para se obter a probabilidade da variável de observação para aquela localização.

$$p(\lambda|s) = \prod_1^k p(\lambda_k|ap_k, s)$$

Tanto o algoritmo *Multiple Nearest Neighbor* quanto o método do *Histograma* são representantes clássicos de algoritmos determinísticos e probabilísticos. Esses dois algoritmos são básicos para o entendimento de outros algoritmos, a maioria deles tendo apenas poucas variações em relação a estes, como poderemos observar no próximo Capítulo (3), onde realizamos uma pes-

quisa sobre abordagens para localização de dispositivos móveis dando ênfase aquelas que utilizam o padrão IEEE 802.11 como meio para obter a localização.