CBPF-NT-006/04

CENTRO BRASILEIRO DE PESQUISAS FÍSICAS

Laboratório de Processamento Digital de Sinais e Imagens
(LPDSI/CAT/CBPF)

Introdução ao Método do Máximo de Entropia

Érica Marques da Silva Santos
Marcelo P. de Albuquerque
Márcio P. de Albuquerque
Aline da Rocha G. Mello
Eugênio S. Caner

RIO DE JANEIRO, DEZEMBRO DE 2004

Sumário

1 Introdução **3**

2 Entropia e Entropia Relativa **4**

2.1 Entropia 4

2.2 Entropia Relativa 4

3 O método de Maximo de entropia **5**

3.1 Exemplo 7

Referências **8**

Introdução ao Método do Máximo de Entropia

Érica M. Silva, Marcelo P. de Albuquerque, Márcio P. de Albuquerque,
Aline da Rocha G. Mello, Eugênio S. Caner

21 de dezembro de 2004

Resumo

Esta nota técnica apresenta uma introdução ao Método do Máximo de Entropia. Inicialmente abordamos os conceitos de Entropia e Entropia Re lativa. Em seguida, apresentamos o Método do Máximo de Entropia, destacando seus princípios. Por fim, apresentamos um exemplo simples é apresentado a fim de ilustrar sua utilização.Método Bayesiano de máximo de entropia, aplicado à imagens com dois pixels

1 Introdução

Em investigações cientificas, nós projetamos instrumentos, fazemos medidas e processamos estas medidas. Terminado este processo, utilizamos as medidas experimentais para fazer inferência sobre os possíveis resultados. A reconstrução de sinais a partir de um conjunto de dados experimentais incompletos e ruidosos pode ser realizada por meio do cálculo de suas probabilidades condicionais e da entropia.

Em física experimetnal temos de uma lado um modelo teórico que descreve um experimento, e do outro lado temos dados experimentais obtidos com uma precisão estatística, devido a limitações no instrumento de medida. O que se faz é uma comparação entre modelos teóricos e dados experimentais.

Em situações onde os dados experimentais são componentes de Fourier (obtidos atrvés de espectômetros, difratômetros entre outros), é comum utilizar a TF^{-1} (Transformada de Fourier Inversa), como uma técnica para a reconstrução do sinal no espaço recíproco. O objetivo da reconstrução é exatamente tentar corrigir a fa lha do instrumento de medida. Quando a TF^{-1} é aplicada, valores iguais a zero são atribuídos as componentes de frequência que não puderam ser medidas, com isto algumas oscilações aprarecem na reconstrução do sinal (Fenômeno de Gibbs). Este fenômeno surge uma vez que se tenta estimar erros de uma função que possui pontos de descontinuidade, ocorre uma salto nos pontos, e este salto resulta na presença de oscilições.

Dados medidos experimentalmente possuem associado a eles um erro expe rimental, este pode ser fixo σ ou proporcional, sendo que este último é igual a $\sigma = F(k).err$. A informação contida no erro expe rimental é relevante sobre os resultados.

Ao se utilizar a TF^{-1}, toda informação contida nos erros experimentais está sendo ignorada. Com o objetivo de melhorar a qualidade da recosntrução do sinal foi desenvolvida uma técnica baseada no princípio do Máximo de Entropia, na qual informações como erros experimentais são levadas em consideração.

2 Entropia e Entropia Relativa

2.1 Entropia

O termo entropia apareceu pela primeira vez na literatura, por volta de 1865 na Alemanha, proposto por Rudolf Clausius para representar uma medida da energia num sistema termodinâmico como função da temperatura e das transferências de calor. Posteriormente , em 1877, foi-lhe dada uma interpretação probabilística no contexto da mecânica estatística de Boltzmann. Mais tarde, em 1948 Shannon, estabeleceu a ligação entre a entropia e informação. A entropia está ligada ao grau de desordem de uma fonte de informação. Quando maior a desordem, maior é a quantidade de informação da fonte. Uma fonte que responda apenas com uma única mensagem, não transmite informação, uma vez que não há redução de incerteza. Esta relação levou a resolução de um numero significativo de problemas básicos na teoria da codificação e na transmissão de dados.

O conceito de entropia raramente é tratado em livros de probabilidades e estatística.

Entropia é uma medida da incerteza de uma variável aleatória. Suponha que X seja uma variável aleatória discreta com uma função de probabilidade p(X).

A entropia H(X) de uma variável aleatória discreta é definida como

Onde $p(x_i)$ é probabilidade de um evento

Geralmente a entropia é calculada com $\log$ na base 2 e é expressa em bits. Veja que a entropia não depende do valor de $X = x_i$, mas apenas das probabilidades $p(X)$.

A entropia descrita por Shannon é experessa da seguinte forma:

$$S = -\sum_i p_i.logp_i \tag{1}$$

2.2 Entropia Relativa

A entropia relativa é a medida de uma distância estatística entre duas distribuições definidas sobre um mesmo alfabeto. Em estatística, isto significa o valor esperado do logaritmo da relação entre as probabilidades. A entropia relativa é definida como sendo

$$D_{KL}(p : p') = \sum_{i=1}^{k} p_i.log\frac{p_i}{p_i^{'}} \tag{2}$$

e é também conhecida como *Distância Kullback-Leibler* ou *Divergência I*.

Apesar da entropia relatica ser chamada de distância Kullback-Leibler, esta não pode ser considerada uma distância verdadeira entre duas distribuições. A entropia relativa não apresenta simetria,

$$\sum_{i=1}^{k}\sum_{i=1}^{k} p_i.log\frac{p_i}{p_i^{'}} \neq \sum_{i=1}^{k} p_i.log\frac{p_i^{'}}{p_i} \tag{3}$$

e portanto não pode ser considerada uma distância métrica. Mesmo assim, é usual pensar na entropia relativa como uma medida de "distância" entre duas distribuições estatísticas.

3 O método de Maximo de entropia

O método do máximo de entropia é baseado no Teorema de Bayes.

O Teorema de Bayes pode ser entendido como um processo de aprendizagem, e é descrito pela seguinte expressão:

$$p(f \mid D) \propto p(D \mid f)p(f) \tag{4}$$

sendo f : modelos teóricos e D : dados experimentais

O ajuste entre os coeficientes de Fourier calculados e medidos é feito através da verossimilhança.

Quando duas informações (no caso, modelo e dados experimentais), diferem somente no valor de algum parâmetro ajustável, a probabilidade de obter este resultado particular pode ser calculada individualmente e é conhecida como a verossimilhança para o modelo ou valores de parâmetros, conhecido os dados.

a verossimilhança possui a seguinte expressão matemática:

$$\exp \frac{-(\bar{x} - x)^2}{2\sigma^2} \tag{5}$$

Utilizando o conceito de verossimilhança é possível calcular $p(D/f)$, então:

$$p(D \mid f) = \prod_i (2\pi\sigma_i^2)^{-1/2} \exp\left(\frac{-1}{2}\chi^2\right) \tag{6}$$

Sendo assim,resta apenas encontrar $p(f)$(probabilidade do modelo).

Vamos supor que um sinal *i* tem a mesma probabilidade de ter um dado valor, isto significa que o modelo é uniforme. Agora imagine uma situação na qual temos uma caixa de tiras. Ao tentarmos estimar o tamanho das tiras, surgem três casos:

1. Caso 1: Todas as tiras possuem o mesmo tamanho,

2. Caso 2: O tamanho das tiras estão dentro de um intervalo (Ex: entre 1m e 2m),

3. Caso 3: O tamanho das tiras se encontram dentro de um intervalo ainda menor.

Quando trabalhamos com probabilidades é possível associarmos a estas uma entropia.

Na teoria da informação, maximizar a entropia significa determinar a distribuição de pro babilidade que represente o máximo de incerteza, dadas certas restrições (condições de não negatividade e soma igual a 1 para as probabilidades).

Ou seja, significa determinar uma distribuição que tenha o maior grau de similaridade de suas probabilidades com uma distribuição de referência ou modelo. Ou ainda que seja mais parecida com a distribuição uniforme (modelo uniforme: $\mathbf{m}^t = [\mathbf{m}_1/N, ..., \mathbf{m}_M/N]$) e diferindo dela apenas devido às restrições.

As distribuições modelo, por sua vez, refletem algum tipo de informação prévia sobre o fenômeno probabilístico de interesse.

O princípio de maximizar a entropia através da medida de Shannon, dado um conjunto de restrições, foi introduzido por Jaynes.

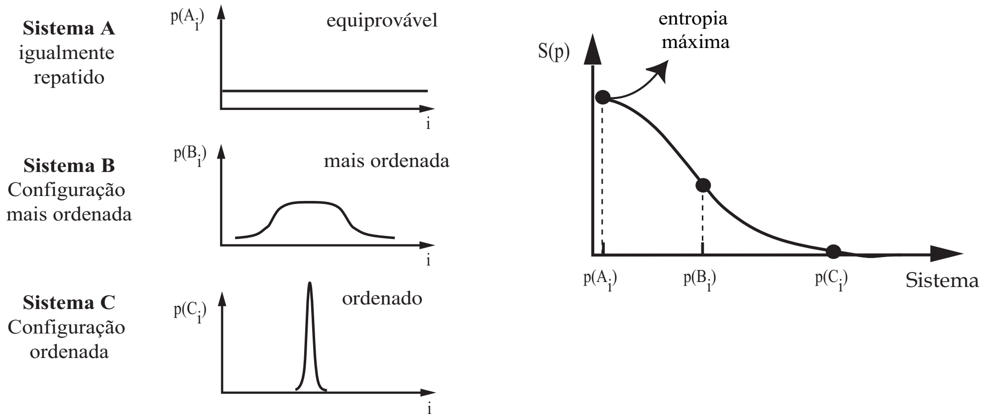

Figura 1: A entropia é a medida de desordem de um sistema fechado. A entropia é máxima numa configuração equiprovável deste sistema. Em configurações mais ordenadas, a entropia toma valores menores. Na realidade, existem diversas configurações mas as probabilidades não as mesmas e portanto, a mais provável é a que possui maior entropia.

$$S(\mathbf{f}, \mathbf{m}) = \sum_{i=0}^{M-1} \mathbf{f}_i - \mathbf{m}_i - \mathbf{f}_i \ln \left(\frac{\mathbf{f}_i}{\mathbf{m}_i} \right) \tag{7}$$

onde **f** é o valor do sinal calculado e **m** é valor atribuído ao modelo inicial. Desta forma, esta análise pode ser generalizada para incluir um conhecimento anterior.

Em sua forma mais simples, a análise de dados com o máximo de entropia esquiva-se da complexidade da formulçao bayesiana correta, forçando o daod experimental dentro de uma prova definitiva, forma com a qual rejeitam quase que completamente um "provável" conjunto de f, e então, selecionando o f mais provável e aplicando diretamente o princípio de máximo de entropia obtêm-se a maior entropia S

$$p(f \mid D) \propto \exp \left(\alpha S(f) - \frac{1}{2}\chi^2 \right) \tag{8}$$

A maximização da entropia é feita respeitando-se a condição de $\chi^2 \leq N$.

As soluções f para vários valores de α define a "trajetória do máximo de entropia", iniciando em $f = me\alpha = \infty$ e terminando em $\alpha = 0$ e com χ^2 mínimo.

Chamamos de $\chi^2 = N$ a técnica de "Máximo de Entropia". Embora tais recosntruções f sejam freqüentemente claras e bastante informativas do que as obtidas com outras técnicas, o máximo de entropia não é bayesiano e conseqüentemente não é perfeito. A limitação $\chi^2 = N$ é apenas uma aproximação da verdadeira verossimilhança $p(D \mid f)$, nenhum f selecionado pode inteiramente representar a probabilidade posterior $p(f \mid D$ como a teoria exige.

3.1 Exemplo

Para exemplificarmos o Método Bayesiano de Máximo de Entropia, vamos apresentar a solução de um problema simples, que envolve uma imagem com apenas dois pixels, $f = (f_1, f_2)$ com os seguintes dados ruidosos:

$$0.56 f_1 + 0.83 f_2 = 5.32 \pm 0.48. \tag{9}$$

$$0.83 f_1 - 0.56 f_2 = 4.24 \pm 2.00. \tag{10}$$

e,

$$R = \begin{bmatrix} 0.56 & 0.83 \\ 0.83 & -0.56 \end{bmatrix}, \quad D = \begin{bmatrix} 5.32 \\ 4.24 \end{bmatrix} \sigma = \begin{bmatrix} 0.48 \\ 2.00 \end{bmatrix}$$

$$[f^-1] = \begin{bmatrix} \frac{1}{f_1} & 0 \\ 0 & \frac{1}{f_2} \end{bmatrix}, \quad \sigma^{-2} = \begin{bmatrix} \frac{1}{0.48^2} & 0 \\ 0 & \frac{1}{4} \end{bmatrix},$$

A entropia e o χ^2 são escritos da seguinte forma:

$$S(f) = f_1 + f_2 - m_1 - m_2 - f_1 \log\left(\frac{f_1}{m_1}\right) - f_2 \log\left(\frac{f_2}{m_2}\right) \tag{11}$$

$$\chi^2(f) = (D - Rf)^t[\sigma^-2](D - Rf) \tag{12}$$

$$\chi^2(f) = \frac{(0.56 f_1 + 0.83 f_2 - 5.32)^2}{0.48^2} + \frac{(0.83 f_1 - 0.56 f_2 - 4.24)^2}{2.00^2} \tag{13}$$

O $\chi^2(f) = C^t e$, descreve a equação de uma elipse no espaço (f_1, f_2). Existe um programa que calcula a trajetória destas soluções $\hat{f}$, com α assumindo dois valores, 100 ou 1000, uma vez que a condição $\chi^2 \leq N$ seja satisfeita. O calculo realizado segue o seguinte caso de $m = (1, 1)$. A figura 1 representa as trajetórias suaves de $\hat{f}$ para este caso. Os traços de contorno das entropias correspondentes $(S(f, m))$ e das elipses $\chi^2(f) = C^t e$ são também representadas.

O programa acima mencionado, foi eleborado em liguagem de programação C. Porém, os resultados apresentados neste trabalho são provinientes de um outro programa, o qual foi escrito em Matlab. Este programa calcula o máximo de entropia e o χ^2 em função de f_1 e f_2.

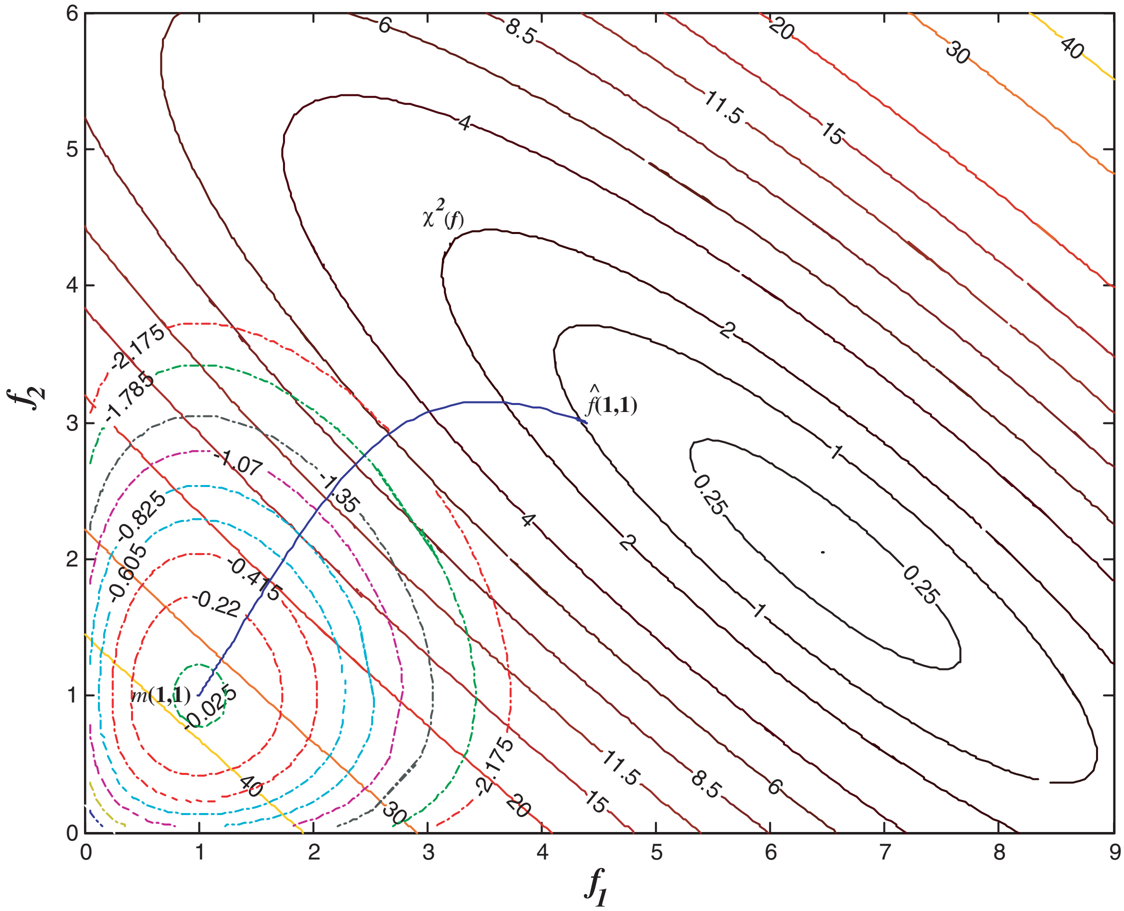

Figura 2: Exemplo da utilização do método do máximo de entropia por um sistema linear com duas incógnitas. Os cálculos são realizados a partir de um "modelo": $\mathbf{m} = (1,1)$, os gráficos de superfícies representa o $\chi^2(\mathbf{f})$ e $S(\mathbf{f},\mathbf{m})$. As curvas em traços fortes são as trajetórias das soluções aproximadas; a solução final a para a qual $\chi^2 \leq 1 = 2$. Todos os valores contidos na elipse são soluções aceitáveis. A solução obtida é $\hat{\mathbf{f}}_{1,1} = (4.34, 3.02)$. A cada iteração, a nova solução aproximada é a que possui maior entropia

Referências

[1] B. Buck and V.A. Macaulay, "Maximum Entropy in Action",Clarendon Press - Oxford, 1991.

[2] Djafari A.M. and G. Demoment, "Maximum Entropy in Bayesian approach in tomographic image reconstruction and restoration.", Kluwer Academic Publisher, page 195, 1989.

[3] Djafari A.M., "Synthése de Fourier Multivariable à Maximum d'Entropie - Aplication à la Reconstruction Tomographique d'Images.", PhD Thesis, Université de Paris-Sud - Centre d'Orsay,1987.