

DECODIFICAÇÃO USANDO BAYES EMPÍRICO  
PARA FONTES COM DISTRIBUIÇÃO DESCONHECIDA

por

HÉLIO MAGALHÃES DE OLIVEIRA

Coordenação do Mestrado em Engenharia Elétrica

COMENE - UFPE

Cidade Universitária - 50.000 Recife-PE BRASIL

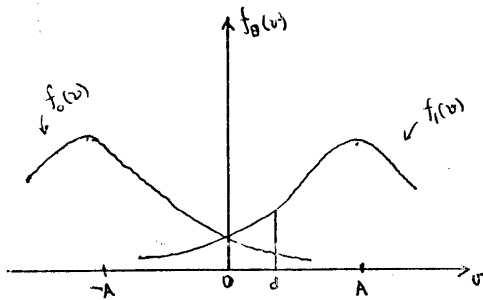
RESUMO

Um algoritmo para decodificação de símbolos de uma fonte com distribuição desconhecida é introduzido, o qual emprega limiares de decisão adaptativos. O método baseia-se no aprendizado Bayes Empírico sem professor, e uma aplicação para canal com ruído aditivo Gaussiano é apresentada. A convergência do Algoritmo é examinada, demonstrando que a probabilidade de erro na decodificação converge a.s. para o valor que seria obtido no caso do conhecimento da distribuição a priori dos símbolos da fonte.

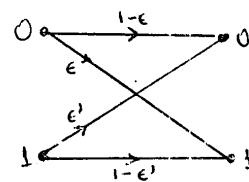
## INTRODUÇÃO

um problema muito discutido em telecomunicações é como proceder na decodificação de fontes onde a distribuição de probabilidades de ocorrência dos símbolos existe mas é desconhecida. Uma técnica muito adotada neste caso é o "approach" Neyman-Pearson (FERGUSON 1967). Recentemente, Wolfenson e Rocha Jr. propuseram um método de decodificação suave aplicável quando códigos de bloco são utilizados (WOLFENSON and ROCHA Jr. 1980). Contudo, tradicionalmente são consideradas fontes discretas sem memória (DMS), transmitindo símbolos  $\omega \in \Theta$  cuja distribuição a priori  $\{p(\omega)\}$  é conhecida. Em muitas situações na prática, entretanto, este não é o caso. O algoritmo introduzido a seguir tem aplicação em situações onde o canal de comunicações é utilizado para transmitir informações de DMS que possuem distribuição a priori desconhecida, e faz uso de canais discretos sem memória (DMC).

Estes modelos de canais resultam do emprego de procedimentos de decisão abrupta para a decodificação da informação recebida. Um exemplo típico desta situação é o uso do Canal Binário na decodificação de fontes transmitindo dígitos binários em um canal Gaussiano, como ilustrado nas figuras 1 e 2. Os símbolos  $\{0, 1\}$  correspondem à transmissão de pulsos de  $\mp A$  volts, respectivamente, e é suposto que o ruído tem distribuição normal  $N(0, \sigma)$ .



Funções Densidades Condicionais  
Fig. 1



Modelo Discreto do Canal  
Fig. 2

Pode ser observado que o estabelecimento de um limiar de decisão  $d$  para efetuar a decisão abrupta determina a matriz de probabilidades de transição  $[P(x|b)]$  do modelo de canal usado.

$$\epsilon = \int_d^{+\infty} f_0(v) dv = Q\left(\frac{d+A}{\sigma}\right)$$

$$\epsilon' = \int_{-\infty}^d f_1(v) dv = 1 - Q\left(\frac{d-A}{\sigma}\right)$$

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-\xi^2/2} d\xi$$

O procedimento usual (SCHWARTZ et al 1966) é recorrer a uma escolha de um limiar ótimo no sentido de minimizar a probabilidade de erro. Para o exemplo acima, este limiar é dado por

$$d^* = \frac{\sigma^2}{2A} \ln \frac{\pi(\omega)}{\pi(\lambda)}$$

Em alguns casos todavia, são estabelecidos  $D$  limiares de decisão diferentes  $d^*(i)$   $i=1,2,\dots,D$ . No caso binário, por exemplo,  $D$  é o número de interseções entre as curvas de densidades condicionais e ademais, os limiares coincidem com os valores correspondentes aos pontos de interseção quando a distribuição a priori é uniforme.

Obviamente, este procedimento faz uso da distribuição a priori dos símbolos, de modo que não pode ser aplicado no caso ora discutido. Uma regra de decodificação baseada na escolha de um limiar ótimo não pode ser estabelecida. Uma maneira de contornar este problema é discutida na seção seguinte, usando a estratégia do aprendizado Bayes Empírico sem professor (SPRAGINS, 1966), resultando em limiares de decisão adaptativos a cada passo da decodificação. A aplicação deste método ao problema discutido só pode ser feita no caso em que os limiares de decisão permanecem fixos.

Admitindo o uso de limiares de decisão fixos, a matriz de probabilidades de transição  $[P_0(x|b)]$  não varia e se a fonte transmite símbolos com probabilidades  $\{\pi(\omega)\}$ , então os símbolos de saída ocorrem com probabilidades dadas por  $P(x) = [P_0(x|b)] [\pi(\theta)]$ . Neste caso, o emprego de estimadores de frequência relativa  $P_n(x)$  das probabilidades  $P(x)$  permite uma atualização no conhecimento da distribuição a priori dos parâmetros, provido que a família  $\{P_0(\cdot|\theta)\}$  seja identificável (YAKOWITZ and SPRAGINS 1968). Isto pode ser feito calculando o vetor  $[\pi_n(\theta)]$ , de acordo com:

$$[\pi_n(\theta)] = [P_n(x|\theta)]^{-1} [P_n(x)]$$

Ademais, para que o resultado obtido seja uma distribuição de probabilidade, se necessário, o resultado é truncado em zero e normalizado, obtendo uma distribuição  $\{\hat{\pi}_n(\theta)\}$ , onde  $\hat{\pi}_n(\theta) \rightarrow \pi(\theta)$  a.s.

Definindo-se os conjuntos  $L_n = \{\theta \in \Theta \rightarrow \pi_n(\theta) < 0\}$  e  $G_n = \{\theta \in \Theta \rightarrow \pi_n(\theta) > 0\}$  o truncamento é feito de acordo com

$$\hat{\pi}_n(\theta) = \begin{cases} 0 & \text{se } \theta \in L_n \\ \frac{\pi_n(\theta)}{\sum_{\theta' \in G_n} \pi_n(\theta')} & \text{se } \theta \in G_n \end{cases}$$

### O ALGORÍTMO DE DECODIFICAÇÃO

Inicialmente, o problema será formulado em termos de teoria da decisão (FERGUSON 1967), caracterizando o espaço de parâmetros, de observações e de ações  $(\Theta, \mathcal{X}, \mathcal{A})$  por  $\Theta = \mathcal{X} = \mathcal{A} = \{0, 1, 2, \dots, q-1\}$ . A função perda empregada é a do observador ideal, i.e.,  $L(\theta, a) = 1 - \delta_{\theta a}$  onde  $\delta_{\theta a}$  é o delta de Kronecker. Adicionalmente, considera-se o emprego de uma única regra de decisão pura  $d: \mathcal{X} \rightarrow \mathcal{A}$ ,  $x \mapsto d(x)$  e uma matriz de probabilidades de transição (Canal)  $[P_n(x|\theta)]$  determinada pelo estabelecimento de limiares de decisão  $d_n(i)$ . Deste modo, o risco associado a uma decisão é dado por:

$$R_n(\theta) = \sum_{x \in \mathcal{X}} L(\theta, d(x)) P_n(x|\theta) = \sum_{x=0}^{q-1} (1 - \delta_{\theta x}) P_n(x|\theta)$$

Denotando por  $\{\pi(\theta)\}$  as probabilidades a priori da fonte, o risco de Bayes associado à regra  $d$ ,

$$R_n = E_{\theta} \{R_n(\theta)\} = \sum_{\theta=0}^{q-1} \pi(\theta) R_n(\theta)$$

representa a probabilidade de erro na decodificação quando é realizada uma decisão abrupta empregando os limiares de decisão  $d_n(i)$ .

Da forma como o problema está estabelecido, não há liberdade para a escolha de uma regra de decisão adequada, como é feito tradicionalmente. Contudo é possível ajustar o limiar de decisão para reduzir a probabilidade de erro, e tal ajuste pode ser interpretado como uma mudan

ça na matriz  $[P_n(x|\theta)]$  do DMC, e não na regra de decisão. Em particular, sendo  $d^*(i)$  os limiares ótimos obtidos com o conhecimento da distribuição a priori  $\{\pi(\theta)\}$  e  $[P^*(x|\theta)]$  a matriz de probabilidades de transição calculada empregando estes limiares, tem-se

$$R^*(\theta) = \sum_{x=0}^{q-1} (1 - \delta_{x, d^*(i)}) f^*(x|\theta)$$

e conseqüentemente o risco de Bayes neste caso é dado por

$$R^* = E_{\theta} \{ R^*(\theta) \} = \sum_{\theta=0}^{q-1} \pi(\theta) R^*(\theta)$$

assim,  $R^*$  representa a probabilidade de erro mínima na decodificação, quando decisão abrupta é empregada.

O fundamento básico do algoritmo consiste em considerar duas classes de limiares de decisão diferentes, no processo de decodificação. Na primeira delas, os limiares são denotados  $d_0^*(i)$  e permanecem inalterados no procedimento de decisão. A eles está associada uma matriz de transição  $[P_0(x|\theta)]$ , e o símbolo de saída  $x_n$  obtido pelo seu uso serve apenas para estimar a <sup>DISTRIBUIÇÃO</sup> a priori. A segunda classe é constituída por limiares adaptativos  $d_n^*(i)$  que variam a cada passo da decodificação, e o dígito obtido  $x_n^*$  corresponde a saída do decodificador associada ao n-ésimo dígito transmitido.

O estabelecimento da condição inicial pode ser feito em geral com base na escolha da distribuição  $\{\pi_0(\theta)\}$  como uniforme. Isto determina limiares ótimos  $d_0^*(i)$  e uma matriz de transição  $[P_0(x|\theta)]$ , uma vez conhecidas as densidades condicionais  $f_{\theta}(v) \theta \in \mathcal{C}$ . Se esta escolha implica na não identificabilidade da família  $\{P_0(\cdot|\theta)\}$ , então uma outra distribuição deve ser escolhida. Considerando o estimador  $P_n(x)$  para os símbolos  $x_j, j=1,2,\dots,n$  obtidos usando  $d_0^*(i)$ , e aplicando o aprendizado Bayes empírico sem professor, é possível determinar um estimador  $\hat{\pi}_n(\theta)$  para a distribuição a priori dos símbolos da fonte, verificando

$$\hat{\pi}_n(\theta) \rightarrow \pi(\theta) \quad \text{a.s.}$$

$$\hat{\pi}_n(\theta) = \frac{1}{n} \sum_{j=1}^n \delta_{x_j, \theta}$$

A idéia então é, baseado na distribuição estimada  $\{\hat{\pi}_n(\theta)\}$ , efetuar a escolha dos limiares ótimos de decisão  $d_n^*(i)$  que minimizem a probabilidade de erro calculada a partir desta distribuição estimada. As amostras são decodificadas abruptamente usando estes limiares adaptativos e entre-

gues ao destinatário.

Considerando o exemplo anterior, isto resultaria em escolher um limiar variável

$$d_n^* = \frac{\sigma^2}{2A} \ln \frac{\hat{\pi}_n(v)}{\hat{\pi}_n(1)}$$

#### O ALGORÍTMO

Passo 1.

Estabelecimento das condições iniciais:  $0 \leftarrow n$ ;  
Calcular os limiares fixos  $d_0^*(i)$  e a matriz de probabilidades de transição  $[P_0(x|\theta)]$ .

Passo 2.

Fazer  $n \leftarrow n+1$  e decodificar a  $n$ -ésima amostra empregando os limiares  $d_0^*(i)$ , obtendo a observação  $x_n$ .

Passo 3.

Atualizar o estimador de frequência relativa  $P_n(x)$  da probabilidade de ocorrência dos símbolos de saída, usando a observação  $x_n$ .

Passo 4.

Calcular a estimativa  $\{\hat{\pi}_n(\theta)\}$  da distribuição a priori empregando a relação matricial  $[\hat{\pi}_n(\theta)] = [P_0(x|\theta)]^{-1} [P_n(x)]$ . Truncar em zero e normalizar, obtendo uma distribuição de probabilidade  $\{\hat{\pi}_n(\theta)\}$ .

Passo 5.

Usando a distribuição  $\{\hat{\pi}_n(\theta)\}$ , calcular novos limiares de decisão  $d_n^*(i)$  e decodificar a amostra usando estes limiares, obtendo a saída do decodificador  $x_n^*$ . A seguir, retornar ao passo 2.

#### ANÁLISE DA CONVERGÊNCIA

As condições para garantir um desempenho de probabilidade de erro assintoticamente ótimo, são analisadas a seguir, com atenção particular para o exemplo citado anteriormente.

O primeiro passo na análise é verificar se há convergência forte entre os valores dos limiares ótimos obtidos, i.e., se  $d_n^*(i) \rightarrow d^*(i)$  a.s.. Isto deve ser feito para cada canal em particular, com o conhecimento das densidades  $f_\theta(v)$   $\theta \in \Theta$ .

Para o exemplo estudado, é demonstrado que o procedimento for -

nece uma sequência de limiares de decisão que converge a.s. para o valor do limiar ótimo  $d^*$ , calculado com o conhecimento da distribuição a priori. No caso de símbolos binários transmitidos em um canal com ruído aditivo Gaussiano com SNR não nula, o estabelecimento da condição inicial é feito usando a hipótese  $\pi_0(v) = \pi_0(1)$ . Isto acarreta em um limiar fixo  $d_0^* = 0$  e uma matriz do canal  $[P_0(x|\theta)]$  caracterizada por  $\epsilon = \epsilon' = Q(A/\sigma) < 0.5$ . Daí pode-se investigar a identificabilidade da família  $\{P_0(\cdot|\theta)\}$ . A condição  $\det [P_0(x|\theta)] = 0$  implica em  $\epsilon + \epsilon' = 1$ , entretanto tem-se que  $\epsilon + \epsilon' < 1$  de modo que a identificabilidade é garantida e o algoritmo pode ser utilizado. Obviamente a aplicação do algoritmo no caso binário é assegurada para qualquer tipo de canal, uma vez admitido que as probabilidades de transição obedecem  $\epsilon < 0.5$  e  $\epsilon' < 0.5$ .

LEMA 1.

Para o canal com ruído aditivo Gaussiano e o sistema binário descrito na introdução,  $d_n^* \rightarrow d^*$  a.s.

PROVA.

tem-se que  $|d_n^* - d^*| = \frac{\sigma^2}{2A} \left| \ln \frac{\hat{\pi}_n(v)}{\hat{\pi}_n(1)} - \ln \frac{\pi(v)}{\pi(1)} \right|$  e portanto  $0 \leq |d_n^* - d^*| \leq \frac{\sigma^2}{2A} \sum_{\theta=0}^1 \left| \ln \hat{\pi}_n(\theta) - \ln \pi(\theta) \right|$

Usando os resultados do aprendizado Bayes empírico sem professor,  $\hat{\pi}_n(\theta) \rightarrow \pi(\theta)$  a.s., e levando em conta que  $\log$  é uma função contínua, segue-se que  $\ln \hat{\pi}_n(\theta) \rightarrow \ln \pi(\theta)$  a.s., provido que  $\forall \theta \in \Theta, \pi(\theta) > 0$ .

Logo  $|\ln \hat{\pi}_n(\theta) - \ln \pi(\theta)| \rightarrow 0$  a.s., donde conclui-se que  $d_n^* \rightarrow d^*$  a.s.

Q.E.D.

A demonstração em casos onde o ruído não é Gaussiano e/ou o sistema empregado é multinível, pode ser feita seguindo o mesmo raciocínio.

Denotando por  $[P_n^*(x|\theta)]$  a matriz de probabilidades de transição calculada a partir dos limiares ótimos adaptativos  $d_n^{*(i)}$ , o seguinte teorema pode ser enunciado:

TEOREMA 1.

Se  $d_n^{*(i)} \rightarrow d^{*(i)}$  a.s.  $i=1,2,\dots,D$  e  $f_\theta(v)$  é uma função contínua e limitada, então para cada  $\theta \in \Theta$  tem-se  $P_n^*(x|\theta) \rightarrow P^*(x|\theta)$  a.s.

PROVA.

$$\text{Dado } \theta \in \Theta, \left| P_n^*(x|\theta) - P^*(x|\theta) \right| \leq \sum_{i=1}^D \left| \int_{d^{*(i)}}^{d_n^{*(i)}} f_\theta(v) dv \right|$$

Lembrando que  $f_\theta(v)$  é contínua e usando o teorema do valor principal obtém-se que

$$\exists \xi_i \ni \int_{d^{*}(i)}^{d_n^{*}(i)} f_{\theta}(v) dv = f_{\theta}(\xi_i) [d_n^{*}(i) - d^{*}(i)] \quad i=1, 2, \dots, D.$$

Por outro lado, como  $f_{\theta}(v)$  é limitada,  $\exists M_{\theta} \ni (\forall v) f_{\theta}(v) \leq M_{\theta}$  e portanto

$$|P_n^{*}(x|\theta) - P^{*}(x|\theta)| \leq M_{\theta} \sum_{i=1}^D |d_n^{*}(i) - d^{*}(i)| \rightarrow 0 \text{ a.s.}$$

Logo  $P_n^{*}(x|\theta) \rightarrow P^{*}(x|\theta)$  a.s.

Q.E.D.

A probabilidade de erro na decodificação do  $n$ -ésimo bit usando o algoritmo proposto é dada pelo risco de Bayes  $r_n^{*}$ , expresso por

$$r_n^{*} = E_{\theta} \{ R_n^{*}(\theta) \} = \sum_{\theta=0}^{q-1} \pi(\theta) R_n^{*}(\theta)$$

onde

$$R_n^{*}(\theta) = \sum_{x=0}^{q-1} (1 - \delta_{\theta x}) P_n^{*}(x|\theta)$$

LEMA 2.

Para o algoritmo proposto a probabilidade de erro na decodificação converge a.s. para o valor mínimo, ou seja,  $r_n^{*} \rightarrow r^{*}$  a.s.

PROVA.

$$\text{Tem-se que } R_n^{*}(\theta) - R^{*}(\theta) = \sum_{x=0}^{q-1} (1 - \delta_{\theta x}) [P_n^{*}(x|\theta) - P^{*}(x|\theta)]$$

$$\text{Logo } |R_n^{*}(\theta) - R^{*}(\theta)| \leq \sum_{x=0}^{q-1} |P_n^{*}(x|\theta) - P^{*}(x|\theta)|$$

Usando os resultados do teorema e lema anteriores, segue-se  $|R_n^{*}(\theta) - R^{*}(\theta)| \rightarrow 0$  a.s., ou seja,  $R_n^{*}(\theta) \rightarrow R^{*}(\theta)$  a.s.

Por outro lado,  $|r_n^{*} - r^{*}| \leq \sum_{\theta=0}^{q-1} \pi(\theta) |R_n^{*}(\theta) - R^{*}(\theta)|$ , de modo que  $|r_n^{*} - r^{*}| \rightarrow 0$  a.s. completando a prova.

Q.E.D.



## CONCLUSÃO:

O algoritmo introduzido presta-se para decodificação de fontes com distribuição desconhecida, com uma vantagem adicional que a técnica é aplicável para fontes onde a distribuição não se mantém constante, i.e., a fonte opera com uma distribuição  $\{p_i^1(\theta)\}$  durante um certo intervalo de tempo  $t_1$ , depois altera a distribuição para  $\{p_i^2(\theta)\}$  por um outro período de tempo  $t_2$ , e assim sucessivamente. A rapidez da convergência, entretanto, não está examinada. Também é interessante observar o fato que a probabilidade de erro é relativamente insensível com relação ao valor do limiar quando o canal é Gaussiano (SCHAWARTZ et al 1966); todavia para outros tipos de canais ou mesmo para baixa Relação Sinal/Ruído (SNR) isto poderá não acontecer, principalmente quando a entropia da fonte é baixa. Nestes casos, uma melhoria considerável é obtida com relação a um sistema que faça uso de limiares obtidos supondo símbolos equiprováveis para representar o desconhecimento da distribuição da fonte. Fica óbvio que a aplicação desta técnica é indicada para canais com SRN baixa, uma vez que para a grande maioria dos canais com SRN elevada, a sensibilidade do desempenho com relação a escolha do limiar é diminuída drasticamente, tornando-se praticamente indiferente o uso dos limiares ótimos ou o uso dos limiares obtidos supondo que a fonte tem entropia máxima. Adicionalmente, códigos corretores de erros podem ser empregados na saída do decodificador com toda sua potencialidade, e o uso combinado deste algoritmo com os códigos corretores de erros pode ser bem eficiente, de modo a tornar possível taxas de erros aceitáveis usando pouca potência no transmissor. O "approach" usado não é convencional, visto que a regra de decisão permanece fixa e uma mudança no limiar de decisão implica em modificação na matriz de probabilidades de transição.

O método apresentado baseia-se no aprendizado Bayes Empírico <sup>de Professor</sup> e pode ser largamente aplicado para vários tipos de canais contínuos, e em diversas situações práticas. O valor da probabilidade de erro empregando o algoritmo converge a.s. para o valor obtido quando conhecida a distribuição da fonte, o que torna o procedimento bastante potente e atrativo.

#### AGRADECIMENTOS.

The author wish to thank profs F. M. CAMPELLO and V. C. ROCHA Jr. for a number of helpful discussions.

#### REFERENCES.

- FERGUSON, T. S., 1967, Mathematical Statistics, Academic Press.
- SCHWARTZ, M., BENNETT, W. R., and STEIN, S., 1956, Communication Systems and Techniques, McGraw-Hill Book Company.
- WOLFENSON, M. and ROCHA Jr., V. C., 1980, Soft Decision Decoding With Unknow Source Distribution, Electronics Letters, 16, 25-26.
- SPRAGINS, J., 1966, Learning Without a Teacher, IEEE Trans. Info. Theory, IT-12, 2, 223-230.
- ROBINS, H., 1964, The Empirical Bayes Approach To Statistical Decision Problems, Ann. Math. Stat., Vol 35, 1-20.
- YAKOWITZ, S.J. and SPRAGINS, J.D., On the Identifiability of finite Mixtures, Ann. Math. Stat., vol 39, 209-214. 1968