Tratamento de Restrições Estocásticas no Algoritmo PSO para Estimação ML de Modelos HMM

E. Marques, N. Maciel e E. Pinto

Resumo— **Uma nova modificação do algoritmo PSO convencional é proposta neste trabalho, a fim de lidar com as restrições estocásticas da estimação ML de modelos HMM. Um critério de parada para o PSO baseado na invariância da função custo é também proposto. Estas propostas são aplicadas com êxito na modelagem de canais com erros em surtos.**

Palavras-Chave— **HMM, Erros em Surto, PSO.**

Abstract— **A new modification of the conventional PSO algorithm is proposed in this work, in order to deal with the stochastic constraints in the ML estimation of HMM models. A PSO stopping criterion based on the invariance of the cost function is also proposed. These proposals are successfully applied to burst-error channel modeling.**

Keywords— **HMM, Burst Error, PSO.**

I. INTRODUÇÃO

Os modelos HMM (*Hidden Markov Model*) têm uma ampla área de aplicação, que inclui os problemas de reconhecimento de locutor [1] e modelagem de canais com erros em surtos [2]. O critério de máxima verossimilhança (ML, de "Maximum Likelihood") tem sido frequentemente utilizado para o ajuste dos parâmetros destes modelos, e o algoritmo Baum-Welch vem sendo a ferramenta mais empregada para tal [3]. Entretanto, este algoritmo não garante convergência para uma solução de máximo global e tem custo computacional elevado.

Com isso, o estudo de algoritmos mais simples, com menor custo computacional e maior facilidade de alcançar um máximo global é de grande interesse neste contexto [4]. Diversos algoritmos de otimização bio-inspirados têm sido investigados como alternativas para este problema e alguns trabalhos recentes têm destacado as potencialidades do método de Otimizacão por Enxames de Particulas (PSO, de "Particle Swarm Optimization") [5], [6].

No entanto, o emprego do algotimo PSO no ajuste de modelos HMM requer cuidados específicos devido às restrições estocásticas dos parâmetros destes modelos. Algumas modificações do PSO têm sido propostas, visando principalmente o seu emprego na otimização de modelos HMM aplicados em processamento de sinais [5], [6].

Este trabalho se concentra na aplicação do algoritmo PSO na estimação ML de modelos HMM para erros em surtos e nele se propõe um novo método para o tratamento das restrições estocásticas acima mencionadas. Além disso, também se introduz um novo critério de parada para o algoritmo PSO e se avalia brevemente os efeitos de sua aplicação.

E. Marques, N. Maciel e E. Pinto, Instituto Militar de Engenharia, Departamento de Engenharia Elétrica, Laboratório de Comunicações Digitais, Emails: {elainecmarques,nilsonmpj,ernesto}@ime.eb.br. Este trabalho foi parcialmente apoiado pelo CNPQ.

II. A TÉCNICA PSO

O método PSO visa a otimização de funções não-lineares e tem por inspiração o comportamento social dos animais [7]. Nele um número pré-fixado de soluções tentativas, chamada "partícula", são atualizadas recursivamente. A cada iteração os vetores de posição $\overrightarrow{x_i}(t)$ e velocidade instantânea $\overrightarrow{v_i}(t)$ de cada partícula são calculados como se segue:

$$\overrightarrow{x_i}(t+1) = \overrightarrow{x_i}(t) + \alpha\overrightarrow{v_i}(t+1) \quad (1)$$

$$\overrightarrow{v_i}(t+1) = W\,\overrightarrow{v_i}(t) + c_1 U_1(\overrightarrow{p_i}(t) - \overrightarrow{x_i}(t)) + c_2 U_2(\overrightarrow{p_g}(t) - \overrightarrow{x_i}(t)) \quad (2)$$

onde α e W são constantes que ponderam os efeitos da velocidade e da inércia nas atualizações da posição e da velocidade, respectivamente. U_1 e U_2 são amostras de variáveis aleatórias independentes com distribuição uniforme no intervalo $[0\ 1]$, e os parâmetros c_1 e c_2 ponderam a influência da experiência própria passada e social, respectivamente. Estas influências estão caracterizadas nas posições $\overrightarrow{p_i}$ e $\overrightarrow{p_g}$, nas quais a i-ésima partícula e todo o enxame, respectivamente, geraram os melhores valores da função custo.

III. RESTRIÇÕES ESTOCÁSTICAS DOS MODELOS HMM

Os parâmetros de um modelo HMM são aqui representados por $\lambda = \{A, B, \Pi\}$, onde A e B são as matrizes das probabilidades condicionais de transição dos estados e de observação, respectivamente, e Π é o vetor de probabilidades do estado em que a operação do modelo é iniciada.

Por restrições estocásticas entende-se que A, B e Π devem ter elementos não negativos e que a soma dos elementos de cada linha destas matrizes, bem como a soma dos elementos do vetor Π, devem ser unitárias.

Observa-se nas Equações (1) e (2) que, para atender a segunda condição acima, é suficiente atribuir $\overrightarrow{v_i}(0)$ que atendam a ambas as condições supracitadas. No entanto, a versão usual do algoritmo não garante o atendimento da segunda condição, o que tem motivado o surgimento de algumas modificações.

A proposta apresentada em [6] consiste em atribuir o valor 0 a todos os parâmetros negativos e, em seguida, normalizar os parâmetros λ (por exemplo: $a'_{ij} = a_{ij}/\sum_{j=1}^{N} a_{ij}$). A nosso ver, o problema com este método é que pode-se perder muita informação útil para a otimização ao se truncar os parâmetros negativos resultantes da aplicação das Equações (1) e (2).

Outra modificação do PSO encontrada na literatura, aqui denominada Método de Referência (MP), consiste em associar um valor muito pequeno da função custo às partículas com pelo menos um elemento negativo, de forma que elas não contribuam para a atualização das posições do enxame [6] na iteração seguinte. Este recurso abre porém a possibilidade de grandes desperdícios de esforço computacional com partículas que não serão utilizadas para melhoria da função custo.

IV. Método Proposto (MP)

Propõe-se aqui, fazer o ajuste do valor de α da Equação (1). Primeiro é atribuído $\alpha = 1$ e se verifica se há algum parâmetro com valor negativo. Caso haja, atribui-se o maior valor de $\alpha < 1$ capaz de produzir um vetor de posição atualizado com elementos não negativos. Com isso, garante-se o movimento da partícula na direção da velocidade indicada na Equação (2) sem violar as restrições estocásticas dos parâmetros.

Como exemplo, suponha que a uma determinada linha da matriz A estão associados os vetores $\vec{x_i}(t) = \{0,1 \;\; 0,15 \;\; 0,75\}$ e $\vec{v_i}(t) = \{-0,3 \;\; -0,6 \;\; 0,9\}$[1]. Com $\alpha = 1$ viola-se as restrições estocásticas pois teria-se $\vec{x_i}(t+1) = \{-0,2 \;\; -0,45 \;\; 1,65\}$. Com $\alpha_1 = 1/3$, de modo que o primeiro elemento de $\vec{x_i}(t+1)$ seja igual a 0, verifica-se que este não seria adequado pois levaria a $\vec{x_i}(t+1) = \{0 \;\; -0,05 \;\; 1,05\}$. Testa-se então o valor de α que torna o segundo elemento da linha igual a 0 ($\alpha = 0,25$), o qual gera $\vec{x_i}(t+1) = \{0,025 \;\; 0 \;\; 0,975\}$, atendendo às restrições.

A. *Critério de Parada*

Em diversos trabalhos publicados, tais como [5], é comum se estabelecer a parada do PSO por um número máximo de iterações, ou então o critério de que a diferença entre dois valores consecutivos da função custo seja menor que um valor pré-estabelecido. No entanto, este critério não nos parece muito adequado, pois verificamos ser comum a ocorrência de situações em que o valor da função custo permanece constante por várias iterações consecutivas e depois volta a mudar siginificativamente. A fim de ter uma alternativa mais interessante, propõe-se neste trabalho considerar que o PSO convergiu quando o valor da função custo permanecer inalterado durante um número pré-fixado de iterações consecutivas, aqui denominado *parâmetro de convergência* (N_{FIX}).

V. Resultados

Adotou-se como função custo o logaritmo neperiano da função de verossimilhança ($\ln P(O|\lambda)$) de modelos HMM de 3 estados, usando o algoritmo PSO com 30 partículas. Procurou-se o ajuste deste modelo a uma amostra de erros com 10^6 bits gerada por simulação em computador, correspondente a um sistema de transmissão QPSK em canal com desvanecimento Rayleigh, espectro Doppler de Jakes e espalhamento Doppler normalizado $f_D T = 10^{-4}$. Admitiu-se sincronização perfeita e $E_b/N_0 = 20dB$ na entrada do receptor.

Foi feita inicialmente uma comparação entre o Método Proposto (MP) e o Método de Referência (MR), usando os parâmetros $W = 0,7$ e $c_1 = c_2 = 2,0$. Foram realizadas 5000 execuções de cada método com inicializações independentes e 2 minutos de processamento, o que correspondeu em média a 38 e 50 iterações, respectivamente[2]. Verificou-se que em 95,75% das inicializações obteve-se maior valor da função custo com o MP. Além disso, observou-se que o cálculo de 75,67% das partículas foi desperdiçado no MR, pelo motivo apresentado na seção III. Ou seja, apesar de possibilitar a execução de mais iterações no tempo de processamento estabelecido, o MR teve desempenho significativamente inferior, o que pode ser atribuído ao desperdício de processamento neste método, por um lado, e ao fato de que o MP consegue atender as restriçoes estocásticas sem modificar as direções de atualização de todas as partículas, por outro.

[1] Note-se que cada partícula compreende todos os parâmetros do modelo. Para efeito de concisão este exemplo se restringe a uma linha da matriz A.

[2] Diversos cuidados foram tomados como o intuito de fazer com que o tempo de processamento refletisse apenas características dos métodos testados.

TABELA I

Efeito do Parâmetro de Convergência N_{FIX}.

		N_{FIX}						
		50	100	150	200	250	300	350
C	50	65,0%	61,2%	59,2%	58,7%	57,9%	57,9%	57,9%
	100	84.6%	81,7%	79,2%	78,7%	77,5%	77,5%	77,5%
	150	93.7%	91,7%	89,6%	89,1%	87,9%	87,9%	87,9%
	200	97.1%	95,8%	93,7%	93,7%	92,5%	92,5%	87,9%
	250	98.7%	97,5%	96,2%	96,2%	95,8%	92,9%	88,3%
	300	99,2%	98,7%	98,3%	98,3%	95,8%	92,9%	88,3%
	350	100%	99,6 %	98,7%	98,3%	95,8%	92,9%	88,3%

A fim de avaliar a relação entre o número de iterações até a convergência e o valor de N_{FIX} associado ao critério aqui proposto, foram realizadas 240 execuções independentes do algoritmo PSO com 500 iterações, usando em cada uma delas trincas de parâmetros (W, c_1, c_2) aleatoriamente escolhidas.

A Tabela I apresenta as porcentagens de casos em que a convergência se deu até a iteração C, para diferentes valores de N_{FIX}. Nota-se que o valor de C varia significativamente com a escolha de N_{FIX}. Vê-se, por exemplo, que para $N_{FIX} = 100$, em $97,5\%$ dos casos o critério de convergência foi atendido com 250 iterações ou menos. Por outro lado, com $N_{FIX} = 300$ o percentual de casos que convergiu não muda quando o valor de C é variado de 250 a 350. Como era de se esperar, para um dado valor de C, o percentual de casos em que ocorre convergência diminui ou permanece constante com o aumento de N_{FIX}. Vale ressaltar que resultados similares foram obtidos com outras sequências de erros.

As constatações acima são de grande valia para a escolha adequada do número máximo de iterações, a fim de evitar desperdícios de processamento.

VI. Conclusões

Foi proposto um novo método para o tratamento das restrições estocásticas na aplicação do algoritmo PSO para estimação ML de modelos HMM. Além disso, foi proposto e brevemente avaliado um novo critério de parada para este algoritmo. Em trabalhos futuros pretende-se investigar em profundidade a escolha de parâmetros do algoritmo, com as modificações aqui propostas, tendo como foco a modelagem de canais com erros em surtos.

Referências

[1] Yeh-Huann Goh and P. Raveendran. HMM-based Speech Recognition Using Adaptive Framing. In *IEEE Region*, pages 1 –5, 2009.

[2] M.S. Fernandes, E.L. Pinto, and M.A. Grivet. A New Structured HMM Model Applied to Burst Errors in Viterbi Decoding. In *IEEE International Conference on Wireless Information Technology and Systems*, 28 2010.

[3] J. Bilmes. A Gentle Tutorial on the EM Algorithm Including Gaussian Mixtures and Baum-Welch. In *ICSI Technical Report TR-97-021*, 1997.

[4] S. Phon-Amnuaisuk. Estimating HMM Parameters Using Particle Swarm Optimisation. In *Proceedings of the EvoWorkshops 2009 on Applications of Evolutionary Computing*, 2009.

[5] L. Xue, J. Yin, Z. Ji, and L. Jiang. A Particle Swarm Optimization for Hidden Markov Model Training. In *8th International Conference on Signal Processing*, 16 2006.

[6] M. Maca, D. Novak, and L. Lhotska. Constraints in Particle Swarm Optimization of Hidden Markov Models. In *Intelligent Data Engineering and Automated Learning (IDEAL)*, volume 4224, pages 1399–1406, 2006.

[7] J. Kennedy and R. Eberhart. Particle Swarm Optimization. In *Proceedings IEEE International Conference on Neural Networks*, volume 4, pages 1942 –1948 vol.4, 1995.