

# Computação Evolucionária Aplicada à Estimação de Parâmetros de Modelos GMM

Jorge Frederico Vieira Campos Flores, Ernesto Leite Pinto, Juraci Ferreira Galdino e Dirceu Gonzaga da Silva

**Resumo**— Este artigo apresenta uma nova técnica de estimação ML do modelo GMM (algoritmo EM) baseada na redução de dimensão utilizando matrizes de projeção selecionadas pela implementação de algoritmos genéticos. Alguns resultados numéricos de avaliação de desempenho da técnica proposta, obtidos a partir de bases de dados de repositórios públicos, são apresentados, e confirmam o ganho de desempenho do método proposto em comparação ao algoritmo EM convencional e a uma técnica anteriormente proposta que emprega projeção aleatória.

**Palavras-Chave**— Modelo de mistura de gaussianas, estimação ML, projeção aleatória, algoritmo genético, estimação de ordem de modelo.

**Abstract**— This article presents a new GMM model ML estimation technique (EM algorithm) based on dimensionality reduction employing projection matrices selected by genetic algorithm implementation. Some performance evaluation numerical results concerning to proposed technique, achieved from public repository databases, are shown, and confirm proposed method performance gain compared to conventional EM algorithm and compared to an early proposed technique which make use of random projections.

**Keywords**— Gaussian mixture model, maximum-likelihood estimation, random projection, genetic algorithm, model-order selection

## I. INTRODUÇÃO

O emprego do algoritmo EM (*Expectation-Maximization*) [1] tem sido comum para a estimação em máxima verossimilhança (*Maximum Likelihood* ou ML) de parâmetros de modelos de misturas de gaussianas (*Gaussian Mixture Models* ou GMM)[2], [3]. Como os modelos GMM são amplamente utilizados em algoritmos de classificação de dados, é importante que seja feito um ajuste eficaz dos parâmetros envolvidos – pesos, vetores de médias e matrizes de covariâncias. Neste sentido, alguns trabalhos [4], [5] focam na melhoria deste ajuste, visando se aproximar da estimação ML global para compensar o fato de que o algoritmo EM inicializado pelo algoritmo k-means [6] ou pelo algoritmo proposto por Linde, Buzo e Gray [7] atinge valores máximos locais de verossimilhança [2]. Uma dessas técnicas, proposta por Dasgupta, [4] introduz o aprendizado de modelos GMM por meio de técnicas de redução de dimensionalidade.

O presente artigo propõe uma nova técnica de estimação ML com melhores características de desempenho do que o

algoritmo EM com ou sem projeções aleatórias. Esta nova técnica se baseia no uso de um algoritmo genético para escolher (em vez de arbitrar aleatoriamente) uma matriz de redução de dimensionalidade [4] visando, nessa escolha, aproximar-se da solução ML global. O enfoque da técnica proposta é diferente da técnica descrita em [5], pois esta estima todos os parâmetros do modelo GMM; já o algoritmo proposto apenas escolhe a matriz de redução de dimensionalidade (designada por  $W$ ), implicando em menor número de parâmetros. Doravante, a técnica proposta é denominada por EM-GA (GA de *Genetic Algorithm*).

O restante deste artigo é organizado em quatro Seções. A Seção II define o problema que serve de motivação à proposta do presente trabalho; a Seção III propõe o algoritmo EM-GA, baseado na busca evolutiva da matriz de redução de dimensionalidade; a Seção IV avalia o algoritmo proposto; por fim, a Seção V tece as conclusões sobre o algoritmo proposto, também sugerindo trabalhos futuros.

## II. DEFINIÇÃO DO PROBLEMA

A expressão matemática da função de densidade de probabilidade (*f.d.p.*)  $f(\mathbf{x})$  de um modelo GMM em função de um vetor  $\mathbf{x}$ , em termos do número  $M$  de gaussianas, dos parâmetros peso, vetor-médio e matriz-covariância  $\lambda = \{\pi_j, \mu_j, \Sigma_j \mid j = 1, \dots, M\}$ , é dada por

$$f(\mathbf{x}; \lambda) = \sum_{j=1}^M \pi_j N(\mathbf{x} \mid \mu_j, \Sigma_j) \quad (1)$$

A verossimilhança logarítmica  $\ell$ , associada ao modelo GMM, para um conjunto de  $n$  vetores  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ , é definida por

$$\ell(X; \lambda) = \sum_{i=1}^n \log f(\mathbf{x}_i; \lambda) \quad (2)$$

O algoritmo EM para o modelo GMM [2], [3] reestima os parâmetros envolvidos em (1) através da estimação ML e, associado a eles, a verossimilhança em (2) que, na realidade, atinge um valor máximo local.

### A. Conceito de Projeção Aleatória

Com o intuito de estimar parâmetros do modelo GMM cuja verossimilhança seja maior – com parâmetros mais próximos da estimação ML global – conjuga-se o algoritmo EM [2], [3] ao método de projeções aleatórias descrito abaixo, efetuado sobre um conjunto  $X$  de  $N$  vetores  $D$ -dimensionais.

J. F. V. C. Flores, E. L. Pinto, J. F. Galdino Seção de Engenharia Elétrica, Instituto Militar de Engenharia, Praça General Tibúrcio, 80, Rio de Janeiro, Brasil, 22290-270. E-mails: jorge.frederico@gmail.com; {ernesto.galdino}@ime.eb.br D. G. da Silva, Centro de Estudos de Telecomunicações da Pontifícia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente, 225, Rio de Janeiro, Brasil, 22453-900, E-mail: dirceu@cetuc.puc-rio.br

### B. Algoritmo EM com Projeção Aleatória [4]

Para cada conjunto de vetores, o algoritmo EM com projeção aleatória consiste na seqüência de passos abaixo:

- 1) Escolher os elementos de  $W$  aleatoriamente ( $W \in \mathbb{R}^{d \times D}$ ,  $d < D$ ) a partir de uma função de densidade de probabilidade gaussiana de média zero e variância unitária;
- 2) Ortonormalizar as linhas da matriz  $W$  através do método de Gram-Schmidt;
- 3) Calcular o conjunto de vetores  $Y$  tal que  $y_i = Wx_i$ ,  $1 \leq i \leq N$ ;
- 4) Calcular, via algoritmo EM, os parâmetros do modelo GMM para o conjunto de vetores  $Y$ ;
- 5) Calcular o vetor de pertinência de  $Y$  em relação aos componentes gaussianos do modelo GMM acima, ou seja, construir um vetor  $\varphi = [k_1, k_2, \dots, k_N]$  tal que  $k_j = \operatorname{argmax}_\theta (\ell(x_j; \lambda_\theta))$ ;  $j \in \mathbb{N}$ ;  $\theta \in \mathbb{N}$ ;  $1 \leq j \leq N$ ;  $1 \leq \theta \leq M$ ;
- 6) Tomar o vetor de pertinência do modelo GMM de  $Y$  e repeti-lo na inicialização do algoritmo EM sobre  $X$ . Ou seja, computar os pesos, vetores de médias e matrizes de covariância empiricamente, com base no conjunto de vetores  $X$  separados por índice;
- 7) Aplicar o algoritmo EM aos vetores de  $X$  para o cálculo dos parâmetros do modelo GMM;
- 8) Armazenar o valor final da verossimilhança;

### C. Comentários sobre o algoritmo de projeção aleatória

Em comparação ao algoritmo EM, as vantagens do algoritmo EM com projeção aleatória se resumem ao menor tempo computacional, aos clusters dos dados projetados menos excêntricos que os clusters dos dados em sua dimensão original (maior dimensão), e à geração de modelos GMM comparáveis ou melhores que os modelos GMM calculados pelo algoritmo EM (sem projeções aleatórias) [4]. Entretanto, como mostra a Figura 1, nem sempre é garantido que a matriz  $W$  calculada pela projeção aleatória leve a uma redução de excentricidade (melhor modelo GMM), pois o procedimento do cálculo da matriz não é sistemático (é, na verdade, um sorteio.) [4].

Caso se queira tentar obter valores de verossimilhança maiores que os computados com o algoritmo EM e que sejam maiores do que os calculados por apenas uma execução do algoritmo acima, pode-se efetuar uma repetição de projeções aleatórias conforme abaixo, doravante chamada de EM-RP (RP de *Random Projection*), tornando o mecanismo de projeção aleatória mais sistemático:

- 1) Inicializar o número de execuções do algoritmo na Seção II-B;
- 2) Repetir  $n_{RP}$  vezes, para cada execução, os itens de 1 a 7 da Seção II-B;
- 3) Para cada repetição do passo 2 – computar o valor da verossimilhança;
- 4) Para cada repetição do passo 2 – armazenar, caso o valor da verossimilhança aumente, os parâmetros da mistura de gaussianas e o valor da verossimilhança; caso contrário, manter os parâmetros existentes.

### III. TÉCNICA PROPOSTA

Devido ao último comentário da Seção II, torna-se necessário definir uma estratégia de escolha da matriz  $W$ , tendo em vista melhorar a eficácia da estimação ML. É proposta a escolha da matriz  $W$  através do algoritmo genético que, de forma diferente da busca aleatória comentada na Seção II, busca otimizar a verossimilhança (função-custo) por meio de um esquema evolucionário análogo à teoria da evolução dos seres vivos. Esses algoritmos empregam operadores que imitam os processos de seleção natural de indivíduos, migração populacional, mutação e recombinação de genes. A terminologia dada às variáveis do processo também segue essa idéia – costuma-se dar à variável vetorial da função-custo o nome de cromossomo; cada elemento constituinte deste cromossomo recebe o nome de gene. No algoritmo proposto, o cromossomo  $V$  é um vetor composto pelos elementos (genes) da matriz  $W$ .

O algoritmo EM-GA proposto se encontra descrito abaixo:

- 1) Inicializar os parâmetros peculiares do algoritmo genético: número de gerações ( $n_g$ ), quantidade de elementos da população inicial ( $n_W$ ), tipo de seleção dos indivíduos, tipo de codificação do cromossomo, fator de recombinação ou *crossover* ( $k_c$ ). Neste caso, a fração de elementos que podem sofrer mutação é dada por  $(1 - k_c)$ . Devem ser escolhidos também, para efeitos de variância inicial, o valor inicial e o fator de redução da variância ao decorrer das gerações;
- 2) Escolher a dimensão de projeção e computar a população inicial:  $n_W$  matrizes de redução de dimensionalidade (vide itens 1 e 2 da Seção II-B);
- 3) Computar o algoritmo genético regido pela população inicial e pelos parâmetros  $n_g$ ,  $n_W$  e  $k_c$  (vide itens acima). A função-objetivo verossimilhança será otimizada em função da variável vetorial  $V$  e calculada de acordo com os itens de 2 a 7 da Seção II-B;
- 4) Armazenar os valores finais da estimação ML e da verossimilhança.

### IV. AVALIAÇÃO DE DESEMPENHO DA TÉCNICA PROPOSTA

Para comprovar a eficácia da técnica proposta na melhoria da estimação ML, foram utilizadas bases de dados reais, extraídas do repositório público UCI Machine Learning Repository<sup>1</sup>, citadas abaixo. As matrizes  $X$  de vetores foram montadas a partir de cada classe das bases SatImage (informações de tipo de terreno referentes a pixels de imagens de satélite), Pima-indians-diabetes (informações relativas à não-incidência ou incidência de diabetes em uma localidade) e Statlog-Segment (informações de classificação de tipo de imagem) chamadas nesse trabalho, respectivamente, de LANDSAT, PIMA e SEGMENT. A Tabela I fornece informações sobre as bases de dados.

Foram comparadas três técnicas diferentes – o algoritmo EM [1] aplicado diretamente aos dados originais, o algoritmo EM-RP e o algoritmo EM-GA proposto. Em todos os casos, são fornecidas tabelas com os valores finais de

<sup>1</sup><http://www.ics.uci.edu/~mllearn/MLRepository.html>

TABELA I  
BASES DE DADOS

	Dimensões ( $D$ )	Total de vetores	Classes
LANDSAT	36	4453	6
PIMA	8	768	2
SEGMENT	19	2310	7

verossimilhança. Os maiores valores de verossimilhança de cada conjunto de vetores (linha da tabela) são representados em negrito. Inicialmente, foi estimado um modelo GMM ( $M = 5$ ) para cada conjunto de vetores. Posteriormente, os resultados comparativos foram decorrentes da seleção da ordem de modelo GMM pela Medida BIC (*Bayesian Information Criterion*) [8], [9], [10].

O algoritmo EM-RP comprime os conjuntos de vetores da sua dimensão original a sub-espacos de dimensões menores com  $n_{RP} = 28002$ . A razão deste número de projeções será vista adiante.

De acordo com itens 1 e 2 da Seção III, o treinamento do algoritmo EM-GA foi implementado, em ambos os casos, com  $k_c = 0,8$  (é interessante manter a fração de mutação muito mais baixa que 1,0 – no caso 0,2 – de forma a retirar do algoritmo proposto a tendência pela busca aleatória pura) e  $n_W = 30$ . Foi empregada mutação gaussiana cuja variância  $\sigma_i^2$  foi reduzida de forma controlada, para cada geração  $i$ , pela equação (3)

$$\sigma_i^2 = \sigma_{i-1}^2 \left(1 - S \frac{i}{n_g}\right); \sigma_0^2 = \frac{1}{2}; 1 \leq i \leq n_g \quad (3)$$

Foi usada seleção estocástica uniforme. Todos os cromossomos utilizados foram codificados como números reais. Para o número fixo de gaussianas e para a estimação da ordem do modelo pela Medida BIC, o valor de  $n_g$  foi fixado em 1000, ou seja, adotaram-se mil gerações. Foram considerados 2 elementos vencedores em cada geração. Isto forma um total de  $30 + (30 - 2)(1000 - 1) = 28002$  projeções no total (justificando o valor  $n_{RP}$  acima).

Além do acima exposto, vale lembrar que o algoritmo genético possui outras estratégias de mutação e recombinação. Porém, para efeito de investigação inicial de desempenho do algoritmo proposto, foi considerada a equação 3, que é um caso particular. Nessa situação empregou-se  $S = 0,75$ .

É importante mencionar que os operadores genéticos de mutação e de cruzamento não produzem uma geração de matrizes ortonormais, mesmo que a geração anterior seja composta de matrizes ortonormais. Nessa situação, poder-se-ia obter uma geração de matrizes ortonormais mediante uso do procedimento de Gram-Schmidt. Isso, porém, prejudicaria a filosofia de evolução que caracteriza os algoritmos genéticos. Assim sendo, foi relaxada a restrição de as matrizes  $W$  serem ortonormais.

Caso o algoritmo proposto usasse a estratégia descrita por Lin [5], o cromossomo possuiria  $(2D + 1)M$  elementos ( $M$  pesos,  $MD$  elementos de vetores de médias e  $MD$  elementos de matrizes de covariâncias diagonais). Com a otimização da função-objetivo do algoritmo EM-GA feita apenas em função da matriz  $W$ , empregam-se apenas  $dD$  elementos.

TABELA II  
CONDIÇÃO PARA ECONOMIA DE PARÂMETROS

Base de dados	Condição	Valor de $M$ mínimo
PIMA – $d = 2$	$M > 16/17$	1
PIMA – $d = 3$	$M > 24/17$	2
SEGMENT – $d = 2$	$M > 38/39$	1
SEGMENT – $d = 3$	$M > 57/39$	2
LANDSAT – $d = 2$	$M > 72/73$	1
LANDSAT – $d = 3$	$M > 108/73$	2

TABELA III  
VEROSSIMILHANÇA - GMM A 5 GAUSSIANAS - PIMA  
 $d = 2$ 

Classe	EM	EM-RP	EM-GA
1	-12568	<b>-11871</b>	<b>-11871</b>
2	-6980	-6432	<b>-6420</b>
Todas	-19716	-18671	<b>-18466</b>

O número de elementos da variável vetorial do algoritmo proposto sempre será menor que o do algoritmo em [5] se  $(2D + 1)M > dD$ , que leva a  $M > \frac{dD}{2D+1}$ . Substituindo o valor de  $D$  para as bases PIMA, SEGMENT e LANDSAT (vide Tabela I) tem-se, para  $d = 2$  e  $d = 3$ , os resultados da tabela II.

#### A. Número fixo de gaussianas

Nota-se nas Tabelas de III a V que, para todas as bases de dados testadas, o algoritmo proposto EM-GA foi tão ou mais eficaz do que o algoritmo EM na estimação ML. O mesmo não ocorreu com o EM-RP, pois este não superou o algoritmo EM em um dos 18 conjuntos de vetores analisados, revelando uma maior fragilidade da busca aleatória. Além disso, nenhum valor de verossimilhança calculada pelo EM-RP superou os valores calculados pelo algoritmo EM-GA.

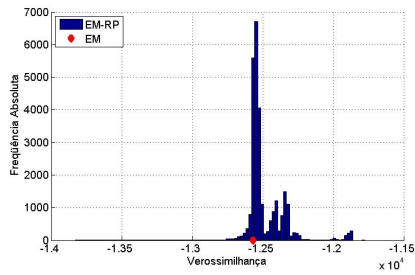
Outra qualidade do algoritmo EM-GA sobre o EM-RP pode ser observada na Figura 2. Nela, percebe-se que, na média entre um número de execuções dos algoritmos EM-RP e EM-GA muito maior do que 1, o algoritmo EM-GA apresenta maior vantagem em termos dos valores de verossimilhança apresentados e maior velocidade de convergência do que o EM-RP. Cabe destacar que a complexidade computacional do algoritmo genético é governada basicamente pelo ajuste dos indivíduos da população na função-custo, tendo em vista que os operadores genéticos são extremamente simples. Como esse ajuste é realizado tanto para o algoritmo genético quanto para a busca aleatória, a complexidade computacional do algoritmo EM-GA é comparável à do EM-RP.

#### B. Estimação do modelo pela Medida BIC [8], [9], [10]

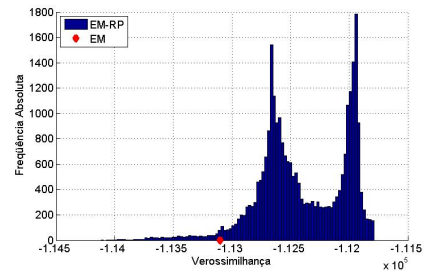
O valor da Medida BIC [8] em função da verossimilhança (2) é dado por

$$BIC(M) = -2\ell(X; \lambda) + \gamma\nu \log(n) \quad (4)$$

A segunda parcela em (4) representa a penalidade aplicada em função do número de vetores  $n$  de cada conjunto de vetores

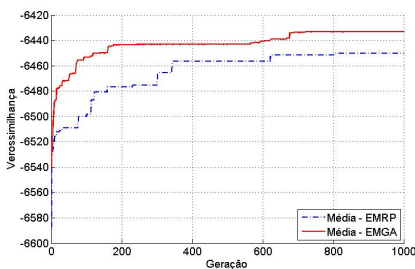


(a) Histograma – PIMA (classe 1)

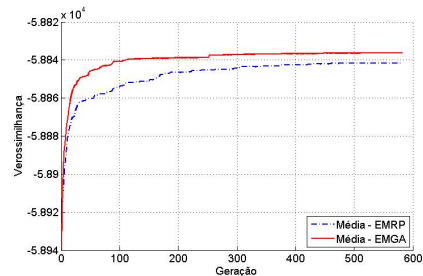


(b) Histograma – LANDSAT (classe 6)

Fig. 1. Histograma das verossimilhanças associadas a todos os modelos estimados pelo algoritmo EM-RP para os conjuntos de vetores indicados. Nem todas as projeções levam a verossimilhanças maiores que a do algoritmo EM.

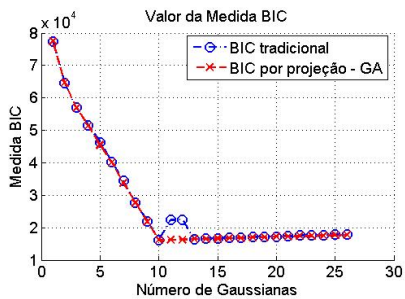


(a) Função objetivo média – PIMA (classe 2)

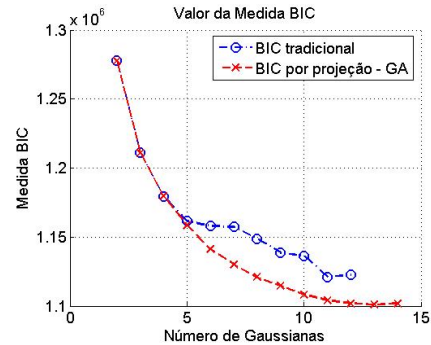


(b) Função objetivo média – LANDSAT (classe 2)

Fig. 2. Comparação entre as médias das verossimilhanças calculadas pelos algoritmos EM-RP e EM-GA. Foi computada a média de 5 e 30 realizações para as bases PIMA e LANDSAT, respectivamente. Pode ser percebido que o algoritmo EM-GA é mais eficiente na média das realizações.



(a) Medida BIC – modelo sintético ( $d = 2$ )



(b) Medida BIC – LANDSAT (todas as classes,  $d = 2$ )

Fig. 3. Comparação entre as Medidas BIC calculadas pelo algoritmo EM e pelo algoritmo genético (EM-GA) até 25 gaussianas. Em (a) a Medida BIC foi calculada sobre um modelo GMM sintético de 10 gaussianas (4000 vetores do  $\mathbb{R}^{10}$ ). Notar que os valores de Medida BIC são mais baixos e de comportamento mais suave devido à implementação do algoritmo proposto. Além disso, o mínimo local [8] (coincidente com o mínimo global) da Medida BIC em (a) é de 10 gaussianas para ambos os algoritmos, que foi exatamente a ordem arbitrada para o modelo sintético.

originais e do coeficiente  $\nu$ , expresso (5) em termos do número de gaussianas  $M$  da mistura e da dimensão  $D$  do conjunto de vetores de dados originais. A segunda parcela é adequada ao caso do modelo GMM com matrizes de covariância diagonais diferentes para cada gaussiana da mistura [8]. A constante  $\gamma$  pode ser ajustada um maior ou um menor valor [10] caso se queira penalizar a quantidade  $n$  de vetores com maior ou menor intensidade, respectivamente.

$$\nu = 2MD + M - 1 \quad (5)$$

Da equação (4) resulta que, para dois algoritmos que estimem, a partir do mesmo conjunto de vetores  $X$ , modelos GMM distintos com verossimilhanças  $\ell_1$  e  $\ell_2$  ( $M_1 = M_2 = M$ ), tem-se

$$\ell_1 > \ell_2 \Leftrightarrow BIC_1(M) < BIC_2(M) \quad (6)$$

TABELA IV  
VEROSSIMILHANÇA - GMM A 5 GAUSSIANAS - LANDSAT  
 $d = 8$

Classe	EM	EM-RP	EM-GA
1	-118843	-118757	<b>-118756</b>
2	-59181	-58841	<b>-58840</b>
3	-104547	<b>-104008</b>	<b>-104008</b>
4	-45632	<b>-45089</b>	<b>-45089</b>
5	-57151	<b>-57122</b>	<b>-57122</b>
6	-113098	<b>-111790</b>	<b>-111790</b>
Todas	<b>-566986</b>	-566988	<b>-566986</b>

TABELA V  
VEROSSIMILHANÇA - GMM A 5 GAUSSIANAS - SEGMENT  
 $d = 4$

Classe	EM	EM-RP	EM-GA
1	-6740	<b>-6474</b>	<b>-6474</b>
2	-9008	<b>-8808</b>	<b>-8808</b>
3	-11296	<b>-9820</b>	<b>-9820</b>
4	-12872	<b>-12357</b>	<b>-12357</b>
5	-8801	<b>-8341</b>	<b>-8341</b>
6	-9743	<b>-9349</b>	<b>-9349</b>
7	-10247	<b>-9562</b>	<b>-9562</b>
Todas	-99438	-96065	<b>-96041</b>

TABELA VI  
VEROSSIMILHANÇA - MEDIDA BIC - PIMA  
 $d = 2 / d = 3$

Classe	EM	EM-RP	EM-GA	$n_g$	$\gamma$
1	-12977(2)	-11904(5) -12658(3)	<b>-11873(5)</b> <b>-11873(5)</b>	2	2
2	-7121(2)	-6887(3) -7121(2)	<b>-6593(4)</b> <b>-6965(2)</b>	5	2
Todas	-19778(4)	-19525(3) -19314(4)	<b>-18321(6)</b> <b>-18354(6)</b>	30	2

Em termos práticos, a equação (6) mostra que, para um mesmo número de gaussianas, e conjunto de vetores, uma estimação ML mais bem ajustada implica em menor Medida BIC.

Nesta subseção foram calculados, além das verossimilhanças, os números de gaussianas resultantes da seleção de ordem de modelo realizada pela Medida BIC. Estes dados constam das tabelas VI, VII e VIII (as ordens dos modelos GMM selecionadas pela Medida BIC se encontram entre parênteses; todas obedecem às condições da Tabela II).

O algoritmo EM-GA proposto foi, individualmente, o que apresentou os melhores resultados. Em 34 das 36 rodadas realizadas – 94% dos casos – (vide Tabelas VI, VII e VIII), o EM-GA superou o desempenho do algoritmo EM-RP (individualmente mais eficaz 4 vezes – 11% dos casos). Em relação ao algoritmo EM convencional, o algoritmo proposto foi mais eficaz em 100% dos casos. O EM-GA, comparado apenas com o EM-RP, obteve maior eficácia 32 vezes – 89% dos casos. As tabelas e a Figura 3 mostram que o algoritmo EM-GA produz

TABELA VII  
VEROSSIMILHANÇA - MEDIDA BIC - LANDSAT  
 $d = 2 / d = 3$

Classe	EM	EM-RP	EM-GA	$n_g$	$\gamma$
1	-117686(4)	-117355(6) -117288(6)	<b>-117283(6)</b> <b>-117284(6)</b>	5	4
2	-59820(4)	-59811(4) -59814(4)	<b>-58901(5)</b> <b>-58901(5)</b>	2	4
3	-105269(4)	-102952(6) -102960(6)	<b>-102920(6)</b> <b>-102927(6)</b>	5	4
4	-46102(4)	-45852(4) -45850(4)	<b>-45850(4)</b> <b>-45849(4)</b>	5	4
5	-59177(3)	-57127(5) <b>-57122(5)</b>	<b>-57123(5)</b> <b>-57122(5)</b>	5	4
6	-115639(3)	-110773(6) -110725(6)	<b>-110734(6)</b> <b>-110708(6)</b>	5	4
Todas	-533642(11)	-518954(13) <b>-518636(13)</b>	<b>-518677(13)</b> -518779(13)	30	8

TABELA VIII  
VEROSSIMILHANÇA - MEDIDA BIC - SEGMENT  
 $d = 2 / d = 3$

Classe	EM	EM-RP	EM-GA	$n_g$	$\gamma$
1	-6640(5)	-5915(7) -5908(7)	<b>-5910(7)</b> <b>-5905(10)</b>	5	2
2	-8833(5)	-8554(6) -8596(6)	<b>-8553(6)</b> <b>-8330(7)</b>	2	2
3	-11673(3)	<b>-8690(9)</b> -9942(5)	-8933(7) <b>-8746(8)</b>	30	2
4	-12745(4)	-11798(7) -12051(6)	<b>-11744(7)</b> <b>-11530(8)</b>	5	2
5	-8001(6)	-7641(7) -7295(8)	<b>-7274(8)</b> <b>-6784(10)</b>	10	2
6	-9789(4)	<b>-8744(7)</b> -9018(6)	<b>-8744(7)</b> <b>-8752(7)</b>	10	2
7	-9337(6)	-9369(6) -9577(5)	<b>-9314(6)</b> <b>-9318(6)</b>	5	2
Todas	-89466(8)	-87929(8) -84255(10)	<b>-85734(9)</b> <b>-82034(11)</b>	30	8

uma estimação ML com melhor ajuste em comparação com o EM. Esta figura mostra o comportamento da Medida BIC em função do número de gaussianas. Nota-se, Na Figura 3(a), que ocorre redução da Medida BIC e maior suavidade da função quando a mesma é computada pelo algoritmo proposto. Na Figura 3(b), percebe-se que a ordem do modelo selecionado é maior quando estimada pela técnica EM-GA. Além disso, a figura mostra na prática o resultado da equação (6) – ocorre, de fato, a melhor estimação ML para cada valor da ordem do modelo.

## V. CONCLUSÕES

Neste artigo foi proposto um novo algoritmo para treinamento do modelo GMM através da escolha da matriz de redução de dimensionalidade de forma evolutiva, utilizando algoritmos genéticos. Este trabalho constatou que o algoritmo

proposto tende a ser mais eficaz na estimação ML que o algoritmo EM. O algoritmo proposto também se mostrou mais eficaz que o algoritmo de projeções aleatórias proposto anteriormente [4], pois aprimora as técnicas citadas na Seção II-C.

Na continuação deste trabalho serão considerados um estudo mais detalhado dos parâmetros do algoritmo genético, algumas aplicações de classificação e avaliações comparadas da técnica proposta com outras técnicas de otimização, como por exemplo busca direta e recozimento simulado.

#### AGRADECIMENTOS

Os autores gostariam de agradecer aos professores José Antonio Apolinário Jr. e Alberto Gaspar Guimarães e outros membros da Seção de Engenharia Elétrica do IME pelas discussões técnicas e pela assistência durante várias fases do trabalho.

#### REFERÊNCIAS

- [1] A. P. Dempster, N. M. Laird and D. B. Rubin, (1977), "Maximum Likelihood from Incomplete Data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, v. 39, p. 1-38, 1977
- [2] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, v. 3, n. 1, jan. 1995.
- [3] D. A. Reynolds, "Speaker identification and verification using gaussian mixture speaker models," *Speech Communication*, v. 17, p. 91-108, 1995.
- [4] S. Dasgupta, "Experiments with random projection," *Proc. 16<sup>th</sup> Conf. on Uncertainty in Artificial Intelligence*, p. 143-151, jun. 2000.
- [5] L. Lin, S. Wang, "Genetic algorithms and fuzzy approach to gaussian mixture model for speaker recognition," *Proc. 16<sup>th</sup> of NLP-KE'05*, p. 142-146, jun. 2000.
- [6] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, v.1, p. 281-297, 1967.
- [7] Y. Linde, A. Buzo and R. M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Transactions on Communications*, p. 702-710, jan. 1980.
- [8] Z. Halbe and M. Aladjem, "Model-based mixture discriminant analysis – an experimental study," *Pattern Recognition*, v. 38, p. 437-440, Elsevier Ltd., 2005.
- [9] P. Stoica and Y. Selén, "Model order selection – a review of information criterion rules," *IEEE Signal Processing Magazine*, jul. 2004.
- [10] J. Ajmera, I. McCowan and H. Bourlard, "Robust speaker change detection," *IEEE Signal Processing Letters*, v. 11, p. 649-651, ago. 2004.