

CLASSIFICADOR BAYESIANO HIERÁRQUICO UTILIZANDO MODELOS GAUSSIANOS

EDWIN VILLANUEVA TALAVERA* RAFAEL EDUARDO RUVIARO CHRIST† CARLOS DIAS MACIEL‡

* *Universidade de São Paulo*
Escola de Engenharia de São Carlos
São Carlos, São Paulo, Brazil

Email: evillanu@sel.eesc.usp.br rchrist@sel.eesc.usp.br maciel@sel.eesc.usp.br

Abstract— This paper presents the Gaussian Hierarchical Bayesian Clustering algorithm (CHBMG) derived from HBC algorithm. CHBMG has many advantages over traditional agglomerative algorithms. (1) It reduces the limitations due time and memory complexity. (2) It uses a bayesian posterior probability criterion to decide on merging clusters (modeling them as Gaussian distributions) rather than *ad-hoc* distance metrics. (3) It automatically finds the partition that most closely matches the data using the Bayesian Information Criterion (BIC). Finally, experimental results show that it can classify synthetic and real data as the best agglomerative and partitional algorithms.

Keywords— hierarchical clustering, bayesian methods, ART2

Resumo— Este trabalho apresenta um Classificador Hierárquico Bayesiano utilizando Modelos Gaussianos (CHBMG) derivado do algoritmo HBC. Suas características são: (1) Reduzir as limitações devido à complexidade de tempo e memória. (2) Utilizar o critério de probabilidade bayesiana *a posteriori* para escolher os grupos que são fundidos. Modelando os grupos como distribuições Gaussianas em vez de alguma métrica de distância *ad-hoc*. (3) Selecionar automaticamente a partição que melhor define os dados através do critério de informação bayesiana (BIC). Finalmente, resultados experimentais tanto em bases sintéticas como reais mostraram que CHBMG tem desempenho similar aos melhores classificadores aglomerativos e particionais.

Palavras-chave— agrupamento hierárquico, métodos bayesianos, ART2

1 Introdução

Agrupamento de dados pode ser definido como o particionamento de dados em grupos, onde elementos em um grupo são mais similares entre si do que elementos em grupos diferentes (Halkidi et al., 2001). Algoritmos hierárquicos aglomerativos são um dos mais importantes métodos de agrupamento (Everitt et al., 2001). Eles são largamente empregados e de fácil utilização (Jain et al., 1999; Willett, 1988). De forma geral, estes algoritmos iniciam criando grupos unitários, os quais são sucessivamente juntados até formar um único grupo. A saída desses algoritmos é uma árvore binária chamada dendrograma, a qual organiza os dados em uma hierarquia de grupos.

Nos métodos aglomerativos clássicos (Kaufman and Rousseeuw, 1990), a decisão de juntar grupos é baseada em alguma métrica de distância. Frequentemente, é difícil determinar a métrica mais adequada para avaliar a similaridade entre os elementos. Métricas distintas normalmente geram resultados diferentes. Outra limitação nestes métodos, é a dificuldade na determinação da partição ótima ou mais natural no dendrograma.

Métodos baseados em modelos probabilísticos foram propostos para sobrepujar essas limitações. Nesses métodos, algum critério de máxima verossimilhança é usado para fundir os grupos. HBC (*Hierarchical Bayesian Clustering*) destaca-se entre esses métodos (Iwayama and Tokunaga, 1995).

Ele funde os grupos tentando maximizar a probabilidade *a posteriori* das partições. HBC foi empregado com sucesso em classificação de documentos de texto (Iwayama and Tokunaga, 1995). É capaz de reconstruir as classes reais com mais precisão e melhor desempenho que métodos não probabilísticos.

O problema de eficiência é contudo relevante em muitas aplicações (Fraley, 1996). O valor prático dos métodos hierárquicos aglomerativos é limitado por uma complexidade de tempo de pelo menos o quadrado do número de elementos. Também, o uso de memória é proporcional à raiz do número inicial de grupos.

Devido a essas limitações, este trabalho propõe o Classificador Hierárquico Bayesiano utilizando Modelos Gaussianos (CHBMG) derivado do algoritmo HBC. A escolha de modelos Gaussianos é justificada pelo sucesso em várias aplicações (Murtagh and Raftery, 1984; Banfield and Raftery, 1993; Dasgupta and Raftery, 1998) e pela simplicidade na estimativa dos parâmetros. Diferentemente de HBC, o método proposto utiliza modelos Gaussianos para descrever os grupos. Também, tem menor custo computacional devido que CHBMG inicia a aglomeração a partir de grupos não unitários. Finalmente, ele permite o uso do Fator de Bayes aproximado pelo critério BIC (*Bayesian Information Criterion*) para determinar automaticamente a partição que melhor agrupa os dados (Schwarz, 1978).

ART significa *Adaptive Resonance Theory*.

Ela foi introduzida por Carpenter e Grossberg (Grossberg, 1976) como uma teoria sobre o sistema cognitivo. Atualmente, evoluiu como uma série de redes neurais em tempo-real. A rede original ART classifica dados binários com aprendizagem não-supervisionada (Carpenter and Grossberg, 2003). ART2 é sua extensão para dados analógicos (Carpenter and Grossberg, 1987b).

O algoritmo proposto é resumido a seguir: (1) Uma rede ART2 modificada é usada para gerar várias partições. Então, distâncias intra e inter-grupo selecionam uma partição inicial com grupos homogêneos. (2) CHBMG começa o agrupamento a partir dessa partição inicial, modelando os grupos com distribuições de probabilidades Gaussianas. Ele funde em cada passo o par de grupos que resulta na maior probabilidade *a posteriori* da partição. (3) BIC avalia as partições geradas a cada passo para selecionar automaticamente uma partição final ótima.

Este trabalho está organizado da seguinte forma: Seção 2 discute a criação de uma partição inicial para CHBMG usando uma rede ART2 modificada. O uso das distâncias inter e intra para selecionar uma partição com grupos homogêneos. O funcionamento do algoritmo proposto. Finalmente, o uso de BIC para determinar a partição ótima. Seção 3 descreve os bancos de dados usados, os experimentos e seus resultados. Seções 4 e 5 são conclusão e agradecimentos.

2 Materiais e Métodos

É necessário criar uma partição inicial homogênea e granular para CHBMG. Como partição homogênea se entende que os elementos dentro dos grupos possuem uma alta similaridade. Esta homogeneidade é exigida devido a que a qualidade do dendrograma resultante é influenciada pela homogeneidade dos grupos iniciais, uma vez que CHBMG somente funde grupos. Como partição granular se entende que ela está formada por uma grande quantidade de grupos pequenos (com poucos elementos). Isto é necessário para gerar um dendrograma com uma quantidade significativa de partições e assim assegurar a existência da partição ótima dentro do dendrograma. O algoritmo usado para gerar a partição inicial deve prover meios para se selecionar a granularidade e homogeneidade dos grupos.

A rede neural ART2 consegue aprender novos grupos (plasticidade), se mantendo estável em reposta aos grupos já aprendidos (estabilidade) (Chang et al., 2000). Também, permite o controle da homogeneidade dos grupos (Fausett, 1997) ajustando o parâmetro de vigilância (ρ).

Este trabalho usou uma ART2 modificada (ART2P), a qual apresentou melhor desempenho na geração de partições iniciais. A rede ART2 original usa somente o ângulo entre os dados como

medida de similaridade. Já, ART2P considera tanto informação do ângulo quanto da amplitude. Os dados não são normalizados para manter a informação da amplitude. ART2P emprega o coeficiente de correlação de Pearson, eq. 1, na comparação dos dados.

Considere-se o conjunto de dados a agrupar $D = \{d_1, d_2, \dots, d_N\}$, onde cada dado d_i é definido pelo vetor de p atributos $d_i = (x_{i1}, \dots, x_{ip})$. O coeficiente de correlação de Pearson $\phi(d_i, d_j)$ mede o grau no qual os dados d_i e d_j estão relacionados e é definido como (Hon, 2002-2005):

$$\phi(d_i, d_j) = \frac{\sum_{k=1}^p (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^p (x_{ik} - \bar{x}_i)^2 \sum_{k=1}^p (x_{jk} - \bar{x}_j)^2}} \quad (1)$$

no qual \bar{x}_i é a média do vetor d_i calculada como $\bar{x}_i = 1/p \sum_{k=1}^p x_{ik}$. Esse coeficiente pode ser transformado numa medida de distância chamada distância de Pearson: $\delta(d_i, d_j) = 1 - \phi(d_i, d_j)$.

Para avaliar a homogeneidade das partições geradas pela ART2P, são usadas duas distâncias: inter e intra-grupo. A distância intra-grupo de uma partição $C = \{c_1, \dots, c_M\}$ é calculada como a média das distâncias entre os dados e os centróides dos seus respectivos grupos c_i , assim:

$$intra(C) = \frac{\sum_{c_i \in C} \sum_{d_j \in c_i} \delta(d_j, \bar{c}_i)}{M} \quad (2)$$

onde M é o número de grupos da partição e \bar{c}_i é o centróide do grupo c_i . A distância inter-grupo da partição C é definida como a distância média entre os centróides dos grupos, assim:

$$inter(C) = \frac{\sum_{c_i, c_j \in C, c_i \neq c_j} \delta(\bar{c}_i, \bar{c}_j)}{M} \quad (3)$$

O algoritmo 2.1 foi criado para encontrar uma partição inicial para CHBMG e está inspirado no método de busca binária. Este algoritmo considera como partições homogêneas aquelas cujas distâncias inter-grupo são maiores que as intra-grupo. O algoritmo recebe como entrada dois parâmetros: ρ_{min} e ρ_{max} , os quais definem o intervalo de busca do parâmetro ρ que resulte em uma partição homogênea. No início o algoritmo é invocado com um intervalo de busca grande ($[0,1]$). A cada iteração o algoritmo escolhe aleatoriamente um valor de ρ dentro do intervalo de busca e gera uma partição com ART2P, logo calcula as distâncias inter e intra-grupo. Se a partição é homogênea e a restrição do modelo é atendida, então a busca finaliza e devolve a partição, senão, o algoritmo é chamado recursivamente com um intervalo de busca reduzido, o qual depende do valor $\frac{inter()}{intra()}$. A restrição do modelo implica que os grupos devem possuir um número mínimo de elementos para poder aproximá-los mediante modelos Gaussianos. Esta restrição é verificada com a função `testaModeloGaussiano()`, a qual verifica

que as matrizes de covariância, calculadas com os dados em cada grupo, não sejam singulares.

algoritmo 2.1 Busca por uma partição inicial para CHBMG. Ela deve ser homogênea e atender à restrição do modelo Gaussiano

```

particaoInicial( $\rho_{min}, \rho_{max}$ ) {
 $\rho \leftarrow \text{random}(\rho_{min}, \rho_{max})$ 
particao  $\leftarrow$  ART2P( $\rho$ )
if  $\frac{\text{inter}(\text{particao})}{\text{intra}(\text{particao})} \geq 1$  then
 $\rho_{min} \leftarrow \frac{\rho_{min} + \rho_{max}}{2}$ 
else  $\left\{ \frac{\text{inter}(\text{particao})}{\text{intra}(\text{particao})} \leq 1 \right\}$ 
 $\rho_{max} \leftarrow \frac{\rho_{min} + \rho_{max}}{2}$ 
end if
if testaModeloGaussiano(particao) = OK then
return particao
else
particaoInicial( $\rho_{min}, \rho_{max}$ )
end if }

```

CHBMG recebe uma partição inicial C_0 com M grupos. A cada passo, ele funde o par de grupos que resulta na maior probabilidade *a posteriori* $Pr(C|D)$ (a probabilidade de agrupar os dados $D = \{d_1, d_2, \dots, d_N\}$ em uma partição com grupos $\{c_1, c_2, \dots, c_M\}$). No passo k , mais dois grupos são fundidos, gerando a partição C_k . Os grupos selecionados são aqueles que produzem a maior $Pr(C_k|D)$ em relação aos outros pares de grupos.

Assumindo independência condicional dos elementos dados os grupos, tem-se $Pr(C_k|D) = \prod_{c \in C_k} \prod_{d_j \in c} Pr(c|d_j)$. Seguindo (Iwayama and Tokunaga, 1995) para determinar essa probabilidade e aplicando-se o teorema de Bayes, obtem-se:

$$Pr(C_k|D) = \frac{PC(C_k)}{Pr(D)} \prod_{c \in C_k} SC(c) \quad (4)$$

Na eq. 4 são identificados os termos (Iwayama and Tokunaga, 1995): (1) $Pr(D)$, probabilidade marginal dos dados (constante para qualquer partição); (2) $PC(C_k)$, probabilidade *a priori* da partição C_k , definida como $PC(C_k) = \prod_{c \in C_k} Pr(c)^{|c|}$. Onde $|c|$ é o tamanho do grupo c e (3) $\prod_{c \in C_k} SC(c)$, a verossimilhança da partição C_k , a qual é a probabilidade dos dados se a partição é verdadeira. $SC(c)$ é a verossimilhança do grupo c , a qual indica a probabilidade que o grupo c possa gerar todos seus elementos, e é definida como:

$$SC(c) = \prod_{d_j \in c} Pr(d_j|c) \quad (5)$$

Deve-se determinar qual par de grupos será fundido no passo k . Assim, considere o par de grupos candidatos c_x e c_y . O grupo resultante é $c_z = c_x \cup c_y$. A partição depois dessa fusão é atualizada como $C_k = C_{k-1} + c_z - c_x - c_y$.

A $Pr(C_k|D)$ da partição é calculada recursivamente em função da probabilidade *a posteriori* das partições anteriores C_{k-1} . Expressando $Pr(C_{k-1}|D)$ de acordo com a eq. 4 e dividindo-as, tem-se a probabilidade *a posteriori*:

$$Pr(C_k|D) = \frac{PC(C_k)}{PC(C_{k-1})} \frac{SC(c_z)}{SC(c_x)SC(c_y)} Pr(C_{k-1}|D) \quad (6)$$

O termo $\frac{PC(C_k)}{PC(C_{k-1})}$ é o fator no qual a probabilidade *a priori* da partição muda a cada passo. De acordo com o princípio *Minimum Description Length - MDL* (Rissanen, 1989), esse fator pode ser estimado como uma constante A^{-1} ($A > 1$) independente do par fundido. Consequentemente, a maximização da eq. 6 é obtida através do fator $U(c_x, c_y) = \frac{SC(c_x \cup c_y)}{SC(c_x)SC(c_y)}$.

O algoritmo calcula U , a cada passo, para todo par de grupos na partição atual e junta aqueles com os maiores valores de U , isto é:

$$(c_{\hat{x}}, c_{\hat{y}}) = \text{argmax}_{c_x, c_y} (U(c_x, c_y)) \quad (7)$$

A probabilidade elementar $Pr(d_j|c)$, eq. 5 (probabilidade do grupo c gerar seu elemento d_j), é usada para obter SC na eq. 6. Para determiná-la, deve-se assumir que os elementos em cada grupo são produzidos por uma distribuição (Gaussiana) normal multivariada. Especificamente, a distribuição de probabilidade que descreve o elemento d_j pertencente ao grupo c_i é $f_{c_i}(d_j|\Theta_{c_i})$. Os parâmetros Θ_{c_i} consistem do vetor de médias μ_{c_i} e da matriz de covariância Σ_{c_i} .

Os parâmetros de máxima verossimilhança são calculados para cada grupo a partir dos dados ((Everitt et al., 2001) para detalhes). O vetor de médias é obtido como a média dos dados $\mu_{c_i} = \frac{1}{|c_i|} \sum_{d_j \in c_i} d_j$ e a matriz de covariância como a covariância amostral.

A principal diferença entre os algoritmos CHBMG e HBC (Rissanen, 1989) é o cálculo da probabilidade elementar $Pr(d_j|c)$. CHBMG utiliza modelos Gaussianos para descrever os grupos, sendo mais abrangente que HBC já que este último está restrito ao agrupamento de documentos de texto. Além disso, CHBMG inicia o agrupamento a partir de uma partição inicial, ao invés de grupos unitários, reduzindo o número de passos para encontrar a partição ótima. Seu funcionamento é descrito no algoritmo 2.2. Onde, $SC(c) = \prod_{d_j \in c} f(d_j|\mu_c, \Sigma_c)$.

A vantagem de se utilizar uma abordagem probabilística é o uso de fatores Bayesianos na comparação dos modelos (Fraley and Raftery, 1998). Esses fatores são uma alternativa Bayesiana a testes de hipótese clássicos. Uma aproximação confiável do fator Bayesiano é o critério BIC definido como $BIC = 2\log(L) + k\log(n)$ (Schwarz, 1978), onde n é o tamanho dos dados, k é o número de

algoritmo 2.2 CHBMG

```

 $D \leftarrow d_1, d_2, \dots, d_N$  {base de dados}
 $C_0 \leftarrow c_1, c_2, \dots, c_M$  {partição inicial}
for  $k = 1$  to  $M - 1$  do
   $M \leftarrow \left\{ \frac{SC(c_i \cup c_j)}{SC(c_i)SC(c_j)} \right\}, \{c_i, c_j \in C_{k-1}\}, \{c_i \neq c_j\}$ 
   $(c_x, c_y) \leftarrow \text{argmax}_{c_i, c_j}(M)$ 
   $c_z \leftarrow c_x \cup c_y$ 
   $C_k \leftarrow C_{k-1} + c_z - c_x - c_y$ 
end for

```

parâmetros livres a serem estimados e L é o máximo valor da função de verossimilhança para o modelo estimado. A verossimilhança da partição C_k é $L_k = \prod_{c \in C_k} SC(c)$. O BIC é estimado para cada partição durante a aglomeração. No final, aquela com o maior BIC é considerada a partição ótima (Fraley and Raftery, 1998).

CHBMG é avaliado em 2 aspectos: (1) A qualidade do dendrograma, dada por um medida de pureza proposta em (Heller and Ghahramani, 2005). CHBMG e os métodos aglomerativos clássicos *Single*, *Complete*, *Average Linkage* são executados até se obter uma partição com um único grupo. As purezas das partições são então calculadas para avaliar a qualidade das árvores. (2) Qualidade do particionamento. Seleciona-se uma partição ótima no dendrograma, aquela com maior BIC (Fraley and Raftery, 1998). O seu número de grupos é usado para gerar os particionamentos pelos algoritmos particionais *K-Means* (Duda et al., 2001), EM (*Expectation Maximization*) (Dempster et al., 1977) e ART2. Finalmente, compara-se a pureza da partição ótima com esses particionamentos.

3 Experimentos e Resultados

Nesta seção são apresentados os resultados do agrupamento em 1 base real e 2 sintéticas: (1) Iris, 3 classes de 50 elementos com 4 atributos contínuos (Dasarathy, 1980). (2) Sintética-1, 3 atributos com 1000 elementos a partir da mistura de 5 Gaussinas isotrópicas. (3) Sintética-2, 4 atributos com 2000 elementos a partir da mistura de 4 Gaussinas isotrópicas.

Na base Iris, o algoritmo 2.1 executou 5 passos até encontrar a partição inicial para CHBMG. A partição escolhida, 5 na Fig. 1, tem 6 grupos, é homogênea o suficiente e atende às restrições do modelo Gaussiano. Observe-se que as partições 4 e 5 são homogêneas (distância inter-grupo é maior que a intra-grupo). Contudo, a partição 4 não satisfaz as restrições do modelo Gaussiano por possuir grupos com muito poucos elementos.

O CHBMG foi então executado até se obter uma partição com 1 único grupo. A pureza de cada partição no dendrograma foi calculada para avaliar a qualidade da árvore. A Fig. 2 mostra

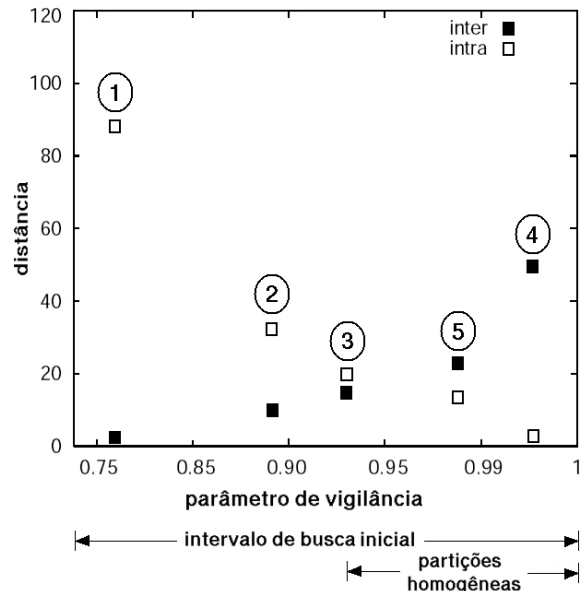


Figura 1: Algoritmo 2.1 busca uma partição inicial para CHBMG na base Iris. Aumentando-se ρ , a rede ART2P gera mais grupos. As partições 4 e 5 são homogêneas. A partição selecionada 5 atende às restrições do modelo Gaussiano.

esta pureza comparada com a pureza das últimas 6 partições dos algoritmos aglomerativos clássicos. Pode-se observar que as purezas das partições de 6 até 3 grupos são superiores em CHBMG com respeito aos outros algoritmos, tendo uma queda abrupta em todos os algoritmos para partições de 2 grupos. Isto porque o número certo de grupos é 3, e em partições com 2 e 1 grupos sempre serão juntados grupos heterogêneos.

Para avaliar o particionamento na base Iris, determinou-se a partição ótima do dendrograma. A Fig. 3 mostra a evolução do BIC da partição inicial com 6 grupos até 1 grupo, encontrando-se que a partição ótima contém 3 grupos, sendo este o número certo de grupos. A pureza nesta partição foi de 100%.

Tabela 1: Qualidade das árvores geradas com algoritmos aglomerativos. Os valores são a média de pureza das partições das árvores. Para CHBMG há o número de grupos da partição inicial.

	Iris	Sintética-1	Sintética-2
CHBMG	83.3% (6)	85.5% (30)	94.3% (35)
<i>Single linkage</i>	61.2%	32.5%	74.2%
<i>Complete linkage</i>	70.3%	86.3%	95.7%
<i>Average linkage</i>	74.4%	87.4%	96.1%

A Tab. 1 mostra de forma resumida a informação da Fig. 2 para cada base. A média das purezas das partições do dendrograma é usada para avaliar sua qualidade. Também, apresenta o nú-

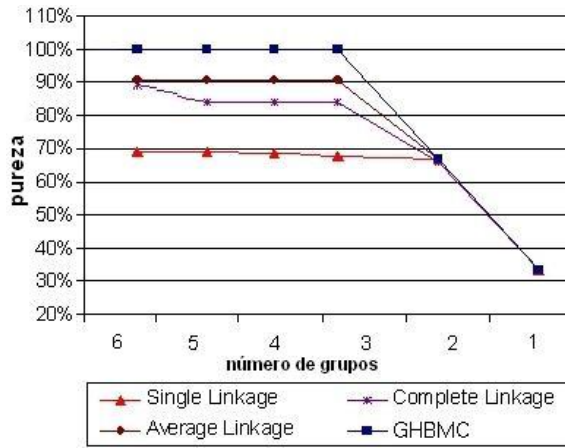


Figura 2: Pureza das últimas 6 partições dos métodos aglomerativos para a base Iris. CHBMG criou árvores de melhor estrutura pois apresentou pureza mais alta para a maioria das partições.

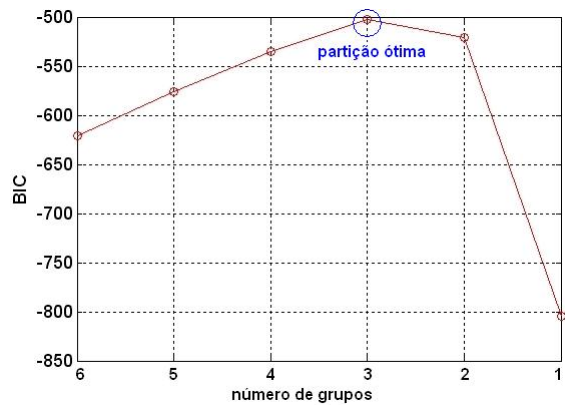


Figura 3: Escolha da partição ótima na árvore para base Iris. BIC em função de cada fusão de 2 grupos por CHBMG. A partição ótima está no passo 3.

mero real de grupos e o sugerido por BIC.

A Tab. 2 apresenta a qualidade do particionamento. Ela compara a pureza da partição ótima do CHBMG com aquelas dos algoritmos particionais.

Tabela 2: Qualidade da classificação com os algoritmos particionais. Número real de grupos e o sugerido por BIC.

	Iris	Sintética-1	Sintética-2
CHBMG	100%	71.6%	98.2%
ART2	96.7%	74.7%	98.8%
<i>K-means</i>	89.3%	78.2%	99.9%
EM	89.3%	77.9%	99.9%
Num. real de grupos	3	5	4
Num. grupos por BIC	3	4	4

4 Conclusões

Este trabalho introduziu um novo algoritmo para realizar agrupamento hierárquico aglomerativo. CHBMG foi comparado com métodos clássicos de agrupamento aglomerativos e particionais em dados reais e sintéticos.

CHBMG tem menor complexidade computacional que métodos aglomerativos clássicos. Para a base Sintética-2 com os algoritmos hierárquicos clássicos foram necessários aproximadamente 2000*2000 iterações para construir o dendrograma e também a mesma quantidade de memória para armazenar as grandes matrizes de distância iniciais. No caso do CHBMG, somente foram necessários 35 * 35 passos na aglomeração e uma quantidade inferior na determinação da partição inicial. Esta vantagem é ainda mais notória em grandes bancos de dados, onde as primeiras fusões são muito exigentes computacionalmente e também desnecessárias, desde que geralmente as partições ótimas são encontradas nos últimos passos da aglomeração, como foi mostrado nos bancos de dados analisados neste trabalho.

CHBMG tende a criar dendrogramas de melhor estrutura em comparação a classificadores aglomerativos clássicos. Ele apresentou pureza mais alta para a maioria das partições. Os resultados foram bons na base Iris e similares nas sintéticas.

CHBMG modela os grupos como distribuições Gaussianas o que possibilita o uso de critérios probabilísticos em vez de alguma métrica de distância *ad-hoc*. Além do modelo Gaussiano ser simples na estimação de parâmetros, sua escolha mostrou-se adequada. Ele apresentou resultados concisos com menor variação que os algoritmos aglomerativos clássicos em relação à mudança de base, Tab. 1. Também, mostrou-se apto a particionar dados. Com resultados similares aos métodos particionais, Tab. 2.

CHBMG determinou corretamente o número de grupos na maioria das bases através do critério BIC. Na base Sintética-1, detectou 4 ao invés de 5 grupos. Apenas nessa base a pureza da partição ótima é menor que a média das purezas das partições da árvore. Também, ela teve o pior particionamento em todos os algoritmos. Isso ocorreu pois esta base tem 2 grupos muito sobrepostos.

CHBMG apresentou um resultado ligeiramente pior nas bases sintéticas. Suas partições iniciais tiveram uma pureza menor que aquelas dos algoritmos aglomerativos com a mesma cardinalidade. Novamente, devido à alta sobreposição dos grupos nessas bases. É possível que o algoritmo usado para criar a partição inicial, rede ART2P, não seja apropriado para esse tipo de base. Entretanto, CHBMG demonstrou fundir os grupos de maneira correta, já que a partição final apresentou pureza comparável com os demais algoritmos.

Em trabalhos futuros poder-se-ia modificar o algoritmo com o objetivo de melhorar seu desempenho em bases com grupos sobrepostos. Finalmente, CHBMG pode satisfazer efetivamente demandas básicas de agrupamento de dados, evitando problemas de performance computacional e encontrando a partição ótima.

Agradecimentos

Os autores agradecem ao CNPq e à Capes pelo suporte financeiro e bolsas de estudo e à Escola de Engenharia de São Carlos (EESC/USP) pela infra-estrutura e pós-graduação.

Referências

- Banfield, J. D. and Raftery, A. E. (1993). Model-based gaussian and non-gaussian clustering, *Biometrics* **49**: 803–821.
- Carpenter, G. A. and Grossberg, S. (1987b). Art2: self-organization of stable category recognition codes for analog input patterns, *Applied Optics* **26**(23): 4919–4930.
- Carpenter, G. A. and Grossberg, S. (2003). Adaptive resonance theory, in A. M.A. (ed.), *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge, pp. 87–90.
- Chang, H., Kopaska-Merkel, D. C., Chen, H. and Durrans, R. D. (2000). Lithofacies identification using multiple adaptative resonance theory neural networks and group decision expert system, *Computers and Geocites* **26**(5): 591–601.
- Dasarathy, B. (1980). Nosing around the neighborhood: A new system structure and classification rule for recognition in partially exposed environments, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2**(1): 67–71.
- Dasgupta, A. and Raftery, A. E. (1998). Detecting features in spatial point processes with clutter via model-based clustering, *American Statistical Association* **93**: 294–302.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood for incomplete data via the em algorithm, *J. R. Stat. Soc.* **7**(39): 1–38.
- Duda, R. O., Hart, P. E. and Stork, D. G. (2001). *Pattern Classification*, second edn, John Wiley, New York.
- Everitt, B. S., Landau, S. and Leese, M. (2001). *Cluster Analysis*, 4 edn, Oxford University Press Inc., New York.
- Fausett, L. V. (1997). *Fundamentals of Neural Networks Architectures, Algorithms, and Applications*, first edn, Prentice Hall International Inc, New Jersey.
- Fraley, C. (1996). Algorithms for model-based gaussian hierarchical clustering, *Technical report*, University of Washington.
- Fraley, C. and Raftery, A. E. (1998). How many clusters? which clustering method? answers via model-based cluster analysis, *The Computer Journal* **41**(8): 578–588.
- Grossberg, S. (1976). Adaptive pattern classification and universal recoding, i: Parallel development and coding of neural feature detectors and ii: Feedback, expectation, olfaction, and illusions, *Biological Cybernetics* **23**(3): 121–134 and 187–202.
- Halkidi, M., Batistakis, Y. and Vazirgiannis, M. (2001). On clustering validation techniques, *Journal of Intelligent Information Systems* **17**(2/3): 107–145.
- Heller, K. A. and Ghahramani, Z. (2005). Bayesian hierarchical clustering, *Appering in Proceedings of the 22nd International Conference on Machine Learning*.
- Hon, M. J. L. (2002-2005). The c clustering library for cdna microarray data. <http://bonsai.ims.u-tokyo.ac.jp/mdehoon/software/cluster>.
- Iwayama, M. and Tokunaga, T. (1995). Hierarchical bayesian clustering for automatic text classification, *International Joint Conference on Artificial Intelligence* **2**: 1322–1327.
- Jain, A. K., Murty, M. N. and Flynn, P. J. (1999). Data clustering: A review, *acm computing surveys*, *ACM Computing Surveys* **31**(3): 264–323.
- Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley.
- Murtagh, F. and Raftery, A. E. (1984). Fitting straight lines to point patterns, *Pattern Recognition* **17**: 479–483.
- Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry*, World Scientific Pub Co Inc.
- Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* **6**(1): 461–464.
- Willett, P. (1988). Recent trends in hierarchic document clustering: A critical review, *Information Processing & Management* **24**(5): 577–597.