

Análise do Desempenho do Algoritmo de Agrupamento Baseado em Colônia de Formigas Modificado

Rosangela Villwock

UFPR - Programa de Pós-Graduação em Métodos Numéricos em Engenharia
Universidade Estadual do Oeste do Paraná - Campus Francisco Beltrão
Francisco Beltrão, PR
e-mail: rosangela@unioeste.br

Maria Teresinha Arns Steiner

Universidade Federal do Paraná
Programa de Pós-Graduação em Métodos Numéricos em Engenharia – PPGMNE
Coordenação de Engenharia de Produção, CP: 19081-CEP: 81531-990, Curitiba, PR
e-mail: tere@ufpr.br

Resumo: No algoritmo de Agrupamento Baseado em Colônia de Formigas os padrões são espalhados numa grade e a cada formiga é atribuído um padrão. As formigas são responsáveis por carregar, transportar e descarregar os padrões na grade. Após a convergência do algoritmo de agrupamento, a recuperação dos grupos é feita usando-se as posições dos padrões na grade. O objetivo do presente trabalho foi avaliar o desempenho do Algoritmo de Agrupamento Baseado em Colônia de Formigas Modificado. Os resultados foram comparados com outra modificação do Algoritmo de Agrupamento Baseado em Colônia de Formigas identificada como ACAM (Ant-based clustering algorithm modified). As principais modificações foram: substituição do padrão carregado pela formiga caso este não seja descarregado em 100 iterações consecutivas; comparação da probabilidade de descarregar um padrão numa determinada posição com a probabilidade de descarregar este padrão em sua posição atual; avaliação da probabilidade de descarregar um padrão para uma nova posição caso o padrão não seja descarregado na posição de sorteio e sim numa posição vizinha. Para a avaliação do desempenho do algoritmo foram utilizados dois exemplos reais. Os resultados mostram que o algoritmo proposto neste trabalho foi melhor do que o ACAM para os dois exemplos.

1. Introdução

O trabalho com métodos baseados em formigas foi iniciado por Marco Dorigo no início dos anos 90. Neste trabalho, um grupo de formigas foi usado para resolver um problema de transporte. Quando cada formiga percorre um caminho, deixa um sinal (feromônio) através do trajeto que usou. Os trajetos mais curtos terão os sinais mais fortes. As formigas seguintes, ao decidir que trajeto fazer, tendem a escolher trajetos com sinais mais fortes com uma maior probabilidade, de modo que os trajetos mais curtos acabam prevalecendo [1].

O algoritmo de agrupamento baseado em Colônia de Formigas foi proposto inicialmente por Deneubourg *et al.* (1991) *apud* Handl *et al.* [5]. As formigas foram representadas como agentes simples que se movem aleatoriamente em uma grade quadrada. Os padrões foram dispersos dentro desta grade e poderiam ser carregados, transportados e descarregados pelos agentes. Estas operações são baseadas na distribuição dos padrões dentro da vizinhança local dos agentes; padrões que foram isolados ou cercados por dissimilares eram mais prováveis de serem carregados e então descarregados numa vizinhança de similares. As decisões de carregar e descarregar padrões são tomadas pelas probabilidades P_p e P_d , respectivamente, dadas pelas equações (1) e (2) a seguir.

$$P_p = \left(\frac{k_1}{k_1 + f} \right)^2 \quad (1)$$

$$P_d = \left(\frac{f}{k_2 + f} \right)^2 \quad (2)$$

Nas equações (1) e (2), f é uma estimativa da fração de padrões localizados na vizinhança que são semelhantes ao padrão atual da formiga e k_1 e k_2 são constantes reais. No trabalho de Deneubourg *et al.* (1991), os autores usaram $k_1 = 0,1$ e $k_2 = 0,3$. Neste trabalho, a estimativa f foi obtida através de uma memória de curto prazo de cada formiga, onde o conteúdo da última célula da grade analisada é armazenado. Esta escolha da função de vizinhança f foi essencialmente motivada pela sua facilidade de realização por robôs simples. Nas equações (1) e (2), f é uma estimativa da fração de padrões localizados na vizinhança, que são semelhantes ao padrão atual da formiga. Neste trabalho, a estimativa f é obtida através de uma memória de curto prazo de cada formiga, onde o conteúdo da última célula da grade analisada é armazenado. Esta escolha da função de vizinhança f foi essencialmente motivada pela sua facilidade de realização por robôs simples.

Lumer & Faieta (1994), *apud* Handl et al. [5], introduziram um número de modificações ao modelo que permitiu a manipulação de dados numéricos e melhorou a qualidade da solução e o tempo de convergência do algoritmo. A idéia era definir uma medida de similaridade ou dissimilaridade entre os objetos (padrões), já que no algoritmo proposto inicialmente, os objetos eram similares se os objetos fossem idênticos e dissimilares se os objetos não fossem idênticos. No referido trabalho aparece pela primeira vez o mapeamento topográfico.

No trabalho de Lumer & Faieta (1994), as decisões de carregar e descarregar padrões são baseadas nas probabilidades P_p e P_d dadas pelas equações (3) e (4) a seguir, onde $f(i)$ é dada pela equação (5). Na equação (5), $d(i, j)$ é uma função de dissimilaridade entre padrões i e j pertencentes ao intervalo $[0, 1]$; α é um parâmetro escalar dependente dos dados e pertencente ao intervalo $[0, 1]$; L é a vizinhança local de tamanho igual a σ^2 (σ é o raio de percepção). Os autores usaram em seu trabalho $k_1 = 0,1$ e $k_2 = 0,15$ e $\alpha = 0,5$.

$$P_p(i) = \left(\frac{k_1}{k_1 + f(i)} \right)^2 \quad (3)$$

$$P_d(i) = \begin{cases} 2f(i) & \text{se } f(i) < k_2 \\ 1 & \text{se } f(i) \geq k_2 \end{cases} \quad (4)$$

$$f(i) = \max \left\{ 0, \frac{1}{\sigma^2} \sum_{j \in L} \left[1 - \frac{d(i, j)}{\alpha} \right] \right\} \quad (5)$$

Mudanças que melhoram a separação espacial dos grupos e permitem que o algoritmo seja mais robusto foram introduzidas por Handl [3]. Uma delas é a restrição na função f dada pela equação (6) a seguir, que serve para penalizar dissimilaridades elevadas.

$$f^*(i) = \begin{cases} \frac{1}{\sigma^2} \sum_{j \in L} \left[1 - \frac{d(i, j)}{\alpha} \right] & \text{se } \forall j \left(1 - \frac{d(i, j)}{\alpha} \right) > 0 \\ 0 & \text{caso contrário} \end{cases} \quad (6)$$

Quanto ao raio de percepção, mais atrativo será empregar vizinhanças maiores para melhorar a qualidade do agrupamento e da distribuição na grade. Porém, além de isso ser mais caro (porque o número das células a serem consideradas para cada ação cresce quadraticamente com o raio), inibe também a formação rápida dos grupos durante a fase de distribuição inicial.

Um raio da percepção que aumenta gradualmente com o tempo acelera a dissolução de grupos pequenos preliminares Handl [3].

As duas principais modificações propostas por Boryczka [2] foram: um novo esquema de adaptação para o valor de α e a substituição do escalar $\frac{1}{\sigma^2}$ na equação (6) pelo escalar $\frac{\sigma_0^2}{\sigma^2}$,

onde σ_0 é o raio de percepção inicial. A autora também testou diferentes medidas de dissimilaridade.

2. Agrupamento Baseado em Colônia de Formigas

2.1. O algoritmo básico

Na fase inicial, todos os padrões são aleatoriamente espalhados na grade. Depois, cada formiga escolhe um padrão para carregar e é colocada em uma posição na grade.

Na segunda fase, uma formiga é selecionada aleatoriamente. Esta formiga se desloca na grade executando um passo de comprimento L , numa direção determinada aleatoriamente. A formiga decide, então, probabilisticamente, se descarrega seu padrão nesta posição.

Se a decisão de descarregar o padrão for negativa, escolhe-se aleatoriamente outra formiga e recomeça-se o processo. No caso de decisão positiva, a formiga descarrega o padrão em sua posição atual na grade. Se esta célula da grade estiver ocupada por outro padrão, o mesmo deve ser descarregado numa célula imediatamente vizinha desta, que esteja livre, por meio de uma procura aleatória.

A formiga procura, então, por um novo padrão para carregar. Dentre os padrões livres na grade, a formiga seleciona aleatoriamente um padrão, pula para a posição do padrão na grade, faz a avaliação da função de vizinhança e decide probabilisticamente se carrega o padrão. Este processo de escolha de um padrão livre na grade é executado até que a formiga encontre um padrão que deva ser carregado.

A segunda fase é reiniciada, escolhendo-se outra formiga até que um critério de parada seja satisfeito.

2.2. Recuperação dos grupos

Segundo Handl [3], o processo inicia com cada padrão formando um grupo. Depois de calcular as distâncias entre todos os grupos, deve-se fundir (ligar) os dois grupos com menor distância. Os tipos de ligações mais comuns são: Ligações Simples, Ligações Completas, Ligações Médias e Método de Ward [7]. As distâncias entre grupos são definidas em termos de distância na grade. Este processo se repete até que um critério de parada seja satisfeito.

Na ligação simples (ou vizinho mais próximo), a distância entre dois grupos é a mínima distância entre todos os pares de padrões i e j , com i pertencente ao primeiro grupo e j ao segundo. Na ligação completa (ou vizinho mais distante), a distância entre dois grupos é a máxima distância entre todos os pares de padrões i e j , com i pertencente ao primeiro grupo e j ao segundo [6]. Na ligação média, a distância entre dois grupos é a distância média entre todos os pares de padrões, sendo que cada padrão do par é de um grupo.

O método de Ward, utilizado neste trabalho, faz a junção de dois grupos baseando-se na “perda de informação”. Considera-se como critério de “perda de informação” a soma do quadrado do erro (SQE). Para cada grupo i , calcula-se a média (ou centróide) do grupo e a soma do quadrado do erro do grupo i (SQE_i) que é a soma do quadrado do erro de cada padrão do grupo em relação à média. Para k grupos têm-se $SQE_1, SQE_2, \dots, SQE_k$ e define-se SQE pela equação (7).

$$SQE = SQE_1 + SQE_2 + \dots + SQE_k \quad (7)$$

Para cada par de grupos m e n , primeiramente, calcula-se a média (ou centróide) do grupo formado (grupo mn). Em seguida, calcula-se a soma do quadrado do erro do grupo mn

(SQE_{mn}) pela equação (8). Os grupos m e n que apresentarem o menor aumento na soma do erro quadrático (SQE) (menor “perda de informação”) são unidos.

$$SQE = SQE_1 + SQE_2 + \dots + SQE_k - SQE_m - SQE_n + SQE_{mn} \quad (8)$$

2.3. Avaliação do agrupamento

As medidas numéricas aplicadas para julgar vários aspectos de avaliação de grupos são classificadas em índices externos, usados para medir até que ponto rótulos de grupos correspondem a rótulos de classes externamente fornecidos, e índices internos, usados para medir quão boa é a estrutura de agrupamento sem relação com informação externa [8].

Neste trabalho foram utilizados três índices externos: Medida F, Índice Aleatório - R (descritas em [4]) e o percentual de classificação errada. A Medida F e o Índice Aleatório - R devem ser maximizadas, enquanto o percentual de classificação errada deve ser minimizado.

3. Metodologia

Neste trabalho foi aplicado o Algoritmo de Agrupamento Baseado em Colônia de Formigas Modificado para o agrupamento dos padrões. Uma das modificações foi a introdução de uma comparação da probabilidade de descarregar um padrão na posição escolhida aleatoriamente com a probabilidade de descarregar este padrão em sua posição atual. A decisão de descarregar um padrão na posição escolhida aleatoriamente só ocorre se esta probabilidade for maior que a probabilidade de descarregar este padrão em sua posição atual.

Outra modificação foi proposta quando a decisão de descarregar um padrão for positiva e a célula em que o padrão deveria ser descarregado estiver ocupada. Neste caso, faz-se uma avaliação da probabilidade para a nova posição, vizinha a posição ocupada, escolhida aleatoriamente. O padrão só é descarregado na célula vizinha se a probabilidade de descarregar o padrão nesta posição for maior que a probabilidade de descarregar este padrão em sua posição atual.

Além destas modificações, foi proposta a substituição do padrão carregado por uma formiga caso este padrão não seja descarregado em 100 iterações consecutivas. Esta modificação foi implementada para evitar formigas inativas. O novo padrão é escolhido por sorteio mas, ele só é carregado pela formiga se a probabilidade de carregar este padrão for maior que 0,13397. O valor 0,13397 foi definido fazendo a probabilidade de carregar (p_p) igual a probabilidade de descarregar (p_d). Caso isto não exista nenhum padrão com probabilidade de carregar maior que 0,13397, o último padrão sorteado é carregado pela formiga.

A medida de dissimilaridade utilizada foi a Distância Euclidiana. Na recuperação dos grupos foi utilizado o Método Ward e foi definido um número máximo de grupos.

Os exemplos abordados foram o IRIS e o WINE, cujos dados podem ser obtidos no endereço <http://mllearn.ics.uci.edu/databases>.

O exemplo IRIS é composto de 150 padrões (plantas). Neste exemplo são conhecidos os grupos a que cada planta pertence. Os 150 padrões são divididos em três grupos com 50 padrões em cada grupo: Íris Setosa, Íris Versicolour e Íris Virginica. Cada padrão é constituído por quatro atributos numéricos: comprimento da pétala, largura da pétala, comprimento da sépala e largura da sépala.

O exemplo WINE é composto de 178 padrões (vinhos). Neste exemplo também são conhecidos os grupos a que cada padrão pertence. Os 178 padrões são divididos em três grupos: 59 padrões pertencem ao grupo 1, 71 padrões pertencem ao grupo 2 e 48 padrões pertencem ao grupo 3. Cada padrão é constituído de 13 atributos numéricos. Os atributos são resultados de uma análise química.

4. Resultados

O algoritmo implementado utilizou como critério de parada o número de iterações. O algoritmo foi executado dez vezes. O número de iterações N_{max} foi definido como $N_{max} = 500.n.m$. A fase inicial foi definida como $t_{inicial} = 0,2.N_{max}$. A matriz Q de entrada dos dados é da ordem $m \times n$, onde m é o número de padrões e n é o número de atributos. A matriz P é a matriz

de dados normalizada. A normalização foi feita por dimensão. A matriz de distâncias também foi normalizada.

Na definição do tamanho da grade, escolheu-se o número de células igual a 10 vezes o número de padrões. Foram utilizadas 10 formigas ($p=10$). Foi utilizada vizinhança quadrada na busca dos padrões vizinhos.

O raio de vizinhança inicial era igual a 1. Foi utilizado o aumento deste valor durante a fase inicial, segundo a equação (9), onde t é a iteração atual da fase inicial. Durante a fase final, este valor decresce em 0,05 a cada 100 substituições do padrão carregado por uma formiga, sendo que isto ocorre quando este padrão não é descarregado em 100 iterações consecutivas. O valor do raio de vizinhança é sempre o valor inteiro menor ou igual ao definido em qualquer uma das fases.

$$\sigma = 4^{\frac{t}{t_{inicial}}} \quad (9)$$

Na definição da vizinhança para o cálculo da probabilidade de descarregar um padrão em sua posição atual e para o cálculo da probabilidade de carregar um padrão, considerou-se o raio de vizinhança igual a 1.

Na busca de uma nova posição, a direção do passo é aleatória. Definida a direção, calcula-se o tamanho máximo possível do passo. Um número aleatório pertencente ao intervalo $[0, 1]$ foi utilizado para determinar o tamanho do passo, multiplicando-se este número pelo tamanho máximo do passo.

As probabilidades de carregar (p_p) e descarregar (p_d) utilizadas são as descritas pelas equações (1) e (2), onde $k_1 = 0,1$ e $k_2 = 0,3$. Um padrão é carregado se a probabilidade p_p for maior que um valor mínimo para carregamento ($pick_{min}$). Um padrão é descarregado se a probabilidade p_d for maior que um valor mínimo para descarregamento ($drop_{min}$). Os valores de $drop_{min}$ e $pick_{min}$ foram definidos como 0,13397 durante a fase inicial. Durante a fase final estes valores foram definidos aleatoriamente a cada iteração com a restrição de serem maiores que 0,13397.

No cálculo da função f (foi utilizada a função f^* definida pela equação (6)), o parâmetro α foi definido como 0,8. Este valor foi atualizado durante a fase inicial segundo a equação (10). Durante a fase final este valor decresce em 0,001 a cada 100 substituições do padrão carregado por uma formiga, sendo que isto ocorre quando este padrão não é descarregado em 100 iterações consecutivas.

$$\alpha = \alpha_0 + \frac{2t}{p \cdot t_{inicial}} - 0,01 \quad (10)$$

Quando um padrão é descarregado na grade, um novo padrão deverá ser carregado. A busca deste padrão é aleatória, porém, cada padrão livre é avaliado somente uma vez, não sendo permitido que ele seja avaliado novamente, até que todos sejam avaliados. Caso nenhum padrão apresente probabilidade p_p maior que $pick_{min}$, o padrão que apresentar a maior probabilidade p_p é carregado.

Quando um padrão não tem vizinhos, definiu-se a função f igual a zero. Isso faz com que a probabilidade p_d seja igual a 0, ou seja, o padrão não deve ser descarregado naquela posição e a probabilidade p_p será igual a 1, ou seja, o dado deverá ser carregado e futuramente deixar esta posição.

O quadro 1, a seguir, apresenta a média e o desvio-padrão para as medidas de avaliação do agrupamento: R, F e classificação errada. Este quadro também apresenta as medidas de avaliação do agrupamento para o melhor resultado.

Quadro 1 – Resultados da avaliação do agrupamento para os exemplos IRIS e WINE.

Resultados		R	F	Classificação errada (%)
IRIS	Média	0,863	0,867	12,9
	Desvio-padrão	0,065	0,076	7,1
	Melhor resultado	0,949	0,960	4,0
WINE	Média	0,845	0,874	12,5
	Desvio-padrão	0,039	0,035	3,5
	Melhor resultado	0,897	0,921	7,9

Já o quadro 2, a seguir, apresenta a comparação das médias das medidas de avaliação do agrupamento para o algoritmo analisado neste trabalho e para o algoritmo ACAM (Ant-based clustering algorithm modified) proposto por Boryczka em [2]. O melhor resultado encontra-se em negrito. Os resultados mostram que o algoritmo com as modificações propostas neste trabalho é melhor que o ACAM para os dois exemplos. Para o exemplo WINE, o ACAM foi melhor que o algoritmo proposto neste trabalho para a medida de avaliação R.

Quadro 2 – Resultados da avaliação do agrupamento para os exemplos IRIS e WINE.

Medidas de Avaliação		ACAM [2]	Algoritmo modificado
IRIS	R	0,819	0,863
	F	0,810	0,867
	Classificação errada (%)	18,7	12,9
WINE	R	0,849	0,845
	F	0,868	0,874
	Classificação errada (%)	13,9	12,5

As figuras 1 e 2, a seguir, apresentam a grade para o melhor resultado (cujas medidas de avaliação foram apresentadas no quadro 1) dos exemplos IRIS e WINE, respectivamente.

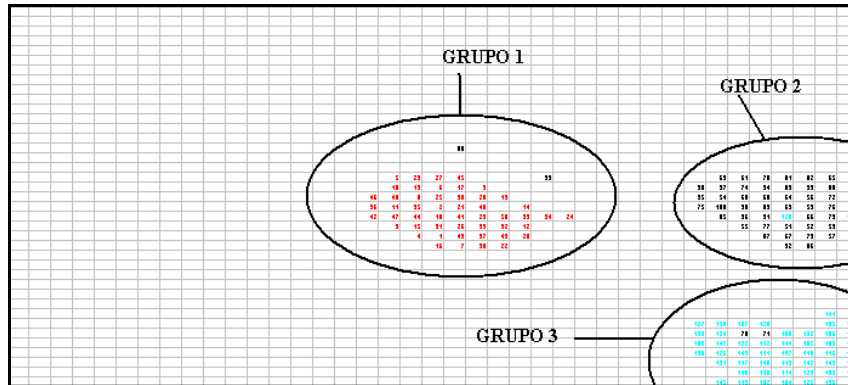


Figura 1 – Resultado do agrupamento para o exemplo IRIS – melhor resultado

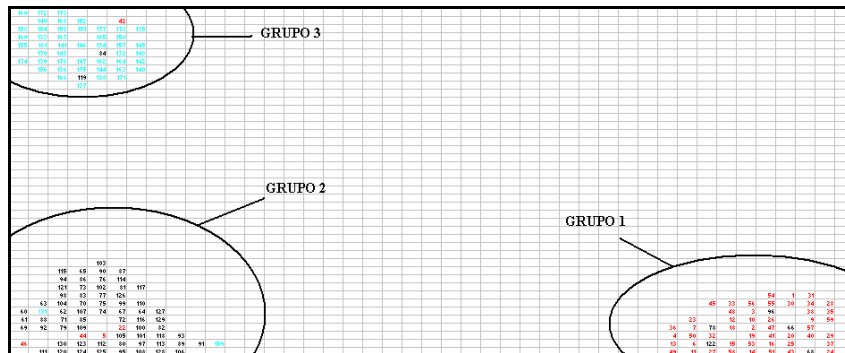


Figura 2 – Resultado do agrupamento para o exemplo WINE – melhor resultado

5. Considerações Finais

Os comportamentos coletivo e auto-organizável de insetos sociais inspiraram cientistas a reproduzir este comportamento. O estudo de colônias de formigas tem oferecido idéias novas para técnicas de agrupamento. O Algoritmos de Agrupamento Baseado em Formigas tem recebido especial atenção já que continuam a exigir muito mais investigação para melhorar seu desempenho.

O objetivo do presente artigo foi avaliar o desempenho do Algoritmo de Agrupamento Baseado em Colônia de Formigas Modificado.

As principais modificações foram: substituição do padrão carregado pela formiga caso este não seja descarregado em 100 iterações consecutivas; comparação da probabilidade de descarregar um padrão numa posição escolhida aleatoriamente com a probabilidade de descarregar este padrão em sua posição atual; avaliação da probabilidade para uma nova posição caso o padrão não seja descarregado na posição de sorteio e sim numa posição vizinha.

Na avaliação do desempenho do algoritmo foram utilizadas duas bases reais, IRIS e WINE. Os resultados mostram que o algoritmo proposto neste trabalho foi melhor que o ACAM para os dois exemplos.

Agradecimentos

A primeira autora agradece à CAPES pela bolsa que vem sendo concedida para o desenvolvimento de seu Curso de Doutorado.

Referências

1. Aranha, C. de C.. “A Survey on using Ant-Based techniques for clustering”. Institute's seminar, 2006. Disponível em <http://www.iba.k.u-tokyo.ac.jp/~caranha/research/ant-cluster-survey.pdf>. Acesso em 05/03/2007.
2. Boryczka, U.. “Finding groups in data: Cluster analysis with ants”. Applied Soft Computing. Volume 9, Issue 1, January 2009, Pages 61-70.
3. Handl, J. “Ant-based methods for tasks of clustering and topographic mapping: extensions, analysis and comparison with alternative methods”. Masters thesis. University of Erlangen-Nuremberg, Germany, 2003.
4. Handl, J.; Knowles, J.; Dorigo, M.. “Ant-based clustering: a comparative study of its relative performance with respect to k-means, average link and 1d-som”. Technical Report TR/IRIDIA/2003-24, IRIDIA, Université Libre de Bruxelles, July 2003.
5. Handl, J.; Knowles, J.; Dorigo, M.. “Ant-Based Clustering and Topographic Mapping”. Artificial Life, V12(1), 2005.
6. Jain A. K., Murty M. N., Flynn P. J.. “Data clustering: a review”. ACM Computing Surveys. v. 31, n. 3, set. 1999.
7. Johnson, R.A. E Wichern, D.W. “Applied Multivariate Statistical Analysis”. 4nd. Edition. Ed. Prentice Hall, 1998.
8. Tan, P. N.; Steinbach, M.; Kumar, V.. “Introduction to Data Mining”. Inc. Boston, MA, USA: Addison-Wesley Longman Publishing Co., 2005.