

ALGORITMO GENÉTICO PARA A CLUSTERIZAÇÃO DE GRUPOS DE APRENDIZES UTILIZANDO O INVENTÁRIO DE ESTILOS DE APRENDIZAGEM - ILS DE FELDER-SOLOMAN

Silvio L. C. Silva, Luiz A. N. Lorena

Laboratório Associado de Computação e Matemática Aplicada (LAC)
Instituto Nacional de Pesquisas Espaciais (INPE) - São José dos Campos, SP - Brasil.

silvioslcs@ififi.cta.br, lorena@lac.inpe.br

Resumo - *Esse trabalho apresenta o Algoritmo Genético (AG) implementado para agrupar estudantes de acordo com seus Estilos de Aprendizagem (EAs), de forma que os estudantes dos grupos formados sejam similares e que sejam diferentes dos estudantes dos outros grupos, de acordo com as características medidas.*

Abstract - *This paper presents the Genetic Algorithm (GA) implemented for clustering of students according to your Learning Styles (LS), so that the students of groups formed are similar and are different of the students other groups, according with your measures characteristics.*

1. Introdução

O paradigma evolutivo utilizado para o agrupamento é encontrado em pesquisas que tentam resolver problemas de *clustering* de grafos orientados [Dias, 2004]; problemas de localização de p -medianas [Lorena e Furtado, 2001]; entre outras, sendo comumente utilizados em problemas de automação e otimização.

Como variáveis de entrada para o método de agrupamento é utilizado o instrumento *Index of Learning Styles (ILS)* na sua versão reduzida [Felder e Soloman, 2010]. Esse instrumento é encontrado em trabalhos de Investigação dos estilos de aprendizagem de estudantes universitários [Lopes, 2002] e em problemas de agrupamento de aprendizes [Zakrzewska, 2009].

2. Problema de *Clustering*

Um conjunto A com n Aprendizes e p -Variáveis de Classificação de Aprendizagem (VCAs) em cada um deles $A_{ij} = \{a_{ij}, \dots, a_{np}\}$, o *clustering* consiste em identificar grupos de aprendizes similares divididos em k *clusters*, $C = \{C_1, C_2, \dots, C_k\}$. Segundo Dias (2004), o conjunto C é considerado um *clustering* com k *clusters*, caso as seguintes condições sejam satisfeitas: $\sum_{i=1}^k C_i = A$, $C_i \neq \emptyset$, para $1 \leq i \leq k$, $C_i \cap C_j = \emptyset$, para $1 \leq i, j \leq k$ e $i \neq j$.

3. Estilos de Aprendizagem

O instrumento para identificar os estilos de aprendizagem dos aprendizes é a versão reduzida do *ILS*, desenvolvido por Felder e Soloman (2010), com o objetivo de investigar e identificar as preferências de aprendizagem dos aprendizes em quatro dimensões de estilos de aprendizagem: *Ativo/Reflexivo*, *Sensorial/Intuitivo*, *Visual/Verbal* e *Sequencial/Global* [Felder e Silverman, 1988].

4. Algoritmo Genético

Proposto por Holland (1975), a implementação do AG iniciasse com a codificação do espaço de soluções do problema, através de um vetor binário ou de números inteiros, essa representação é chamada de *Cromossomo*. Posteriormente é criada uma *População Inicial (PI)* composta por um conjunto de cromossomos aleatoriamente preenchidos. Durante o processo evolutivo essa população é avaliada, através da *Função de Aptidão (FA)*, e cada cromossomo recebe uma nota (*fitness*) que representa a adaptabilidade do cromossomo ao problema. Após esses processos são aplicados os operadores genéticos *Seleção, Cruzamento e Mutação* para a construção de novas gerações mais adaptadas.

O espaço de soluções foi codificado através da representação *group-number* [$e_1 e_2 e_3 \dots e_n$] [Kim et al., 2011], onde cada gene $e_i \in \{1, \dots, k\} \mid k=p$ representa o grupo do aprendiz da i -ésima posição $A = \{a_{ij}, \dots, a_{np}\}$ com suas j -ésimas variáveis de classificação de aprendizagem.

A Função de Aptidão utilizada é representada pelo somatório das distâncias euclidianas das VCAs entre os aprendizes do mesmo grupo, onde $d(a_i, a_s) = [\sum_{j=1}^p (a_{ij} - a_{sj})^2]^{1/2}$, tal que $a_{ij}, a_{sj} \in \{0, 1, \dots, 11\}$, $p=8$ [Mingoti, 2005].

$$\text{Aptidão} = \sum_{i=1}^{n-1} \sum_{s=i+1}^n d(a_i, a_s); \forall e_i = e_s \quad (1)$$

O método de elitismo utilizado consiste na cópia do melhor indivíduo da população anterior para a nova população, garantindo que o melhor indivíduo da população antiga não seja perdido na nova população após a aplicação dos operadores genéticos.

O método de seleção por torneio é utilizado no AG proposto, esse método seleciona m indivíduos aleatoriamente da população corrente, formando uma subpopulação, dessa população intermediária é realizado o torneio, onde é selecionado o melhor indivíduo dos m indivíduos da população intermediária, esse processo é realizado até que a nova população fique completa.

Para o operador cruzamento utilizamos o de um ponto, esse tipo de cruzamento realiza a combinação, após um ponto, de segmentos de dois cromossomos selecionados (Pais) para gerar dois novos cromossomos (Filhos), onde o ponto de cruzamento é obtido de forma aleatória.

Na mutação alteram-se o valor (*fenótipo*) de alguns *genes* dos cromossomos da população corrente, essa diversidade genética é inserida de forma aleatória, tanto para o número de genes quanto para seu valor, com isso a busca é realizada em diferentes áreas do espaço de solução, não permitindo que a população fique estagnada em uma região de ótimo local. [Dias, 2004].

5. Testes Computacionais

5.1 Preparação dos Dados

Utilizou-se uma amostra de 54 alunos dos 120 regularmente matriculados no curso de Ciência da Computação do Centro Universitário Salesiano de São Paulo – Unidade de Lorena. O tamanho da amostra para estimar a proporção de uma população finita com

95% de confiança foi determinado segundo Martins (2008), com os seguintes parâmetros: $N=120$; $Z=1,96$ para 95%; $p'=0,50$; $q'=(p'-1)=0,50$; $d=10\%=0,10$.

5.2 Método de Avaliação do Agrupamento

Na análise da contagem-z ou também conhecida como valor padronizado podemos interpretar $Z_{e_i} = MÁX \left| \frac{a_{sj} - \bar{a}_{e_{ij}}}{S_{e_{ij}}} \right|$; $\forall e_i = e_s$; $i, s \in \{1, \dots, n\}$ e $j \in p$, como o valor de desvios padrão $S_{e_{ij}}$ que o aprendiz, mais distante, do *cluster* e_i , está afastado da média $\bar{a}_{e_{ij}}$ desse mesmo grupo. [Anderson, Sweeney, e Williams, 2008].

Anderson *et al.* (2008), também afirmam que para a detecção de pontos fora da curva (*outliers*) são utilizados valores padronizados, sendo consideradas observações fora da curva àquelas que apresentarem valores z menores que -3 e valores z maiores que +3.

5.3 Resultados Computacionais

Os aprendizes responderam o questionário *ILS on-line*, a máquina responsável pela execução da aplicação possui a seguinte configuração: processador *Intel Atom*, 1,66 GHz; 2 Gb de memória *ram* e sistema operacional *Windows 7 Starter*.

Tabela 1: Agrupamento Realizado pelo AG

		ALUNOS																	
		4	7	15	28	31	47	3	13	21	32	35	42	12	33	40	49	50	1
ESTILOS DE APRENDIZAGEM	ATIIVO	7	9	8	11	9	11	5	4	3	6	4	4	6	6	7	6	4	1
	REFLEXIVO	4	2	3	0	2	0	6	7	8	5	7	7	5	5	4	5	7	10
	SENSORIAL	8	8	8	9	10	7	6	6	6	8	6	6	2	4	5	4	5	2
	INTUITIVO	3	3	3	2	1	4	5	5	5	3	5	5	9	7	6	7	6	9
	VISUAL	8	7	7	8	8	7	4	4	2	4	6	4	7	6	8	6	7	8
	VERBAL	3	4	4	3	3	4	7	7	9	7	5	7	4	5	3	5	4	3
	SEQUENCIAL	4	4	6	5	6	5	4	5	5	3	4	1	1	5	4	4	3	6
	GLOBAL	7	7	5	6	5	6	7	6	6	8	7	10	10	6	7	7	8	5
	GRUPO	0	0	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2
		ALUNOS																	
		17	27	37	10	16	24	41	45	52	5	14	18	29	30	43	44	54	6
ESTILOS DE APRENDIZAGEM	ATIIVO	2	0	4	7	7	6	7	7	6	9	10	9	11	9	10	8	8	9
	REFLEXIVO	9	11	7	4	4	5	4	4	5	2	1	2	0	2	1	3	3	2
	SENSORIAL	6	5	4	9	6	9	8	8	6	7	6	8	6	7	6	5	5	6
	INTUITIVO	5	6	7	2	5	2	3	3	5	4	5	3	5	4	5	6	6	5
	VISUAL	6	9	6	6	6	4	5	6	6	11	10	9	8	8	8	9	10	5
	VERBAL	5	2	5	5	5	7	6	5	5	0	1	2	3	3	3	2	1	6
	SEQUENCIAL	10	4	9	6	5	5	4	7	6	7	8	8	7	8	6	6	8	5
	GLOBAL	1	7	2	5	6	6	7	4	5	4	3	3	4	3	5	5	3	6
	GRUPO	3	3	3	3	4	4	4	4	4	4	5	5	5	5	5	5	5	5
		ALUNOS																	
		9	20	25	36	48	51	2	8	23	26	34	46	11	19	22	38	39	53
ESTILOS DE APRENDIZAGEM	ATIIVO	9	8	8	9	10	8	7	7	6	9	8	10	6	6	6	6	3	4
	REFLEXIVO	2	3	3	2	1	3	4	4	5	2	3	1	5	5	5	5	8	7
	SENSORIAL	5	8	6	5	4	5	10	9	10	9	11	10	6	5	6	8	8	8
	INTUITIVO	6	3	5	6	7	6	1	2	1	2	0	1	5	6	5	3	3	3
	VISUAL	5	5	5	6	6	7	9	11	9	9	10	11	9	8	8	10	8	10
	VERBAL	6	6	6	5	5	4	2	0	2	2	1	0	2	3	3	1	3	1
	SEQUENCIAL	6	9	7	6	9	9	8	8	6	8	8	11	6	9	7	9	7	8
	GLOBAL	5	2	4	5	2	2	3	3	5	3	3	0	5	2	4	2	4	3
	GRUPO	6	6	6	6	6	6	6	7	7	7	7	7	7	8	8	8	8	8

O agrupamento apresentado pela Tabela 1, representa a combinação de parâmetros genéticos mais eficientes para o AG com $FA=56,071383$, após 20 testes realizados. Esse agrupamento possui máximo valor de contagem-z $Z_{e_i}=2,041$, representando assim, um agrupamento válido, sem *outliers*, com um tempo de execução de 4 segundos. Pode-se também enquadrar os resultados da Tabela 1 em uma escala de *score*, com referência

no valor em módulo da subtração entre os EAs de mesma dimensão, esse valor é atribuído ao EA de maior valor e enquadrado na escala: 1 a 3 leve preferência pelo EA; 5 a 7 moderada preferência e de 9 a 11 forte preferência pelo EA [Lopes, 2002].

6. Conclusão

Durante a classificação dos aprendizes foi constatada a predominância pelos Estilos de Aprendizagem: *ATIVO*, *SENSORIAL*, *VISUAL* e *SEQUENCIAL*. Foram realizados 20 testes de eficiência para a escolha dos melhores parâmetros genéticos, sendo escolhidos os seguintes valores: *Taxa de Cruzamento = 80%*, *Taxa de Mutação = 1%*, *Elitismo do Melhor Indivíduo*, *População Inicial = 20* e *Critério de Parada = 500 iterações*. O agrupamento foi validado utilizando o método estatístico contagem-z, onde não foram encontrados *outliers*, dentro dos limites de -3 a +3 desvios padrão da média. Dessa forma, os profissionais da educação podem personalizar aulas e materiais didáticos, de acordo com as preferências de aprendizagem de cada grupo.

Referências Bibliográficas

- Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2008). *Estatística Aplicada à Administração e Economia*. São Paulo: CENGAGE Learning.
- Dias, C. R. (2004). Algoritmos Evolutivos para Problemas de Clusterização de Grafos Orientados: Desenvolvimento e Análise Experimental. 111. Niterói, Rio de Janeiro, Brasil: Universidade Federal Fluminense.
- Felder, R. M., & Silverman, B. A. (1988). Index of Learning Styles In Engineering Education. *Journal of Engineering Education*, 78(7), 674-680.
- Felder, R. M., & Soloman, B. A. (2010). *Index of Learning Styles*. Acesso em 20 de Setembro de 2010, disponível em <http://www.engr.ncsu.edu/learningstyles/ilsweb.html>
- Holland, J. H. (1975). Adaptation in Natural and Artificial Systems. *MIT Press*, 11-147.
- Kim, J., Kim, Y.-H., Hwang, I., & Moon, B.-R. (12-16 de Julho de 2011). Genetic Approaches for Graph Partitioning: A Survey. *GECCO 11*, (p. 8). Dublin, Ireland.
- Lopes, W. M. (2002). *Inventário de Estilos de Aprendizagem de Felder-Soloman: Investigação de sua Validade em Estudantes Universitários de Belo Horizonte*. Engenharia de Produção. Florianópolis: Universidade federal de Santa Catarina.
- Lorena, L. A., & Furtado, J. C. (2001). Constructive Genetic Algorithm for Clustering Problems. *Evolutionary Computation*, 9(3), 309-327.
- Martins, G. A. (2008). *Estatística Geral e Aplicada* (3 ed.). São Paulo: Atlas.
- Mingoti, S. A. (2005). *Análise de Dados Através de Métodos de Estatística Multivariada*. Belo Horizonte: Editora UFMG.
- Zakrzewska, D. (2009). Cluster Analysis in Personalized E-Learning Systems. *Intelligent Systems for Knowledge Management Studies in Computational Intelligence*, 252, 229-250.