

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

Silvio Luiz Castro Silva
Luiz Antônio Nogueira Lorena

Instituto Nacional de Pesquisas Espaciais - INPE/LAC
Laboratório Associado de Computação e Matemática Aplicada
e-mails: silvioslcs@ifi.cta.br, lorena@lac.inpe.br
Av dos Astronautas, 1.758 - Jd. Granja,
CEP: 12227-010, São José dos Campos, SP, Brasil

Resumo- Esse trabalho tem como objetivo apresentar a comparação do Algoritmo Genético – AG desenvolvido para agrupar estudantes de acordo com seus Estilos de Aprendizagem – EA com o método heurístico de agrupamento *K-means*. Esse agrupamento tem o objetivo de concentrar em cada grupo um conjunto de estudantes similares e que os aprendizes de cada grupo sejam diferentes dos elementos dos outros grupos, de acordo com as características medidas.

Com esse agrupamento o profissional de educação (presencial ou à distância) poderá personalizar o material didático proposto para cada um dos grupos identificados, maximizando assim, a assimilação do conteúdo e, conseqüentemente, diminuindo o número de não aprendizes.

Para a classificação dos alunos utilizamos a ferramenta psicológica *ILS- Index of Learning Styles* para identificação dos estilos de aprendizagem de cada estudante. Essa identificação parte da ideia de que os indivíduos têm diferentes maneiras de "perceber" e de "processar a informação" o que caracteriza formas particulares de aprender, cada indivíduo percebe os estímulos do ambiente de uma maneira particular e conseqüentemente como a percepção é a primitiva da aprendizagem cada pessoa aprende de uma forma particular.

Palavras-chave- Estilos de Aprendizagem, Problemas de agrupamento, *Clustering*, Algoritmos Genéticos, Otimização Combinatória.

Abstract -*This work aims to present a comparison of the Genetic Algorithm - AG developed to group students according to their Learning Styles - EA with the heuristic K-means clustering. This group aims to concentrate on each group a set of similar students and learners in each group are different from elements of other groups, according to the characteristics measured.*

With this grouping the professional education (classroom or distance) can customize the learning material proposed for each groups identified, maximizing the assimilation of the content and decreasing the number of non-learners.

For the classification of students used the tools ILS - Index of Learning Styles to identify the learning styles of each student. This identification have the idea that individuals have different ways to "see" and "process information" that characterizes the particular forms of learning, each individual perceives the stimuli of the environment in a particular way and therefore how the perception is the earliest learning each person learns in a particular way.

Keywords - *Learning Styles, Problem grouping, Clustering, Genetic Algorithms, Combinatorial Optimization, K-Means.*

1 Introdução

A análise de agrupamentos, também conhecida como análise de conglomerados, classificação ou cluster, tem como objetivo dividir os elementos da amostra, ou população, em grupos de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si, com respeito às variáveis características que neles foram medidas, e que sejam diferentes dos elementos em grupos diferentes em relação a estas mesmas características. (Mingoti 2005).

Esse artigo utiliza para o agrupamento dos aprendizes uma metaheurística conhecida como Algoritmos Genéticos - (AG). Essa técnica esta presente em pesquisas que tentam resolver problemas de *clustering* de grafos orientados (Dias 2004); problemas de localização de p-medianas (Lorena e Furtado 2001); problemas de programação de operações em um ambiente de produção *Flow Shop*, tendo como objetivo minimizar o estoque em processamento (Ribeiro, Nagano e Lorena 2007); problemas de expansão de sistemas de transmissão (Miasaki e Romero 2007); problemas químicos (Filho e Poppi 1999); entre outros; sendo comumente utilizados em problemas de automação e otimização.

Para problemas de *clustering* as metaheurísticas são encontradas em trabalhos, como: Algoritmo Evolutivo Híbrido para a Formação de Células de Manufatura em Sistemas de Produção (Trindade e Ochi 2006), Algoritmos Evolutivos para o Problema de *Clusterização* de Grafos Orientados: Desenvolvimento e Análise Experimental (Dias 2004), *Constructive Genetic Algorithm for Clustering Problems* (Lorena e Furtado 2001), Uma Metaheurística Híbrida com Busca por Agrupamentos Aplicada a Problemas de Otimização Combinatória (Chaves 2009), *Clustering Students to Help Evaluate Learning* (Merceron e Yacef 2005). Encontram-se também na literatura trabalhos que apresentam outras técnicas para o agrupamento de aprendizes segundo EA, tais como: *K-means* e *Self-organizing Maps* (Pimentel, França e Omar 2003), Agentes Inteligentes (Felix e Tedesco 2006).

Para identificar os EA dos alunos foi utilizado o instrumento denominado *Index of Learning Styles – ILS* na sua versão reduzida, (Felder e Soloman 2010), com o objetivo de investigar e identificar as preferências de aprendizagem em quatro dimensões de estilos de aprendizagem: *Ativo/Reflexivo*, *Sensorial/Intuitivo*, *Visual/Verbal* e *Sequencial/Global*.

O ILS é um instrumento encontrado em trabalhos de investigação dos estilos de aprendizagem de estudantes universitários (Lopes 2002), para identificar os estilos de aprendizagem de alunos de engenharia (Cury 2000), em problemas de agrupamento de aprendizes (Zakrzewska 2009) e na busca de maior eficiência nas ações de capacitação das áreas de treinamento e desenvolvimento de instituições (Machado, et al. 2010). As dimensões utilizadas pelo instrumento ILS, são baseadas no modelo de estilos de aprendizagem de Felder e Silverman (Felder e Silverman 1988).

O agrupamento realizado dividiu os alunos das quatro séries de 2008 do curso de Bacharel em Ciência da Computação do Centro Salesiano Universitário de São Paulo, Unidade de Lorena-SP em oito grupos distintos, de acordo com suas características de aprendizagem, sendo realizada uma avaliação estatística, contagem-z, para validar a eficiência do agrupamento realizado.

O artigo está dividido em 6 (seis) capítulos, o primeiro apresenta a introdução do artigo; o segundo capítulo aborda o problema de *clusterização* e a medida para a determinação da similaridade dos aprendizes; no terceiro capítulo é apresentado o instrumento para a identificação dos estilos de aprendizagem dos aprendizes; no quarto capítulo encontramos a definição da metaheurística Algoritmo Genético que é utilizada para a otimização do agrupamento; no quinto capítulo apresentamos o critério para a definição do tamanho da amostra utilizada, a técnica matemática (contagem-z) para a validação do agrupamento e os testes computacionais realizados; no sexto capítulo concluímos o estudo realizado.

2 Problema de *Clustering*

Com um conjunto de n elementos amostrais (aprendizes) com p -variáveis (variáveis de classificação de aprendizagem) em cada um deles $A_{ij} = \{a_{ij}, \dots, a_{np}\}$, o *clustering* consiste em identificar grupos com características similares divididos em k *clusters*, $C = \{C_1, C_2, \dots, C_k\}$, proporcionando que os elementos de cada cluster C_i tenham a maior similaridade entre os elementos do mesmo grupo do que com os elementos de qualquer um dos outros grupos (*clusters*) do conjunto C .

Segundo Dias (2004), o conjunto C é considerado um *clustering* com k *clusters*, caso as seguintes condições sejam satisfeitas: $\sum_{i=1}^k C_i = A, C_i \neq \emptyset, \text{ para } 1 \leq i \leq k, C_i \cap C_j = \emptyset, \text{ para } i \leq i, j \leq k \text{ e } i \neq j$.

Segundo Mingoti (2005) existem algumas formas para se mensurar a similaridade e dissimilaridade dos elementos de uma amostra, as mais comuns para variáveis quantitativas são: *Distância Euclidiana*, *Distância Generalizada ou Ponderada* e *Distância de Minkowsky*. Nesse trabalho é utilizada a distância euclidiana como medida de similaridade entre os aprendizes,

$d(a_l, a_k) = \left[\sum_{j=1}^p (a_{lj} - a_{kj})^2 \right]^{1/2}$, tal que $a_{ij} \in \{0, 1, \dots, 11\}$, $p = 8$, $l, k \in \{1, \dots, n\}$, $n =$ número de aprendizes. A escolha da distância euclidiana teve como referência o formato dos parâmetros quantitativos obtidos após a aplicação do instrumento ILS.

3 Estilos de Aprendizagem

O estilo de aprendizagem pode ser entendido como o comportamento do aprendiz durante o processo de aprendizado. O conhecimento prévio do estilo de aprendizagem pode, então, maximizar a assimilação do conhecimento, encurtando os laços do material didático apresentado com o aprendiz.

O instrumento para a identificação dos estilos de aprendizagem utilizado neste trabalho é a versão reduzida do *ILS – Index of Learning Styles*, desenvolvido por Richard M. Felder e Bárbara A. Soloman (Felder e Soloman 2010), com o objetivo de investigar e identificar as preferências de aprendizagem em quatro dimensões de estilos de aprendizagem: *Ativo/Reflexivo*, *Sensorial/Intuitivo*, *Visual/Verbal* e *Sequencial/Global*. (Felder e Silverman 1988)

4 K-Means

O método *K-means* é um método heurístico que organiza os elementos de uma amostra em k clusters, através da contínua geração de centroides, com base nas médias dos grupos gerados a cada iteração; os centros desses clusters são comparados,

continuamente, com os elementos da amostra, sendo, os elementos, agrupados ao cluster mais próximo. Dessa forma, são gerados novos clusters, com novos centros, até que os grupos não se alterem mais, minimizando assim, a soma das distâncias totais entre os dados de cada grupo e seu respectivo centro. (Jain, Murty e Flynn 1999)

5 Algoritmo Genético

O grande desafio para a otimização de rotinas de agrupamento está diretamente ligada ao tamanho do espaço de soluções, decorrentes das diferentes formas de agrupamento dos n elementos da amostra em *kclusters*. Existem métodos de otimização que encontram boas soluções para esses problemas, os mais utilizados são os métodos baseados em mecanismos de evolução de populações, sendo os mais utilizados deles os Algoritmos Genéticos proposto por Holland (1975).

A estrutura básica de um 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* ou *Genótipos*. Posteriormente criasse uma *População Inicial(PI)* composta por um conjunto de cromossomos aleatoriamente preenchidos. Durante o processo evolutivo essa população é avaliada e cada cromossomo recebe uma nota (*fitness*) que representa a adaptabilidade do cromossomo ao problema. Após esse processo são aplicados os operadores genéticos *Seleção*, *Cruzamento* e *Mutação* para a construção das novas gerações mais adaptadas.

5.1. Codificação

Cada cromossomo da população deve representar uma solução para o agrupamento dos aprendizes, sendo a representação escolhida para esse trabalho a codificação *group-number* (Kim, et al. 2011).

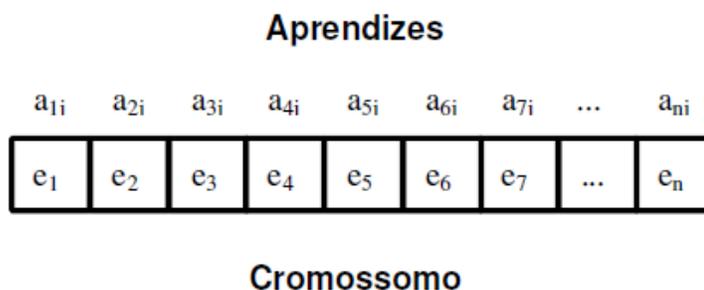


Figura 1: Representação do cromossomo para codificação *Group-Numbers*.

onde:

- $N=\{1, 2, \dots, n\}$, e n representa o número de aprendizes.
- $P=\{1, 2, \dots, p\}$, e $p=8$ representa os 8 (oito) Estilos de Aprendizagem.
- $i \in N$ e representa os índices dos aprendizes.
- $j \in P$ e representa os índices das variáveis de classificação de aprendizagem.
- $e_i \in \{1, 2, \dots, 8\}$ e representa o grupo do aprendiz da i -ésima posição de acordo com os 8 (oito) Estilos de Aprendizagem de Felder e Silverman (1988).
- $A=\{a_{ij}, \dots, a_{np}\}$ e representa o conjunto de aprendizes e suas variáveis de classificação de aprendizagem (Kim, et al. 2011).

5.2. População Inicial

O AG desenvolvido cria aleatoriamente a *PI* com um número pré-definido de soluções. Após testes computacionais definiu-se o melhor tamanho para a *PI*: $Np = 6$ cromossomos, onde os genes $e_i \in \{1, 2, \dots, 8\}$ são escolhidos aleatoriamente criando a *PI*. A criação dos cromossomos iniciais torna-se simples de implementação através da linguagem de programação JAVA na versão 5.0, pela eficiência da classe *Random* para criação de números pseudo-aleatórios, através de um algoritmo gerador de congruência linear. (Oracle, API Specification Java 2 Platform 2011).

5.3. Função de Aptidão

A Função de Aptidão é a forma como o Algoritmo Genético mede a capacidade do indivíduo de se adaptar ao ambiente, isto é, o quanto o cromossomo consegue representar uma solução para o problema proposto. A função de aptidão normalmente corresponde à “função-objetivo” do problema, fornecendo um valor que permite avaliar a qualidade da solução representada pelo indivíduo. (Dias 2004).

A função de aptidão desenvolvida é representada pelo somatório das distâncias euclidianas das variáveis de classificação de estilos de aprendizagem dos aprendizes pertencentes ao mesmo grupo. A função proposta procura minimizar as distâncias entre os aprendizes do mesmo grupo, procurando grupos que tenham um conjunto de aprendizes similares, tendo como referência a distância euclidiana de suas variáveis de aprendizagem.

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

Utilizamos o índice $s \in N$ para possibilitar a comparação entre os aprendizes agrupados no mesmo grupo.

5.4. Seleção

O método de seleção escolhido para a implementação foi o Método da Roleta. A seleção pelo Método da Roleta consiste em representar cada indivíduo da população em uma fatia da roleta e a área cedida para cada indivíduo corresponde, proporcionalmente, ao seu valor de aptidão; dessa forma, os indivíduos com maiores aptidões terão mais chances de serem escolhidos para formar a população intermediária que sofrerá as operações genéticas criando uma nova geração. A execução da seleção pelo método da roleta é realizada quantas vezes forem necessárias para formar a população intermediária com os indivíduos escolhidos.

5.5. Cruzamento

Consiste na combinação de segmentos de 2 (dois) cromossomos selecionados (Pais) para gerar dois novos cromossomos (Filhos). O método de cruzamento escolhido para a implementação foi o método de cruzamento de um ponto, o ponto de cruzamento é obtido de forma aleatória, utilizando a classe *Random* da linguagem JAVA (Oracle 2011), e os genes após o ponto de cruzamento são trocados entre os indivíduos, figura 2.

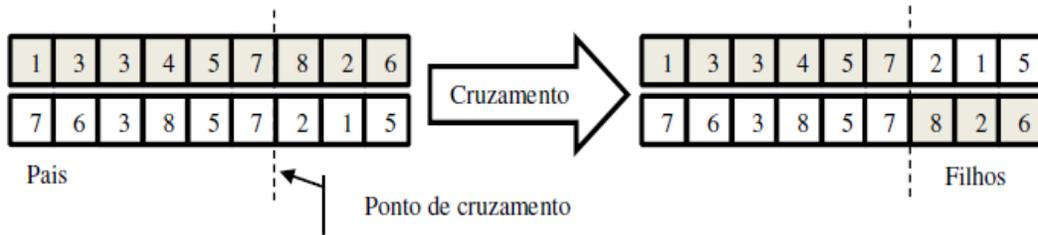


Figura 2: Exemplo de aplicação do operador de cruzamento de um ponto.

De acordo com Carvalho et al. (2003), a taxa de cruzamento (Tc) deve ser maior que a taxa de mutação, tendo usualmente uma probabilidade de $0,6 < Tc < 0,99$. A taxa de cruzamento utilizada no AG desenvolvido é de 0,8 da população.

5.6. Mutação

Na mutação alteram-se os valores dos genes (*alelos*) de um cromossomo, após o operador cruzamento, inserindo diversidade genética em uma população, e com isso, realizando buscas em diferentes áreas do espaço de solução, não permitindo que a população fique estagnada em uma região de ótimo local, figura3.

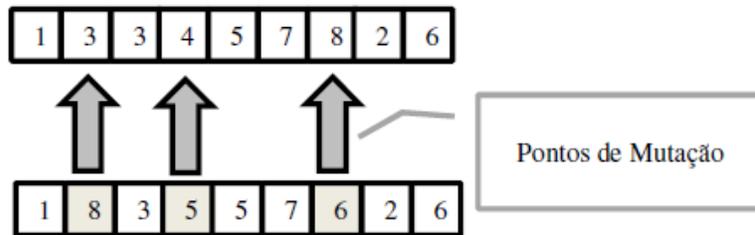


Figura 3: Exemplo de aplicação do operador de mutação

A aplicação do operador de mutação possui uma taxa de mutação (Tm) menor que a taxa de cruzamento(Tc) e essa taxa de probabilidade segundo Carvalho et al. (2003), deve ser de $0,01 < Tm < 0,10$.

A taxa de mutação utilizada na implementação é $Tm=0,3$, sendo aplicada sobre todos os genes e_i do cromossomo. Os *alelos* dos genes e_i selecionados são alterados aleatoriamente com $e_i \in \{1, 2, \dots, 8\}$.

5.7. Substituição da População

O mecanismo de substituição da população atual controla a porcentagem, (λ/μ) “gap geracional”, de μ pais da população atual que serão substituídos por λ filhos da população temporária, gerando assim, uma nova população. Com um intervalo geracional $\lambda(=\mu)$, indivíduos com alta aptidão poderão ser substituídos por indivíduos com baixa aptidão mais facilmente, desestruturando a busca. Já com um intervalo muito pequeno $\lambda (< \mu)$, o algoritmo pode se tornar lento.

O intervalo de geração utilizado nesse artigo é o Intervalo Geracional $\lambda (= \mu)$ acrescido de Elitismo, onde o melhor indivíduo de todas as anteriores gerações substitui o pior indivíduo dessa nova população.

5.8. Critério de Parada

Existem alguns critérios muito utilizados para terminar o critério de parada de um algoritmo genético, tais como: após um número de iterações, com base na convergência, isto é, quando após um dado número de gerações não for encontrado um indivíduo com melhor aptidão que a do melhor indivíduo das gerações anteriores, quando as aptidões se tornarem parecidas ou quando for encontrada a solução ótima para o problema.

O critério de parada utilizado nesse trabalho é o critério que leva em consideração a convergência da população. Nesse critério a execução do AG é interrompida quando ocorrer 10000 gerações sequenciais não houver melhora da melhor solução encontrada nas gerações passadas.

6 Testes Computacionais

6.1. Preparação dos Dados

Utilizou-se uma amostra de 54 alunos universitários regularmente matriculados no curso de Ciência da Computação no Centro Universitário Salesiano de São Paulo – Unidade de Lorena. A pesquisa foi realizada com todas as quatro séries do curso no ano letivo de 2008.

Tendo como população para o experimento os alunos do 1º, 2º, 3º, 4º séries do curso de Ciência da Computação e considerando o número médio de 30 alunos por série, teremos uma população de 120 alunos, e o tamanho da amostra para estimar a proporção de uma população finita para esse experimento é definida pela fórmula. (Martins 2008):

Dados: $N=120$; $Z=1,96$ para 95%; $p'=0,50$; $q'=(p'-1)=0,50$; $d=10\%= 0,10$.

$$Amostra = \frac{Z^2 \cdot p' \cdot q' \cdot N}{d^2(N - 1) + Z^2 \cdot p' \cdot q'} \quad (2)$$

onde:

- N = tamanho da população (número de aprendizes)
- Z = abscissa da distribuição normal padrão, fixado um nível de $(1-\alpha)\%$ de confiança do intervalo de confiança para a média. ($Z=1,96$ para 95%).
- p' = estimativa da verdadeira proporção de um dos níveis da variável escolhida. (Admite-se $p=0,50$, caso não haja estimativa prévia).
- $q' = 1-p'$.
- d = erro amostral expresso em decimais. Máxima diferença que o investigador admite suportar entre a verdadeira proporção (p) e a proporção do evento (p') a ser calculada com base na amostra.

Aplicando a fórmula acima obtemos o número mínimo de 54 alunos para representar a população para um erro amostral de 10%. Dessa forma, como o experimento está sendo realizado com uma amostra de 54 alunos, validamos a representatividade da amostra.

6.2. Método de Avaliação do Agrupamento

Para a validação do agrupamento foi utilizada uma técnica estatística conhecida como: contagem-z. Essa técnica tem a finalidade de avaliar o quanto um elemento está afastado da média dos elementos de seu grupo. Segundo Anderson et al. (2008), usando a média e o desvio padrão, é possível determinar a posição relativa de qualquer observação, pois cada observação além de estar relacionada à média e ao desvio padrão da amostra, também estará associada a um outro valor que se chama contagem-z.

$$Z_{e_i} = \text{M}\ddot{\text{A}}\text{X} \left| \frac{a_{sj} - \bar{a}_{e_{ij}}}{S_{e_{ij}}} \right|; \forall e_i = e_s; i, s \in \{1, \dots, n\} e j \in p \quad (3)$$

onde:

- $S_{e_{ij}}$ - representa o vetor dos desvios padrão amostral das variáveis de classificação de aprendizagem j dos aprendizes do grupo (*cluster*) representado pelo valor de e_i .
- $\bar{a}_{e_{ij}}$ - representa o vetor das médias das variáveis de classificação de aprendizagem j dos aprendizes do grupo (*cluster*) representado pelo valor de e_i .

Na análise da contagem-z ou também conhecida como valor padronizado podemos interpretar Z_{e_i} como o valor de desvios padrão 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).

Foi realizado o teste de *Shapiro-Wilk* para o conjunto de variáveis de aprendizagem (ATI: Ativo, REF:Reflexivo, SEN: Sensitivo, INTU: Intuitivo, VIS: Visual, VER: Verbal, SEQ: Sequencial, GLO: Global) dos 54 alunos, onde se determinou a estatística de teste média $w_{média}=0,9622$ e a probabilidade **p-valor**=0,101601 com um nível de significância de $\alpha=0,05$, validando assim, a normalidade do conjunto de variáveis de aprendizagem.

Anderson et al. (2008), também afirmam que para a detecção de pontos fora da curva podemos utilizar valores padronizados (contagem-z), pois em relação a dados com uma distribuição em forma de sino, quase todos os valores estão contidos em três desvios padrão da média. Sendo consideradas observações fora da curva àquelas que apresentarem valores z menores que -3 e valores z maiores que +3.

6.3. Resultados Computacionais

O AG foi implementado em linguagem de programação JAVA na versão 5.0 (Oracle, Tutorial JAVA EE5.0 2011), foi criado um ambiente web para que os aprendizes pudessem responder o questionário *ILS –Index of Learning Styles* (Felder e Soloman 2010) 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*. Os testes realizados são apresentados na Tabela 1, de acordo com as siglas **nFOB** – número de execuções da função objetivo, **FOB** – valor da função objetivo para a solução encontrada.

Todos os 50 testes foram realizados com os seguintes parâmetros genéticos Taxa de Cruzamento $Tc=0,8$, Taxa de Mutação $Tm=0,3$, Critério de Parada = 10000 gerações sequenciais sem melhora da melhor solução e Intervalo de Geração Geracional $\lambda(=\mu)$ com de Elitismo do melhor indivíduo de todas as anteriores gerações. Esses parâmetros genéticos foram determinados empiricamente, onde combinações aleatórias eram utilizadas até que se encontrasse a melhor, após encontrar essa combinação foram executados os 50 testes e uma dessas execuções foi analisada, Tabela 2, para validar o agrupamento. A análise dos testes realizados encontrou um valor médio para o número de execuções da função objetivo – $nFOB_{médio}=150122$ com um desvio padrão $S_{nFOB}=48025,9814$ e um valor médio para as 50 melhores soluções encontradas – $FOB_{médio}=765,8331$, com desvio padrão – $S_{FOB}=14,6748$.

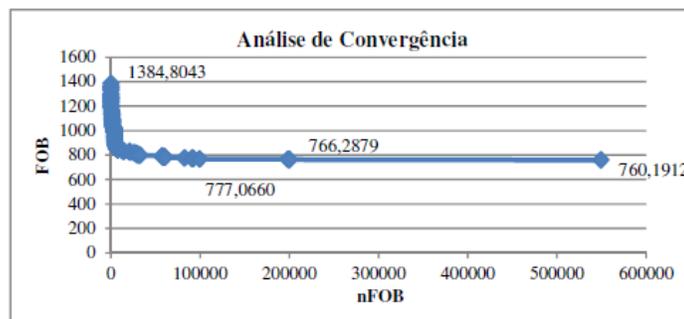


Figura 4: Gráfico da Análise de Convergência.

Tabela 1: Execuções do AG

Execuções	nFOB	FOB	Execuções	nFOB	FOB
1	142836	755,7087	26	104304	759,6803
2	137070	781,3475	27	241974	753,1103
3	118236	776,4259	28	166164	763,5906
4	104910	778,5111	29	112548	757,4209
5	102564	761,6021	30	101640	753,5603
6	181404	747,2395	31	197310	769,5435
7	155112	771,4897	32	137406	767,4463
8	179424	791,6596	33	228606	754,1723
9	139116	807,4047	34	90342	768,9969
10	92916	766,5646	35	116580	769,8406
11	173646	783,1562	36	87210	795,308
12	119010	747,2395	37	199644	762,3616
13	261708	754,0238	38	191034	771,2457
14	148092	758,932	39	92184	801,7921
15	222546	753,1368	40	132054	761,6831
16	189642	741,5339	41	119310	751,9338
17	98820	774,6985	42	109488	761,1869
18	191304	765,437	43	129360	760,0371
19	151362	775,3405	44	109614	775,8424
20	249006	753,4571	45	144522	769,4648
21	155202	751,9619	46	145872	744,091
22	281760	794,5603	47	168918	754,738
23	139854	775,3062	48	92268	765,3691
24	147732	764,2382	49	133716	762,3476
25	111312	753,8145	50	159462	752,1037
	nFOB_{médio}	Desvio Padrão nFOB		FOB_{médio}	Desvio Padrão FOB
MÉDIA	150122	48024,9814		765,8331	14,6748

Verificando o Coeficiente de Variação da amostra FOB, $CV = \frac{s}{\bar{x}} \cdot 100\% = 1,9\% < 15\%$, constatou-se que a amostra consiste de um conjunto de dados homogêneos. (Martins 2008)

Pode-se verificar, para uma execução aleatória com melhor **FOB**=760,1912, na Figura 4, que a convergência do AG se estabiliza, aproximadamente, a partir de **nFOB**=200000 chamadas da função objetivo, tendo **FOB**=766,2870, pois a próxima melhor solução encontrada pelo AG precisou de **nFOB**=350901 chamadas da função objetivo. Comparando o melhor resultado da Figura 4, com o resultado médio da Tabela 1 verificou-se que o valor **FOB**=760,1912 está a -0,3844 desvios padrão do valor médio encontrado, sendo dessa forma, um valor pertencente à população.

A Tabela 2 mostra os *clusters* formados pelo AG durante a execução 42 (um) com **FOB**=761,19 da Tabela 1, também são apresentadas as variáveis de classificação de estilos de aprendizagem de cada aprendiz: ATI: Ativo, REF: Reflexivo, SEN: Sensitivo, INTU: Intuitivo, VIS: Visual, VER: Verbal, SEQ: Sequencial, GLO: Global.

Tabela 2: Agrupamento Realizado pelo AG

GRUPO	APRENDIZ	ATI	REF	SEN.	INTU	VIS	VER	SEG	GLO
1	Aprendiz 4	7	4	8	3	8	3	4	7
	Aprendiz 7	9	2	8	3	7	4	4	7
	Aprendiz 15	8	3	8	3	7	4	6	5
	Aprendiz 23	6	5	10	1	9	2	6	5
	Aprendiz 28	11	0	9	2	8	3	5	6
	Aprendiz 31	9	2	10	1	8	3	6	5
	Aprendiz 47	11	0	7	4	7	4	5	6
MÉDIA		8,7	2,3	8,6	2,4	7,7	3,3	5,1	5,9
DESVPAD		1,89	1,89	1,13	1,13	0,76	0,76	0,9	0,9
Z ₁		1,70							

GRUPO	APRENDIZ	ATI	REF	SEN.	INTU	VIS	VER	SEG	GLO
2	Aprendiz 1	1	10	2	9	8	3	6	5
	Aprendiz 17	2	9	6	5	6	5	10	1
	Aprendiz 27	0	11	5	6	9	2	4	7
	Aprendiz 37	4	7	4	7	6	5	9	2
	Aprendiz 39	3	8	8	3	8	3	7	4
	Aprendiz 53	4	7	8	3	10	1	8	3
MÉDIA		2,3	8,7	5,5	5,5	7,8	3,2	7,3	3,7
DESVPAD		1,63	1,63	2,35	2,35	1,6	1,6	2,16	2,16
Z ₂		1,54							

GRUPO	APRENDIZ	ATI	REF	SEN.	INTU	VIS	VER	SEG	GLO
3	Aprendiz 10	7	4	9	2	6	5	6	5
	Aprendiz 20	8	3	8	3	5	6	9	2
	Aprendiz 24	6	5	9	2	4	7	5	6
	Aprendiz 32	6	5	8	3	4	7	3	8
	Aprendiz 41	7	4	8	3	5	6	4	7
	Aprendiz 45	7	4	8	3	6	5	7	4
MÉDIA		6,8	4,2	8,3	2,7	5	6	5,7	5,3
DESVPAD		0,75	0,75	0,52	0,52	0,89	0,89	2,2	2,2
Z ₃		1,55							

GRUPO	APRENDIZ	ATI	REF	SEN.	INTU	VIS	VER	SEG	GLO
4	Aprendiz 2	2	7	4	10	1	9	2	8
	Aprendiz 8	8	7	4	9	2	11	0	8
	Aprendiz 18	18	9	2	8	3	9	2	8
	Aprendiz 26	26	9	2	9	2	9	2	8
	Aprendiz 34	34	8	3	11	0	10	1	8
	Aprendiz 38	38	6	5	8	3	10	1	9
	Aprendiz 46	46	10	1	10	1	11	0	11
MÉDIA		8	3	9,3	1,7	9,9	1,1	8,6	2,4
DESVPAD		1,41	1,41	1,11	1,11	0,9	0,9	1,1	1,1
Z ₄		2,14							

Tabela 2: Agrupamento Realizado pelo AG (Continuação)

GRUPO	APRENDIZ	ATI	REF	SEN.	INTU	VIS	VER	SEG	GLO
5	Aprendiz 11	6	5	6	5	9	2	6	5
	Aprendiz 16	7	4	6	5	6	5	5	6
	Aprendiz 22	6	5	6	5	8	3	7	4
	Aprendiz 33	6	5	4	7	6	5	5	6
	Aprendiz 40	7	4	5	6	8	3	4	7
	Aprendiz 49	6	5	4	7	6	5	4	7
	Aprendiz 52	6	5	6	5	6	5	6	5
MÉDIA		6,3	4,7	5,3	5,7	7	4	5,3	5,7
DESVPAD		0,49	0,49	0,95	0,95	1,29	1,29	1,1	1,1
Z ₅		1,55							

GRUPO	APRENDIZ	ATI	REF	SEN.	INTU	VIS	VER	SEG	GLO
6	Aprendiz 5	9	2	7	4	11	0	7	4
	Aprendiz 14	10	1	6	5	10	1	8	3
	Aprendiz 19	6	5	5	6	8	3	9	2
	Aprendiz 30	9	2	7	4	8	3	8	3
	Aprendiz 44	8	3	5	6	9	2	6	5
	Aprendiz 51	8	3	5	6	7	4	9	2
	Aprendiz 54	8	3	5	6	10	1	8	3
MÉDIA		8,3	2,7	5,7	5,3	9	2	7,9	3,1
DESVPAD		1,25	1,25	0,95	0,95	1,41	1,41	1,1	1,1
Z ₆		1,82							

GRUPO	APRENDIZ	ATI	REF	SEN.	INTU	VIS	VER	SEG	GLO
7	Aprendiz 6	9	2	6	5	5	6	5	6
	Aprendiz 9	9	2	5	6	5	6	6	5
	Aprendiz 25	8	3	6	5	5	6	7	4
	Aprendiz 29	11	0	6	5	8	3	7	4
	Aprendiz 36	9	2	5	6	6	5	6	5
	Aprendiz 43	10	1	6	5	8	3	6	5
	Aprendiz 48	10	1	4	7	6	5	9	2
MÉDIA		9,4	1,6	5,4	5,6	6,1	4,9	6,6	4,4
DESVPAD		0,98	0,98	0,79	0,79	1,35	1,35	1,3	1,3
Z ₇		1,91							

GRUPO	APRENDIZ	ATI	REF	SEN.	INTU	VIS	VER	SEG	GLO
8	Aprendiz 3	5	6	6	5	4	7	4	7
	Aprendiz 12	6	5	2	9	7	4	1	10
	Aprendiz 13	4	7	6	5	4	7	5	6
	Aprendiz 21	3	8	6	5	2	9	5	6
	Aprendiz 35	4	7	6	5	6	5	4	7
	Aprendiz 42	4	7	6	5	4	7	1	10
	Aprendiz 50	4	7	5	6	7	4	3	8
MÉDIA		4,3	6,7	5,3	5,7	4,9	6,1	3,3	7,7
DESVPAD		0,95	0,95	1,5	1,5	1,86	1,86	1,7	1,7
Z ₈		2,20							

O agrupamento apresentado na Tabela 2 apresentou um agrupamento com o máximo valor de contagem-z $Z_8 = 2,20$, representando assim, um agrupamento sem pontos fora da curva (*outliers*), comprovando estatisticamente a homogeneidade dos grupos formados pelo AG.

Tabela 3: Agrupamento Realizado pelo *K-means*

APRENDIZ	GRUPO	APRENDIZ	GRUPO	APRENDIZ	GRUPO
Aprendiz 1	6	Aprendiz 19	3	Aprendiz 37	4
Aprendiz 2	5	Aprendiz 20	8	Aprendiz 38	5
Aprendiz 3	1	Aprendiz 21	6	Aprendiz 39	6
Aprendiz 4	3	Aprendiz 22	5	Aprendiz 40	7
Aprendiz 5	7	Aprendiz 23	2	Aprendiz 41	3
Aprendiz 6	2	Aprendiz 24	2	Aprendiz 42	4
Aprendiz 7	4	Aprendiz 25	2	Aprendiz 43	3
Aprendiz 8	8	Aprendiz 26	8	Aprendiz 44	3
Aprendiz 9	6	Aprendiz 27	3	Aprendiz 45	6
Aprendiz 10	2	Aprendiz 28	1	Aprendiz 46	2
Aprendiz 11	2	Aprendiz 29	6	Aprendiz 47	8
Aprendiz 12	1	Aprendiz 30	7	Aprendiz 48	4
Aprendiz 13	1	Aprendiz 31	2	Aprendiz 49	1
Aprendiz 14	4	Aprendiz 32	5	Aprendiz 50	7
Aprendiz 15	8	Aprendiz 33	6	Aprendiz 51	2
Aprendiz 16	6	Aprendiz 34	4	Aprendiz 52	8
Aprendiz 17	4	Aprendiz 35	8	Aprendiz 53	3
Aprendiz 18	6	Aprendiz 36	7	Aprendiz 54	5

O agrupamento realizado pela heurística *K-means* apresentou um valor máximo de contagem-z no *cluster* 6 $Z_6 = 2,13$, também representando assim, um agrupamento sem pontos fora da curva (*outliers*), mas o valor da função objetivo, quando utilizada para avaliar a similaridade dos elementos de cada grupo, apresentou um valor maior, **FOB**=1363,7488.

7 Conclusão

Foi realizada a classificação e o agrupamento (*clustering*) de um conjunto de alunos das quatro séries do curso de Ciência da Computação do Centro Universitário Salesiano de São Paulo – Unidade de Lorena, e constatou-se uma predominância do curso pelos estilos de aprendizagem: ATIVO, SENSORIAL, VISUAL e SEQUENCIAL.

O agrupamento foi realizado através de um algoritmo genético proposto. Foram realizados testes de eficiência para a escolha dos melhores parâmetros genéticos, sendo escolhidos os seguintes valores: *Taxa de Cruzamento* - **Tc**=0,8, *Taxa de Mutação* - **Tm**=0,03, *Tamanho da População* - **Np**= 6 indivíduo, *Intervalo de Geração Geracional com elitismo do melhor indivíduo das gerações anteriores*, **Critério de Parada** =10000 gerações sem melhorar o valor da melhor solução.

A heurística *K-means* também foi utilizada para o agrupamento dos aprendizes, apresentando um valor máximo de contagem-z no *cluster* 6 $Z_6 = 2,13$, o que demonstra que o método heurístico prioriza a minimização das distâncias com base nos centroides formados, não levando em consideração as distâncias entre todos os elementos dos clusters formados. Dessa forma, o agrupamento pelo método *K-means* resultou em um agrupamento pior que o realizado pelo AG, pois a similaridade entre os aprendizes de cada cluster é menor sendo mensurada através da função objetivo **FOB**=1363,7488.

Dessa forma, o AG proposto possui maior eficiência na *clusterização* de aprendizes com base no instrumento de classificação de estilos de aprendizagem *Index of Learning Styles - ILS* (Felder e Soloman2010), pois com auxílio do método estatístico contagem-z, foi validada a similaridade interna dos clusters formados, pois não foram encontrados pontos fora da curva (*outliers*) dentro dos limites de -3 a +3 desvios padrão da média, sendo também mais eficiente que o método heurístico *K-means*, proporcionando grupos de aprendizes mais homogêneos.

Agradecimentos

Agradecimentos ao CNPq pelo auxílio nos projetos: 300692/2009_9 e 470813/2010_5.

Referências Bibliográficas

- Anderson, D. R., D. J. Sweeney, e T. A. Williams. *Estatística Aplicada à Administração e Economia*. 2a. edição. São Paulo: CENGAGE Learning, 2008.
- Carvalho, A. C. P. L. F., A. P. Braga, e T. B. Ludemir. “*Computação Evolutiva*.” In: *Sistemas Inteligentes: Fundamentos e Aplicações*, por S. O. Resende, 225-239. Barueri, São Paulo: Manole, 2003.

Cury, Helena Noronha. “*Estilos de Aprendizagem de Alunos de Engenharia.*” XXVIII CONGRESSO BRASILEIRO DE ENSINO DE ENGENHARIA, 2000: 9.

Dias, C. R. *Algoritmos Evolutivos para Problemas de Clusterização de Grafos Orientados: Desenvolvimento e Análise Experimental.* Mestrado, Computação Aplicação, Niterói, Rio de Janeiro: Universidade Federal Fluminense, 2004, 111.

Felder, R. M., e B. A. Silverman. “*Index of Learning Styles In Engineering Education.*” Journal of Engineering Education 78, n. 7 (1988): 674-680.

Felder, R. M., e B. A. Soloman. “*Index of Learning Styles.*” 2010. <http://www.engr.ncsu.edu/learningstyles/ilsweb.html> (acesso em 20 de Setembro de 2010).

Felix, Zildomar, e Patrícia Tedesco. “*Formação de Grupos de Aprendizagem em Ambientes CSCL Ciente de Contexto.*” III SEGeT – Simpósio de Excelência em Gestão e Tecnologia, 2006: 11.

Filho, P. A. C., e R. J. Poppi. “*Algoritmo Genético em Química.*” Química Nova (Química Nova) 22, n. 3(1999): 405-411.

Holland, J. H. “*Adaptation in Natural and Artificial Systems.*” MIT Press, 1975: 11-147.

Jain, A. K, M. N. Murty, e P. J. Flynn. “*Data Clustering: A Review.*” ACM Computing Surveys, Setembro de 1999: 264-323.

Kim, Jin, Yong-Hyuk Kim, Inwood Hwang, e Byung-Ro Moon. “*Genetic Approaches for Graph Partitioning: A Survey.*” GECCO 11. Dublin, Ireland, 2011. 8.

Lopes, W. M. G. *Inventário de Estilos de Aprendizagem de Felder-Soloman: Investigação de sua Validade em Estudantes Universitários de Belo Horizonte.* Dissertação (Mestrado em Engenharia de Produção), Engenharia de Produção, Florianópolis: Universidade Federal de Santa Catarina, 2002, 85.

Lorena, L. A. N., e J. C. Furtado. “*Constructive Genetic Algorithm for Clustering Problems.*” Evolutionary Computation (Massachusetts Institute of Technology) 9, n. 3 (2001): 309-327.

Machado, C. S., M. D. Palhano, M. L. Peconick, e V. A. Ávila. “*Estilos de Aprendizagem – Uma Abordagem Utilizando o ILS - Index of Learning Styles.*” 2010. http://www.abepro.org.br/biblioteca/ENEGEP2001_TR111_0240.pdf (acesso em 19 de Fevereiro de 2010).

Martins, G. A. *Estatística Geral e Aplicada.* 3. São Paulo: Atlas, 2008.

Miasaki, C. T., e R. Romero. “*Um Algoritmo Genético Especializado Aplicado ao Planejamento da Expansão do Sistema de Transmissão com Alocação de Dispositivos de Compensação Série.*” Revista Controle & Automação 18, n. 2 (2007): 210-222.

Mingoti, S. A. *Análise de Dados Através de Métodos de Estatística Multivariada.* Belo Horizonte: Editora UFMG, 2005.

Oracle. “*API Specification Java 2 Platform.*” Vers. 1.4.2. 2011. <http://download.oracle.com/javase/1.4.2/docs/api/java/util/Random.html> (acesso em 15 de Junho de 2011).

—. “*Tutorial JAVA EE 5.0.*” 2011. <http://download.oracle.com/javase/5/tutorial/doc> (acesso em 15 de Junho de 2011).

Pimentel, Edson P, Vilma F. de França, e Nizam Omar. “*A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização.*” XIV Simpósio Brasileiro de Informática na Educação, 2003: 495-504.

Ribeiro, G. F., M. S. Nagano, e L. A. N. Lorena. “*Metaheurística Híbrida Algoritmo Genético-Clustering Search para redução de estoques em processamento em sistemas de produção Flow Shop permutacional.*” XXXIX Simpósio Brasileiro de Pesquisa Operacional. Fortaleza, 2007.

Trindade, Áthila Rocha, e Luiz Satoru Ochi. “*Um Algoritmo Evolutivo Híbrido para a Formação de Células de Manufatura em Sistemas de Produção.*” Pesquisa Operacional, Maio-Agosto de 2006: 255-294.

Zakrzewska, D. “*Cluster Analysis in Personalized E-Learning Systems.*” Intelligent Systems for Knowledge Management, Studies in Computational Intelligence 252 (2009): 229-250.