

## UMA NOVA ABORDAGEM PARA AUMENTAR A DIVERSIDADE GENÉTICA DA POPULAÇÃO INICIAL EM ALGORITMOS GENÉTICOS

Thiago Hilgert Souza  
Universidade Federal da Grande Dourados,  
Faculdade de Ciências Exatas e Tecnologia  
Dourados, MS, Brasil  
*thiagohilgert@gmail.com*

Celso G. Camilo-Junior  
Universidade Federal de Goiás,  
Instituto de Informática  
Goiânia, GO, Brasil  
*celsocamilo@gmail.com*

### RESUMO

A diversidade da população em Algoritmos Genéticos é fundamental para obtenção de uma solução otimizada para o problema a ser resolvido. É interessante que esta diversidade seja estabelecida desde sua população inicial, pois é baseado nela que novas populações serão formadas. Para tal, este trabalho propõe uma nova técnica para a geração da população inicial de Algoritmos Genéticos com foco na diversidade. A técnica proposta foi implementada e comparada com outras duas técnicas, por meio de testes de diversidade. Os resultados obtidos apresentam um bom desempenho.

**Palavras-chave:** Metaheurísticas, Algoritmos Genéticos, População Inicial.

**Área principal:** Metaheurística.

### ABSTRACT

The diversity of the population in Genetic Algorithms is essential to obtaining a optimized solution for the problem. It is interesting that this diversity is established since its initial population, because the others are based on that. Therefore, this paper proposes a new technique to generating the initial population of Genetic Algorithms with a focus on diversity. The proposed technique was implemented and compared with two other techniques for diversity. The results show a good performance.

**Keywords:** Metaheuristics, Genetic Algorithms, Initial Population.

**Main area:** Metaheuristics.

## 1. INTRODUÇÃO

A Computação Evolucionária é uma subárea da Inteligência artificial que propõe modelos, inspirados na Teoria da Evolução de Darwin (Holland, 1975), para a solução de problemas.

Uma característica fundamental para o bom desempenho de um Algoritmo Genético é a diversidade de sua População Inicial (Hill, 1999), pois essa é diretamente responsável por um problema denominado convergência prematura (Castro, 2001).

A implementação de uma estrutura que melhore esta diversidade pode representar um ganho significativo em tempo no processo de busca por uma solução apropriada.

Por isso, este trabalho propõe uma nova técnica que visa melhorar esta diversidade, através de uma estrutura inspirada nas árvores binárias.

A seção 2 apresenta uma breve descrição sobre os Algoritmos Genéticos. A seção 3 dedica-se a expor com maior ênfase o conteúdo relacionado à População Inicial de um Algoritmo Genético, discorrendo sobre a importância da sua diversidade, as principais maneiras de gerá-la e a apresentação de uma nova forma de gerá-la, proposta neste trabalho. Na seção 4, técnicas para calcular a diversidade de uma População são apresentadas. Os testes, com seus cenários e resultados, serão abordados na seção 5. Por fim, a seção 6 conclui este artigo com as análises finais e as possibilidades de trabalhos futuros.

## 2. ALGORITMOS GENÉTICOS

Algoritmos Genéticos (AGs) é uma das técnicas compreendidas na Computação Evolutiva e aplicada a problemas de busca e otimização. Sua fundamentação é baseada nos princípios da Teoria da Evolução, proposta pelo naturalista inglês Charles Darwin, em meados do século XIX. (Holland, 1975)

No AG cada indivíduo da população é um candidato à solução do problema. Esse indivíduo é representado por um cromossomo, que possui uma sequência de genes. No AG clássico, a maneira de representar o indivíduo proposta por Holland (1975) é através de arranjos binários (Figura 1). Entretanto, este tipo de representação em algumas aplicações pode levar a um desempenho insatisfatório, sendo necessário representar o cromossomo de outra maneira, como arranjos de números reais ou inteiros, ou até mesmo através de estruturas mais complexas (Von Zuben, 2000).

1	0	1	0	0	1	0	1	1	0
---	---	---	---	---	---	---	---	---	---

Figura 1. Exemplo de Codificação Binária de um indivíduo da população.

Dentro desta população é aplicado um processo de seleção, onde os indivíduos que tem uma melhor adaptação possuem maiores chances de serem selecionados. Baseado nestes indivíduos os processos evolutivos de cruzamento e mutação são implementados, de forma a se obter uma evolução na qualidade da população com o passar das gerações (Bäck, 1998). A Figura 2 representa um Algoritmo Genético Clássico:

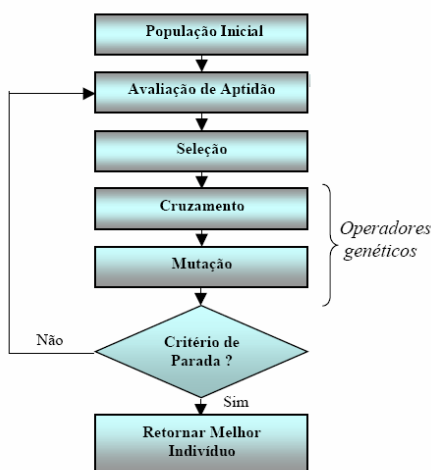


Figura 2. Estrutura básica de um AG clássico (Pozo et. al., 2007)

### 3. POPULAÇÃO INICIAL

A População Inicial é o ponto de partida de um AG, é através dela que se atingirá a solução final. Esta população deve conter diversidade suficiente para permitir ao algoritmo combinar características e produzir novas soluções (Pozo et al. 2007). A baixa diversidade desta população pode significar uma convergência prematura (Azevedo, 1999), onde as soluções para o problema se concentram em uma determinada região do espaço de busca, que pode não representar as melhores soluções (Gao, 2003).

#### 3.1. TÉCNICAS PARA A GERAÇÃO DA POPULAÇÃO INICIAL

Existem várias formas de se gerar a População Inicial de um AG, as mais comuns são as seguintes:

- **Geração Aleatória:** Cada gene do cromossomo recebe um valor gerado aleatoriamente.
- **Geração Baseada em Conhecimento Específico sobre o problema:** Por exemplo, em uma situação onde a codificação do cromossomo é binária e sabe-se que a solução final deverá apresentar mais 0s do que 1s, esta informação pode ser utilizada para a geração (Von Zuben, 2000).
- **Geração Complementar:** Em cromossomos com representação binária, metade da População é gerada aleatoriamente e a outra metade recebe a inversão dos bits da primeira metade. Esta técnica garante que toda posição na cadeia genética da população tenha a mesma quantidade de 0s e 1s (Lacerda, 2008). Isto estabelece uma diversidade genética, a partir da metade aleatória.
- **Geração Híbrida:** Gera-se parte da população de forma aleatória e o restante aplicando alguma das técnicas citadas.

#### 3.2. GERAÇÃO DA POPULAÇÃO INICIAL ATRAVÉS DA ÁRVORE DIVERSIFICADORA POPULACIONAL (ADP)

Em virtude do baixo número, na literatura, de técnicas que sejam aplicadas a problemas genéricos, de representação binária, e que promova uma diversidade de valores no espaço de busca, uma nova técnica, inspirada em Árvores binárias, denominada Árvore Diversificadora Populacional (ADP) é proposta.

Este mecanismo consiste em estabelecer, no processo de geração da população inicial de um AG, duas estruturas do tipo Árvore: uma denominada Árvore Começando com Zero (ACZ) e outra denominada Árvore Começando com Um (ACU). Em cada estrutura há um mecanismo para se calcular a quantidade de indivíduos que cada Árvore contém.

No ADP a construção de cada indivíduo é auxiliado pelas árvores ACZ e ACU. A medida que se insere elementos na árvore preenche-se o indivíduo na população. Para exemplificar o funcionamento, segue um exemplo de geração. Neste será usada uma população de indivíduos com três genes e representação binária. Sendo assim, o tamanho máximo que esta população pode assumir, sem que haja indivíduos iguais, ou seja, totalmente diversificada, é de 8 indivíduos ( $2^3$ ).

A figura 3 apresenta as duas árvores em seu estado inicial e também o estado inicial da população de indivíduos.

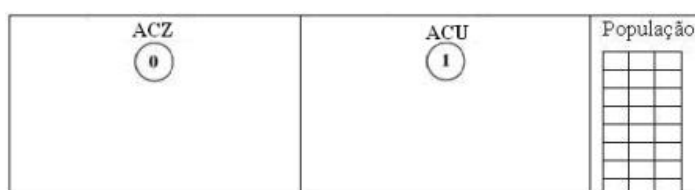


Figura 3. Exemplo das Estruturas do Tipo Árvore em seu estado inicial.

A inserção de elementos na árvore depende do número de indivíduos que a mesma contém, sendo que a com menor quantidade recebe o indivíduo. Como inicialmente ambas as estruturas contém o mesmo número de elementos, aleatoriamente escolhe-se uma das árvores para receber o primeiro indivíduo. Supondo que a árvore sorteada foi a ACZ, chamamos o procedimento de inserção na ACZ, assim o valor do primeiro alelo do cromossomo do primeiro indivíduo assume valor 0, conforme mostra a Figura 4.

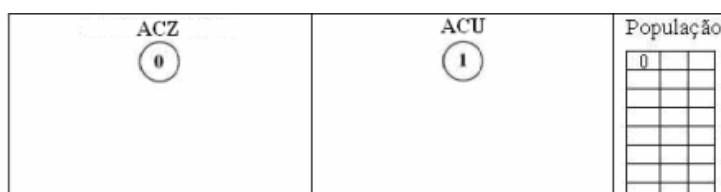


Figura 4. Exemplo das Estruturas do Tipo Árvore em seu processo inicial de povoamento.

Em seguida, verifica-se se as sub-Árvores estão nulas, se ambas estiverem nulas, sorteia-se para qual sub-árvore será chamado o procedimento de inserção. Se apenas uma estiver nula, inserimos o elemento nesta, já se ambas não estiverem nulas, inserimos na que possuir menor número de elementos. No nosso exemplo, a sub-Árvore direita da ACZ é sorteada, logo o valor do segundo gene do primeiro cromossomo assume valor 1, como apresentado na figura 5.

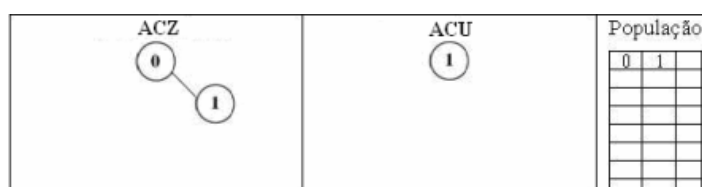


Figura 5. Geração do segundo alelo do primeiro indivíduo.

Posteriormente, chamamos o procedimento de inserção para a sub-Árvore esquerda da ACZ (aleatoriamente sorteada) e o mesmo processo é aplicado, ver figura 6.

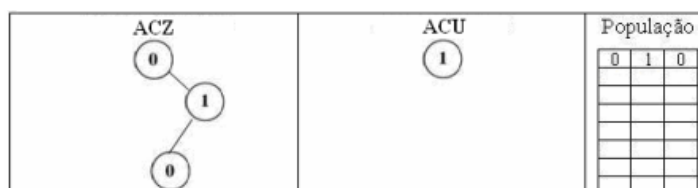


Figura 6. Geração do terceiro alelo do primeiro indivíduo.

Quando o procedimento de inserção retornar o indivíduo calcula-se os elementos de cada sub-árvore, para ser usado como decisão nas próximas inserções. No exemplo, o próximo indivíduo deverá ser inserido na ACU, em virtude dela possuir menor número de elementos em relação a ACZ. A figura 7 ilustra como ficaria, após 24 inserções, as Árvores com a sua carga total de elementos possíveis, no caso 8.

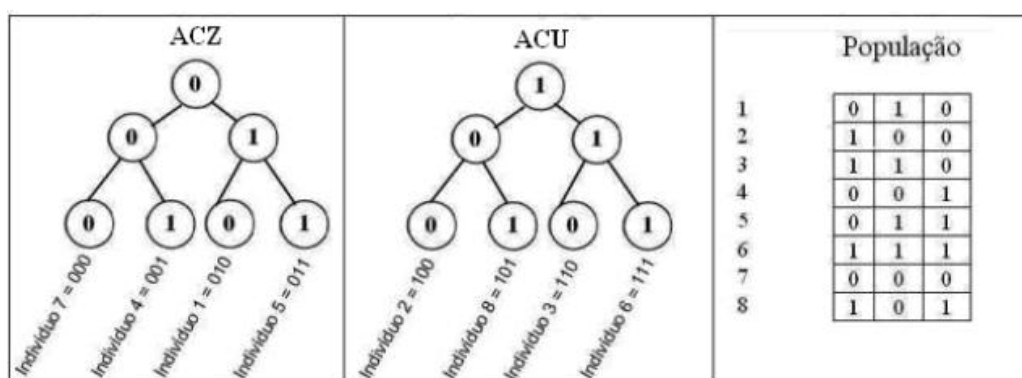


Figura 7. Exemplo das estruturas do tipo árvore totalmente povoadas.

## 4. CÁLCULO DE DIVERSIDADE

No que diz respeito à população dos AGs, duas formas de diversidade devem ser consideradas: a diversidade genética, que é a diversidade codificada expressada pela configuração cromossômica dos indivíduos, e a diversidade decodificada, que possui uma maior relevância em virtude de determinar a aptidão dos indivíduos (Gomez, Hougen, 2007). A seguir, apresenta-se maneiras de se calcular estas diversidades.

### 4.1. CÁLCULO DA DIVERSIDADE CODIFICADA

Uma das formas de calcular a diversidade codificada (ou diversidade genética) de uma população com representação binária é pelo modelo proposto por Gudwin e Von Zuben (2002).

Primeiro, é necessário calcular a frequência  $P$  de ocorrência do alelo 1 para cada gene, ver equação 1.

$$P = \frac{W}{N} \quad (1)$$

onde  $N$  é o número de indivíduos da população e  $W$  é o número de indivíduos que possui o alelo 1 em um determinado gene.

Quanto mais próximo de 0,5 o valor de  $P$ , mais próximo da diversidade máxima o gene se encontra. A partir desta definição, aplicamos este cálculo para todos os genes da cadeia cromossômica da população e obtemos a sua média. Logo, quanto mais próxima a média for de 0,5, mais diversificada geneticamente está o indivíduo e, conseqüentemente, a população.

## 4.2. CÁLCULO DA DIVERSIDADE DECODIFICADA

Uma forma de calcular a diversidade decodificada de uma população é em função da capacidade de exploração do espaço de busca decodificado, específico de cada problema. Para tal, é necessário conhecer os limites deste espaço

Para exemplificar uma maneira de realizar este cálculo, considera-se cromossomos de três genes que assumem valores binários e que decodificados expressam números inteiros. Logo, a faixa do espaço de busca é de 0 a 7. A Figura 7 apresenta a população, que usaremos nos próximos exemplos, em seu estado codificado e decodificado.

População codificada	População decodificada																																
<table><tr><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>1</td><td>1</td></tr><tr><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>0</td></tr><tr><td>0</td><td>1</td><td>1</td></tr></table>	1	0	0	0	0	1	0	1	0	1	1	1	0	1	0	0	0	1	1	1	0	0	1	1	<table><tr><td>4</td></tr><tr><td>1</td></tr><tr><td>2</td></tr><tr><td>7</td></tr><tr><td>2</td></tr><tr><td>1</td></tr><tr><td>5</td></tr><tr><td>3</td></tr></table>	4	1	2	7	2	1	5	3
1	0	0																															
0	0	1																															
0	1	0																															
1	1	1																															
0	1	0																															
0	0	1																															
1	1	0																															
0	1	1																															
4																																	
1																																	
2																																	
7																																	
2																																	
1																																	
5																																	
3																																	

Figura 8. Exemplo de uma população com indivíduos em sua representação codificada e decodificada.

Em seguida, dividi-se este espaço de busca em intervalos. Neste exemplo, como o tamanho da população é pequeno em relação ao espaço de busca, dividiremos esta população em 8 intervalos, sendo eles [0, 1, 2, 3, 4, 5, 6, 7]. Depois de se determinar a quantidade de intervalos que a população assumirá, calcula-se qual seria a distribuição ideal (DI) de indivíduos dentro de cada intervalo, em função do tamanho da população, conforme a equação 2.

$$DI = TPop / Z \quad (2)$$

onde DI determina a distribuição ideal de indivíduos por intervalo, TPop representa o tamanho da população e Z é o número de intervalos que foi criado.

No nosso exemplo, a distribuição ideal de indivíduos por intervalo é de 1. Assim, calcula-se quantos indivíduos há em cada intervalo que foi criado e subtrai pela distribuição ideal de indivíduos no mesmo intervalo, ver equação 3.

$$K = QI - DI \quad (3)$$

onde QI indica a quantidade de indivíduos no intervalo e DI distribuição ideal de indivíduos no mesmo intervalo.

Depois, efetua-se a soma dos valores obtidos nos intervalos e dividi-se pelo número de intervalos. Ver equação 4.

$$DD = \sum_{i=1}^z |K| / z \quad (4)$$

onde z é o número de intervalos.

Assim, DD representa a diversidade decodificada da população. Desta forma, quanto mais próximo de zero estiver o valor de DD, melhor está a distribuição dos indivíduos pelo espaço de busca. O cálculo da diversidade, do exemplo adotado, é mostrado na tabela 1.

Intervalos	QI	K
0	0	-1
1	2	1
2	2	1
3	1	0
4	1	0
5	1	0
6	0	-1
7	1	0
Soma dos  K		4
Diversidade Decodificada (DD)		0,571428571

Tabela 1. Demonstração da distribuição dos indivíduos por intervalos e cálculo da diversidade decodificada da população.

## 5. TESTES

O problema de maximização da função F6 é utilizado para os testes. Esta função é um benchmarking com vários picos e vales, ver equação 5.

$$F6(x,y) = 0,5 * \frac{(\sin \sqrt{x^2 + y^2})^2 * 0,5}{(1,0 + 0,0001(x^2 + y^2))^2} \quad (5)$$

Esta função é difícil de ser otimizada, pois possui uma grande variedade de máximos locais.

Considerando o domínio das variáveis x e y igual a [-100, 100], com 4 a 5 casas de precisão. A codificação do cromossomo para esta função assume 44 bits, onde os 22 primeiros representam os valores de x e os 22 últimos representam os valores de y.

Assim, o objetivo dos testes é avaliar a diversidade dos valores gerados para as variáveis da função.

### 5.1 CENÁRIOS

Como o propósito deste trabalho é tratar da diversidade da população inicial de um AG, os processos evolutivos de cruzamento e mutação não foram aplicados.

As técnicas de geração da população inicial implementadas para a análise dos resultados foram as seguintes: Geração Aleatória, Geração Complementar e a Geração ADP, proposta neste trabalho.

Os testes foram aplicados para populações de 50, 100, 150, 200, 250, 300, 350, 400, 450 e 500 indivíduos, onde para cada população foram feitas dez execuções para cada técnica.

Como os valores de x e y são obtidos de forma igual, ou seja, os 22 primeiros bits são atribuídos a x, e os 22 últimos a y, utiliza-se para os testes apenas os valores de x.

Para calcular a diversidade decodificada do problema testado, foram criados 40 intervalos, com variação de 5 unidades entre cada um, sendo eles: [-100|-95, -95|-90, -90|-85, -85|-80, -80|-75, -75|-70, -70|-65, -65|-60, -60|-55, -55|-50, -50|-45, -45|-40, -40|-35, -35|-30, -30|-

25, -25|-20, -20|-15, -15|-10, -10|-5, -5|0, 0|5, 5|10, 10|15, 15|20, 20|25, 25|30, 30|35, 35|40, 40|45, 45|50, 50|55, 55|60, 60|65, 65|70, 70|75, 75|80, 80|85, 85|90, 90|95, 95|100].

A implementação das estruturas e os testes foram realizados em uma IDE NetBeans 6.5, com compilador Java JDK 1.7.0, e em um processador Pentium IV de 3.0 Ghz e Memória 1 Gb.

## 5.2 RESULTADOS

As tabelas 2, 3 e 4 apresentam as diversidades codificadas e decodificadas dos indivíduos gerados pelas técnicas de geração Aleatória, Complementar e ADP, respectivamente.

ALEATÓRIA		
Tamanho da População	Diversidade Decodificada de X	Diversidade Genética
50	0,895	0,49941
100	1,2875	0,49908
150	1,6575	0,49909
200	1,815	0,50
250	2,0425	0,50
300	2,3425	0,50
350	2,5125	0,50
400	2,7275	0,50
450	2,805	0,50
500	2,59	0,50
<b>Média</b>	<b>2,0675</b>	<b>0,499648091</b>

Tabela 2. Resultados dos cálculos de diversidade obtidos pela geração Aleatória.

COMPLEMENTAR		
Tamanho da População	Diversidade Decodificada de X	Diversidade Genética
50	0,86	0,5
100	1,21	0,5
150	1,59375	0,5
200	1,8525	0,5
250	2,13875	0,5
300	2,1825	0,5
350	2,39375	0,5
400	2,6125	0,5
450	2,81875	0,5
500	2,83	0,5
<b>Média</b>	<b>2,04925</b>	<b>0,5</b>

Tabela 3. Resultados dos cálculos de diversidade obtidos pela geração Complementar.



ADP		
Tamanho da População	Diversidade Decodificada de X	Diversidade Genética
50	0,575	0,45368
100	0,69	0,50155
150	0,66125	0,4993
200	0,6225	0,49985
250	0,68625	0,50
300	0,915	0,50
350	0,90625	0,50
400	0,8	0,50
450	0,8575	0,50
500	0,92	0,50
<b>Média</b>	<b>0,763375</b>	<b>0,495938</b>

Tabela 4. Resultados dos cálculos de diversidade obtidos pela geração ADP.

As figuras 8, 9 e 10 apresentam a distribuição dos valores decodificados de x dentro de cada intervalo criado. Para a representação dos gráficos, foram utilizados os valores obtidos na quinta execução com população de 200 indivíduos. Sendo a DI igual a 5 indivíduos por intervalo.

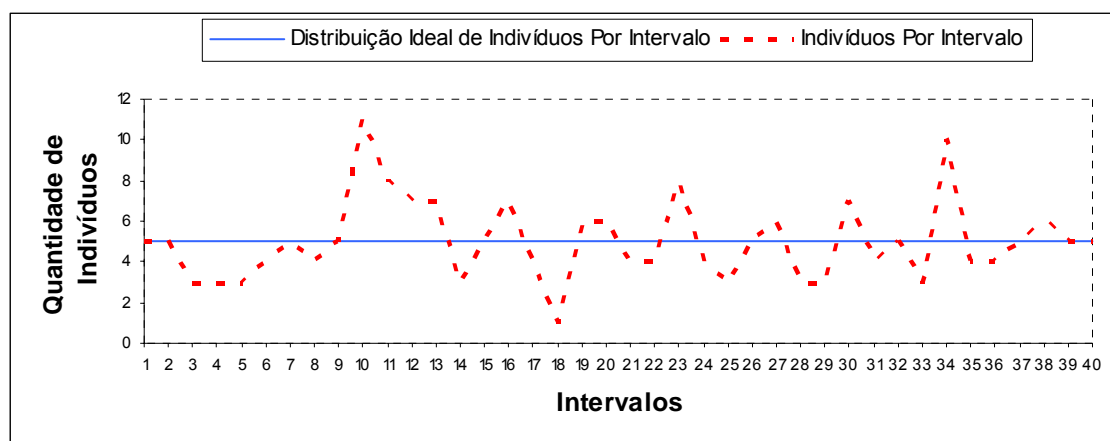


Figura 8. Gráfico da quantidade de indivíduos por intervalo, através da geração Aleatória.

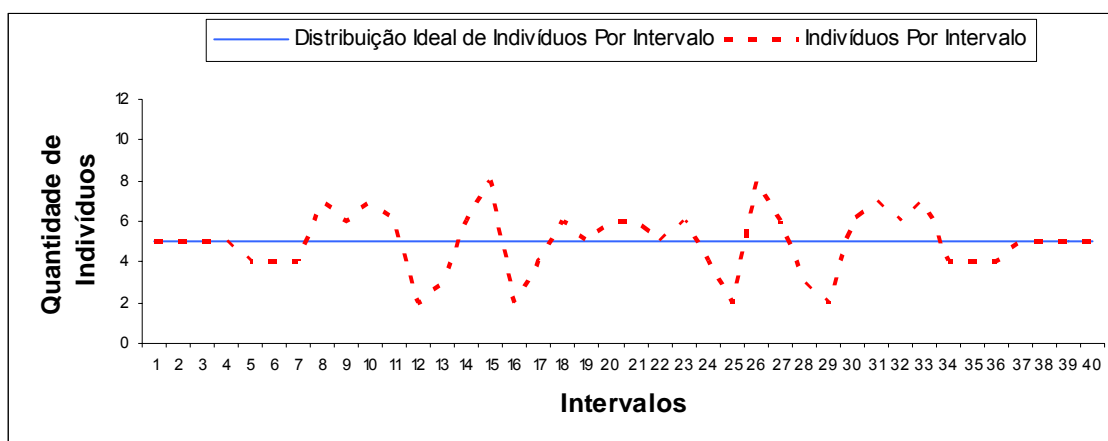


Figura 9. Gráfico da quantidade de indivíduos por intervalo, através da geração Complementar.

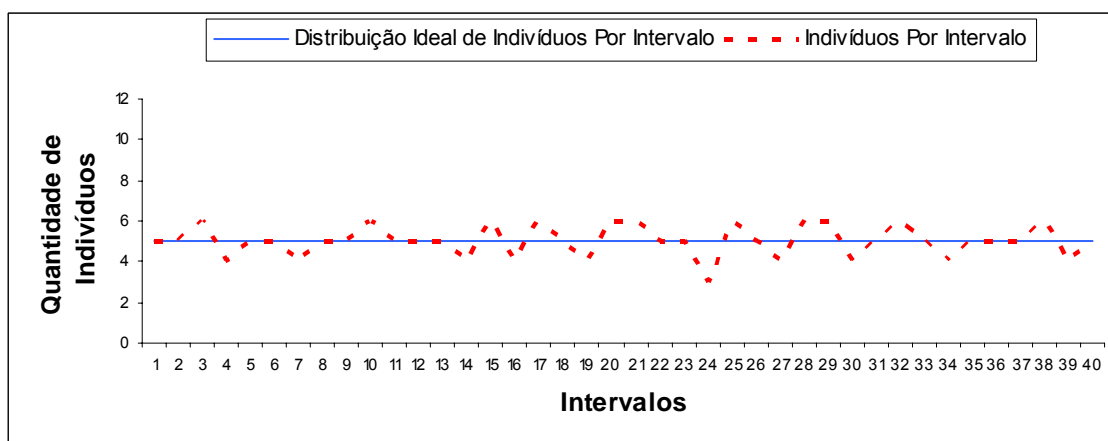


Figura 10. Gráfico da quantidade de indivíduos por intervalo, através da geração ADP.

### 5.3. ANÁLISE DOS RESULTADOS

Os resultados obtidos através dos testes apontaram que a técnica ADP, ofereceu uma diversidade decodificada 63 % melhor em relação as demais técnicas de geração da população inicial testadas, pois o valor médio da sua distribuição decodificada foi de 0,763375 e o das demais técnicas 2,0675 e 2,04925. Já a diversidade codificada, durante os testes a pequena variação existente entre os resultados, obtidos pelas estruturas, não exerceu influência significativa nos resultados.

## 6. CONCLUSÃO

Os AGs são aplicados a problemas de busca e otimização de alta complexidade (Gomez, Hougen, 2006), como: Roteamento de Veículos (Ribeiro e Lorena, 2005), Mineração de Dados (Santos, 2008), Problemas relacionados a Formação de Células de Fabricação (Gonçalves Filho et. al, 2004), Otimização de Alternativas de Desenvolvimento de Campos Petrolíferos (Villanueva, 2008), Balanceamento de Linhas de Produção (Mayler e Santos, 2003), Geração de Tabelas de Horários (Freitas et. al, 2007), etc. Problemas estes que nem sempre se conhece uma solução completamente eficiente (Von Zuben, 2000).

Em virtude disto, em alguns casos, encontrar uma solução adequada pode levar uma quantidade significativa de tempo (Silva, 1998).

Por isso, este trabalho propôs implementar uma estrutura que melhorasse a diversidade, promovendo um ganho significativo na exploração no espaço de busca.

Os resultados dos testes comprovam a melhora na diversidade da população inicial, em relação às demais técnicas testadas, em todos os cenários de testes.

Desta forma, como objeto de trabalhos futuros, a ADP poderia ser avaliada segundo a qualidade final da solução produzida pelo AG.

## 7. REFERÊNCIAS

- Azevedo, Fernando M.** Algoritmos Genéticos em Redes Neurais Artificiais. Relatório técnico sobre Algoritmos Genéticos e Redes Neurais Artificiais. Universidade Federal de Santa Catarina, 1999.
- Bäck, T.** Evolutionary Algorithms in Theory and Practice. Oxford University Press, 1996.
- Castro, R. E.** Otimização de Estruturas com Multi-Objetivos Via Algoritmos Genéticos. Tese para obtenção do grau de Doutor em Ciências Em Engenharia Civil. Universidade Federal do Rio de Janeiro, 2001.
- Freitas, C. et. al.** Uma ferramenta Baseada em Algoritmos Genéticos para Geração de Tabela de Horário Escolar. 7ª Escola Regional de Computação dos Estados da Bahia, Alagoas e Sergipe, 2007.
- Gao, Y.** Population size and sampling complexity in genetic algorithms. In Proceedings of the Bird of a Feather Workshops (GECCO2003), 2003.
- Gomez, P. A. D, e Hougén, D. F.** Pseudo Random Generation of the Initial Population in Genetic Algorithms: an Experiment. 2nd Annual Computer Science Research Conference at the Stephenson Research and Technology Center, University of Oklahoma, 2006.
- Gomez, P. A. D, e Hougén, D. F.** Empirical Study: Initial Population Diversity and Genetic Algorithm Performance. Artificial Intelligence and Pattern Recognition 2007, 2007.
- Gonçalves Filho, E. V. et. al.** Algoritmo Genético Para a Formação de Células de Fabricação. XXIV Encontro Nacional de Engenharia de Produção, 2004.
- Gudwin, e Von Zuben.** Estudo de diversidade populacional: efeito da taxa de mutação. Relatório técnico sobre Diversidade Populacional, Universidade Estadual de Campinas, 2002.
- Hill, R. R.** A Monte Carlo Study of Genetic Algorithm Initial Population Generation Methods. Proceedings of the 1999 Winter Simulation Conference. Air Force Institute of Technology, 1999.
- Holland, J. H.** Adaptation in Natural and Artificial Systems. The University of Michigan Press, 1975.
- Lacerda, E. G. M.** Algoritmos Genéticos: Aspectos Práticos. Relatório técnico sobre Algoritmos Genéticos. Universidade Federal do Rio Grande do Norte, 2008
- Mayler, S. F. e Santos, R. N.** Algoritmo Genético Para o Balanceamento de Linhas de Produção. XXIII Encontro Nacional de Engenharia de Produção, 2003.
- Pozo, et al.** Computação Evolutiva. Relatório técnico disponível na internet, elaborado pelo Grupo de Pesquisas em Computação Evolutiva da Universidade Federal do Paraná. <http://www.inf.ufpr.br/aurora/tutoriais/Ceapostila.pdf>, 3, 2009.
- Ribeiro, M. R. e Lorena, L. A. N.** Roteamento de Veículos Dinâmico Usando Algoritmos Genéticos. XIX Congresso de Ensino e Pesquisa em Transportes, 2005.
- Santos, J. S.** Mineração de Dados Utilizando Algoritmos Genéticos. Monografia apresentada ao curso de graduação em Ciência da Computação da Universidade Federal da Bahia, 2008.
- Silva, T. T., Barbosa.** Inicialização da População em Algoritmos Genéticos com Codificação Real. V Seminário sobre Elementos Finitos e Métodos Numéricos em Engenharia, Universidade Federal de Juiz de Fora, 1998.
- Villanueva, J. M. M.** Otimização de Alternativas de Desenvolvimento de Campos Petrolíferos por Algoritmos Genéticos. Revista Inteligência Computacional Aplicada. PUC-RIO. 2008.
- Von Zuben, F. J.** Computação Evolutiva: Uma Abordagem Pragmática. Manual sobre Computação Evolutiva. Universidade Estadual de Campinas, 2000.