

Desempenho do Algoritmo Genético com Iteração Retroviral para otimização de funções com representação real

Performance of the Genetic Algorithm with Retroviral Iteration to real representation functions optimization

Dielle Corrêa Franco Pereira, SUPDE/DEBML/DE204, Serviço Federal de Processamento de Dados - SERPRO, diellefranco@yahoo.com.br

Renato Simões Moreira, Pró-Reitoria de Pesquisa e Inovação, Instituto Federal de Brasília - IFB, renatosm@gmail.com

Otávio Noura Teixeira, Laboratório de Computação Natural, Centro Universitário do Pará - CESUPA, onoura@gmail.com

Roberto Célio Limão de Oliveira, Faculdade de Eng. da Comp. e de Telecomunicações, Universidade Federal do Pará - UFPA, limao@ufpa.br



Resumo: Este trabalho apresenta uma análise do efeito da variação dos parâmetros no desempenho do Algoritmo Genético Retroviral Iterativo (AGRI), utilizado para o problema de otimização de funções reais. São analisadas várias configurações do Algoritmo Genético infectado com retrovírus na busca do máximo das funções *benchmarking* F1 (Shifted Sphere Function), F2 (Shifted Schwefel's Problem), F3 (Shifted Rotated High Conditioned Elliptic Function) e F5 (Schwefel's Problem 2.6 with Global Optimum on Bounds). Os resultados comprovam o melhor desempenho do AGRI frente ao AG clássico e ao AG com alta taxa de mutação utilizando as configurações com infecção fraca, média e agressiva do AGRI.

Palavras-chave: Algoritmo Genético; Infecção por Retrovírus; Função Real; Maximização.

Abstract: *This work presents an analysis of the effect for varying the parameters on the Retroviral Iterative Genetic Algorithm's performance (RIGA), used for real functions optimization problem. Various Genetic Algorithm infected with retroviruses' settings are analyzed intending the best of benchmarking functions F1 (Shifted Sphere Function), F2 (Shifted Schwefel's Problem), F3 (Shifted Rotated High Conditioned Elliptic Function) e F5 (Schwefel's Problem 2.6 with Global Optimum on Bounds). The results showed the AGRI best performance, against the classic AG and high mutation rate AG, using low, medium and aggressive infection settings.*

Keywords: Genetic Algorithm; Retroviruses' Infection; Real Function; Maximization.

A habilidade de combinar bons pedaços (*building blocks*) de soluções diferentes e juntá-los em uma única super-solução, é o fator do sucesso dos AGs (Algoritmos Genéticos) clássicos. Apesar disso, surgem problemas quando os pedaços de composição utilizados não se encontram agrupados dentro do cromossomo (Silva, Gouvêa e Freitas, 2001, p. 127-136). O uso das teorias da evolução biológica, as quais estudam as características de evolução dos vírus, permite uma série de recursos muito úteis ao projeto de algoritmos computacionais evolutivos, pois envolve processos que procuram encontrar um equilíbrio entre os esforços de intensificação e diversificação da busca.

Os Algoritmos Genéticos utilizam, de modo geral, como ferramentas de operacionalização, os chamados mecanismos de seleção, cruzamento e mutação. O processo imita a chamada evolução Darwiniana e pode se tratar de cruzamento entre células. Convém notar que, no processo reprodutivo biológico, a seleção pertence a um contexto diferente daquele próprio

ao cruzamento e mutação dos cromossomos. A seleção reprodutiva ocorre dividida em duas partes: no contexto ambiental, quando da escolha do parceiro de reprodução, e na da corrida das células reprodutivas. Enquanto a seleção ocorre em um universo macroscópico ou, no mínimo, multicelular, o cruzamento e as mutações ocorrem no contexto da célula embrião, da microbiologia; pode-se dizer que seu fluxo de operacionalização da manipulação genética é extracelular (Kargupta e Sterfford, 1998).

É possível abordar a resolução de problemas considerando aspectos e paradigmas da evolução biológica que não somente a chamada seleção natural. Uma forma de utilizar os conhecimentos da evolução biológica, diferente da Teoria da Seleção Natural, é incorporar o fluxo intracelular no contexto computacional da metáfora evolucionária. Com isto tem-se a possibilidade da incorporação de informações de forma planejada no processo evolucionário e a manipulação de genes, com o objetivo de melhorar o código genético. As manipulações podem ser

realizadas, em analogia ao modelo biológico, por meio de agentes lógicos apropriados à codificação computacional. Os principais agentes da manipulação genética são os Vírus, Enzimas Mensageiras e as Partículas Genéticas Móveis (Goldbarg e Gouvêa, 2000). Neste trabalho, o agente transgenético utilizado para a manipulação é o vírus da família Retroviridae, devido ao fato de os mesmos serem os seres que evoluem mais rapidamente na natureza (Carter e Saunders, 2007).

A família Retrovírus compreende vírus que utilizam a RNA-dependente-DNA-polimerase, ou seja, uma transcriptase reversa (TR), como particularidade de sua estratégia replicativa no organismo hospedeiro (Baltimore, 1970, pp. 1209-1211). Os retrovírus normalmente infectam células somáticas e possuem uma fase denominada "pró-vírus", representada pela integração do genoma viral no genoma do hospedeiro (Coffin e Hughes, 1997). ERVs, do inglês *endogenous retroviruses*, representam a fase pró-viral dos respectivos retrovírus exógenos que, em algum momento, se integraram em células germinativas do seu hospedeiro. Verticalmente transmitidos aos descendentes, os pró-vírus colonizaram definitivamente seus genomas (Li e Graur, 1991).

O papel de ERVs na geração de novos genes, bem como no aumento da variabilidade entre os organismos, talvez seja um dos mais relevantes efeitos positivos da presença desses elementos no hospedeiro (Romano, Ramalho e Zanotto, 2006). Retroelementos com integração polimórfica, assim como LTRs solitárias (do inglês, Long Terminal Repeats) contribuem com a variação alélica na população (Barbulescu *et al.*, 1999). A origem de novos genes, que ocorre por meio de rearranjos e duplicação gênica, é fundamental para a evolução dos genomas, possibilitando a variabilidade fenotípica (Le Rouzic e Capy, 2005). Já é consenso que inversões cromossômicas, rearranjos, inserções e deleções são os principais responsáveis pelas diferenças observadas entre as diferentes linhagens de primatas (Sverdlov, 1990), (Frazer *et al.*, 2003). Visto que retroelementos atuam como *hotspots* para esses eventos, sua participação na evolução dos genomas de primatas parece evidente (Sverdlov, 1990).

Desta maneira, as referências anteriores mostram que o uso do conceito do vírus, para manter a diversidade em uma população, é uma metáfora biológica que pode ser utilizada na Computação Evolucionária.

A ideia da infecção viral vem sendo utilizada em algoritmo genético desde meados dos anos 90, quase sempre em substituição ao operador de mutação, como em: (Kano e Nakamura, 2000), (Kano, Hasegawa e Kato, 1997, pp. 531–537), (Kano, Kondo e Sugimoto, 2002), (Fontes *et al.*, 2007), entre outros. E segundo (Kano, Kondo e Sugimoto, 2002), o uso da Infecção viral no Algoritmo Genético melhora a taxa de busca do algoritmo.

Este trabalho mostra um estudo da variação de parâmetros em uma nova meta-heurística evolutiva inspirada na estrutura viral da família dos retrovírus em conjunto com o Algoritmo Genético, denominado de 'Algoritmo Genético Retroviral Iterativo' (em inglês RIGA - *Retroviral Iterative Genetic Algorithm*), com o objetivo de avaliar a influência dos parâmetros na variabilidade genética da população e adaptação dos indivíduos ao ambiente, de modo a direcionar o modelo de configuração dos parâmetros do Algoritmo Genético Retroviral Iterativo diante do desempenho obtido nos ensaios (Moreira *et al.*, 2011).

AGRI - ALGORITMO GENÉTICO RETROVIRAL ITERATIVO

Os AGs - Algoritmos Genéticos - são essencialmente algoritmos de busca, direcionada pelo operador seleção, de um mínimo ou de um máximo de uma função. O melhor indivíduo é encontrado por meio de uma função de avaliação (*fitness*) que deve representar o ambiente no qual uma população evolui, passando por fases de cruzamentos, mutações e seleção natural (Linden, 2012).

O princípio fundamental é explorar o espaço de busca por meio de uma população, representada por cromossomos, cuja evolução depende das operações de cruzamento e mutação, bem como do processo de seleção, seguindo a inspiração da regra da evolução das espécies (Sonkar, 2012). O AG tradicional, denominado de canônico, tem a dificuldade de manter a diversidade da população, levando a uma convergência prematura das soluções, o que indica um ponto ótimo local (Rudolph, 1994). Alguns estudos mostram o enfrentamento da convergência prematura por meio do controle da diversidade da população (Tragante e Tinós, 2009), (Zhu, 2003). Uma forma de se manter a diversidade em uma população, manipulada pelo AG, é a introdução de infecção viral (Silva *et al.*, 2001), (Goldbarg e Gouvêa, 2000), (Kano, 1997), (Moreira *et al.*, 2011), (Villarreal, 2009).

A principal razão para o uso de uma estrutura viral na manipulação genética do AG é o fato de que vírus estão associados com a fonte de inovação genética, a qual é influenciada pela alta taxa de replicação e mudança que estas estruturas apresentam (Romano, 2009). Desta forma, o AGRI apresenta-se como uma ferramenta de troca de material genético entre os cromossomos da população, e não apenas como ferramenta de mutação.

No próximo subitem serão apresentados os componentes do AGRI (Moreira *et al.*, 2011).

Vírus no AGRI

Os vírus no AGRI são estruturas que têm o mesmo tamanho do cromossomo, no entanto, com alguns espaços vazios, visto que, a ideia consiste em compartilhar o material genético e evitar outra população de cromossomos trabalhando em paralelo. A quantidade de espaços vazios e as suas disposições são determinadas aleatoriamente. Deste modo, para um problema que requer uma representação binária de oito posições, alguns vírus têm a informação como mostrado na Figura 1.

(A)		1		0		1	0	
(B)			1		1			1
(C)	0			1		1		

Figura 1: Possíveis vírus (A),(B),(C) para um cromossomo binário de oito posições (Moreira et al.,2011).

Para a criação de novos vírus, que irão compor a população viral, o AGRI foi inspirado no processo natural do retrovírus chamado de transcrição reversa. O processo consiste basicamente dos passos mostrados na Figura 2.

(A)		1		1		1	0	
(B)	0	1	1	1	1	0	0	1
(C)	0	1	1	1	1	1	0	1
(D)	0	1		1				1

Figura 2: Processo de criação de um novo vírus: (A) Vírus aleatório; (B) Cromossomo da população; (C) Cromossomo auxiliar contendo uma mistura do material genético de ambos (vírus e cromossomo); (D) Novo vírus contendo material do vírus A e do cromossomo B. Depois que o material genético é misturado em (C), espaços em branco são gerados aleatoriamente formando o novo vírus (D) (Moreira et al.,2011).

Infecção

A infecção é o processo de inclusão do material genético viral no cromossomo hospedeiro, o que requer um vírus e um cromossomo. O cromossomo alvo terá o seu material genético modificado na mesma posição onde os genes estão arranjados no vírus. Toda a informação genética viral irá ser copiada para o cromossomo alvo, excetuando os espaços vazios, o qual irá ser preenchido pelo cromossomo hospedeiro. A infecção no AGRI está representada na Figura 3.

(A)		1		1		1	0	
(B)	0	1	1	1	1	0	0	1
(C)	0	1	1	1	1	1	0	1

Figura 3: A infecção no cromossomo: (A) Vírus; (B) Cromossomo; (C) Cromossomo infectado (Moreira et al.,2011).

O processo de infecção depende exclusivamente de um fator: aumento da aptidão do cromossomo. A infecção tem sucesso quando o cromossomo infectado tem um aumento na sua aptidão e fracassa quando essa aptidão diminui. Assim, os piores vírus, que têm menor taxa de infecção, serão extintos. A razão desse procedimento é impedir que um bom cromossomo, ao ser infectado, torne-se um péssimo cromossomo. Desse modo, somente as infecções com sucesso são importantes para o AGRI.

Dessa maneira, o algoritmo AGRI tem apenas uma etapa a mais do que o AG tradicional ('Aplicação Viral'), como visto no algoritmo a seguir:

1. Inicializa população aleatória

2. Aplicação Viral

I - Gera uma população viral aleatória, caso seja primeira vez

II - Gera novos vírus baseados na população

III - Propaga a infecção nos indivíduos

a) Infeciona cada cromossomo escolhido com cada um dos Vírus existentes.

b) Para cada cromossomo infectado com sucesso, ou seja, quando houver aumento na fitness, incrementar 1 para o Vírus, caso contrário, decrementar 1.

c) Verificar quais foram os Vírus com maiores graus de infecção e manter de acordo com a porcentagem os melhores para a próxima geração viral.

3. Efetua Seleção, Reprodução e Mutação.

4. Avalia os indivíduos infectados e verifica a solução, se solução não for encontrada voltar ao passo 2.

PARÂMETROS DO AGRI

O AGRI utiliza os mesmos parâmetros do AG clássico (número de indivíduos, taxa de mutação, taxa de cruzamento, elitismo, o tipo de seleção e o tipo de cruzamento). Entretanto, ao aplicar os conceitos de infecção viral por retrovírus, faz-se necessário definir outros parâmetros, conforme (Moreira et al.,2011):

1. **Taxa de infecção da população:** é a taxa de cromossomos que irão ser infectados na geração.

2. **Taxa de Elitismo Viral:** é o percentual de vírus que será mantido na população viral da próxima geração. Se esse valor foi maior do que 50%, a evolução fica comprometida.
3. **Número de vírus:** é o tamanho da população viral. Se esse número for grande, o tempo de execução do algoritmo também é grande.
4. **Infecção dos mais fracos:** esse parâmetro força a infecção dos piores cromossomos favorecendo uma rápida evolução na população cromossômica.
5. **Infecção Simples:** esse parâmetro força uma única infecção por cromossomo em cada geração.
6. **Taxa de infecção interna:** este parâmetro indica o percentual máximo de material genético de qualquer cromossomo que irá formar um novo vírus.

METODOLOGIA

Com o objetivo de analisar como os parâmetros afetam, contribuem e se complementam para obtenção de um melhor desempenho, isto é, para alcançar o ótimo global nos algoritmos AGRI, AG Clássico e AG com alta taxa de mutação, serão estudadas as seguintes características da população: Melhor *Fitness*, *Fitness* Médio, Pior *Fitness* e Diversidade Genética da População.

Os fatores que influenciam nessas características são os parâmetros do Algoritmo Genético canônico e os parâmetros da infecção viral, conforme apresentados na seção 3.

Para a realização dos experimentos foi utilizada uma população inicial com 100 indivíduos, representados por cromossomos, cujos genes são números reais gerados, a partir de uma função de distribuição de probabilidade uniforme.

Um experimento comparativo foi realizado para analisar a contribuição de cada um dos parâmetros da infecção viral no desempenho do AGRI. Um parâmetro por vez foi alterado, enquanto os demais se mantinham fixos. Após essa análise

inicial, foram realizados experimentos fatoriais considerando a combinação dos parâmetros que apresentaram maior contribuição no desempenho do AGRI, individualmente. Tanto os experimentos simples quanto os experimentos fatoriais tiveram o mesmo número de repetições, a saber 60 execuções, e cada ensaio com 500 gerações.

Definidas as configurações virais, novos experimentos foram realizados para definir a configuração dos parâmetros do Algoritmo Genético canônico que, aliada à melhor configuração da infecção viral, aumentaria o desempenho do AGRI. Seguiu-se uma série de experimentos para avaliar cada um dos parâmetros do AG clássico individualmente e, posteriormente, a combinação dos parâmetros que melhor contribuíram para a evolução do *fitness* da população. Cada ensaio teve 500 gerações e foram realizadas 60 repetições para cada experimento.

O resultado das configurações obtidas, em relação ao desempenho do AGRI, é apresentado na seção de resultados.

Os comparativos de desempenho entre as configurações delineadas para o AGRI, AG clássico e AG com alta taxa de mutação foram realizados com experimentos que utilizaram uma população inicial fixa, como controle local e experimentos que utilizaram populações iniciais variáveis, considerando a casualização para lograr sua validade. Ambos os tipos de população têm 100 cromossomos, gerados por uma função de distribuição de probabilidade uniforme. Esses experimentos – objeto de estudo desta pesquisa – são apresentados na Seção de Resultados.

RESULTADOS

Para os testes do AGRI, utilizou-se como problema encontrar o máximo das funções *benchmarking* F1 (Shifted Sphere Function), F2 (Shifted Schwefel's Problem), F3 (Shifted Rotated High Conditioned Elliptic Function) e F5 (Schwefel's Problem 2.6 with Global Optimum on Bounds), que tratam do problema de otimização de variáveis reais (Suganthan, 2005). As funções são mostradas nas equações (1), (2), (3) e (4):

$$F_1(x) = \sum_{i=1}^D z_i^2 + f - bias_1, z = X - 0, X = [x_1, x_2, \dots, x_D] \quad (1)$$

$$F_2 = \sum_{i=1}^D \left(\sum_{j=1}^i z_j \right) + f - bias_2, Z = X - 0, X = [x_1, x_2, \dots, x_D] \quad (2)$$

$$F_3 = \sum_{i=1}^D (10^6)^{\frac{i-1}{D-1}} z_i^2 + f - bias, Z = (X - 0) * M, X = [x_1, x_2, \dots, x_D] \quad (3)$$

$$F_5 = \max \{ (A_i x - B_i) + f - bias_5, i = 1, \dots, D, x = [x_1, x_2, \dots, x_D] \} \quad (4)$$

Os testes foram realizados para algumas configurações do AGRI, a princípio denominadas de melhor, média e pior. A primeira representa uma configuração que apresentou os piores resultados para uso do AGRI, a segunda representa um comportamento médio do AGRI e a última é aquela considerada com o melhor desempenho do AGRI. Para se conseguirem os parâmetros dessas três configurações, foram realizados ensaios em que os parâmetros do AG permaneceram com valores fixos e os parâmetros da infecção viral foram alterados, individualmente, para avaliação do desempenho da infecção viral. Encontrada a melhor definição das taxas de cada parâmetro da infecção viral, novos ensaios foram realizados para determinar a melhor configuração do AG em relação à melhor configuração viral. Ensaios com o AG clássico, tendo uma alta taxa de mutação, também foram realizados como critério de comparação de desempenho com o AGRI. Abaixo são exibidas, na Tabela 1, as configurações inicialmente classificadas como melhor, média e pior do AGRI, e na Tabela 2 a parametrização do AG clássico.

Tabela 1: Definição dos parâmetros de configuração das infecções virais com desempenhos considerados melhor, médio e pior nos ensaios.

Parâmetros da Infecção Viral	Configurações		
	Melhor	Média	Pior
Taxa de Infecção Populacional (%)	60	50	20
Taxa de Elitismo Viral (%)	10	20	10
Quantidade de Vírus	60	50	40
Taxa de Infecção Interna (%)	40	30	20
Infecção dos mais fracos	Sim	Sim	Sim

Tabela 2: Definição dos parâmetros do AG clássico com melhor desempenho no AGRI

Parâmetros do AG	Melhor Configuração
Taxa de Elitismo (%)	5
Taxa de Mutação (%)	3
Taxa de Cruzamento (%)	90
Tipo de Seleção	Torneio
Tipo de Cruzamento	Ponto de Corte

Para os ensaios com o AG clássico configurado com alta taxa de mutação, estipulou-se essa taxa com o valor 50%. Os demais parâmetros foram configurados conforme a Tabela 2.

Esse estudo revelou que a diversidade genética da população cromossômica é promovida à medida que o elitismo viral, que segue o mesmo princípio do elitismo do AG, é baixo; a infecção viral em indivíduos mais fracos é ativada e o nível de infecção populacional é alto. Configurando a infecção viral dessa forma, a variabilidade da população viral que irá infectar a população cromossômica é maior; a probabilidade de os indivíduos menos aptos, com características importantes para a obtenção do ótimo global, participarem das operações genéticas aumenta, assim como a quantidade de cromossomos da população aptos à infecção.

O acréscimo na quantidade de vírus melhora o desempenho do AGRI por aumentar o número da população viral que irá infectar a população cromossômica, propiciando maior variedade genética viral exploratória. Verifica-se que a taxa de infecção interna muito baixa e a infecção única de um cromossomo por geração, diminuem a performance do AGRI, pois são gerados vírus com material genético reduzido, ou seja, trechos de solução menores com pouca contribuição significativa e os cromossomos podem não receber um material genético de qualidade.

Considerando o resultado dos ensaios realizados (50 para o AGRI, 50 para o AG clássico, que também utiliza os parâmetros da Tabela 2, e 50 ensaios para o AG com alta taxa de mutação, todos com a mesma população inicial), a melhor configuração do AGRI teve um desempenho superior ao AG clássico e ao AG com alta taxa de mutação em todos os experimentos. Nesses 50 ensaios o AGRI, na sua melhor configuração, apresentou um comportamento semelhante em praticamente todos eles.

Os gráficos das Figuras (4), (5), (6) e (7) mostram o desempenho do AGRI em comparação com o AG clássico e com o AG configurado com alta taxa de mutação para os melhores *fitness* considerando a mesma população inicial com 100 indivíduos, no decorrer de 500 gerações, em 50 ensaios para as funções F1, F2, F3 e F5. Para cada ensaio, o melhor *fitness* foi extraído totalizando 50 melhores cromossomos para o AGRI, para o AG clássico e para o AG com Alta Taxa de Mutação, os quais foram relacionados à geração em que cada um ocorreu. No decorrer desta seção de resultados, o AGRI é representado pela curva de linha azul, o AG clássico pela curva de linha verde e o AG com alta taxa de mutação pela curva de linha vermelha.

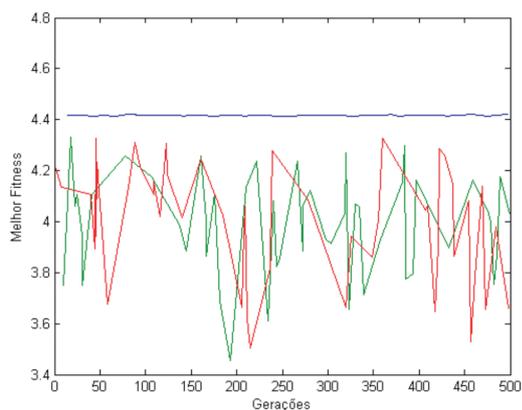


Figura 4: Evolução do melhor fitness para função F1 em 50 ensaios com 500 gerações cada. Comparação entre a melhor configuração viral, o AG clássico e o AG com alta taxa de mutação.

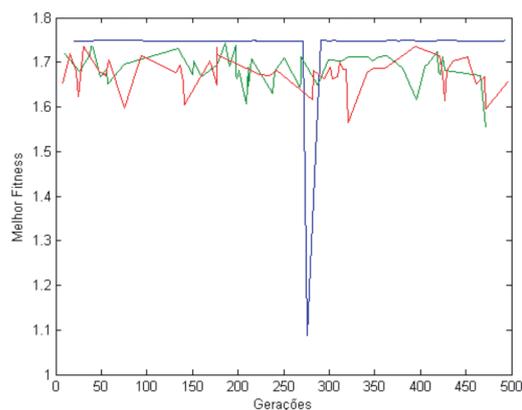


Figura 7: Evolução do melhor fitness para função F5 em 50 ensaios com 500 gerações cada. Melhor configuração viral, o AG clássico e o AG com alta taxa de mutação.

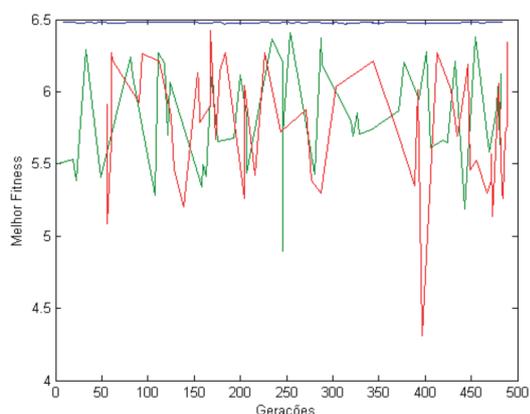


Figura 5: Evolução do melhor fitness para função F2 em 50 ensaios com 500 gerações cada. Melhor configuração viral, o AG clássico e o AG com alta taxa de mutação.

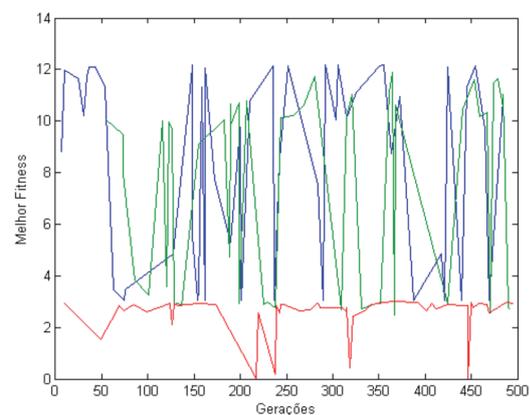


Figura 6: Evolução do melhor fitness para função F3 em 50 ensaios com 500 gerações cada. Melhor configuração viral, o AG clássico e o AG com alta taxa de mutação.

A seguir, serão apresentados os resultados de um ensaio com 5000 gerações para a função F1 com uma mesma população inicial fixa (o desempenho do AGRI para as outras funções, F2, F3 e F5, é muito semelhante). Visto que, o comportamento da configuração média é bastante próximo da melhor configuração do AGRI, será apresentada a comparação das configurações consideradas melhor e pior. O resultado é expresso nas Figuras (8), (9), (10) e (11) onde a curva azul representa o indivíduo com melhor *fitness*, a curva verde representa a *fitness* médio da população e a curva vermelha representa o indivíduo com pior *fitness*.

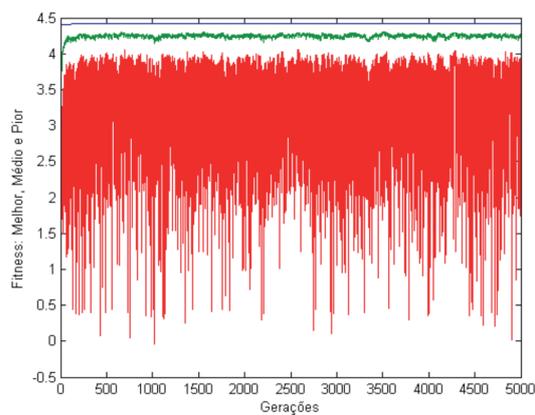


Figura 8: F1- Melhor, média e pior fitness para melhor configuração do AGRI em 1 ensaio com 5000 gerações.

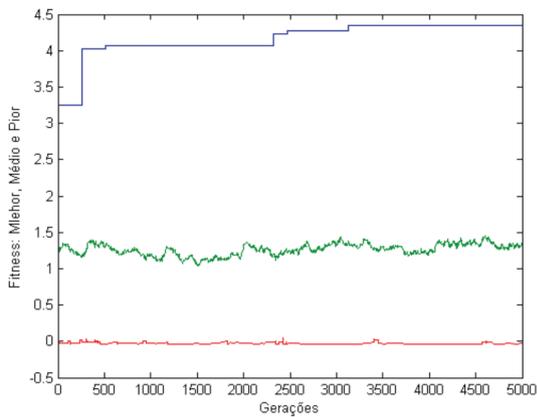


Figura 9. F1: Melhor, média e pior fitness para a pior configuração do AGRI em 1 ensaio com 5000 gerações.

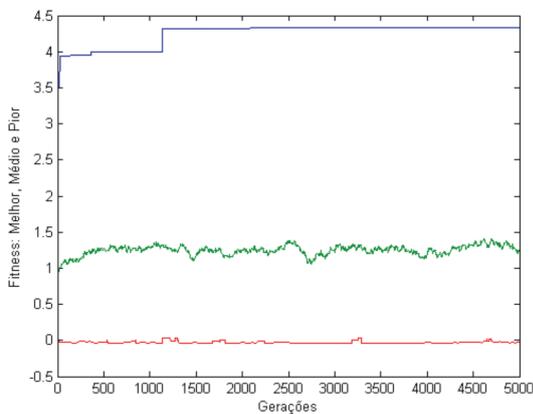


Figura 10: F1- Melhor, média e pior fitness para o AG clássico em 1 ensaio com 5000 gerações.

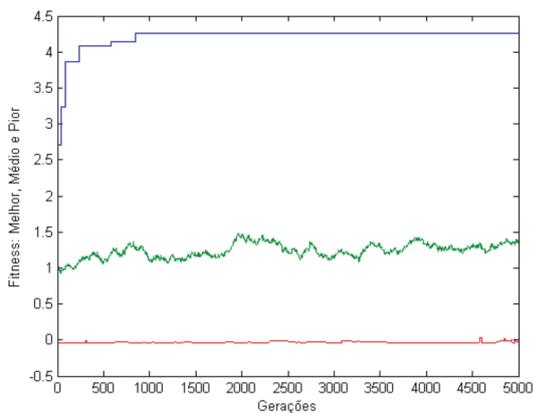


Figura 11: F1- Melhor, média e pior fitness para o AG com alta taxa de mutação em 1 ensaio com 5000 gerações.

Observa-se a tendência a uma convergência prematura logo nas primeiras gerações, quando a infecção viral é potencializada na população cromossômica, conforme a Figura 8. Quando o nível da infecção viral é reduzido, mantém-se o progresso no aumento do *fitness* até que o mesmo se aproxima do melhor *fitness* da melhor configuração do AGRI, conforme a Figura 9.

Notadamente, a manutenção da diversidade genética com o uso do AGRI ficou bastante reduzida para a melhor configuração viral em relação ao AG com alta taxa de mutação e ao AG Clássico, como mostram as Figuras 8, 10 e 11. Isso ocorre devido à infecção viral ter maior influência nas primeiras gerações do AGRI, conduzindo ao ótimo logo no início do processamento, conforme observado na grande quantidade de indivíduos considerados fracos, pior *fitness*, nesse ensaio.

Conclui-se que, quanto mais a população cromossômica evolui, menor é a capacidade de infecção dos vírus, isso acontece porque com o passar das gerações o material genético dos cromossomos fica similar ao material genético da população de vírus, devido ao próprio mecanismo de infecção e criação de vírus do AGRI, que envolve troca de material genético. Assim, pode-se dizer que os cromossomos tornam-se “vacinados”, ou seja, o *fitness* não tem nenhuma mudança após uma tentativa de infecção, denominado código viral inativo, visto que uma infecção só é concretizada quando há um aumento no *fitness* do cromossomo.

No entanto, para a pior configuração viral observou-se um aumento na diversidade populacional o que conduziu o AGRI a um desempenho mais lento.

Para verificar a diversidade média da população nesse ensaio, foi feito o cálculo da diversidade utilizando a MDG (Medida de Diversidade Genética), que consiste na razão entre o *fitness* médio (*Fmed*) e o *fitness* máximo (*Fmax*) da população.

$$MDG = \frac{Fmed}{Fmax} \quad (5)$$

Quanto mais essa medida se aproxima de 1, menor a diversidade da população, e quanto mais próxima de 0, maior é a diversidade cromossômica.

As Figuras (12), (13) comparam a diversidade média da população cromossômica no AGRI, no AG clássico e no AG com alta taxa de mutação para a função F1, considerando a melhor e a pior configuração viral em relação à configuração do AG clássico, da Tabela 2, para um ensaio com 5000 gerações. Nessas Figuras, o AGRI é representado pela linha azul, o AG clássico pela linha verde e o AG alterado para uma alta taxa de mutação pela linha vermelha (o comportamento para as outras funções F2, F3 e F5 é muito semelhante).

Observa-se que, infectar uma grande porcentagem dos cromossomos torna a população muito parecida geneticamente no decorrer das gerações, devido à permanência dos melhores indivíduos e melhores vírus no processo evolutivo. No entanto, uma infecção viral fraca não mantém cromossomos competitivos no AGRI, aumentando o tempo de processamento e reduzindo momentaneamente a *fitness* de toda população. Quanto à alta taxa de mutação no AG clássico, podemos notar que o espaço de busca aumenta bastante e o comportamento do AG para este tipo de configuração pode levar a um mínimo local.

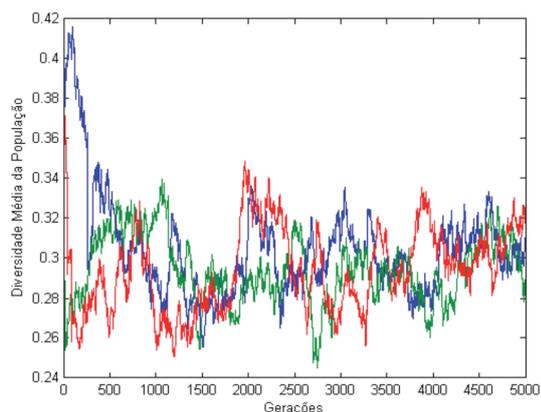


Figura 12: Diversidade populacional comparando a melhor configuração do AGRI, o AG clássico e o AG com alta taxa de mutação para a função F1 para um ensaio com 5000 gerações.

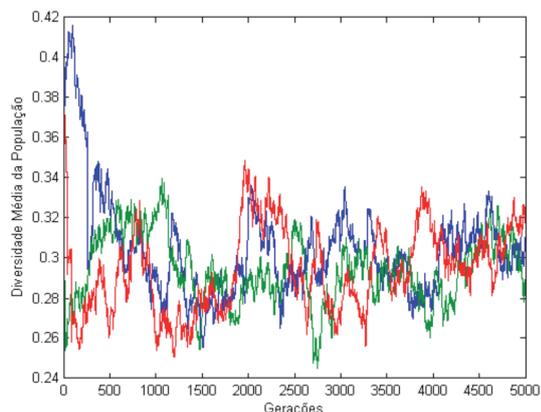


Figura 13: Diversidade populacional comparando a pior configuração do AGRI, o AG clássico e o AG com alta taxa de mutação para a função F1 para um ensaio com 5000 gerações.

O estudo das medidas de diversidade média da população conduz à conclusão de que o AGRI é uma ferramenta de comportamento estocástico reduzido, diferente do AG, com alta taxa de mutação. Essa conclusão é feita com base na análise dos

parâmetros da infecção viral como a taxa de infecção interna, que determina a quantidade máxima de genes que compõem os vírus que serão criados, ou seja, a quantidade de material genético que será trocado com os cromossomos da população, e o parâmetro taxa de infecção populacional, que indica a quantidade de indivíduos aptos a essa infecção.

Com a carga de material genético para troca definido tanto na quantidade quanto nas posições que afetarão o cromossomo, diminui-se o comportamento estocástico característico de um AG com alta taxa de mutação, em que a posição do gene afetado é aleatória.

O comportamento do AGRI foi analisado ainda quanto ao número de gerações necessário para que o ótimo seja encontrado ou chegue-se perto dele. Foram realizados ensaios com uma população inicial fixa de 100 indivíduos e com 10000 gerações. A comparação restringiu-se à configuração de menor desempenho do AGRI (linha azul), o AG clássico (linha verde) e o AG com elevada taxa de mutação (linha vermelha), visto que já foi observado que o AGRI supera, rapidamente, os demais quando configurado com uma infecção viral agressiva.

As Figuras (14), (15), (16) e (17) mostram que a infecção viral considerada com pior desempenho no AGRI precisa de mais gerações para se recuperar, superar o AG clássico em alguns casos, e se aproximar da melhor configuração do AGRI. Isso ocorre devido essa configuração viral fraca manter a diversidade populacional e a infecção viral forte encontrar a melhor solução logo nas gerações iniciais, devido à baixa diversidade populacional reduzindo o tempo de processamento.

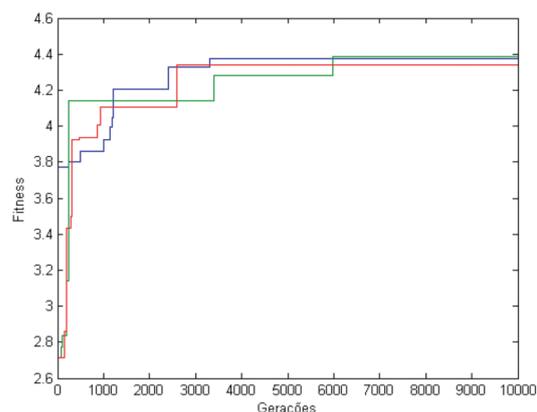


Figura 14: Comparação de desempenho da pior configuração do AGRI, AG clássico e AG com alta taxa de infecção para a função F1 em um ensaio com 10000 gerações.

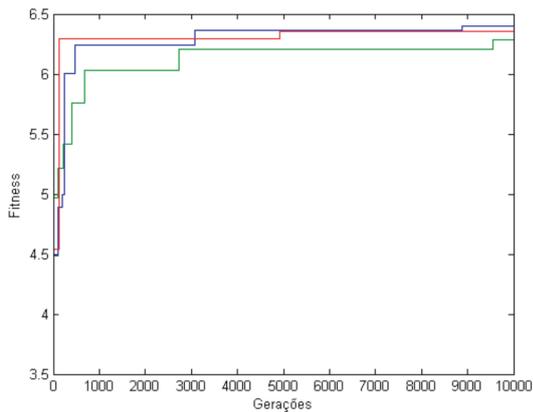


Figura 15: Comparação de desempenho da pior configuração do AGRI, AG clássico e AG com alta taxa de infecção para a função F2 em um ensaio com 10000 gerações.

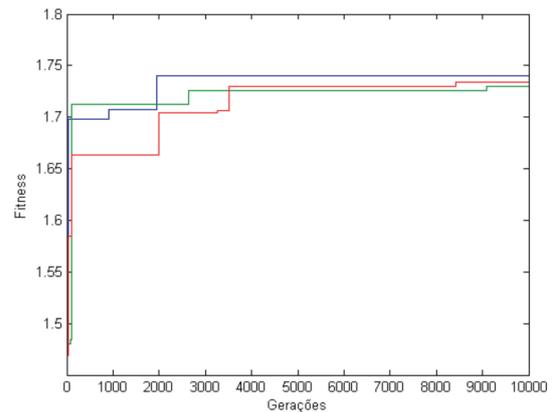


Figura 17: Comparação de desempenho da pior configuração do AGRI, AG clássico e AG com alta taxa de infecção para a função F5 em um ensaio com 10000 gerações.

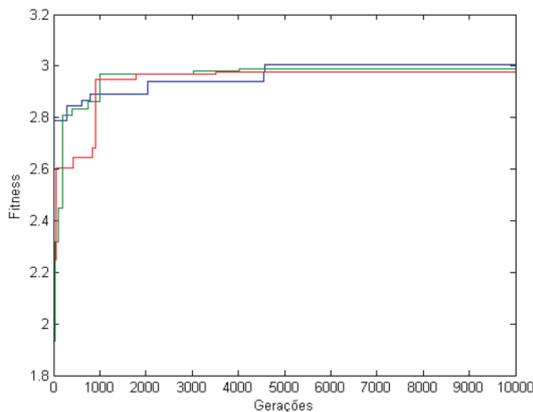


Figura 16: Comparação de desempenho da pior configuração do AGRI, AG clássico e AG com alta taxa de infecção para a função F3 em um ensaio com 10000 gerações.

Para observar o desempenho global do AGRI, foram realizados alguns ensaios específicos. Primeiramente, para uma dada população inicial, realizaram-se 50 ensaios diferentes. Para essa situação, tem-se o *label* 'População Inicial Fixa' na Tabela 3. Depois, realizaram-se experimentos a partir de 50 populações iniciais diferentes. Para cada uma dessas populações iniciais foram realizados 50 ensaios diferentes. Para essa segunda situação, tem-se o *label* 'População Inicial Variável' na mesma Tabela.

Esse comportamento global do AGRI, com a sua configuração denominada de melhor, para as quatro funções *benchmarking* é comparado com o AG Clássico e com o AG com alta taxa de mutação, conforme a Tabela 3, em que os valores das funções considerados ótimos globais estão normalizados para potencia de base decimal quatro (104).

Função	Max (f(x))	Adaptação dos Indivíduos ao Ambiente $f(x)$					
		AG Clássico		AG com alta taxa de mutação		AGRI	
		Melhor <i>Fitness</i>	Média do <i>Fitness</i> Médio	Melhor <i>Fitness</i>	Média do <i>Fitness</i> Médio	Melhor <i>Fitness</i>	Média do <i>Fitness</i> Médio
População Inicial Fixa							
F1	4,4	4,332	1,196	4,329	1,157	4,4	4,061
F2	6,5	6,411	1,498	6,426	1,480	6,484	5,759
F3	3,8	2,997	2,412	2,989	0,724	3,472	2,426
F5	1,8	1,743	0,723	1,735	0,714	1,748	1,601
População Inicial Variável							
F1	4,4	4,368	1,197	4,399	3,672	4,4	3,954
F2	6,5	6,411	1,492	6,442	6,324	6,484	5,844
F3	3,8	2,997	0,729	3,50	2,715	3,042	2,496
F5	1,8	1,80	1,469	1,668	1,594	1,748	1,742

Tabela 3: Comportamento global do AGRI (com melhor configuração), AG Clássico e AG com Alta Taxa de Mutação para F1, F2, F3 e F5, considerando ensaios com uma população inicial fixa e ensaios com populações iniciais diferentes.

Considerando uma infecção viral agressiva (equivalente a configuração denominada de melhor), pode-se afirmar que o AGRI apresenta um melhor desempenho para o problema de otimização com variáveis reais das funções F1, F2, F3 e F5, conforme a Tabela 3. Para a condição de infecção viral fraca (equivalente a configuração denominada de pior) pode-se afirmar que o AGRI também apresenta um melhor desempenho para o problema citado em comparação com o AG clássico e o AG com alta taxa de mutação; apenas se observa que, nesta outra condição, se leva mais tempo de processamento.

CONCLUSÕES

Este trabalho apresentou a avaliação de desempenho do AGRI, considerando a variação dos parâmetros da infecção viral, em comparação com a melhor configuração de um AG clássico e de um AG com elevada taxa de mutação. O resultado mostrou que o AGRI teve um desempenho superior em todos os casos para as funções *benchmarking* F1, F2, F3 e F5 mesmo considerando a pior configuração dos parâmetros da infecção viral. Este resultado mostra que para alcançar a melhor solução, em menos gerações, basta configurar o AGRI com uma infecção forte e que, a configuração da infecção viral de modo mais sutil alcança um resultado similar, com maior número de gerações. Assim, parece razoável nomear as infecções do AGRI apresentadas na Tabela 1 para infecção "agressiva", "moderada" e "fraca".

REFERÊNCIAS

BALTIMORE, D. *VIRAL. RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses*, Nature, vol. 226, pp. 1209-1211, 1970.

BARBULESCU, M.; TURNER, G.; SEAMAN, M. I.; DEINARD, A. S.; KIDD, K. K.; LENZ, J. *Many human endogenous retrovirus k (KERV-K) proviruses are unique to humans*, Curr. Biol., vol. 9, pp. 861-866, 1999.

CARTER, J.; SAUNDERS V. *Virology Principles And Applications*. John Wiley & Sons Ltd., England, 2007.

COFFIN, J. M.; HUGHES, S. H. E. H. *Varmus, Retrovíruses. Plainview (NY): Cold Spring Harbor Laboratory Press, 1997.*

FRAZER, K. A.; CHEN, X.; HINDS, D. A.; PANT, P. V.; PATIL, N.; COX, D. R. *Genomic DNA insertions and deletions occur frequently between humans and nonhumans primates*, Genome Res., vol. 13 (3), pp. 341-346, 2003.

GOLDBARG, M. C.; GOUVÊA, E. F. *Metáfora transgenética: uma nova abordagem para computação evolucionária*. Relatório Técnico COPPE-PEP, 2000.

KANO, H., KONDO, M.; SUGIMOTO, M. *Solving Timetabling Problems Using Genetic Algorithms Based on Minimizing Conflict Heuristics*, Barcelona: CIMNE, 2002.

KANO, H.; HASEGAWA, K.; KATO, N. *Solving Constraint Satisfaction Problems by a Genetic Algorithm Adopting Viral Infection*. Engineering Applications of Artificial Intelligence, Volume 10, Issue 6, Pages 531–537, December 1997.

KANO, H.; NAKAMURA, N. *Route Guidance with Unspecified Staging Posts Using Genetic Algorithm for Car Navigation Systems*. IEEE Intelligent Transportation Systems, 2000.

KARGUPTA, H.; STERFFORD, B. *Gene expression: intro, variable length representation, and all that*, Computational Science Methods Group. Los Alamos: National Laboratory, 1998.

LE ROUZIC, A.; CAPY, P. *The first steps of transposable elements invasion: parasitic strategy vs. genetic drift*, Genetics, vol. 169, pp. 1033-1043, 2005.

LINDEN, R. *Algoritmos Genéticos*, 3. Ed., Ciência Moderna, 2012.

MOREIRA, R. S. ; MONTEIRO, G. D. ; TEIXEIRA, O. N.; OLIVEIRA, R. C. L. *Retroviral Iterative Genetic Algorithm for Real Parameter Function Optimization Problems*. Lecture Notes in Computer Science, v. 6382, p. 220-228, 2011.

ROMANO, C. M. *Caracterização e dinâmica evolutiva de retrovírus endógenos da família K (ERV-K) em genoma de primatas*, Tese de Doutorado, Depart. de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, Dezembro, 2009.

ROMANO, M.; RAMALHO, R. F.; ZANOTTO, P. M. *Time and mode of ERV-K evolution in human and chimpanzee genomes*, Arch Virol., vol. 151, pp. 2215-2228, 2006.

RUDOLPH, G. *Convergence Analysis of Canonical Genetic Algorithms*. IEEE Trans. on Neural Networks, 1994, Vo1.5(1), p96-101.

SILVA, L. M. M.; GOUVÊA, E. F.; FREITAS, J. E. L. *Uma aplicação do algoritmo transgenético extra intracelular para solução do problema da coloração de grafos*, R. FARN, Natal, vol. 1, n. 1, pp. 127-136, jul/dez. 2001.

SONKAR, S. K. *et al. Software Testing using Genetic Algorithm. International Journal of Computer Science & technology (IJCSST) -vol 3 issue 1, ISSN: 0976-8491 (Online), 2012.*

SUGANTHAN, P. *et al. Problem Definitions And Evaluation Criteria For The CEC 2005 Special Session On Real-Parameter Optimization. Technical Report, Nanyang Technological University, Singapore And Kangal Report Number 2005005 (Kanpur Genetic Algorithms Laboratory, IIT Kanpur) (2005).*

SVERDLOV, E. D. *Retroviruses and primate evolution*, BioEssays, vol. 22, pp. 161-171, 1990.

TRAGANTE, V.; TINÓS, R. Diversity Control in Genetic Algorithms for Protein Structure Prediction. In: *VII Encontro Nacional de Inteligência Artificial (ENIA'2009)*, 2009, Bento Gonçalves. Anais do XXIX Congresso da Sociedade Brasileira de Computação (CSBC 2009), 1:727-737, 2009.

VILLARREAL, L. *Virus São Seres Vivos?*. Scientific American Brasil, Edição Especial, n. 28, p. 21-24. São Paulo, 2009.

W-H. LI, GRAUR, D. *Fundamentals of molecular evolution*, 2nd ed., Sunderland: Sinauer Associates, 1991.

ZHU, K.Q. A Diversity-Controlling Adaptive Genetic Algorithm for the Vehicle Routing Problem with Time Windows. In: *Proceedings of the 15th IEEE international Conference on Tools with Artificial intelligence. ICTAI. IEEE Computer Society, Washington, DC, 176, 2003.*