

Proposta de um algoritmo genético para a resolução de sudokus

Thiago Costa Rizuti da Rocha

DCC067 - Computação Evolucionista

Professor: Hélio José Corrêa Barbosa

Universidade Federal de Juiz de Fora

I. INTRODUÇÃO

Sudoku é um tipo popular de quebra-cabeças baseado em lógica e combinações numéricas. Este trabalho objetiva investigar a viabilidade da utilização de um algoritmo genético para encontrar a solução de sudokus. Propõe-se analisar a convergência para diferentes instâncias e verificar se existe alguma correlação entre o número de gerações necessárias e a classificação de dificuldade dada aos sudokus.

O sudoku consiste em um tabuleiro de nove linhas e nove colunas dividido em blocos de três linhas e três colunas. Cada célula pode conter um valor entre 1 e 9. Algumas células têm valores predefinidos que não podem ser alterados. O objetivo do problema é preencher as demais células de forma que não existam valores repetidos em cada linha, coluna e bloco.

8	3		1		6		5	
						8		
			7		9			
5			1	7				
	3				2			
			3	4			1	
	4			8				
9								
3	2			6		4	7	

Figura 1: Exemplo de um sudoku N = 3.

II. ALGORITMO PROPOSTO

Para propor este algoritmo genético foi utilizado como principal referência o trabalho de (Mantere e Koljonen, 2007).

Para que um sudoku esteja resolvido as seguintes restrições devem ser atendidas:

- 1) Não deve existir repetição de valores nas linhas.
- 2) Não deve existir repetição de valores nas colunas.
- 3) Não deve existir repetição de valores nos blocos.
- 4) Os valores predefinidos não podem ser alterados.

No algoritmo genético proposto, o cromossomo é uma possível solução do sudoku representado como um vetor de inteiros. Cada posição do vetor representa uma célula do tabuleiro percorrido linha por linha. Um vetor auxiliar é utilizado para indicar se uma posição é um valor predefinido, ou seja, se ele pode ser alterado ou não. Uma posição no vetor do cromossomo pode ser alterada se a mesma posição no vetor auxiliar contém o valor 0 e não pode ser alterada se a mesma posição no vetor auxiliar contém o valor predefinido. Desta forma, a restrição (4) é satisfeita.

A população inicial é criada de forma aleatória, porém garantindo que os valores sejam únicos em cada uma das linhas, satisfazendo assim a restrição (1).

830100408	000000080	000700900	050017000	003000200	000340010	004008000	090000000	302006047
837194028	549623781	621785934	286817493	413569278	978342516	164278359	795431862	302956147

Figura 2: Vetor auxiliar e vetor de solução.

A função de aptidão é calculada de forma que meça o quão correto está uma solução. É atribuída uma penalização de 1 a cada repetição de valores em cada coluna e em cada bloco. Sendo 0 a aptidão da solução correta e o objetivo minimizar este valor.

A seleção de indivíduos para o cruzamento foi feita através da seleção de roleta (aleatório proporcional a aptidão). Como a função de aptidão é de minimização, a chance de um indivíduo ser selecionado foi calculada com a equação (1).

$$Chance_i = \frac{1}{Fitness_i} \quad (1)$$

Para o cruzamento é utilizado a recombinação uniforme. Os pontos de recombinação são cada uma das linhas do tabuleiro. Sejam dois cromossomos, as linhas de cada um são intercambiadas com uma dada probabilidade, dando origem a dois novos indivíduos. Apenas as mesmas linhas podem ser trocadas entre si. Este operador mantém as restrições (1) e (4) satisfeitas.

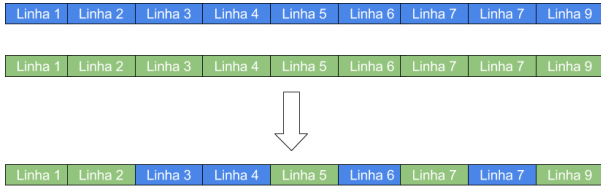


Figura 3: Recombinação uniforme.

O operador de mutação realiza troca de valores dentro de uma mesma linha desde que não sejam valores predefinidos, mantendo satisfeitas as restrições (1) e (4). Uma dentre as N linhas é escolhida e em seguida dois valores dessa linha são selecionados para terem seus valores trocados. Ambas as escolhas são feitas de forma aleatória. Esta escolha e troca de valores é feita cinco vezes dada a probabilidade de acontecer mutação.

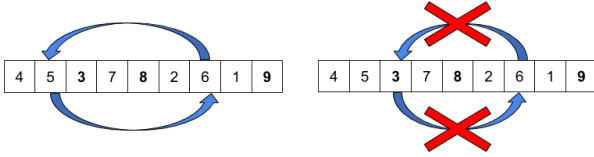


Figura 4: Exemplos de mutação válida e não válida. Os valores em negrito são os predefinidos e não podem ser alterados.

Para a substituição dos novos indivíduos na população foi aplicada uma política de elitismo, mantendo fixo o tamanho da população. Seja uma população de M indivíduos, são gerados M novos indivíduos através da recombinação. Estes novos indivíduos podem sofrer mutação e são ordenados de acordo com a aptidão. 10% dos antigos indivíduos com melhor valor de aptidão são escolhidos para sobreviver até a próxima geração. Os outros 90% são aqueles com melhor valor de aptidão dentre os novos indivíduos.

III. EXPERIMENTOS

Para a realização dos testes foi utilizada uma população de 500 indivíduos, taxa de recombinação de 50% e taxa de mutação de 30%. Os critérios de parada foram a obtenção da solução, 500 gerações realizadas ou 100 gerações sem aumento do melhor valor de aptidão.

Para cada uma das instâncias de teste foram feitas 10 execuções com diferentes sementes de randomização.

Foram utilizados quinze sudokus do jornal finlandês Helsingin Sanomat, sendo três de cada um dos cinco níveis de dificuldade em que são classificados. Também foram realizados testes com o sudoku aberto, ou seja, sem valores predefinidos e com sudokus de tamanho 4, 5 e 6.

As figuras (6) e (7) mostram os valores de aptidão para cada uma das instâncias testadas. Somente em duas instâncias da classe de dificuldade mais baixa e no sudoku aberto foi possível chegar à solução ótima.

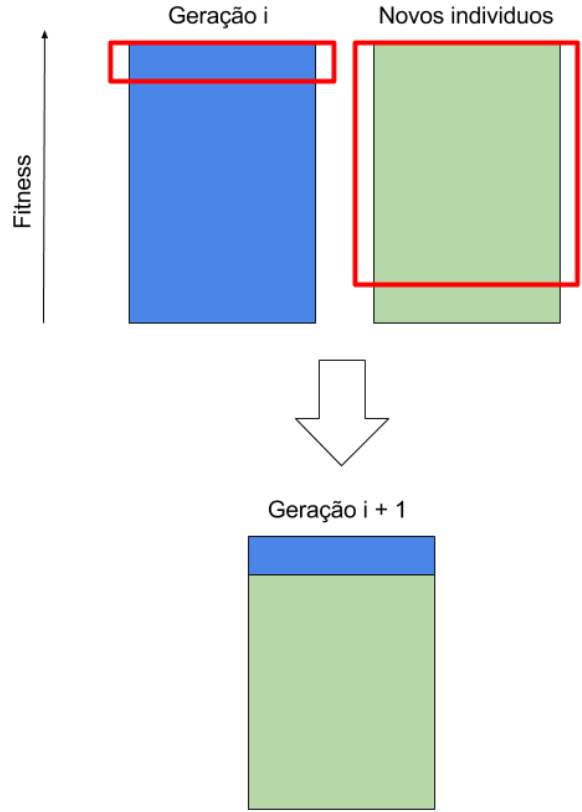


Figura 5: Substituição dos novos indivíduos na população.

As figuras (8) e (9) mostram, respectivamente, a análise de convergência para a média de cada uma das classes de dificuldade e para as instâncias de dimensões maiores.

Na figura (8) pode-se observar claramente uma maior facilidade para resolver os problemas de dificuldade 1. As relações entre as outras dificuldades não fica clara.

A figura (9), como esperado, mostra uma convergência menor a medida que a dimensão da instância cresce.

Instância	Dificuldade	Min	Max	Média	Desvio Padrão
s01a	1	0	6	4,2	1,66
s01b	1	0	2	0,8	0,98
s01c	1	4	8	6,7	1,49
s02a	2	2	6	4,2	1,66
s02b	2	4	8	5,7	1,42
s02c	2	2	7	3,8	1,60
s03a	3	2	7	5,8	1,33
s03b	3	2	7	4,6	1,62
s03c	3	2	6	4,2	1,25
s04a	4	3	8	4,7	1,49
s04b	4	2	6	3,9	1,14
s04c	4	2	8	4,4	2,01
s05a	5	2	7	3,8	1,60
s05b	5	2	10	6,4	1,74
s05c	5	4	7	4,8	1,08
Open	-	0	6	3,0	1,84

Figura 6: Melhor valor de aptidão obtido.

Instância	Tamanho	Min	Max	Média	Desvio Padrão
size4	4	24	39	31,00	4,43
size5	5	84	106	91,55	6,29
size6	6	213	235	22,21	8,21

Figura 7: Melhor valor de aptidão obtido para instâncias maiores.

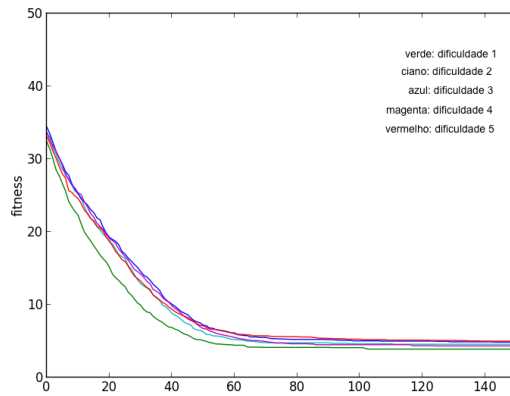


Figura 8: Análise de convergência.

IV. CONCLUSÕES

Através da observação dos resultados, é possível concluir que o algoritmo genético não obteve um excelente desempenho. A convergência do valor de aptidão foi rápida durante as gerações. Porém esta ocorre com frequência para um mínimo local, onde o algoritmo estaciona e nenhuma melhor solução é gerada. Sobre este comportamento levanta-se duas hipóteses.

A primeira hipótese é de que a convergência prematura deve-se a diminuição de variabilidade genética no decorrer das gerações. Esta poderia ser resolvida através de técnicas de aumento de variabilidade, como a inserção de indivíduos aleatórios na população a cada geração e a restrição de inserir apenas indivíduos distintos.

A segunda hipótese é da falta de lógica específica do problema. A função de aptidão não consegue guiar a busca a partir de um certo ponto devido a sua simplicidade. Esta poderia ser resolvida com uma função de aptidão mais especializada, ou com uma abordagem multi objetiva (sendo as repetições em linhas e colunas tratadas em funções distintas). Outra possibilidade é hibridizar o algoritmo com um algum tipo de busca que ajude o algoritmo genético a caminhar até a ótima.

REFERÊNCIAS

- [1] Mantere, T., Koljonen, J. (2007). Solving, rating and generating sudoku puzzles with GA. 2007 IEEE Congress on Evolutionary Computation, CEC 2007, 1382–1389.

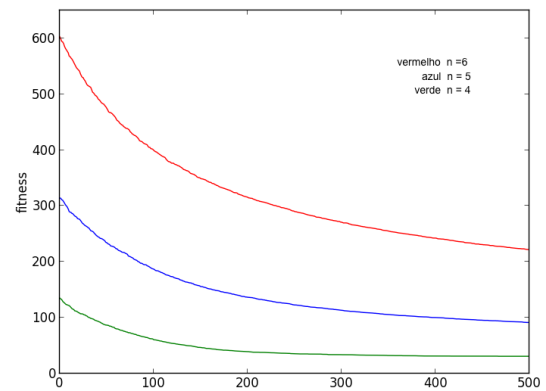


Figura 9: Análise de convergência para instâncias maiores.