

- I.1 EVOLUÇÃO COMO INSPIRAÇÃO
- I.2 TERMINOLOGIA BIOLÓGICA
- I.3 ESPAÇOS DE BUSCA E CENÁRIOS DE FITNESS
- I.4 ELEMENTOS DE ALGORITMOS GENÉTICOS
- I.5 UM ALGORITMO GENÉTICO BÁSICO
- I.6 ALGORITMOS GENÉTICOS E MÉTODOS DE BUSCA TRADICIONAIS
- I.7 APLICAÇÕES DE ALGORITMOS GENÉTICOS

I.1 EVOLUÇÃO COMO INSPIRAÇÃO

(PARA RESOLVER PROBLEMAS COMPUTACIONAIS)

PROBLEMAS COMPUTACIONAIS
REQUEREM:

BUSCA AO LONGO DE UM nº MUITO
GRANDE DE POSSÍVEIS SOLUÇÕES

A EVOLUÇÃO PERMITE:

- Nº POSSIBILIDADES:
CONJUNTO DE SEQUÊNCIAS GENÉTICAS
- SOLUÇÕES DESEJADAS:
ORGANISMOS ALTAMENTE AJUSTADOS
(ADEQUADOS), PERFEITAMENTE
CAPAZES A SOBREVIVER E SE
REPRODUZIR EM SEUS AMBIENTES.

CAPACIDADE DE INNOVAR

- A EVOLUÇÃO PERMITE PROJETER
NOVAS SOLUÇÕES P/ PROBLEMAS
COMPLEXOS.
(SISTEMA IMUNOLÓGICO DOS MAMÍFEROS
FACE A CONSTANTE
INVASÃO DO CORPO POR GERMES.)

ADAPTATIVIDADE (CONTINUAR A
DESEMPENHAR ADEQUADAMENTE
EM AMBIENTES NÃO-ESTACIONÁRIOS)

- A ADEQUAÇÃO DE UM ORGANISMO
BIOLÓGICO DEPENDE DE VÁRIOS
FATORES: COMO RESISTE AS
CARACTERÍSTICAS FÍSICAS DO
AMBIENTE, COMO COMPETE OU
COOPERA COM OUTROS ORGANISMOS.
- O CRITÉRIO DE ADEQUAÇÃO
MUDA CONTINUAMENTE, À MEDIDA
QUE A CRIAÇÃO EVOLUI.

SOLUÇÕES COMPLEXAS, DIFÍCEIS DE
PROGRAMAR: INTELIGÊNCIA
ARTIFICIAL → REGRAS MUITO
SIMPLES PERMITEM A EMERGÊNCIA
DE COMPORTAMENTOS COMPLEXOS
(COMO INTELIGÊNCIA) ATRAVÉS
DA APLICAÇÃO FORTEMENTE
PARALELA DESTAS REGRAS E
DA INTERAÇÃO ENTRE ELAS.

- A EVOLUÇÃO NÃO SE OCUPA DE
UMA ESPÉCIE A CADA VEZ, DE
FORMA FORTEMENTE PARALELA
TESTA E MUDA MILHÕES DE
ESPÉCIES EM PARALELO.

REGRAS DE EVOLUÇÃO :

AS ESPÉCIES EVOLUEM POR MEIO DE VARIAÇÕES ALEATÓRIAS (VIA MUTAÇÃO, RECOMBINAÇÃO E OUTROS OPERADORES), SEGUIDAS POR SELEÇÃO NATURAL.

POR MEIO DA SELEÇÃO NATURAL AS ESPÉCIES MAIS "AJUSTADAS" TÊM A SOBREVIVER E SE REPRODUZIR, PROPAGANDO, DESTA FORMA SEU MATERIAL GENÉTICO PARA FUTURAS GERAÇÕES.

CROMOSSOMOS

TODOS OS ORGANISMOS VIVOS CONSISTEM DE CÉLULAS, E CADA CÉLULA CONTÉM UM CONJUNTO DE UM OU MAIS CROMOSSOMOS, QUE SÃO CADEIAS DE DNA.

GENES

UM CROMOSSOMO PODE SER CONCEITUALMENTE DIVIDIDO EM GENES, QUE SÃO OS BLOCOS FUNCIONAIS DO DNA, CADA UM DELES CODIFICANDO UMA PARTICULAR PROTEÍNA. AS DIFERENTES POSSIBILIDADES QUE UM GENE PODE ASSUMIR SÃO CHAMADAS ALELOS. CADA GENE ESTÁ LOCALIZADO EM UMA PARTICULAR POSIÇÃO, NO CROMOSSOMO.

GENOMA

MUITOS ORGANISMOS TÊM MÚLTIPLOS CROMOSSOMOS EM CADA CÉLULA. A COLEÇÃO COMPLETA DO MATERIAL GENÉTICO (TODOS OS CROMOSSOMOS) É CHAMADA DE GENOMA DO ORGANISMO.

GENÓTIPO

É O CONJUNTO DE GENES CONTIDOS EM UM GENOMA. DOIS INDIVÍDUOS QUE POSSUEM GENOMAS IDÊNTICOS POSSUEM O MESMO GENÓTIPO.

FENÓTIPO

CARACTERÍSTICAS FÍSICAS E MENTAIS DE UM INDIVÍDUO, TAIS COMO: COR DOS OLHOS, ALTURA, ... DECORRENTES DO GENÓTIPO.

DIPLÓIDES E HAPLÓIDES

ORGANISMOS CUJOS CROMOSSOMOS SÃO ALINHADOS AOS PARES SÃO CHAMADOS DIPLÓIDES.

ORGANISMOS CUJOS CROMOSSOMOS NÃO SÃO ALINHADOS AOS PARES SÃO DITOS HAPLÓIDES.

NA NATUREZA, A MAIOR PARTE DAS ESPÉCIES, INCLUINDO SERES HUMANOS, SÃO DIPLÓIDES (SERES HUMANOS POSSUEM 23 PARES DE CROMOSSOMOS EM CADA CÉLULA).

CRUZAMENTO

DURANTE A REPRODUÇÃO OCORRE A RECOMBINAÇÃO (OU CRUZAMENTO), OPERAÇÃO NA QUAL, EM CADA INDIVÍDUO PAI, GENES SÃO TROCADOS ENTRE CADA PAR DE CROMOSSOMOS PARA FORMAR UM GAMETA (UM ÚNICO CROMOSSOMO). ENTÃO, OS GAMETAS PROVENIENTES DOS DOIS PAIS SÃO USADOS PARA CRIAR UM CONJUNTO COMPLETO DE CROMOSSOMOS DIPLÓIDES. NA REPRODUÇÃO HALÓIDE, GENES SÃO TROCADOS ENTRE OS CROMOSSOMOS (NÃO PARES) DOS DOIS INDIVÍDUOS PAIS.

MUTAÇÃO

NOVAS GERAÇÕES SÃO SUJEITAS À MUTAÇÃO, OPERAÇÃO NA QUAL AS UNIDADES ELEMENTARES DO DNA (NUCLEOTÍDEOS) SÃO TROCADAS NOS PAIS, PARA GERAR OS CROMOSSOMOS FILHOS.

FITNESS

É A PROBABILIDADE DE QUE O ORGANISMO VIVA E POSSA SE REPRODUZIR (VIABILIDADE). TAMBÉM PODE SER UMA FUNÇÃO DO NÚMERO DE INDIVÍDUOS FILHOS QUE O ORGANISMO POSSUI (FERTILIDADE).

ANALOGIA ENTRE A TERMINOLOGIA UTILIZADA EM ALGORITMOS GENÉTICOS E A TERMINOLOGIA BIOLÓGICA

MCFC
GA
I4

CROMOSSOMOS

SÃO SOLUÇÕES CANDIDATAS AO PROBLEMA EM QUESTÃO, MUITAS VEZES CODIFICADAS EM CADEIAS DE BITS.

GENES

SÃO BITS OU CURTOS BLOCOS DE BITS ADJACENTES QUE CODIFICAM UM PARTICULAR ELEMENTO DE UMA SOLUÇÃO CANDIDATA.

ALELOS

UM ALELO EM UM GENE PODE SER ϕ OU 1. EM CASOS DE BLOCOS DE MAIS DO QUE 1 BIT, MUITOS ALELOS SÃO POSSÍVEIS EM UM GENE.

CRUZAMENTOS

CONSISTEM EM TROCAR MATERIAL GENÉTICO ENTRE DOIS INDIVÍDUOS PAIS HAPLÓIDES (DE UM ÚNICO CROMOSSOMO).

MUTAÇÃO

CONSISTE EM VIRAR O BIT EM UMA POSIÇÃO ALEATORIAMENTE ESCOLHIDA (OU, PARA BLOCOS DE MAIS DO QUE 1 BIT, SUBSTITUIR O SÍMBOLO PARA UMA POSIÇÃO ALEATORIAMENTE ESCOLHIDA POR UM NOVO SÍMBOLO ALEATORIAMENTE ESCOLHIDO).

GENÓTIPO

AS APLICAÇÕES DE GAS EMPREGAM, NA SUA MAIORIA, INDIVÍDUOS HAPLÓIDES, DE UM ÚNICO CROMOSSOMO.

O GENÓTIPO DE UM INDIVÍDUO EM UM GA QUE UTILIZA CADEIAS DE BITS É SIMPLEMENTE A CONFIGURAÇÃO DE BITS NO CROMOSSOMO DO INDIVÍDUO.

FENÓTIPO

GERALMENTE NÃO HÁ NOÇÃO DE FENÓTIPO NO CONTEXTO DE GAS.

(NO ENTANTO, RECENTEMENTE, TÊM SIDO CONSIDERADOS GAS NOS NÍVEIS DE GENÓTIPO E FENÓTIPO)

CODIFICAÇÃO DA
CADEIA DE BITS
DE UMA RNA

↳ A PRÓPRIA
RNA

I.3 ESPAÇOS DE BUSCA E CENÁRIOS DE FITNESS

MCFC
GA
I.5

ESPAÇOS DE BUSCA

O TERMO "SEARCH SPACE" SE REFERE A ALGUMA COLEÇÃO DE SOLUÇÕES CANDIDATAS PARA O PROBLEMA ALVO E A UMA NOÇÃO DE DISTÂNCIA ENTRE SOLUÇÕES CANDIDATAS.

DESEJAMOS REALIZAR UMA BUSCA COMPUTACIONAL POR UMA PROTEÍNA - UMA SEQUÊNCIA DE AMINO ÁCIDOS - QUE ASSUME UMA FORMA TRIDIMENSIONAL, DE TAL FORMA QUE POSSA SER USADA PARA COMBATER UM VIRUS ESPECÍFICO.

ESPAÇO DE BUSCA: É A COLEÇÃO DE TODAS AS POSSÍVEIS SEQUÊNCIAS DE PROTEÍNAS - UM CONJUNTO INFINITO DE POSSIBILIDADES.

COMO RESTRIÇÃO, O ESPAÇO DE BUSCA SERÁ REDUZIDO A TODAS AS POSSÍVEIS SEQUÊNCIAS DE TAMANHO ≤ 100 , O QUE AINDA CONSTITUI UM IMENSO ESPAÇO DE BUSCA, JÁ QUE HÁ 20 POSSÍVEIS AMINO ÁCIDOS PARA CADA POSIÇÃO NA SEQUÊNCIA.

SE OS 20 AMINO ÁCIDOS FOREM REPRESENTADOS POR LETRAS DO ALFABETO, AS SOLUÇÕES CANDIDATAS SERÃO DO TIPO:

A G G M C G B L ...

DISTÂNCIA ENTRE DUAS SEQUÊNCIAS:

MCFC
GA
I.6

É O N.º DE POSIÇÕES NAS QUAIS AS LETRAS EM POSIÇÕES CORRESPONDENTES DIFEREM.

A G G M C G B L
M G G M C G B L
↑

DISTÂNCIA = 1

A G G M C G B L
L B M P A F G A
↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑

DISTÂNCIA = 8

UM ALGORITMO PARA BUSCA NESTE ESPAÇO CONSISTE EM ESCOLHER QUAIS SOLUÇÕES CANDIDATAS SERÃO TESTADAS, A CADA ESTÁGIO DA BUSCA.

NA MAIOR PARTE DOS CASOS A PRÓXIMA SOLUÇÃO CANDIDATA A SER TESTADA DEPENDERÁ DOS RESULTADOS DE PRÉVIAS SEQUÊNCIAS TESTADAS.

A MAIOR PARTE DOS ALGORITMOS ASSUME QUE HAVERÁ ALGUMA CORRELAÇÃO ENTRE A QUALIDADE DAS SOLUÇÕES CANDIDATAS VIZINHAS - AQUELAS PRÓXIMAS NO ESPAÇO DE BUSCA.

AS ASSUMEM QUE INDIVÍDUOS PAIS DE ALTA QUALIDADE (SOLUÇÕES CANDIDATAS DE ALTA QUALIDADE) LOCALIZADAS EM DIFERENTES REGIÕES NO ESPAÇO DE BUSCA PODEM SER COMBINADAS POR MEIO DE CRUZAMENTO PARA PRODUIR SOLUÇÕES CANDIDATAS (INDIVÍDUOS FILHOS) DE ALTA QUALIDADE.

CENÁRIOS DE FITNESS

→ DEFINIDO (EM 1931) PELO BIÓLOGO S. WRIGHT NO CONTEXTO DE GENÉTICA POPULACIONAL.

É UMA REPRESENTAÇÃO DO ESPAÇO DE TODOS OS POSSÍVEIS GENÓTIPOS, ASSOCIADOS A SEUS PARTICULARES FITNESS.

SUPONDO QUE:

→ CADA GENÓTIPO É UMA CADEIA DE BITS DE TAMANHO l .

→ A DISTÂNCIA ENTRE 2 GENÓTIPOS É MEDIDA PELA DISTÂNCIA DE HAMMING (QUE EQUIVALE AO N.º DE POSIÇÕES NAS QUAIS BITS CORRESPONDENTES DIFEREM).

→ A CADA GENÓTIPO É ATRIBUÍDO UM VALOR DE FITNESS.

UM CENÁRIO DE FITNESS PODE SER DESCRITO COMO UM GRÁFICO $(l+1)$ DIMENSIONAL NO QUAL CADA GENÓTIPO É UM PONTO EM l DIMENSÕES E SEU FITNESS É PLOTADO AO LONGO DO EIXO DE ORDEM $(l+1)$.

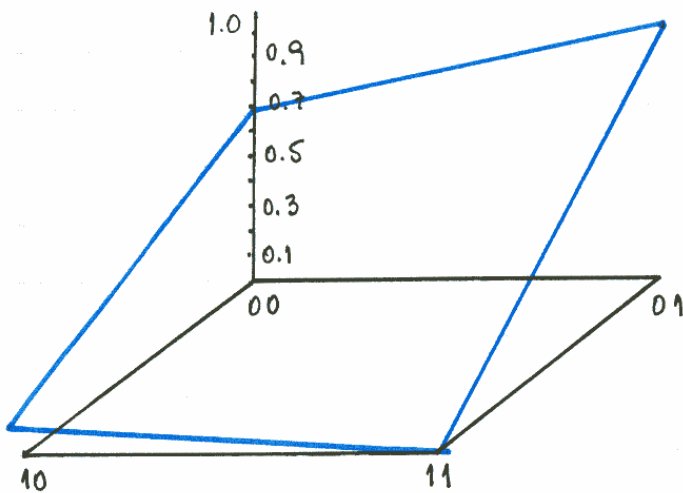


FIGURA I.1: UM SIMPLES CENÁRIO (OU PAISAGEM) DE FITNESS f / $l=2$.
 $f(00)=0.7$; $f(01)=1.0$; $f(10)=0.1$; $f(11)=0.0$.

DE ACORDO COM A FORMULAÇÃO DE WRIGHT, A EVOLUÇÃO FAZ COM QUE AS POPULAÇÕES SE MOVAM AO LONGO DOS CENÁRIOS POR CAMINHOS PARTICULARES E

→ ADAPTAÇÃO PODE SER VISTA COMO

UM MOVIMENTO EM DIREÇÃO A PICOS LOCAIS.

* UM PICO LOCAL (OU ÓTIMO LOCAL) NÃO É NECESSARIAMENTE O PONTO MAIS ALTO NO CENÁRIO, MAS QUALQUER PEQUENO MOVIMENTO NO SENTIDO DE NOS AFASTARMOS DO PICO LOCAL CONDUZIRÁ A UMA DIMINUIÇÃO EM FITNESS.

→ OPERAÇÕES DE CRUZAMENTO E MUTAÇÃO PODEM SER VISTAS COMO FORMAS DE MOVER UMA POPULAÇÃO EM UM CENÁRIO DEFINIDO PELA FUNÇÃO DE FITNESS.

CENÁRIOS FIXOS: NÃO PODEMOS ATRIBUIR UM VALOR DE FITNESS A UM ORGANISMO DE FORMA INDEPENDENTE DE OUTROS ORGANISMOS EM SEU AMBIENTE. ASSIM, À MEDIDA QUE A POPULAÇÃO MUDA, O FITNESS DE PARTICULARES GENÓTIPOS TAMBÉM MUDARÁ.

A MAIOR PARTE DOS MÉTODOS CHAMADOS GAS POSSUEM, NO MÍNIMO, OS SEGUINTE ELEMENTOS Em comum:

- POPULAÇÕES DE CROMOSSOMOS
- SELEÇÃO DE ACORDO COM O FITNESS
- CRUZAMENTOS PARA PRODUZIR UMA NOVA GERAÇÃO (PROLE)
- MUTAÇÃO ALEATÓRIA DA NOVA PROLE

- OS CROMOSSOMOS EM UMA POPULAÇÃO TÍPICAMENTE CONSTITUEM CADEIAS DE BITS.
- CADA POSIÇÃO NO CROMOSSOMO TEM 2 POSSÍVEIS ALELOS: 0 E 1.
- CADA CROMOSSOMO PODE SER VISTO COMO UM PONTO NO ESPAÇO DE BUSCA DE SOLUÇÕES CANDIDATAS.
- GAS PROCESSAM POPULAÇÕES DE CROMOSSOMOS, SUBSTITUINDO SUCESSIVAMENTE UMA POPULAÇÃO POR OUTRA.
- GAS REQUEREM UMA FUNÇÃO DE FITNESS QUE ATRIBUA UM VALOR (FITNESS) A CADA CROMOSSOMO NA POPULAÇÃO EM QUESTÃO.
- O FITNESS DE UM CROMOSSOMO DEPENDE DE O QUÃO BEM AQUELE CROMOSSOMO RESOLVE O PROBLEMA EM QUESTÃO.

EXEMPLOS DE FUNÇÕES DE FITNESS E DOS OPERADORES SELEÇÃO, CRUZAMENTO E MUTAÇÃO:

FUNÇÕES DE FITNESS

UMA APLICAÇÃO COMUM DE GAS É EM OTIMIZAÇÃO DE FUNÇÕES, EM QUE O OBJETIVO É ENCONTRAR UM CONJUNTO DE VALORES QUE MAXIMIZE, POR EXEMPLO, UMA FUNÇÃO MULTIPARAMÉTRICA COMPLEXA. PARA O CASO EM QUE DESEJAMOS MAXIMIZAR A FUNÇÃO DE VALORES REAIS

$$f(y) = y + |\sin(32y)|, \quad 0 \leq y < \pi$$

NESTE CASO, AS SOLUÇÕES CANDIDATAS SÃO VALORES PARA y QUE POSSAM SER CODIFICADOS COMO CADEIAS DE BITS REPRESENTANDO NÓS REAIS.

O CÁLCULO DO FITNESS "TRADUZ" UMA DADA CADEIA DE BITS x EM UM N.º REAL y E ENTÃO SUBSTITUI ESTE VALOR PARA y EM $f(y)$.

O VALOR DO FITNESS DE UMA CADEIA DE BITS É O VALOR DA FUNÇÃO NESTE PONTO

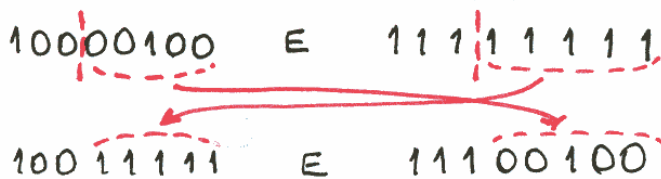
OPERADOR SELEÇÃO

SELECIONA CROMOSSOMOS NA POPULAÇÃO P/ REPRODUÇÃO.

QUANTO MAIOR O FITNESS DO CROMOSSOMO, MAIS VEZES ELE SERÁ SELECIONADO PARA REPRODUÇÃO.

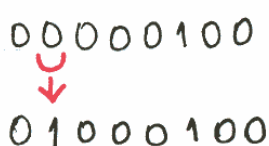
OPERADOR CRUZAMENTO

ESCOLHE ALEATORIAMENTE UMA POSIÇÃO E TROCA OS BITS NAS POSIÇÕES ANTERIORES E POSTERIORES ÀQUELA ENTRE 2 CROMOSSOMOS PARA CRIAR 2 INDIVÍDUOS FILHOS.

10000100 E 11111111

 10011111 E 11100100

OPERADOR MUTAÇÃO

"VIRA" ALEATORIAMENTE ALGUNS DOS BITS EM UM CROMOSSOMO.

00000100

 01000100

MUTAÇÃO PODE OCORRER EM CADA POSIÇÃO EM UMA CADEIA DE BITS, COM ALGUMA PROBABILIDADE, USUALMENTE MUITO PEQUENA (DA ORDEM DE 0.001).

SEMDO DADOS:

- * UM PROBLEMA CLARAMENTE DEFINIDO P/ SER RESOLVIDO E
- * CADEIAS DE BITS REPRESENTANDO SOLUÇÕES CANDIDATAS,
UM ALGORITMO GENÉTICO BÁSICO IRÁ OPERAR ASSIM:

1. INICIE COM UMA POPULAÇÃO GERADA ALEATORIAMENTE DE n CROMOSSOMOS DE l BITS (SOLUÇÕES CANDIDATAS).
2. CALCULE O FITNESS $f(x)$ DE CADA CROMOSSOMO x PRESENTE NA POPULAÇÃO.
3. REPITA OS SEGUINTESS PASSOS ATÉ QUE n INDIVÍDUOS TERHAM SIDO CRIADOS:
 - a. SELECIONE UM PAR DE CROMOSSOMOS PERTENCENTE À POPULAÇÃO ATUAL. A PROBABILIDADE DE SELEÇÃO DEVE SER UMA FUNÇÃO CRESCENTE DO FITNESS. NO PROCESSO DE SELEÇÃO O MESMO CROMOSSOMO PODE SER SELECIONADO PARA SER PAI MAIS DO QUE UMA VEZ.
 - b. COM PROBABILIDADE p_c (PROBABILIDADE DE CRUZAMENTO OU TAXA DE CRUZAMENTO) EXECUTE O CRUZAMENTO A PARTIR DE UM PONTO ALEATORIAMENTE ESCOLHIDO (ESCOLHIDO COM PROBABILIDADE UNIFORME) PARA FORMAR 2 INDIVÍDUOS FILHOS. SE NÃO OCORRER CRUZAMENTO OS 2 INDIVÍDUOS GERADOS SERÃO CÓPIAS EXATAS DOS PAIS.
(CRUZAMENTO EM UM ÚNICO PONTO X CRUZAMENTO MULTI-PONTO)
 - c. EXECUTE MUTAÇÃO NOS 2 INDIVÍDUOS FILHOS COM PROBABILIDADE p_m (PROBABILIDADE DE MUTAÇÃO OU TAXA DE MUTAÇÃO) E COLOQUE OS CROMOSSOMOS RESULTANTES NA NOVA POPULAÇÃO.
SE n FOR ÍMPAR, UM NOVO MEMBRO SERÁ DESCARTADO, ALEATORIAMENTE.
4. SUBSTITUA A POPULAÇÃO ATUAL PELA NOVA POPULAÇÃO.
5. VOLTE AO PASSO 2.
CADA ITERAÇÃO É CHAMADA UMA GERAÇÃO.

- ⇒ CADA ITERAÇÃO DO PROCESSO DESCRITO CONSTITUI UMA GERAÇÃO.
- ⇒ UM GA É TÍPICAMENTE ITERADO DE 50 A 500 GERAÇÕES, PODENDO O N.º DE ITERAÇÕES SER MAIOR.
- ⇒ O CONJUNTO DE GERAÇÕES É CHAMADO "RODADA".
- ⇒ AO FINAL DE UMA RODADA HA' FREQUENTEMENTE UM OU MAIS CROMOSSOMOS ALTAMENTE ADEQUADOS NO SENTIDO DO FITNESS, PERTENCENTES À POPULAÇÃO.
- ⇒ DUAS RODADAS QUE UTILIZEM SEMENTES DIFERENTES NA HEURÍSTICA RESPONSÁVEL POR GERAR ALEATORIEDADE IRÃO PRODUIR DIFERENTES COMPORTAMENTOS EVOLUTIVOS.
- ⇒ COSTUMA-SE REGISTRAR E REPORTAR ESTATÍSTICAS (TAIS COMO: MELHOR FITNESS EM UMA RODADA E EM QUAL GERAÇÃO O INDIVÍDUO COM MELHOR FITNESS FOI GERADO) EXTRAÍDAS APÓS MUITAS DIFERENTES RODADAS DO GA SOBRE O MESMO PROBLEMA.
- ⇒ O ALGORITMO DESCRITO, EMBORA SIMPLES, É A BASE PARA A MAIORIA DAS APLICAÇÕES DE GAs.
- ⇒ HA' UM N.º DE DETALHES A SEREM DISCUTIDOS, TAIS COMO:
 - TAMANHO DAS POPULAÇÕES
 - PROBABILIDADES DE CRUZAMENTO
 - PROBABILIDADES DE MUTAÇÃO
 DOS QUAIS O SUCESSO DO ALGORITMO DEPENDE GRANDEMENTE.
- ⇒ HA' VERSÕES MAIS COMPLICADAS DE GAS QUE, POR EXEMPLO, TRABALHEM SOBRE REPRESENTAÇÕES QUE NÃO SEJAM CADEIAS DE BITS, OU UTILIZEM DIFERENTES TIPOS DE OPERAÇÕES DE CRUZAMENTO E MUTAÇÃO (DEFININDO DIFERENTES OPERADORES PARA TAL).

EXEMPLO:

TAMANHO DA CADEIA DE BITS DO CROMOSSOMO $l=8$
 FUNÇÃO DE FITNESS $f(x) = \text{N.º DE BITS "1" NA CADEIA DE BITS } x$
 TAMANHO DA POPULAÇÃO $n=4$
 PROBABILIDADE DE CRUZAMENTO $p_c = 0.7$
 PROBABILIDADE DE MUTAÇÃO $p_m = 0.001$

- * $f(x)$, l E n FORAM ESCOLHIDOS POR SIMPLICIDADE.
- * VALORES TÍPICOS PARA l E n ESTÃO ENTRE 50 E 1000.
- * VALORES ATRIBUÍDOS PARA p_c E p_m SÃO TÍPICOS.

A TABELA I.1 APRESENTA UMA POSSÍVEL POPULAÇÃO INICIAL, GERADA ALEATORIAMENTE.

IDENTIFICAÇÃO DO CROMOSSOMO	CADEIA DE BITS DO CROMOSSOMO	FITNESS DO CROMOSSOMO
A	00000110	2
B	11101110	6
C	00100000	1
D	00110100	3

TABELA I.1 : POPULAÇÃO INICIAL DO GA EXEMPLO

MÉTODO UTILIZADO PARA SELEÇÃO → SELEÇÃO PROPORCIONADA P/ FITNESS

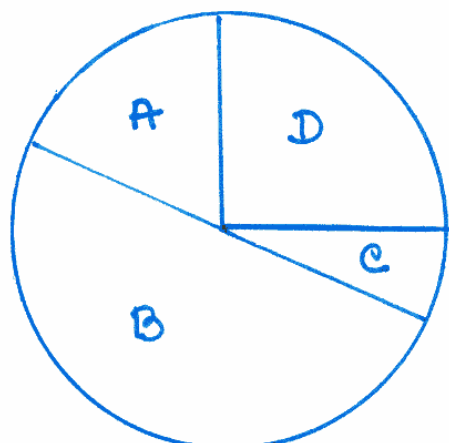
NESTE TIPO DE SELEÇÃO, O N.º

DE VEZES QUE UM INDIVÍDUO

PODERÁ SE REPRODUZIR É DADO POR:

ESTE CRITÉRIO É EQUIVALENTE À "SELEÇÃO POR VIABILIDADE", EM BIOLOGIA.

$$\frac{\text{FITNESS DO INDIVÍDUO}}{\text{FITNESS DA POPULAÇÃO}}$$

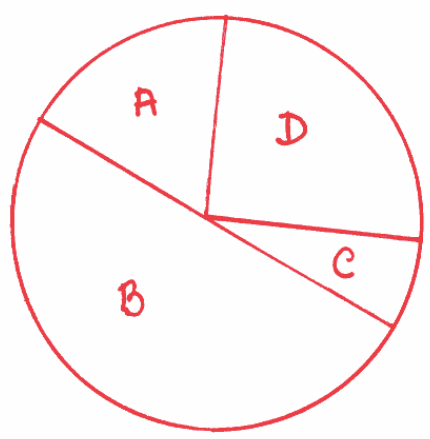


- O CRITÉRIO DE SELEÇÃO PROPORCIONADA PELO FITNESS PODE SER IMPLEMENTADO CONFORME FIGURA I.2.
- CONSISTE EM ATRIBUIR A CADA INDIVÍDUO UMA "FATIA" (DE UMA ROLETA) QUE CORRESPONDA EM ÁREA AO FITNESS DO INDIVÍDUO.

FIGURA I.2: "ROULETTE-WHEEL SAMPLING" (GOLDBERG, 1989).

SELEÇÃO DO INDIVÍDUO → SORTEIO :

A ROLETA É GIRADA E A BOLINHA REPOUSA SOBRE UMA DAS "FATIAS".



COMO O TAMANHO DA POPULAÇÃO É $n=4$, A ROLETA SERÁ GIRADA 4 VEZES, PODENDO ESCOLHER, POR EXEMPLO:

- | | | | |
|---------|------------|---|-----------------------------|
| 1º GIRO | CROMOSSOMO | B | } INDIVÍDUOS PAIS SORTEADOS |
| 2º GIRO | CROMOSSOMO | D | |
| 3º GIRO | CROMOSSOMO | B | } INDIVÍDUOS PAIS SORTEADOS |
| 4º GIRO | CROMOSSOMO | C | |

- * O CROMOSSOMO A NÃO FOI SORTEADO POR ACASO.
- * SE A ROLETA FOR GIRADA (SORTEIO REALIZADO) MUITAS VEZES, OS RESULTADOS MÉDIOS ESTARÃO PRÓXIMOS AOS ESPERADOS.

CRUZAMENTO : OS CROMOSSOMOS PAIS SELECIONADOS SOFRERÃO A OPERAÇÃO DE CRUZAMENTO COM $p_c=0.7$: SE FOREM CRUZADOS IRÃO GERAR 2 INDIVÍDUOS FILHOS. SE NÃO FOREM CRUZADOS, SUA NOVA GERAÇÃO (PROLE) SERÁ UMA CÓPIA EXATA DOS PAIS (CLONES).

- * SUPONDO QUE O CRUZAMENTO ENTRE B E D OCORRA APÓS O 1º BIT, A PROLE GERADA SERÁ:



- * SUPONDO QUE O CRUZAMENTO ENTRE B E C NÃO OCORRA, A PROLE GERADA SERÁ:

$B = 11101110$ $C = 00100000$

MUTAÇÃO : CADA CROMOSSOMO GERADO NA ETAPA DE CRUZAMENTO SERÁ SUJEITO À MUTAÇÃO EM CADA POSIÇÃO, com $p_m = 0.001$.

A POPULAÇÃO (INTERMEDIÁRIA) GERADA APÓS AS OPERAÇÕES DE CRUZAMENTO É :

E = 10110100 B = 11101110
F = 01101110 C = 00100000

SUPONDO QUE O INDIVÍDUO E SEJA MUTADO NA 6ª POSIÇÃO:

E = 10110100 FORMARA' E' = 10110000

SUPONDO QUE O INDIVÍDUO B SEJA MUTADO NA 1ª POSIÇÃO:

B = 11101110 FORMARA' B' = 01101110

SUPONDO QUE OS INDIVÍDUOS F E C NÃO SOFRAM MUTAÇÃO:

F = 01101110 C = 00100000

* APÓS AS OPERAÇÕES DE SELEÇÃO, CRUZAMENTO E MUTAÇÃO, A NOVA POPULAÇÃO SERÁ CONFORME TABELA I.2.

* OBSERVE NA TABELA I.2 QUE, EMBORA A NOVA POPULAÇÃO NÃO CONTENHA O ELEMENTO DE FITNESS = 6 (MAIOR FITNESS PRESENTE NA POPULAÇÃO INICIAL, CORRESPONDENTE AO INDIVÍDUO B), O FITNESS MÉDIO DA NOVA POPULAÇÃO AUMENTOU PARA 14/4.

* MAIS ITERAÇÕES DESTES PROCEDIMENTO PODERÃO CONDUZIR A CADEIAS DE BITS FORMADAS UNICAMENTE POR BITS "1".

IDENTIFICAÇÃO	CADEIA DE BITS	FITNESS	
A	0 0 0 0 0 1 1 0	2	FITNESS MÉDIO POP. INICIAL = $= \frac{2+6+1+3}{4} = \frac{12}{4}$
B	1 1 1 0 1 1 1 0	6	
C	0 0 1 0 0 0 0 0	1	
D	0 0 1 1 0 1 0 0	3	
E'	1 0 1 1 0 0 0 0	3	FITNESS MÉDIO POP. FINAL = $= \frac{3+5+1+5}{4} = \frac{14}{4}$
F	0 1 1 0 1 1 1 0	5	
C	0 0 1 0 0 0 0 0	1	
B'	0 1 1 0 1 1 1 0	5	

TABELA I.2: POPULAÇÕES INICIAL E FINAL DO GA-EXEMPLO

I.6 ALGORITMOS GENÉTICOS E MÉTODOS DE BUSCA TRADICIONAIS

(BUSCA NO CONTEXTO DE GAS X OUTROS CONTEXTOS)

1. BUSCA POR DADOS ARMAZENADOS
2. BUSCA POR CAMINHOS QUE CONDUZAM A ALVOS
3. BUSCA POR SOLUÇÕES

BUSCA POR DADOS ARMAZENADOS

- OBJETIVA RECUPERAR DE FORMA EFICIENTE, INFORMAÇÕES ARMAZENADAS EM MEMÓRIAS DE COMPUTADORES.
- EXEMPLO: QUAL A MELHOR MANEIRA DE RECUPERAR UM DADO SOBRENOME EM UMA GRANDE BASE DE DADOS DE NOMES E ENDEREÇOS ARMAZENADOS DE ACORDO COM ALGUM CRITÉRIO DE ORDEMÇÃO?
- ALGORITMOS: BUSCA BINÁRIA (KNUTH, 1973).

BUSCA POR CAMINHOS QUE CONDUZAM A ALVOS

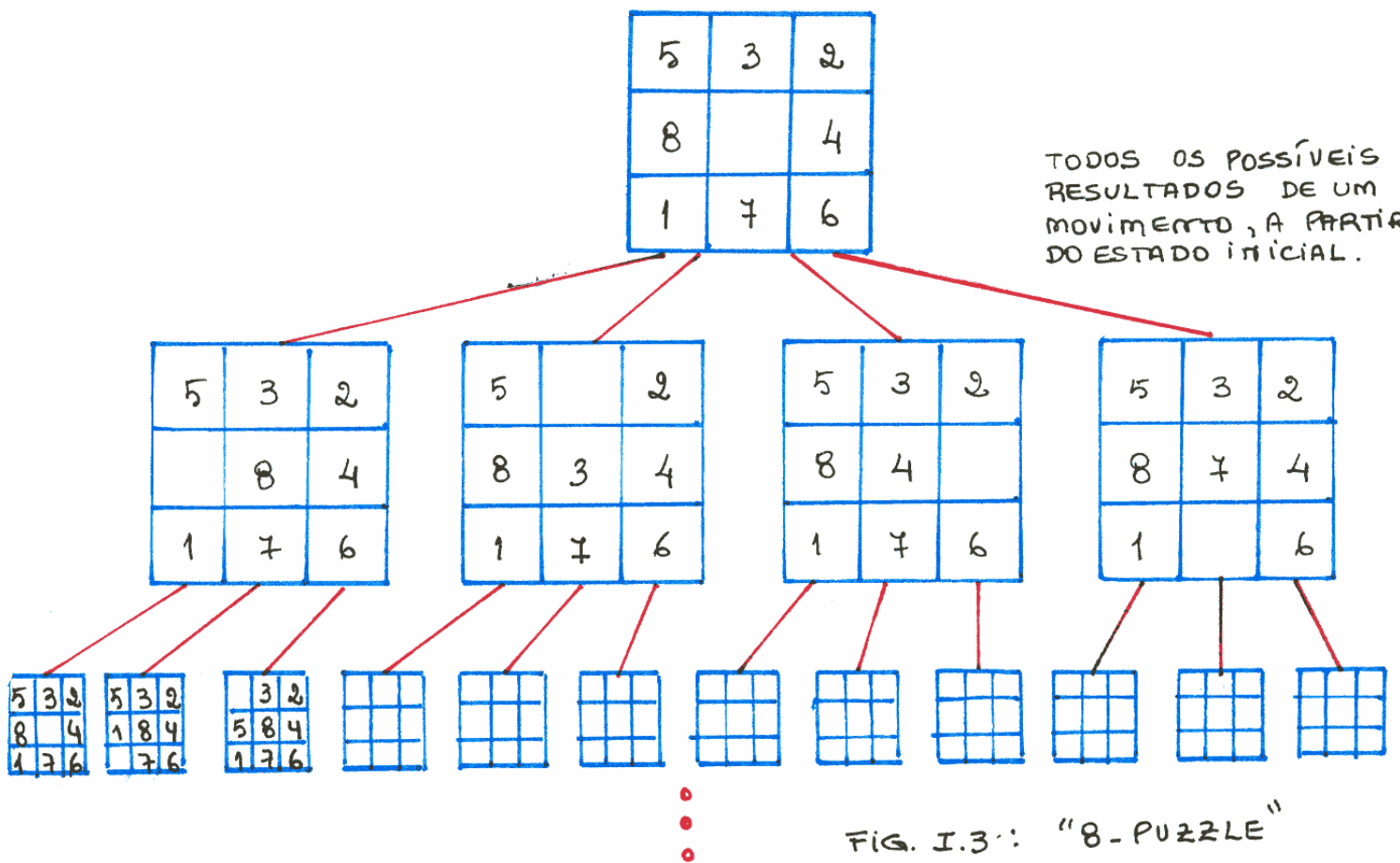
- OBJETIVA ENCONTRAR DE FORMA EFICIENTE UM CONJUNTO DE AÇÕES QUE CONDUZIRÃO A UM DETERMINADO ALVO, PARTINDO DE UM ESTADO INICIAL.
- EXEMPLO: QUAL O MELHOR CONJUNTO DE MOVIMENTOS QUE, A PARTIR DE UM ESTADO INICIAL CONDUZIRÃO AO ESTADO EM QUE TODAS AS PEÇAS ESTARÃO ORDENADAS EM UM QUEBRA-CABEÇAS** DE 8 PEÇAS? ("8-PUZZLE" ESTUDADO EM INTELIG. ARTIFICIAL)
- ALGORITMOS: "DEPTH-FIRST SEARCH", "BRANCH AND BOUND" E "A*".

** VER FIGURA I.3

5	3	2
8		4
1	7	6



1	2	3
8		4
7	6	5



OS ALGORITMOS DE BUSCA, NO CONTEXTO DE INTELIGÊNCIA ARTIFICIAL, SÃO MÉTODOS PARA ENCONTRAR, DE FORMA EFICIENTE, O MELHOR (NO CASO, O MEMOR) CAMINHO NA ÁRVORE A PARTIR DO ESTADO INICIAL, ATÉ ATINGIR O ESTADO ALVO.

BUSCA POR SOLUÇÕES

- OBJETIVA ENCONTRAR DE FORMA EFICIENTE UMA SOLUÇÃO PARA UM PROBLEMA EM UM GRANDE ESPAÇO DE SOLUÇÕES CANDIDATAS.
- * É UMA CLASSE MAIS GERAL DE BUSCA DO QUE A "BUSCA POR CAMINHOS QUE CONDUZAM A ALVOS".
- * SÃO PROBLEMAS DE BUSCA EM QUE SÃO USADOS GAS.

QUANDO SE BUSCA UM ESPAÇO DE SOLUÇÕES CANDIDATAS COM UM GA, NEM TODAS AS POSSÍVEIS SOLUÇÕES CANDIDATAS SÃO CRIADAS PRIMEIRO E DEPOIS AVALIADAS;

NA VERDADE, O GA É UM MÉTODO PARA ENCONTRAR SOLUÇÕES ÓTIMAS OU BOAS EXAMINANDO APENAS UMA PEQUENA FRAÇÃO DOS POSSÍVEIS CANDIDATOS.

UMA "BUSCA POR CAMINHOS QUE CONDUZAM A ALVOS" PODE SER CONSIDERADA UMA "BUSCA POR SOLUÇÕES, JÁ QUE UM CAMINHO ATRAVÉS DE UMA ÁRVORE DE BUSCA PODE SER CODIFICADO COMO UMA SOLUÇÃO CANDIDATA. (EXPL: "8-PUZZLE" → AS SOLUÇÕES CANDIDATAS PODEM SER LISTAS DE MOVIMENTOS A PARTIR DO ESTADO INICIAL A ALGUM OUTRO ESTADO, E QUE CHEGUEM NO ESTADO FINAL QUE SEJA O ESTADO ALVO.)

A COMBINAÇÃO PARTICULAR DE ELEMENTOS EM GAS, TAIS COMO BUSCA BASEADA NA POPULAÇÃO EM PARALELO COM SELEÇÃO ESTOCASTICA DE MUITOS INDIVÍDUOS, CRUZAMENTO ESTOCASTICO E MUTAÇÃO DISTINGUEM OS GAS DE OUTROS MÉTODOS DE BUSCA.

MUITOS OUTROS MÉTODOS DE BUSCA TÊM ALGUNS DESTES ELEMENTOS, MAS NÃO ESTA PARTICULAR COMBINAÇÃO.

I.7 APLICAÇÕES DE GAS

UMA VERSÃO SIMPLES DE UM GA PODE SER DESCRITA COMO:

- (1) GERAR UM CONJUNTO DE SOLUÇÕES CANDIDATAS, QUE CONSTITUEM A POPULAÇÃO INICIAL.
- (2) AVALIAR A ADEQUAÇÃO DAS SOLUÇÕES CANDIDATAS DE ACORDO COM ALGUM CRITÉRIO DE FITNESS.
- (3) DECIDIR, DE ACORDO COM A AVALIAÇÃO DO FITNESS, QUAIS INDIVÍDUOS SERÃO MANTIDOS E QUAIS SERÃO DESEARTADOS.
- (4) PRODUZIR VARIANTES DA POPULAÇÃO (NOVOS INDIVÍDUOS) ATRAVÉS DO USO DE ALGUNS TIPOS DE OPERADORES SOBRE A POPULAÇÃO DE CANDIDATOS SOBREVIVENTES.

VARIAÇÕES DA VERSÃO SIMPLES DE UM GA SÃO USADAS EM UM GRANDE NÚMERO DE MODELOS E PROBLEMAS CIENTÍFICOS E DE ENGENHARIA, TAIS COMO:

- PROGRAMAÇÃO AUTOMÁTICA;
- MACHINE LEARNING;
- ECONOMIA;
- SISTEMAS IMUNOLÓGICOS;
- ECOLOGIA;
- GENÉTICA POPULACIONAL;
- EVOLUÇÃO E APRENDIZADO;
- SISTEMAS SOCIAIS E
- OTIMIZAÇÃO.

PROGRAMAÇÃO AUTOMÁTICA: GAS SÃO USADOS PARA EVOLUIR PROGRAMAS COMPUTACIONAIS PARA TAREFAS ESPECÍFICAS, E PARA PROJETAR OUTRAS ESTRUTURAS COMPUTACIONAIS COMO AUTÔMATOS E REDES CLASSIFICADORAS.

MACHINE LEARNING: GAS TÊM SIDO USADOS EM MUITAS APLICAÇÕES, INCLUINDO TAREFAS DE CLASSIFICAÇÃO E PREDIÇÃO (TAIS COMO PREDIÇÕES DE CLIMA E DE ESTRUTURAS DE PROTEÍNAS). GAS TAMBÉM TÊM SIDO USADOS PARA EVOLUIR ASPECTOS DE SISTEMAS DE APRENDIZADO DE MÁQUINAS, TAIS COMO PESOS DE REDES NEURAIS ARTIFICIAIS, SISTEMAS DE PRODUÇÃO SIMBÓLICA E SENSORES PARA ROBÔS.

ECONOMIA: GAS TÊM SIDO USADOS PARA MODELAR PROCESSOS DE INOVAÇÃO E A EMERGÊNCIA DE MERCADOS ECONÔMICOS.

SISTEMAS IMUNOLÓGICOS: GAS TÊM SIDO USADOS PARA MODELAR VÁRIOS ASPECTOS DE SISTEMAS IMUNOLÓGICOS NATURAIS, INCLUINDO MUTAÇÕES SOMÁTICAS DURANTE O TEMPO DE VIDA DE UM INDIVÍDUO E A DESCOBERTA DE FAMÍLIAS MULTI-GENES DURANTE EVOLUÇÕES.

ECOLOGIA: GAS TÊM SIDO USADOS PARA MODELAR FENÔMENOS ECOLÓGICOS COMO COEVOLUÇÃO HOSPEDEIRO-PARASITA E SIMBIOSSES.

GENÉTICA POPULACIONAL: GAS TÊM SIDO USADOS PARA ESTUDAR QUESTÕES COMO "SOB QUAIS CONDIÇÕES IRÁ UM GENE PARA RECOMBINAÇÃO SER EVOLUCIONARIAMENTE VIÁVEL?".

EVOLUÇÃO E APRENDIZAGEM: GAS TÊM SIDO USADOS PARA ESTUDAR COMO APRENDIZAGEM INDIVIDUAL E EVOLUÇÃO DE ESPÉCIES AFETAM UMA A OUTRA.

SISTEMAS SOCIAIS: GAS TÊM SIDO USADOS PARA ESTUDAR ASPECTOS EVOLUCIONÁRIOS DE SISTEMAS SOCIAIS, TAIS COMO A EVOLUÇÃO DO COMPORTAMENTO SOCIAL EM COLÔNIAS DE INSETOS E, DE FORMA MAIS GERAL, A EVOLUÇÃO DA COOPERAÇÃO E DA COMUNICAÇÃO EM SISTEMAS MULTI-AGENTES.

OTIMIZAÇÃO: GAS TÊM SIDO USADOS EM UMA GRANDE VARIEDADE DE TAREFAS DE OTIMIZAÇÃO, INCLUINDO OTIMIZAÇÃO NUMÉRICA E PROBLEMAS DE OTIMIZAÇÃO COMBINATORIAL, COMO, POR EXEMPLO, LAYOUT DE CIRCUITOS.