Computação Evolutiva

Cardinalidade de espaços de busca discretos

1. Introdução 2
2. Listas compostas por atributos discretos 6
   2.1 Problema de bipartição 7
   2.2 Problema de tripartição 8
   2.3 Caixeiro viajante 8
   2.4 Listas com restrições 9
3. Programação genética 10
4. Árvores binárias com raiz 10
5. Árvores binárias sem raiz 13
6. Máquinas de estados finitos 13
7. Binomiais 15
8. Conjuntos 16
   8.1 Conjunto potência 17
   8.2 Probabilidades e combinações 18
   8.3 Combinação: exemplo 1 19
   8.4 Combinação: exemplo 2 19
   8.5 Combinação: exemplo 3 20
   8.6 Multiconjuntos 21
   8.7 Permutação de multiconjuntos 22
9. Resumo dos principais resultados 24
10. Número de Stirling de tipo 2 25
11. Uma aproximação para a função fatorial 26
12. Referências bibliográficas 26

1. Introdução

- A representação computacional utilizada para codificar soluções candidatas em computação evolutiva, associada ao genótipo de cada indivíduo da população, deve atender, ao menos, dois aspectos básicos:
  - ✓ Garantia de que o mapeamento genótipo $\rightarrow$ fenótipo conduza a um único fenótipo;
  - ✓ Garantia de contemplar todas as soluções candidatas, ou seja, garantia de que existe pelo menos um genótipo para cada fenótipo possível.

- Podem até existir algoritmos evolutivos que empregam representações computacionais em que essas duas condições (ou uma delas) não sejam atendidas. No entanto, esses casos se configuram como excepcionais e devem ser evitados em aplicações práticas.
- De qualquer modo, é a representação computacional que define o espaço de busca em que se dará o processo evolutivo.

- Os espaços de busca podem ser contínuos, como o $\Re^n$, ou discretos, como o $\{0,1\}^n$, para um $n \geq 1$ inteiro.
- A capacidade de busca de um algoritmo evolutivo está vinculada a diversos fatores, como:
  - ✓ Dimensão do espaço de busca contínuo ou cardinalidade do espaço de busca discreto;
  - ✓ Adequação/eficácia dos operadores genéticos;
  - ✓ Adequação/eficácia dos operadores de busca local;
  - ✓ Acidentalidade da superfície de *fitness*.
- Neste tópico do curso, serão estudadas técnicas de matemática discreta para determinar a cardinalidade de espaços de busca discretos.
- A **cardinalidade** de um conjunto contável de elementos é dada pelo número de elementos do conjunto.

- Em teoria de complexidade computacional, existe a complexidade intrínseca de um problema e existe a complexidade dos algoritmos que foram concebidos para resolver esse problema. A complexidade dos algoritmos será sempre maior ou igual à complexidade intrínseca do problema, e cada algoritmo alternativo para resolver o mesmo problema pode apresentar complexidades computacionais distintas. Quando um algoritmo apresenta complexidade igual à complexidade intrínseca do problema, então ele é dito ter complexidade mínima.
- De forma análoga, em termos de representação computacional, cada problema pode admitir várias alternativas de representação de suas soluções candidatas. E cada representação vai definir um espaço de busca distinto. Logo, para um mesmo problema, podem existir espaços de busca completamente diferentes, tanto em termos de dimensão (caso contínuo) ou cardinalidade (caso discreto), como em termos de características (acidentalidade) da superfície de *fitness*.

- Representações computacionais de soluções candidatas que conduzam a espaços de busca de menor dimensão (ou cardinalidade) devem ser buscadas, mas nem sempre conduzem a algoritmos evolutivos mais eficientes. Um espaço de busca de maior dimensão (ou cardinalidade) que algum outro pode permitir que a evolução ocorra de forma mais rápida, talvez pela presença de operadores genéticos mais adequados, ou operadores de busca local mais eficazes, ou então por causa da maior suavidade (menor acidentalidade) da superfície de *fitness*.
- Mesmo assim, o conhecimento do número de soluções candidatas em todo o espaço de busca, ou então em uma vizinhança de uma dada proposta de solução, pode auxiliar em vários aspectos do projeto de um algoritmo evolutivo.
- A seguir, serão considerados alguns problemas e propostas de representações computacionais, seguidas da determinação da cardinalidade dos espaços de busca associados. Serão apresentados também cálculos de cardinalidade de conjuntos, mesmo que estes não correspondam necessariamente a espaços de busca.

2. Listas compostas por atributos discretos

- Uma quantidade expressiva de soluções candidatas em problemas passíveis de tratamento por processos evolutivos pode ser representada na forma de uma lista (ou vetor, ou arranjo ordenado) de atributos.
- Particularmente no caso de listas de atributos discretos, o número de soluções candidatas, dentre as que admitem o tipo de representação computacional definida, define a cardinalidade do espaço de busca. É nesse sentido que se fala aqui de cardinalidade de listas.
- Qualquer lista de um conjunto de n objetos é chamada de **permutação** de n objetos.
- Qualquer lista de um conjunto de $k \leq n$ objetos é chamada de k-permutação, ou permutação de n objetos tomados k de cada vez.
- O número de listas distintas de comprimento k cujos elementos são escolhidos de um conjunto de n elementos possíveis (portanto, um conjunto discreto), é:
  - ✓ n^k, caso se permitam repetições;

✓ $(n)_k$, caso não se permitam repetições.

- $(n)_k = P(n,k) = {}_nP_k = \prod_{j=0}^{k-1}(n-j) = \dfrac{n!}{(n-k)!} = n\cdot(n-1)\cdots(n-k+2)\cdot(n-k+1)$
- $(n)_n = \prod_{j=0}^{n-1}(n-j) = \prod_{j=1}^{n} j = n!$
- Quando se permitem repetições, n e k podem apresentar quaisquer relações entre si. Por outro lado, quando não se permitem repetições, é necessário que $n \geq k$.

2.1 Problema de bipartição

- Um conjunto de N números positivos deve ser dividido em 2 subconjuntos de modo a produzir 2 somas que, entre si, tenham distância mínima. Qual é o número de soluções candidatas?

Solução:

Cada número pode estar em um de 2 subconjuntos. Conclusão: $\boxed{2^N}$.

Uma representação binária permite um mapeamento direto do genótipo para o fenótipo.

2.2 Problema de tripartição

- Um conjunto de N números positivos deve ser dividido em 3 subconjuntos de modo a produzir 3 somas que, entre si, tenham distância mínima.

Solução:

Cada número pode estar em um de 3 subconjuntos. Conclusão: $\boxed{3^N}$.

Uma representação ternária permite um mapeamento direto do genótipo para o fenótipo.

2.3 Caixeiro viajante

- Qual é o número de soluções candidatas para o problema do caixeiro viajante com N cidades? Não importa a cidade inicial e nem o sentido de percurso.

Solução:

Número de permutações das N cidades: $N!$.

Como qualquer cidade pode ser a cidade de origem, então existem N soluções idênticas para cada percurso possível, o que reduz a cardinalidade do espaço de busca a: $\frac{N!}{N} = (N-1)!$.

Como o sentido de percurso não importa, então ainda existem 2 soluções idênticas para cada percurso possível, o que reduz a cardinalidade do espaço de busca a: $\frac{(N-1)!}{2}$.

2.4 Listas com restrições

- Dado que se dispõe de n livros diferentes de Inglês, de k livros diferentes de Espanhol e de p livros diferentes de Japonês, quantas alternativas possíveis existem para ordená-los numa prateleira?

$$(n + k + p)!$$

- E se os livros de cada língua precisam ficar juntos?

$$3!n!k!p!$$

3. Programação genética

- Na abordagem de um problema de programação genética, com representação computacional restrita a árvores binárias com N folhas, (N–1) nós internos, T topologias distintas, S símbolos não-terminais candidatos e P símbolos terminais candidatos, qual seria a cardinalidade do espaço de busca?

Solução:

Cada folha pode ter um de P símbolos.

Cada nó interno pode ter um de S símbolos.

Cada topologia tem N folhas e N–1 nós internos (T será calculado na próxima seção).

Conclusão: $T \times S^{N-1} \times P^{N}$.

4. Árvores binárias com raiz

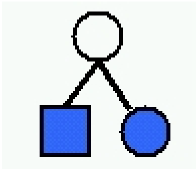
- Qual é o número de topologias possíveis para árvores binárias com raiz e N folhas?

Solução:

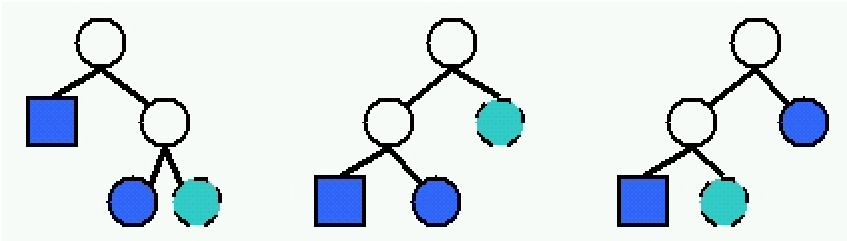
O número de nós internos de árvores binárias com N folhas é $N-1$.

O número total de nós de árvores binárias com N folhas é $2N-1$.

Para 2 nós-folha, existe apenas uma topologia possível:



Para considerar o terceiro nó, este pode se combinar com qualquer um dos três nós já existentes, requerendo sempre a criação de um nó interno adicional:



Para considerar o quarto nó, este pode se combinar com qualquer um dos cinco nós já existentes, e assim por diante, até o N-ésimo nó-folha, o qual poderá se combinar com cada um dos $2(N-1)-1 = 2N-3$ nós. Assim, o número total de topologias possíveis é:

$$1\times 3\times 5\times\cdots\times(2N-5)\times(2N-3)=\prod_{i=2}^{N}(2i-3)$$

Multiplicando e dividindo a expressão acima por:

$$\prod_{i=2}^{N-1}(2i-2)=2\times 4\times 6\times\cdots\times(2(N-2)-2)\times(2(N-1)-2)$$

resulta:

$$\frac{1\times 2\times 3\times 4\times 5\times 6\times\cdots\times(2N-6)\times(2N-5)\times(2N-4)\times(2N-3)}{2\times 4\times 6\times\cdots\times(2N-6)\times(2N-4)}$$

O numerador já pode ser convertido em fatorial. Para o denominador, repare que existem N–2 fatores pares no produto, sendo que é possível dividir todos por 2 e multiplicar por 2^{N-2}, produzindo, finalmente:

$$\boxed{\frac{(2N-3)!}{2^{N-2}(N-2)!}}$$

5. Árvores binárias sem raiz

Qual é o número de topologias possíveis para árvores binárias sem raiz e N folhas?

Solução:

Como a raiz pode se combinar com qualquer nó de uma árvore sem raiz, então pode-se concluir que o número de topologias para árvores binárias sem raiz e N folhas é igual ao caso com raiz, mas considerando um nó a menos. Assim resulta:

$$\boxed{\frac{(2N-5)!}{2^{N-3}(N-3)!}}$$

6. Máquinas de estados finitos

- Qual é o número de máquinas de estados finitos com N estados, estado inicial a determinar, P símbolos de entrada e Q símbolos de saída. Suponha que não haja estado final.

Solução:

Cada um dos N estados pode ser o estado inicial.

Para cada estado:

- ✓ Para cada símbolo de entrada, existem N próximos estados possíveis;
- ✓ Para cada próximo estado, existem Q símbolos de saída possíveis.

Conclusão: $N \times \left(N^P \times Q^P\right)^N$.

Solução alternativa:

Cada um dos N estados pode ser o estado inicial. De cada um dos N estados saem P arcos, de modo que existem $P \times N$ arcos. Cada arco pode chegar a qualquer um dos N estados. Logo, o número de topologias distintas para a máquina de estado finito é $N^{P\times N}$. Cada arco pode ter associado a si um dentre os Q símbolos de saída. Logo, o número de conjuntos distintos de arcos para uma mesma topologia é $Q^{P\times N}$.

Conclusão: $N \times N^{P\times N} \times Q^{P\times N}$.

7. Binomiais

- Seja n um número natural. Então é sabido que:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot x^{n-k} \cdot y^k$$

- Esta é a razão pela qual $\binom{n}{k}$ é chamado de coeficiente binomial.
- $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$, com k um número natural tal que $k \le n$.
- É possível provar que: $\binom{n}{k} = \binom{n}{n-k}$.
- É possível provar que: $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$, com $0 < k < n$.
- É possível provar que: $k \cdot \binom{n}{k} = n \cdot \binom{n-1}{k-1}$, com $0 < k < n$.

- É possível provar que: $\binom{n}{k}\cdot\binom{k}{m}=\binom{n}{m}\cdot\binom{n-m}{k-m}$, com $0\leq m\leq k\leq n$.
- É possível provar que: $\binom{n}{3}=\binom{2}{2}+\binom{3}{2}+\binom{4}{2}+\cdots+\binom{n-1}{2}$.

8. Conjuntos

- Um conjunto pode ser definido como um **arranjo não-ordenado** de elementos, ou uma **combinação** de elementos.
- $C(n,k)=\binom{n}{k}=\dfrac{n!}{k!(n-k)!}$ é o número de subconjuntos (combinações, arranjos não-ordenados) de *k* elementos que podem ser formados a partir de *n* elementos.
- Esse número é distinto do número de listas (arranjos ordenados) de *k* elementos que podem ser formados a partir de *n* elementos, pois no caso de conjuntos a ordem não importa e não se admite a repetição de elementos.

- Assim, como já visto, quando a ordem importa, que é o caso de listas, o número de listas de k elementos que podem ser formados a partir de n elementos é dado por $(n)_k = P(n,k) = \dfrac{n!}{(n-k)!}$, sem repetição, e n^k, com repetição.

8.1 Conjunto potência

- O conjunto potência de um conjunto A é o conjunto de todos os subconjuntos de A. Logo, qual é a cardinalidade do conjunto potência de A?
- Exemplo: O conjunto potência do conjunto {1, 2, 3} é dado por

{ {}, {1}, {2}, {3}, {1, 2}, {1, 3}, {2, 3}, {1, 2, 3} }

Solução:

Nos conjuntos que representam os elementos do conjunto potência, cada elemento do conjunto original pode estar presente ou não. A cardinalidade é igual ao número de subconjuntos de A, o qual é dado por $2^{|A|}$, onde $|A|$ é a cardinalidade do conjunto A.

8.2 Probabilidades e combinações

- Algumas decisões de projeto em algoritmos evolutivos também envolvem probabilidades, sendo que o processo evolutivo pode ser interpretado como um processo estocástico.
- Na inicialização de um cromossomo binário de N bits, sabe-se que cada bit é escolhido por um gerador pseudo-aleatório de bits que obedece à seguinte regra de probabilidade: p para o bit 1 e $(1–p)$ para o bit 0. Qual é a probabilidade do cromossomo ter R bits com o valor 1?

Solução:

É evidente que deve-se considerar que $R \leq N$.

Probabilidade do gerador produzir R bits com o valor 1 e $N–R$ bits com o valor 0:

$$(p)^R(1-p)^{N-R}$$

Número de combinações possíveis de ocorrência dos R bits 1 em N posições:

$$\binom{N}{R} = \frac{N!}{R!(N-R)!}$$

Logo, a resposta é dada na forma:

$$\boxed{\binom{N}{R}(p)^R(1-p)^{N-R}}$$

8.3 Combinação: exemplo 1

- Existem 20 cromossomos em uma população. Se cada um fizer *crossover* com todos os outros cromossomos exatamente uma vez, quantos *crossovers* terão sido feitos?

$$\binom{20}{2} = \frac{20!}{2!(20-2)!} = 10 \cdot 19 = 190$$

8.4 Combinação: exemplo 2

- 50 soluções candidatas distintas vão competir para ocupar as 10 primeiras posições num processo seletivo de um algoritmo evolutivo. Sem saber a priori o *fitness* dessas soluções candidatas, quantas são as possibilidades de ocupação das 10 primeiras posições?

$$\binom{50}{10} = \frac{50!}{10!(50-10)!} = 1.027.227.817$$

- E se for levada em conta também a ordem das 10 primeiras colocadas?

$$\binom{50}{10} \cdot 10!$$

que nada mais é que o número de permutações de 50 objetos tomados 10 de cada vez.

8.5 Combinação: exemplo 3

- Qual é o número de combinações possíveis de 9 operadores genéticos que devem ser atribuídos a 4 populações, sendo que a primeira população deve receber 3 operadores e as demais 2?

$$\binom{9}{3}\cdot\binom{6}{2}\cdot\binom{4}{2}\cdot\binom{2}{2}=\frac{9!}{3!\cdot 2!\cdot 2!\cdot 2!}=7.560$$

- Repare que aqui estamos particionando 9 operadores, de forma ordenada, em 4 conjuntos disjuntos, ou seja, estamos realizando uma **partição ordenada**.

8.6 Multiconjuntos

- Dado um conjunto, um elemento pertence ou não a ele. Um mesmo elemento não pode figurar duas vezes em um conjunto.
- Um multiconjunto é uma generalização de um conjunto em que, além de não haver ordem entre seus elementos (como num conjunto), os elementos podem aparecer em multiplicidade.
- Exemplo: Apresente todos os multiconjuntos possíveis de 3 elementos, onde os elementos são escolhidos do conjunto {1, 2, 3}.

$$\langle 1\quad 1\quad 1\rangle,\ \langle 1\quad 1\quad 2\rangle,\ \langle 1\quad 1\quad 3\rangle,\ \langle 1\quad 2\quad 2\rangle,\ \langle 1\quad 2\quad 3\rangle$$
$$\langle 1\quad 3\quad 3\rangle,\ \langle 2\quad 2\quad 2\rangle,\ \langle 2\quad 2\quad 3\rangle,\ \langle 3\quad 3\quad 2\rangle,\ \langle 3\quad 3\quad 3\rangle$$

- O número de multiconjuntos de k elementos que podem ser formados pelos inteiros de 1 a n é dado por:

$$\left(\!\!\binom{n}{k}\!\!\right) = \binom{n+k-1}{k}$$

- Exemplo em algoritmos evolutivos: No processo de seleção por torneio, considerando 4 participantes e 20 indivíduos na população, quantas são as possibilidades de grupos?

$$\left(\!\!\binom{20}{4}\!\!\right) = \binom{20+4-1}{4} = \binom{23}{4} = 8.855$$

8.7 Permutação de multiconjuntos

- As permutações de um multiconjunto são também conhecidas como arranjos com repetição de elementos. Que fique claro que o multiconjunto não tem ordem, mas os seus elementos podem ser dispostos em uma lista (arranjo ordenado de elementos).
- Sem repetição de elementos, o número de arranjos possíveis de n elementos é $n!$.

- Dado um multiconjunto composto por k elementos distintos, sendo $n_1, n_2, \cdots, n_k$ as multiplicidades respectivas desses elementos, então o número de permutações (listas distintas) que se pode obter a partir desses elementos é dada por:

$$\frac{(n_1 + n_2 + \cdots + n_k)!}{(n_1)! \cdot (n_2)! \cdots (n_k)!}$$

- Exemplo: Se um multiconjunto tem 8 elementos ao todo, com 3 repetições de um elemento e 2 repetições de um outro elemento, com os demais elementos não se repetindo, então o número total de arranjos diferentes é dado por:

$$\frac{8!}{3!2!} = 3.360$$

- Isso é utilizado para se calcular o número de anagramas de uma palavra, que é dado pelo número de palavras diferentes que podem ser formadas com as letras da palavra. O exemplo numérico acima pode ser aplicado à palavra PAPAGAIO. É claro que poucas das 3.360 palavras diferentes fazem algum sentido em alguma língua.

9. Resumo dos principais resultados

- A tabela a seguir também é conhecida como contagem de coleções.
- Tamanho do universo: n
- Tamanho da coleção: k
- Quantas coleções de k elementos existem, sabendo que existem n elementos candidatos?

	Repetição permitida	**Repetição proibida**
Ordenado	n^k	$(n)_k = \frac{n!}{(n-k)!}$
Não-ordenado	$\left(\binom{n}{k}\right) = \binom{n+k-1}{k}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

- A última coluna desta tabela mostra a diferença entre o número de permutações e combinações de n objetos tomados k de cada vez: $C(n,k) = \frac{P(n,k)}{k!}$.

10. Número de Stirling de tipo 2

- O número de Stirling de tipo 2 realiza a contagem do número de **diferentes partições de um conjunto** de p elementos em k conjuntos não-vazios e não-ordenados. Ele é dado pela seguinte fórmula (com versões alternativas):

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \frac{1}{k!} \sum_{j=0}^{k} (-1)^{j} \binom{k}{j} (k-j)^{n} = \sum_{j=1}^{k} (-1)^{k-j} \frac{j^{n-1}}{(j-1)! \cdot (k-j)!} = \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^{n}$$

- Para chegar a esta fórmula, James Stirling (1692-1770) aplicou o princípio da inclusão-exclusão, usado para contar o número de elementos resultante da união de conjuntos que possivelmente apresentam elementos em comum.
- Para dois conjuntos A e B, sabe-se que:

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

- Para três conjuntos A, B e C, sabe-se que:

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$

11. Uma aproximação para a função fatorial

- James Stirling (1692-1770) provou que:

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} = 1$$

- Logo, para valores elevados de *n*, vale a aproximação:

$$n! \cong \sqrt{2\pi n}\left(\frac{n}{e}\right)^n = \sqrt{2\pi n} \cdot n^n \cdot e^{-n}$$

12. Referências bibliográficas

LIPSCHUTZ, S. & LIPSON, M. "Discrete Mathematics", 2nd. Edition, Schaum's Outline Series, 1997.

SCHEINERMAN, E.R. "Mathematics: A Discrete Introduction", Brooks Cole, 2nd. Edition, 2005.