MAC323 EXERCÍCIO-PROGRAMA 5

O ESPAÇO DAS ÁRVORES BINÁRIAS

Y. KOHAYAKAWA

Data de entrega: 20/7/2012 (23:55)

Introdução. Neste EP, você investigará *árvores binárias com n nós internos.* Podemos munir o conjunto dessas árvores com uma *distância*, declarando a distância $d(T, T')$ entre tais duas árvores T e T' como sendo o número mínimo de rotações que levam T a T'. Vamos denotar esse *espaço métrico* por $\mathcal{T}_n$. Seu programa deverá ser capaz de executar algumas operações básicas envolvendo $\mathcal{T}_n$, como, por exemplo, determinar $d(T, T')$ para quaisquer T, $T \in \mathcal{T}_n$. Ademais, além de determinar $d(T, T')$, seu programa deverá ser capaz de fornecer uma seqüência $T_0, T_1, \ldots, T_d \in \mathcal{T}_n$ com $T_0 = T$, $T_d = T'$, $d = d(T, T')$ e com cada T_i resultante de uma rotação de T_{i-1} $(1 \le i \le d)$.

Computação em $\mathcal{T}_n$. Além do cálculo da distância entre duas árvores dadas, seu programa deverá ser capaz de determinar ou estimar certas outras quantidades associadas a $\mathcal{T}_n$, definidas a seguir.

Diâmetro de $\mathcal{T}_n$. O *diâmetro* de $\mathcal{T}_n$ é definido como

$$\operatorname{diam}(\mathcal{T}_n) = \max\{d(T, T') \colon T,\ T' \in \mathcal{T}_n\}, \tag{1}$$

isto é, o diâmetro de $\mathcal{T}_n$ é a distância máxima entre dois membros de $\mathcal{T}_n$.

A distância média entre membros de $\mathcal{T}_n$. Já encontramos a noção de árvores binárias aleatórias: se munimos $\mathcal{T}_n$ da distribuição uniforme, isto é, se consideramos todos os membros de $\mathcal{T}_n$ equiprováveis, obtemos o espaço de probabilidade das árvores aleatórias U_n, já vistas anteriormente. Também vimos uma outra distribuição de probabilidade sobre $\mathcal{T}_n$: podemos gerar uma árvore aleatória $T(\sigma) \in \mathcal{T}_n$, inserindo n chaves distintas em uma árvore binária de busca inicialmente vazia, sorteando uma ordenação $\sigma = (\sigma(1), \ldots, \sigma(n))$ uniformemente ao acaso dentre todas as $n!$ possíveis ordenações das n chaves. Temos denotado tais árvores aleatórias por B_n.

Definimos a *distância média uniforme* em $\mathcal{T}_n$ como sendo $d_n^{\mathrm{U}} = \mathbb{E}(d(U_n', U_n''))$, onde U_n' e U_n'' são cópias independentes de U_n e definimos a *distância média permutacional* em $\mathcal{T}_n$ (não consegui inventar um nome melhor) como sendo $d_n^{\mathrm{B}} = \mathbb{E}(d(B_n', B_n''))$, onde B_n' e B_n'' são cópias independentes de B_n.

Codificação de árvores. Será conveniente usarmos uma codificação compacta de árvores binárias. Usaremos palavras sobre o alfabeto $\{$`.`, `x`$\}$ (`.` representa um nó interno e `x` representa um nó externo). A árvore vazia, com 0 nós internos e 1 nó externo, é representada por `x`. Se as subárvores esquerda e direita de T são representadas por `s` e `t`, então `.st` representa T. Por exemplo, as 5 árvores com 3 nós internos (e consequentemente 4 nós externos) são

`.x.x.xx`, `.x..xxx`, `..xx.xx`, `..x.xxx`, `...xxxx`. (2)

Versão de 3 de julho de 2012, 17:41.

Seu programa. A entrada de seu programa, que deve ser lida no `stdin`, deve ser uma seqüência de linhas, cada uma delas expressando um "comando". Cada linha pode ter uma das seguintes formas.

n <n>

Define o valor de n como sendo <n>. Um comando dessa forma deve ocorrer no início da entrada. Durante a execução de seu programa, o usuário poderá mudar o valor de n, emitindo outros comandos dessa forma.

d <s> <t>

Imprime o valor de $d($<s>$,$<t>$)$, onde <s> e <t> são representações de árvores em $\mathcal{T}_n$ (com o valor atual de n).

p <s> <t>

Imprime uma seqüência de árvores $T_0, \ldots, T_d$, como especificada na introdução desse enunciado (mais precisamente, seu programa deve imprimir as representações de $T_0, T_1, \ldots$).

diam

Imprime o valor de $\mathrm{diam}(\mathcal{T}_n)$.

dU

Imprime o valor de d_n^{U}.

dB

Imprime o valor de d_n^{B}.

Estimativas de d_n^{U} e d_n^{B}. Uma forma de *estimar* o valor de d_n^{U} é gerar vários pares (T, T') de acordo com a distribuição U_n, e usar a média das distâncias $d(T, T')$ assim calculadas. Seu programa deverá aceitar o comando

dULGN <M>

que significa o seguinte: seu programa deve gerar <M> pares (T, T') como acima, e deve devolver a média dos números $d(T, T')$. (Explique por que este comando contém `LGN` no nome.) Seu programa deve também estimar d_n^{B} de forma análoga. O comando correspondente deve ser

dBLGN <M>

Relatório. Você deve preparar um relatório, considerando as seguintes instruções/perguntas.

(*i*) Inclua uma breve descrição de como seu programa determina a distância entre duas árvores.

(*ii*) Faça uma tabela com os valores de $\mathrm{diam}(\mathcal{T}_n)$ para valores pequenos de n (quanto maior for sua tabela, melhor).

(*iii*) Faça o mesmo com d_n^{U} e d_n^{B}. (Não deixe de explicitar se os valores nessas tabelas foram estimados estatisticamente ou foram calculados precisamente (a menos de erros de arredondamento).)

(*iv*) Quanto tempo seu programa leva para calcular as entradas das tabelas em (*ii*) e (*iii*)?

(*v*) Você montou as tabelas em (*ii*) e (*iii*) até um certo valor de n. Quanto tempo você estima que levaria para seu programa calcular as entradas dessas tabelas para $n+1$? Justifique sua resposta, levando em conta seus algoritmos, estruturas de dados e fatos sobre árvores binárias.

Observações

1. *Este EP é estritamente individual.* Programas semelhantes receberão nota 0.
2. Seja cuidadoso com sua programação (correção, documentação, apresentação, clareza do código, etc), dando especial atenção a suas estruturas de dados. A correção será feita levando isso em conta.
3. Comparem entre vocês o desempenho de seus programas.
4. Entregue seu EP no Paca.
5. Não deixe de incluir em seu código um *relatório* para discutir seu EP: discuta as estruturas de dados usadas, os algoritmos usados, etc. *Se você escrever claramente como funciona seu EP, o monitor terá pouca dificuldade em corrigi-lo, e assim você terá uma nota mais alta.* (Se o monitor sofrer para entender seu código, sua nota será baixa.)

Observação final. Enviem dúvidas para a lista de discussão da disciplina.

Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–090 São Paulo, SP

Endereço eletrônico: `yoshi@ime.usp.br`