Codificação de Fonte
(Código de Huffman Truncado)

Teoria da Informação - AULA 14
Profª. Verusca Severo

Universidade de Pernambuco
Escola Politécnica de Pernambuco

18 de agosto de 2021

Introdução

- O código de Huffman é uma das técnicas mais populares de codificação de fonte;

- A ideia básica do algoritmo é atribuir palavras código de menor comprimento para os símbolos mais frequentes, e palavras código mais longas para os símbolos mais raros.

- Em suma, a codificação de Huffman possui duas etapas:
  1. Cria-se uma série de reduções dos símbolos através da junção dos dois de menores probabilidades a cada iteração.
  2. Codificam-se todos os símbolos que foram reduzidos. Consiste em codificar cada fonte reduzida, iniciando pela menor fonte e caminhando em direção à fonte original.

Introdução

- Em suma, a codificação de Huffman possui duas etapas:
  1. Cria-se uma série de reduções dos símbolos através da junção dos dois de menores probabilidades a cada iteração.

Fonte Original		Reduções de fonte			
Símbolo	Probabilidade	1	2	3	4
a_2	0,4	0,4	0,4	0,4	0,6
a_6	0,3	0,3	0,3	0,3	0,4
a_1	0,1	0,1	0,2	0,3	
a_4	0,1	0,1	0,1		
a_3	0,06	0,1			
a_5	0,04				

- Na 1a redução, os 2 símbolos de menor probabilidade são unidos, formando um "símbolo composto" com $P = 0,06 + 0,04 = 0,1$. Este símbolo e sua respectiva probabilidade são posicionados na coluna correspondente à 1a redução de forma que todos os valores da coluna estejam em ordem decrescente.

Introdução

2. Consiste em codificar cada fonte reduzida, iniciando pela menor fonte e caminhando em direção à fonte original.

Fonte Original			Reduções de fonte							
Símbolo	Prob.	Código	1		2		3		4	
a_2	0,4	1	0,4	1	0,4	1	0,4	1	0,6	0
a_6	0,3	00	0,3	00	0,3	00	0,3	00	0,4	1
a_1	0,1	011	0,1	011	0,2	010	0,3	01		
a_4	0,1	0100	0,1	0100	0,1	011				
a_3	0,06	01010	0,1	0101						
a_5	0,04	01011								

- O símbolo com $P = 0,6$ foi gerado a partir da junção de dois outros símbolos na fonte reduzida à sua esquerda, o 0 usado para codificá-lo é agora atribuído a ambos os símbolos que lhe deram origem, colocando-se um 0 ou 1 à direita de cada um para distingui-los.

Introdução

- Note que cada símbolo é codificado individualmente, ou seja, a codificação ocorre para um símbolo de cada vez;

- Os dois símbolos com menor probabilidade de ocorrência têm palavras código com o mesmo comprimento, diferindo apenas no *bit* menos significativo.

Introdução

- Quando a codificação tem de ser aplicada a um elevado número de símbolos, a construção de um código Huffman pode tornar-se uma tarefa relativamente complexa do ponto de vista computacional.
  - Além disso, aos símbolos menos prováveis poderão ser atribuídas palavras código muito longas.
  - Neste caso, é preferível sacrificar a eficiência da codificação de modo a reduzir a complexidade computacional.
- Para o caso geral de K símbolos, são necessárias:
  - $K-2$ reduções de fonte
  - $K-2$ atribuições de código

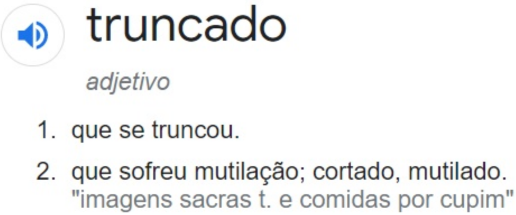
- Uma possível modificação sobre o código de Huffman original consiste em se codificar somente os M símbolos mais prováveis, dentre os K símbolos da fonte (com $M < K$).

- Para os demais símbolos, utiliza-se uma palavra código de prefixo seguida de um código de comprimento fixo adequado.

- Esta modificação do algoritmo original de codificação por Huffman é denominada **código de Huffman truncado**.

Huffman Truncado

- O código de Huffman truncado, como o nome sugere, é uma variação da codificação de Huffman tradicional.



Huffman Truncado

- No código de Huffman truncado, os primeiros M símbolos mais prováveis da fonte (de todos os K símbolos) unidos com um símbolo hipotético cuja probabilidade é igual à soma das probabilidades dos $K - M$ símbolos da fonte menos prováveis são codificados com o código de Huffman tradicional.

- Os $K - M$ símbolos menos prováveis são asscoiados ao código dado ao símbolo hipotético concatenado com o código binário natural de comprimento $\log_2(K - M)$.

  - Obs.: O $\log_2(K - M)$ permite saber a quantidade de *bits* necessária para representar os $K - M$ símbolos.

Huffman Truncado

S	$P(S = A_i)$
A0	0,3
A1	0,2
A2	0,15
A3	0,1
A4	0,08
A5	0,06
A6	0,05
A7	0,04
A8	0,02

⇨

S	$P(S = A_i)$
A0	0,3
A1	0,2
A2	0,15
A3	0,1
A4	0,08
A5	0,06
A6	0,05
A7	0,04
A8	0,02

⇨

S	$P(S = A_i)$
AX	0,35
A0	0,3
A1	0,2
A2	0,15
A3	0,1
A4	0,08
A5	0,06
A6	0,05
A7	0,04
A8	0,02

K = 3 + 1 hipotético = 4
(Huffman tradicional)

código dado ao símbolo hipotético concatenado com os binários de comprimento $\log_2(9-3) = \log_2(6) = 3$ *bits*.

Huffman Truncado

- A constante $M < K$ pode ser escolhida arbitrariamente.

- A codificação de Huffman truncado faz $M - 1$ reduções de fonte, levando menos tempo, pelo custo de maior comprimento médio de código e menos eficiência.

Huffman Truncado

- Passo a passo para o Código de Huffman truncado:
  1. Reordene os K símbolos da fonte em ordem decrescente de probabilidade;
  2. Divida o número total de símbolos em dois grupos: GRUPO 1$= M$ mais prováveis e GRUPO 2$= K - M$ menos prováveis;
  3. Adicione um símbolo hipotético ao GRUPO 1, cuja probabilidade é a soma dos símbolos do GRUPO 2;
  4. Reordene os símbolos do GRUPO 1 em ordem decrescente;
  5. Codifique os símbolos do GRUPO 1 usando a codificação de Huffman tradicional;
  6. Codifique os símbolos do GRUPO 2 associado-os ao código dado ao símbolo hipotético concatenado com os binários naturais de comprimento $\log_2(K - M)$.

Huffman Truncado

- **EXEMPLO 1:** Construa um código de Huffman truncado para a fonte de informação com $K = 9$ símbolos cujo alfabeto e a respectiva distribuição de probabilidade são apresentados na tabela abaixo.

Fonte de Informação	Probabilidade (P_i)
A0	0,3
A1	0,2
A2	0,15
A3	0,1
A4	0,08
A5	0,06
A6	0,05
A7	0,04
A8	0,02

- **EXEMPLO 1-SOLUÇÃO**
  1. Reordene os $K = 9$ símbolos da fonte em ordem decrescente de probabilidade;

Fonte de Informação	Probabilidade (P_i)
A0	0,3
A1	0,2
A2	0,15
A3	0,1
A4	0,08
A5	0,06
A6	0,05
A7	0,04
A8	0,02

- **EXEMPLO 1-SOLUÇÃO**
  2. Divida o número total de símbolos em dois grupos: GRUPO 1$= M = 3$ mais prováveis e GRUPO 2$= K - M = 9 - 3 = 6$ menos prováveis;

Fonte de Informação	Probabilidade (P_i)
A0	0,3
A1	0,2
A2	0,15
A3	0,1
A4	0,08
A5	0,06
A6	0,05
A7	0,04
A8	0,02

- **EXEMPLO 1-SOLUÇÃO**
  3. Adicione um símbolo hipotético ao GRUPO 1, cuja probabilidade é a soma dos símbolos do GRUPO 2;
  4. Reordene os símbolos do GRUPO 1 em ordem decrescente;

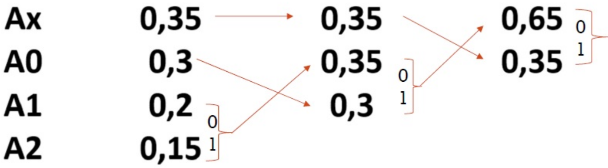
Fonte de Informação	Probabilidade (P_i)
Ax	**0,35**
A0	0,3
A1	0,2
A2	0,15

Soma das probabilidades dos símbolos menos prováveis!

Huffman Truncado

- **EXEMPLO 1-SOLUÇÃO**
  5. Codifique os símbolos do GRUPO 1 usando a codificação de Huffman tradicional;

Símbolo	Código
Ax	1
A0	01
A1	000
A2	001



- **EXEMPLO 1-SOLUÇÃO**
  6. Codifique os símbolos do GRUPO 2 associado-os ao código dado ao símbolo hipotético concatenado com os binários naturais de comprimento $\log_2(9-3) = \log_2(6) = 3$.

Fonte de Informação	Probabilidade (P_i)	Huffman Truncado
A0	0,3	01
A1	0,2	000
A2	0,15	001
A3	0,1	1000
A4	0,08	1001
A5	0,06	1010
A6	0,05	1011
A7	0,04	1100
A8	0,02	1101

- **EXEMPLO 1-SOLUÇÃO**
  - O código encontrado tem uma eficiência de:

$$L = \sum_{i=0}^{8} P(A_i) l_{A_i} = 3,05 \ bits/\text{símbolo}$$

$$H = \sum_{i=0}^{8} P(A_i) \log_2 \left[\frac{1}{P(A_i)} \right] = 2,778 \ bits/\text{símbolo}$$

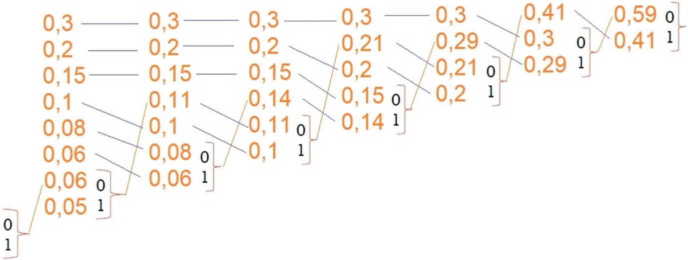
Logo:

$$\eta = \frac{H}{L} = \frac{2,778}{3,05} = 91,08\%$$

Huffman Truncado

- Por curiosidade, vamos construir o código de Huffman tradicional para a fonte do Exemplo 1, e vamos comparar a eficiência obtida com a do código de Huffman truncado.

00	A0	0,3
11	A1	0,2
010	A2	0,15
101	A3	0,1
0110	A4	0,08
0111	A5	0,06
1001	A6	0,05
10000	A7	0,04
10001	A8	0,02



Huffman Truncado

- Por curiosidade, vamos construir o código de Huffman tradicional para a fonte do Exemplo 1, e vamos comparar a eficiência obtida com a do código de Huffman truncado.
  - O código encontrado tem uma eficiência de:

$$L = \sum_{i=0}^{8} P(A_i) l_{A_i} = 2,81 \ bits/\text{símbolo}$$

Logo:

$$\eta = \frac{H}{L} = \frac{2,778}{2,81} = 98,86\%$$

Huffman Truncado

- Por curiosidade, vamos construir o código de Huffman tradicional para a fonte do Exemplo 1, e vamos comparar a eficiência obtida com a do código de Huffman truncado.
  - Comparando as eficiências, temos:
    1. Huffman truncado: $\eta = 91,08\%$
    2. Huffman tradicional: $\eta = 98,86\%$
  - Sacrificamos a eficiência pela redução da complexidade computacional;
  - Ou seja, o tempo é reduzido conforme a necessidade de apenas $4 - 2 = 2$ estágios de redução de fonte, enquanto o Huffman tradicional precisa de $9 - 2 = 7$ estágios de redução.

- **Exercício 1:** A fonte de informação F gera os símbolos mostrados na tabela abaixo. Codifique os símbolos utilizando o codificador de Huffman e o codificador de Huffman truncado para $M = 3$, $M = 4$ e $M = 5$. Compare as eficiências dos códigos obtidos.

F	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7
$P(f_i)$	0,35	0,24	0,16	0,1	0,1	0,02	0,02	0,01

Solução (Exercício 1):

VER MATERIAL EM ANEXO!

Huffman Truncado

- **Exercício 2:** Defina, independente da distribuição de probabilidade do símbolos, para que valor de M a codificação de Huffman truncado sempre será equivalente à codificação de Huffman tradicional.

Solução (Exercício 2):

VER MATERIAL EM ANEXO!

Huffman Truncado

- **Exercício 3:** Construa o código de Huffman truncado para o valor de M definido no Exercício 2 (valor para o qual a codificação de Huffman truncado é equivalente à codificação de Huffman tradicional) para a fonte do Exercício 1.

Solução (Exercício 3):

VER MATERIAL EM ANEXO!