

Código de Huffman

Ewerton Carlos de Araújo Assis

26 de Junho de 2013

Resumo

O presente trabalho, que comporá nota parcial na disciplina de Análise e Projeto de Algoritmos, ministrada pelo professor Leonardo Alves, no Instituto de Informática da Universidade Federal de Goiás, tem como objetivo apresentar o método de compressão de dados Código de Huffman.

1 Introdução

Compressão de dados é uma disciplina dentro da ciência da computação e da teoria da informação que tem por finalidade, a partir de uma sequência de dados original, obter uma nova sequência de dados, menor ou igual à sequência original, que mantenha a mesma integridade das informações armazenadas nessa sequência de dados. Dentre os vários mecanismos de compressão de dados, existem duas possibilidades de distinção entre esses mecanismos: compressão com perda e compressão sem perda.

O Código de Huffman, publicado em artigo em 1952 sob o título “A Method for the Construction of Minimum-Redundancy Codes”, por David A. Huffman, é um método de compressão sem perda de dados, garantindo que a informação obtida após a descompressão é idêntica à informação original, antes de ser comprimida.

2 Funcionamento do código de Huffman

A codificação é feita a partir da probabilidade de ocorrência de símbolos no conjunto de dados a ser comprimido, com o fim de determinar códigos de tamanho variável para cada símbolo. Uma árvore binária completa, denominada árvore de Huffman, é construída recursivamente a partir da junção dos dois símbolos de menor probabilidade, que são posteriormente somados em símbolos auxiliares e recolocados no conjunto de símbolos. O processo termina quando todos os símbolos forem unidos em símbolos auxiliares, formando assim uma árvore binária completa. A árvore é por fim percorrida, atribuindo-se valores binários (0 ou 1) para cada aresta, auxiliando na geração de código ao longo desse percurso.

Na árvore binária construída, as folhas representam os símbolos presentes na sequência de dados, associados às suas respectivas probabilidades de ocorrência; os nós intermediários representam a soma das probabilidades de ocorrência presentes em suas ramificações; e a raiz representa a soma da probabilidade de todos os símbolos no conjunto. O algoritmo para a construção da árvore é apresentado no algoritmo 1.

Algorithm 1: Código de Huffman: construção da árvore binária

Assuma: $\Sigma : \{ \text{Alfabeto de símbolos de entrada, com probabilidades associadas} \}$

```

begin
  while  $|\Sigma| > 1$  do
     $S_0 \leftarrow$  Retira menor probabilidade do alfabeto  $\Sigma$ 
     $S_1 \leftarrow$  Retira menor probabilidade do alfabeto  $\Sigma$ 
     $X \leftarrow$  Novo nó
     $X.\text{filho}_0 \leftarrow S_0$ 
     $X.\text{filho}_1 \leftarrow S_1$ 
     $X.\text{probabilidade} \leftarrow S_0.\text{probabilidade} + S_1.\text{probabilidade}$ 
    Insere a tupla  $(\Sigma, X)$  na árvore  $\bar{X}$ 
  end
   $X \leftarrow$  Retira menor probabilidade do alfabeto  $\Sigma$ 
  for all  $f \in F(\bar{X}) : \{ \text{Conjunto de folhas em } \bar{X} \}$  do
    código[ $f$ ]  $\leftarrow$  Percorre raiz até a folha  $f$ 
  end
end

```

3 Análise do código de Huffman

A codificação tem desempenho ótimo quando as probabilidades dos símbolos são potências negativas de dois ($2^{-1}, 2^{-2}, \dots$), além de garantir que a nova sequência de dados gerada não é ambígua, pois nenhum código poder ser o prefixo de outro código.

O algoritmo para construção da árvore binária, apresentado no algoritmo 1, tem complexidade de tempo de ordem $\mathcal{O}(n \log n)$.

4 Conclusão

O Código de Huffman é um método relativamente simples de compressão de dados, que usa estratégia guloso para comprimir uma sequência de dados. A vantagem mas evidente é por tratar-se de um método sem perda de dados, mantendo a integridade das informações comprimidas.

Referências

- [1] CORMEN, T. H.; LEISERSON, C. E.; RIVEST, R. L.; STEIN, C. **Introduction to Algorithms**. The MIT Press, 2nd edition, 2001.
- [2] WIKIPÉDIA, A. E. L. **Codificação de huffman**. http://pt.wikipedia.org/wiki/Codifica%C3%A7%C3%A3o_de_Huffman, Junho 2013.
- [3] WIKIPEDIA, T. F. E. **Huffman coding**. http://en.wikipedia.org/wiki/Huffman_coding, Junho 2013.