

MODELAGEM PARA MÉTODOS ESTATÍSTICOS DE COMPRESSÃO DE DADOS

ROGÉRIO T. BRITO
JOSÉ A. R. SOARES (ORIENTADOR)

1. INTRODUÇÃO

O processo de compressão de dados pode ser observado como o resultado de dois processos, a *modelagem* e a *codificação*. A modelagem é a parte que coleta e analisa informações da entrada, enquanto a codificação é a parte que gera a saída, baseando-se nos dados obtidos pela modelagem.

Dentre os dois tipos básicos de compressão de dados existentes, *compressão baseada em dicionários* e *compressão estatística*, a divisão em etapas fica mais evidente nos métodos estatísticos. Embora os métodos baseados em dicionário sejam muito rápidos e, por esse motivo, sejam amplamente divulgados¹, eles não atingem taxas de compressão tão boas quanto os métodos estatísticos.

Nosso estudo baseia-se, pelo motivo exposto acima, em métodos estatísticos. Mais especificamente, nosso trabalho concentra-se em técnicas de modelagem para codificação aritmética.

2. MODELAGEM

Nos métodos estatísticos, a modelagem consiste em compilar uma tabela contendo frequências de cada símbolo da entrada. O codificador toma estas probabilidades e “transforma-as” em bits na saída.

Para obter maior eficiência na compressão, é usual, na construção do modelo, analisarmos as probabilidades de seqüências de caracteres, não só de símbolos isolados. Modelos construídos desta maneira são chamados modelos de contexto finito de ordem n , em que analisam-se as probabilidades de seqüências de $n + 1$ caracteres da entrada (os n primeiros caracteres constituem o *contexto* do $(n + 1)$ -ésimo caracter).

Um cuidado especial que deve ser tomado durante a construção de modelos é que, para o correto funcionamento dos codificadores, nenhuma seqüência de caracteres pode ter probabilidade igual a zero em nosso modelo. Esta situação indesejável é chamada *Problema da Probabilidade Zero*. Para evitar este problema, uma técnica que apresenta bons resultados na prática é chamada *blending* e consiste em colocar, em um mesmo modelo, modelos de ordens $-1, 0, 1, \dots, n$, para n fixado.

Neste trabalho, estamos implementando um gerador de modelos blended para serem usados com codificadores aritméticos. O codificador aritmético que estamos usando é implementado por [1].

REFERÊNCIAS

- [1] J. G. Cleary T. C. Bell and I. H. Witten. *Text Compression*. Prentice-Hall Inc., 1990.

¹Basicamente, todas as implementações de compressores comerciais (e.g., PKZIP, ARJ) são implementações de métodos baseados em dicionário.