Caio Magno Aguiar de Carvalho

Estudo e Desenvolvimento de Algoritmos de Compressão Sem Perda sobre Dados Uniformemente Distribuídos

São Luís, Maranhão, Brasil

08/2022

Caio Magno Aguiar de Carvalho

Estudo e Desenvolvimento de Algoritmos de Compressão Sem Perda sobre Dados Uniformemente Distribuídos

Tese de Doutorado submetida ao Programa de Pós-Graduação em Engenharia de Eletricidade da Universidade Federal do Maranhão como parte dos requisitos necessários à obtenção do grau de Doutor em Engenharia Elétrica.

Universidade Federal do Maranhão - UFMA

Centro de Ciências Exatas e Tecnologia

Programa de Pós-Graduação em Engenharia Elétrica

Orientador: Allan Kardec Barros Filho

São Luís, Maranhão, Brasil

08/2022

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

de Carvalho, Caio Magno Aguiar.
Estudo e Desenvolvimento de Algoritmos de Compressão Sem Perda sobre Dados Uniformemente Distribuídos / Caio Magno Aguiar de Carvalho. - 2022.
80 f.

Orientador(a): Allan Kardec Duailibe Barros Filho.
Tese (Doutorado) - Programa de Pós-graduação em Engenharia Elétrica/ccet, Universidade Federal do Maranhão, São Luís, Maranhão, 2022.

1. Compressão de dados. 2. Concatenação. 3. Dados uniformes. 4. Teoria da informação. I. Barros Filho, Allan Kardec Duailibe. II. Título.

Caio Magno Aguiar de Carvalho

Estudo e Desenvolvimento de Algoritmos de Compressão Sem Perda sobre Dados Uniformemente Distribuídos

Tese de Doutorado submetida ao Programa de Pós-Graduação em Engenharia de Eletricidade da Universidade Federal do Maranhão como parte dos requisitos necessários à obtenção do grau de Doutor em Engenharia Elétrica.

Trabalho aprovado. São Luís, Maranhão, Brasil, 16/09/2022:

Allan Kardec Barros Filho
Orientador

Ewaldo Éder Carvalho Santana
Convidado 1

Francisco da Chagas de Souza
Convidado 2

Hugo Valadares Siqueira
Convidado 3

São Luís, Maranhão, Brasil
08/2022

Dedico esse trabalho primeiramente ao Deus do Céu, que "dá sabedoria e ciência"(Dn 2.21), à minha esposa Aline Kessia e meu filho Raul que são o tempero e o consolo da minha vida, à minha mãe Ayla Aguiar e meu pai Hermes Alberto (in memorian), pela educação que não se pode aprender na escola.

Agradecimentos

Direciono meus agradecimentos primeiramente ao professor Allan Kardec, que me propôs tal desafio, por sua disposição e paixão pelo trabalho. Sua dedicação certamente será um ideal a ser alcançado em minha carreira.

Aos amigos do laboratório de Processamento da Informação Biológica (PIB): Jonathan Queiroz, Letícia Correia, Marta Barreiros, Felipe Gomes, George Vagner e Daniel Luna. Obrigado pelo apoio manifestado em críticas, contribuições, ideias e revisões.

"Quem encerrou o mar com portas (...) e disse: Até aqui virás, e não mais adiante(...)?"Jó 38: 8, 11

Resumo

A alta produção e consumo de informação digital em ritmo cada vez mais acelerado não acompanha as atuais ofertas de armazenamento e transmissão de dados, ou seja, produzimos mais conteúdo digital do que podemos armazenar e comunicar, e essa corrida aparentemente não será equilibrada facilmente. As técnicas de compressão de dados foram desenvolvidas afim de otimizar os mecanismos de armazenamento e comunicação de forma que a informação ocupe o mínimo de espaço em um sistema de gerenciamento de arquivos ou o mínimo de largura de banda em um canal de comunicação. Tais técnicas estão baseadas na Teoria da Informação proposta por Shannon, nas quais as estatísticas do sinal a ser comprimido desempenham papel fundamental na representação eficiente da informação. Repetição e estrutura são características fundamentalmente exploradas por algoritmos de compressão. Entretanto sequências de dados uniformemente distribuídos, independentes e identicamente distribuídos (i.i.d) rompem esses dois pilares que fundamentam a compressão estatística. É sabido também que idealmente a saída codificada de um algoritmo de compressão é uniformemente distribuída, portanto, estudar a possibilidade de compressão de distribuições uniformes é abrir a possibilidade de compressão recursiva. O presente trabalho tem como objetivo explorar essa possibilidade através da observação do problema da compressão fora do campo estatístico, mas a partir da redundância inerente da codificação binária padrão, proposta pelo algoritmo da Concatenação e da perspectiva geométrica através do método SVD-esfera-espiral. O algoritmo da Concatenação aproveita as frações de bits não utilizadas na representação binária padrão, tendo o seu desempenho máximo quando o tamanho do alfabeto dos dados comprimidos é $2^N + 1$. Os experimentos foram conduzidos sobre os dados da RAND Corporation, os quais são dados uniformes produzidos por processos físicos com alfabeto de tamanho 10. Os resultados mostraram que é possível obter até 12,5% de compressão sobre esse conjunto.

Palavras-chaves: compressão de dados. Dados uniformes. teoria da informação. concatenação.

Lista de ilustrações

Lista de tabelas

Sumário