Abstract

Arithmetic coding is used in many compression techniques during the entropy encoding stage. Further compression is not possible without changing the data model and increasing redundancy in the data set. To increase the redundancy, we have applied index based byte-pair transformation (BPT-I) as a pre-processing to arithmetic coding. BPT-I transforms most frequent byte-pairs (2-byte integers). Here, most frequent byte-pairs are sorted in the order of their frequency and groups consisting of 256 byte-pairs are formed. Each byte-pair in a group is then encoded using two tokens: group number and the location in a group. Group number is denoted using variable length prefix codeword; whereas location within a group is denoted using 8-bit index. BPT-I is designed to be applied on any type of source; not necessarily text. More the number of groups considered during transformation, better is the compression. Experimental results have shown around 4.30% additional reduction in compressed file size when arithmetic coding is applied after byte-pair data transformation BPT-I.


Senthil S, Robert L. "IIDBE: A Lossless Text Transform for Better
Achieving Better Compression Applying Index-based Byte-Pair Transformation before Arithmetic Coding


Index Terms

Computer Science
Circuits And Systems

Keywords

Data Compression  better compression rate  index based byte-pair data transformation
a pre-processing to arithmetic coding

arithmetic coding