Universal Data Compression Algorithms
by Using Full Tree Models

Soo-Young Chang and John J. Metzner

The Pennsylvania State University
Department of Electrical and Computer Engineering
University Park, PA 16802

Universal data compression algorithms are algorithms by which source data can be coded without a priori knowledge of source statistics. The objective of this work is to devise universal data compression algorithms by using full tree models. A brief description and results of analysis of splitting and merging for constructing and updating full trees is provided. Since all the leaves are assumed to have equal probability, by using these two tree-updating procedures the variable-to-fixed length codes are realized.

Two algorithms are suggested. The first one is an algorithm which is devised by modifying the Ziv-Lempel algorithm. In this algorithm the number of leaves varies but is always a power of the size of an encoded output alphabet. In the second algorithm trees always have a fixed number of leaves which is the power of the size of the encoded output alphabet. In this paper universal source models for these two algorithms are described by using full tree models. Also, implementations are proposed.

The two proposed algorithms are evaluated by three parameters - the normalized average code length per source output symbol, required memory space to store information about trees, and encoding complexity. Performances of the two proposed algorithms are compared with those of the Ziv-Lempel algorithm and the Rissanen algorithm. The algorithms proposed in this paper are more efficient than the Z-L algorithm and the Rissanen algorithm.