

Data Compression Theory

Abstract

Data compression technology is necessary in today’s multimedia society. Using this technology, information can be transmitted in a shorter amount of time and the storage space can be made smaller. Since data compression in broad terms means “expressing things concisely,” it is applied to many fields, such as learning and inference, and so on. We are researching theoretical studies on limits and algorithms for data compression systems. Our main results are as follows.

The complexity-distortion theory and the semifaitful data compression algorithm

It is impossible to transmit continuous information, such as sound and images, faithfully on digital communication systems. Semifaitful data compression is a practical solution to this problem, where we permit distortion between the original data and the reproduced data. Since many faithful data compression algorithms are available, we study the problem of whether these algorithms are applicable to semifaitful ones. We first focus on the common properties of optimal faithful data compression algorithms to define the notion of the complexity. Next, we define the notion of the complexity at a distortion level and prove a basic theorem on its property [1][3]. The construction of our semifaitful data compression algorithm involves the following two steps: firstly, a string is chosen within a fixed distortion to minimize the complexity, and secondly, the distorted string is compressed by a faithful data compression algorithm (Fig. 1). So far, we have needed specific complexities and faithful data compression algorithms to achieve the minimum coding rate. Our theory guarantees that we can use any complexity and faithful data compression algorithm to construct our semifaitful data compression algorithm with the minimum coding rate (Table 1).

Data compression with common side information

We consider the coding problems of a source when the encoder and the decoder have access to a common source which is correlated with the target source. This framework is called ‘data compression with common side information.’ The asymptotically optimal coding rate has been known, but not a concrete algorithm. We construct a concrete algorithm for faithful data compression with common side information, and then we apply the complexity-distortion theory to construct a semifaitful data compression algorithm [2][3].

(Contact: Jun Muramatsu; Email: pure@cslab.kecl.ntt.co.jp)

- [1] Muramatsu, J. and Kanaya, F.: Distortion-complexity and rate-distortion function, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E77-A, No. 8, pp. 1224–1229 (1994).
- [2] Muramatsu, J. and Kanaya, F.: Universal data compression with common side information, *Proc. of the 1997 IEEE International Symposium on Information Theory*, p. 183 (1997).
- [3] Muramatsu, J. : Universal data compression algorithms for stationary ergodic sources based on the complexity of sequences, Ph. D. thesis, Nagoya University (1998).

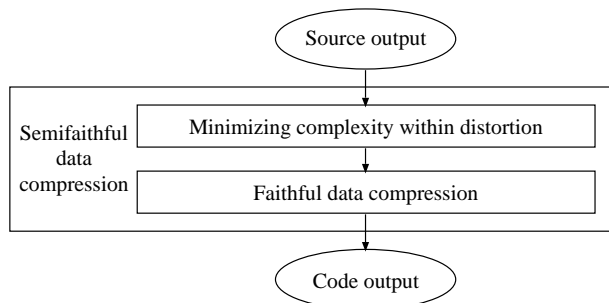


Table 1: Complexities and faithful data compression which can be used to construct a semifaitful data compression algorithm

	Old method	Our method
Complexities (faithful data compression)	Kolmogorov complexity (program length) Lempel-Ziv code (incremental parsing)	Any (clarifying condition)

Fig. 1: Our construction of a semifaitful data compression algorithm