Available online at www.sciencedirect.com
ScienceDirect
Communications in Nonlinear Science and Numerical Simulation 14 (2009) 1013–1020
Communications in Nonlinear Science and Numerical Simulation
www.elsevier.com/locate/cnsns

Short communication

Arithmetic coding as a non-linear dynamical system

Nithin Nagaraj *, Prabhakar G. Vaidya, Kishor G. Bhat

School of Natural Sciences and Engineering, National Institute of Advanced Studies, Indian Institute of Science Campus, Bangalore 560 012, India

Received 27 June 2007; received in revised form 29 October 2007; accepted 3 December 2007
Available online 8 December 2007

Abstract

In order to perform source coding (data compression), we treat messages emitted by independent and identically distributed sources as imprecise measurements (symbolic sequence) of a chaotic, ergodic, Lebesgue measure preserving, non-linear dynamical system known as Generalized Luröth Series (GLS). GLS achieves Shannon's entropy bound and turns out to be a generalization of arithmetic coding, a popular source coding algorithm, used in international compression standards such as JPEG2000 and H.264. We further generalize GLS to piecewise non-linear maps (Skewed-nGLS). We motivate the use of Skewed-nGLS as a framework for joint source coding and encryption.

PACS: 05.45.Vx; 05.45.Ac

Keywords: Data compression; Encryption; Chaotic map; Arithmetic coding; Source coding

1. Introduction

Chaos theory has shown much promise for communications in recent years, especially to cryptography owing to desirable properties of non-linear dynamical systems such as ergodicity, sensitive dependence on initial conditions and good pseudo-random properties in spite of low algorithmic complexity [1]. However, applications to source coding (data compression) and joint source coding and encryption have received very little or no attention.

The source coding problem is stated as follows: given an independent and identically distributed (i.i.d or discrete memoryless source) binary source X emitting bits of information in the absence of noise, how do we obtain the shortest lossless representation of this information? Source coding is also known as entropy coding or data compression and is an important part of most communication systems [2]. Multimedia contents used in storage (compact discs) and in network transmissions (entertainment, tele-medicine, etc.) need both

* Corresponding author.
E-mail addresses: nithin_nagaraj@yahoo.com (N. Nagaraj), pgvaidya@yahoo.com (P.G. Vaidya), kishor@nias.iisc.ernet.in (K.G. Bhat).
URL: http://nithin.nagaraj.googlepages.com (N. Nagaraj).

1007-5704/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.cnsns.2007.12.001

compression and encryption. The idea of joint source coding and encryption is very recent [3] where the goal is to simultaneously achieve both compression and encryption and thus avoid a separate external encryption block. This is highly beneficial for scenarios where computational resources are limited (for e.g. wireless networks used in deep space communications).

Shannon in his 1948 masterpiece [4] defined the most important concept of Information Theory, namely 'Entropy'. Shannon's Entropy of a discrete memoryless source $H(X)$ is defined as the amount of information content or the amount of uncertainty associated with the source, or equivalently the least number of bits required to represent the information content of a source without any loss. Shannon proposed a method (Shannon–Fano coding [5]) that achieves this limit as the block-length (number of symbols taken together) for coding increases asymptotically to infinity. Huffman [6] proposed what are called minimum-redundancy codes with integer code-word lengths that achieve Shannon's entropy in the limit of the block-length tending to infinity. Huffman codes are quite popular and are also used in the old international standard for image compression (JPEG). However, there are problems associated with both Shannon–Fano coding and Huffman coding (and other similar techniques). As the block-length increases, the number of alphabets exponentially increase, thereby increasing the memory needed for storing and handling. Also, the complexity of the encoding algorithm increases since these methods build code-words for all possible messages for a given length instead of designing the codes for a particular message at hand. Another disadvantage of all such methods is that they do not lend themselves easily to an adaptive coding technique. For non-stationary sources, the idea of adaptive coding is to update the probability model of the source during the coding process. Unfortunately, for both Huffman and Shannon–Fano coding, the updating of the probability model would result in re-computation of the code-words for all the symbols which is a computationally expensive process.

In this paper, we address the source coding problem from a dynamical systems perspective. We model the information bits of the source X as measurements of a non-linear dynamical system. Since measurement is rarely accurate, we treat these measured bits of information as a symbolic sequence of the Tent map [7] and their skewed cousins (measurement of any kind can be thought of as a symbolic sequence [8]). Source coding is then seen as determination of the initial condition that has generated the given symbolic sequence. Subsequently, we establish that such an approach leads us to a well known entropy coding technique (arithmetic coding) which is optimal for compression. Furthermore, this new approach enables a robust framework for joint source coding and encryption.

2. Entropy coding using the chaotic Tent map

We shall now employ the Tent map to encode a message. Consider the message $M = 'AABABBABAA'$ of length $N = 10$ bits (from an i.i.d binary source X). We consider two Markov partitions – the one pertaining to

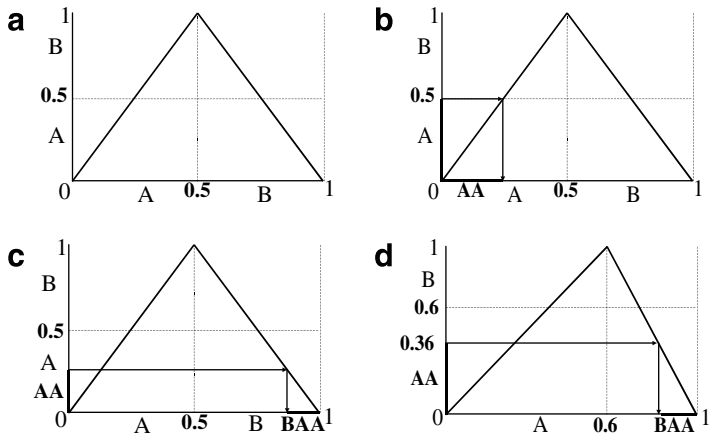

Fig. 1. (a–c): source coding using the Tent map. (d) Skewed-Tent map achieves Shannon's optimality.