Identification via Compressed Data

Rudolf Ahlswede, * Enhui Yang, † and Zhen Zhang‡

July 7, 1998

*Rudolf Ahlswede is with Fakultät für Mathematik, Bielefeld Universität, Bielefeld, Germany, Postfach 100131

†En-hui Yang is with Dept. of Mathematics, Nankai University, Tianjin 300071, P.R. China. He is currently visiting Communication Sciences Institute, Dept. of EE-Systems, University of Southern California, Los Angeles, CA 90089-2565.

‡Zhen Zhang is with Communication Sciences Institute, Electrical Engineering–Systems, University of Southern California, Los Angeles, CA. 90089-2565. Research for this author is supported in part by NSF Grant NCR-9205265.

Abstract

We introduce and analyze a new coding problem for a correlated source $(X^n, Y^n)_{n=1}^{\infty}$. The observer of X^n can transmit data depending on X^n at a prescribed rate R. Based on these data the observer of Y^n tries to identify whether for some distortion measure ρ (like the Hamming distance) $\frac{1}{n}\rho(X^n, Y^n) \leq d$, a prescribed fidelity criterion. We investigate as functions of R and d the exponents of two error probabilities, the probabilities for misacceptance and the probabilities for misrejection. Our analysis has led to a new method for proving converses. Its basis is "The Inherently Typical Subset Lemma". It goes considerably beyond the "Entropy Characterisation" of [2], the "Image Size Characterisation" of [3], and its extensions in [5]. It is conceivable that it has a strong impact on Multi-user Information Theory.

Key words:
Correlated source, identification with fidelity, misacceptance and misrejection error probabilities.

I Introduction and Formulation of Problem

A Introduction

In this paper, we consider a new model: identification via compressed data. To put it in perspective, let us first review the traditional problems in source coding theory. Consider the following diagram, where $\{X_n\}_{n=1}^{\infty}$ is an i.i.d. source taking values in a finite alphabet $\mathcal{X}$. The encoder output is a binary sequence which appears at a rate of R bits per symbol. The decoder output is a sequence $\{\hat{X}_n\}_{n=1}^{\infty}$ which takes values in a finite reproduction alphabet $\mathcal{Y}$. In traditional source coding theory, the decoder is required to recover $\{X_n\}_{n=1}^{\infty}$ either completely or with some allowable distortion. That is, the output sequence $\{\hat{X}_n\}_{n=1}^{\infty}$ of the decoder must satisfy

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{E}\rho(X_i, \hat{X}_i) \leq d \,, \tag{1}$$

for sufficiently large n, where $\mathbf{E}$ devotes the expected value,

$$\rho : \mathcal{X} \times \mathcal{Y} \to [0, +\infty) \tag{2}$$

is a distortion measure, and d is the allowable distortion between the source sequence and the reproduction sequence. The problem is then to determine the infimum of the rate R such