Universal Coding for Arbitrarily Varying Sources

Meir Feder and Neri Merhav

Department of Electrical Engineering - Systems, Tel Aviv University, Tel Aviv, 69978, ISRAEL
Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, ISRAEL

Abstract — **The minimum universal coding redundancy for finite-state arbitrarily varying sources, is investigated. If the space of all possible underlying state sequences is partitioned into types, then this minimum can be essentially lower bounded by the sum of two terms. The first is the minimum redundancy within the type class and the second is the minimum redundancy associated with a class of sources that can be thought of as "representatives" of the different types. While the first term is attributed to the cost of uncertainty within the type, the second term corresponds to the type itself. The bound is achievable by a Shannon code w.r.t an appropriate two-stage mixture of all arbitrarily varying sources in the class.**

We investigate the minimum attainable redundancy in universal coding for arbitrarily varying sources (AVS's). An AVS is a nonstationary memoryless source characterized by the probability mass function (PMF),

$$P(\mathbf{x}|\mathbf{s}) = \prod_{i=1}^{n} p(x_i|s_i), \tag{1}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ is an observed data sequence to be encoded, x_i taking on values in a finite set $\mathcal{X}$, and $\mathbf{s} = (s_1, \ldots, s_n)$ is an unknown arbitrary sequence of states corresponding to $\mathbf{x}$, where each s_i takes on values in a set $\mathcal{S}$. We shall assume, for the sake of simplicity, that the parameters of the AVS $\{p(x|s)\}_{x\in\mathcal{X}, s\in\mathcal{S}}$ are known.

The problem of universal coding for AVS's has relatively received only little attention. Berger [1, Sect. 6.1.2] and Csiszár and Körner [2, Theorem 4.3] have characterized the best attainable rate-distortion tradeoff for block-to-block (BB) codes where the average distortion is required to be within a prescribed level D for the *worst* possible state sequence. For the distortionless case ($D = 0$) the best attainable rate in this sense is given by the entropy of the worst memoryless source in the convex closure of $\{p(\cdot|s), s \in \mathcal{S}\}$, that is the maximum entropy attained among all mixtures $m(x) = \int_{\mathcal{S}} w(ds)p(x|s)$, w being a probability measure on $\mathcal{S}$. The reason for this worst case result is that both the rate is held fixed at each block and the distortion constraint must be met for every possible state sequence.

We show that one can improve upon this pessimistic result if variable-rate codes are allowed because then there is a potential freedom to "adapt" the rate to the underlying state sequence in some sense. Specifically, we show that for finite-state AVS's there exists lossless a block-to-variable (BV) code whose compression ratio is essentially the entropy of the memoryless source $m_{\mathbf{s}}(x) = \sum_{s\in\mathcal{S}} w_{\mathbf{s}}(s)p(x|s)$, where $w_{\mathbf{s}}(s)$ is the empirical probability (i.e., relative frequency) of $s \in \mathcal{S}$ along the underlying state sequence $\mathbf{s}$. This entropy is of course never larger than the maximum entropy mentioned above. It is therefore easy to see that the redundancy, namely, the excess rate beyond the per-letter entropy of the AVS given $\mathbf{s}$, is essentially equal to the mutual information $I_{w_{\mathbf{s}}}(S;X)$ associated with the joint PMF $w_{\mathbf{s}}(s)p(x|s)$. This quantity in turn agrees with that of [1] and [2] only if $\mathbf{s}$ maximizes the entropy.

Furthermore, $I_{w_{\mathbf{s}}}(S;X)$ is essentially a lower bound on the redundancy in a fairly strong sense. If we consider the set of all state sequences of a certain type class (i.e., the same empirical PMF $w_{\mathbf{s}}$) and hence yield the same $m_{\mathbf{s}}$, then by a direct application of [3, Theorem 1], for any uniquely decipherable code that is independent of $\mathbf{s}$, the redundancy is essentially never less than $I_{w_{\mathbf{s}}}(S;X)$ for *most* state sequences in this type class.

This bound is valid even if the type class in known a-priori. But if the type class is *not* known in advance intuition suggests that there must be an additional cost. We next demonstrate a coding scheme that is optimal in the sense of yielding the minimum attainable extra term, which in turn can be thought of as the redundancy associated with universal coding for a class of auxiliary sources that are "representing" the different type classes in a certain sense. Specifically, The proposed coding scheme can be interpreted as an hierarchical, two-step universal code, where the first step is to construct the best universal code within each type, and the second is to optimally integrate these codes by constructing another universal code for the class of the above mentioned auxiliary sources. The optimality of the proposed hierarchical code is in the sense that for any other code, most type classes have the property that except for a small minority of state sequences in the type class, the redundancy is essentially never less than the redundancy of the proposed code.

Finally, we point out that a natural subdivision of a class Λ of sources into subclasses $\Lambda_1, \Lambda_2, \ldots$, takes place in other situations as well. Another example is the class of all Markov sources, where Λ_i is the class of ith order Markov sources. The hierarchical universal coding approach demonstrated here, extends in the general case to a Shannon code w.r.t the double mixture, first over each Λ_i and then over $\{i\}$. Such a code was called "twice universal" in [4]. Similarly to Theorem 2, it can be shown that any other code cannot outperform the twice universal code, for "most" points in every Λ_i, except for a minority of classes Λ_i. Here by "most" we mean with high probabilty as measured by the mixture weights.

References

[1] T. Berger, *Rate Distortion Theory*. Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1971.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, 1981.

[3] N. Merhav and M. Feder, "A Strong Version of the Redundancy-Capacity Theorem of Universal Coding," to appear in *IEEE Trans. Inform. Theory*, May 1995.

[4] B. Ya. Ryabko, "Twice-universal coding," *Problems of Information Transmission*, pp. 173-177, July-Sept., 1984.