Minimum Redundancy Zero-error Source Coding with Side Information

Prashant Koulgi, Ertem Tuncel, Shankar Regunathan, Kenneth Rose[1]
Dept. of ECE, University of California, Santa Barbara, CA-93106, rose@ece.ucsb.edu

Abstract — **We consider the design of *instantaneous variable-length zero-error codes* where the decoder has access to side information about the source, which is not available to the encoder. We show that finding the optimal code, which minimizes the expected codeword length, is an NP-Hard problem. We derive an exponential-time optimal design algorithm. We also provide polynomial-time approximate algorithms, and discuss their average-case and worst-case performance guarantees.**

Problem Definition: (X, Y) is a pair of random variables distributed over a finite product set $\mathcal{X} \times \mathcal{Y}$ according to a probability distribution $P(x, y)$. A sender knows X while a receiver knows Y and wants to learn X *without error*. The characteristic graph $G = (\mathcal{X}, E)$ for the source pair (X, Y) was defined by Witsenhausen in [1]: G is defined on the vertex set $\mathcal{X}$, and for distinct $x_1, x_2 \in \mathcal{X}$, $(x_1, x_2) \in E$ if there exists $y \in \mathcal{Y}$ such that $P(x_1, y) > 0$ and $P(x_2, y) > 0$. We study the design of the following two families of k-ary ($k \geq 2$) codes $\phi : \mathcal{X} \to \{0, 1, \cdots, k-1\}^*$ for (G, P) (for motivation, see [2]):

1. ϕ is a *"Restricted Inputs (RI) code"* if $(x_1, x_2) \in E$ implies $\phi(x_1)$ is not a prefix of $\phi(x_2)$.
2. ϕ is an *"Unrestricted Inputs (UI) code"* if, for every distinct pair $x_1, x_2 \in \mathcal{X}$, $\phi(x_1)$ is not a proper prefix of $\phi(x_2)$, and, for $(x_1, x_2) \in E$, $\phi(x_1) \neq \phi(x_2)$.

We consider the design of minimum-rate RI and UI codes, where the rate of a code is given by its expected codeword length. It was shown in [2] that every RI (UI) code over G may be thought of as a coloring of G followed by one-to-one encoding (prefix-free encoding) of the colors. But simple examples show that the minimum-rate RI/UI code for G may *not* correspond to the minimum coloring of G.

For the sake of conciseness, we will illustrate our results for the example of binary codes. All our proofs and algorithms can be generalized to k-ary codes, $k \geq 2$.

Design of Optimal Codes is NP-Hard: We show that the following problem is NP-Hard: A) Given an arbitrary graph G with an arbitrary distribution P over its vertices, find the optimal RI (UI) code. Here we give a sketch of the proof for the simpler example of UI codes. Let an arbitrary graph G be given as input, and fix $0 < \epsilon < 1/12$. It is well-known that the following problem is NP-Complete: Is G colorable with 3 colors? Via a polynomial-time procedure, we derive the secondary graph G' with distribution P' on its nodes such that G is colorable with 3 colors iff there exists a UI code of rate $< 5/3 + \epsilon$ bits for (G', P'). Thus the following coding problem is NP-Complete: Given G and P, is there a UI code of rate $< 5/3 + \epsilon$ bits? This implies that problem A is NP-Hard for UI codes.

[1]This work was supported in part by the NSF under grant no. MIP-9707764, EIA 9986057, the UC MICRO program, Conexant Systems, Inc., Lucent Technologies, Inc., Medio Stream, Inc., and Qualcomm, Inc.

For UI codes we also prove a stronger claim. By building on the previous NP-Completeness result, we show that fast design algorithms to find even almost-optimal UI codes may not exist. In particular, we prove that the following problem is NP-Hard: "given arbitrary G and P, find a UI code of rate within $1/3 - 4\epsilon$ bits of the optimal UI code."

Exponential-time Optimal Design Algorithms: We will now briefly sketch the recursions on which our optimal design algorithms for RI and UI codes are based. Let $V' \subseteq \mathcal{X}$, and let $G' = (V', E')$ be the subgraph of G induced by V'. Let I' be the set of the isolated vertices of G'. Write $P(G') = \sum_{i \in V'} P(i)$, and define the distribution $P_{G'}$ on V' as $P_{G'}(i) = \frac{P(i)}{P(G')}$ for $i \in V'$. Let $\overline{L}_2(G')$ and $\overline{\mathcal{L}}_2(G')$ denote, respectively, the rates of the optimal RI and UI codes for $(G', P_{G'})$. We write $\overline{L}_2^*(G') = P(G')\overline{L}_2(G')$ and $\overline{\mathcal{L}}_2^*(G') = P(G')\overline{L}_2(G')$. The following equations recursively relate the optimal RI (UI) code for $(G', P_{G'})$ to those of the induced subgraphs of G':

$$\bar{L}_2^*(G') = \min_{D \subseteq G' - I'} [\bar{L}_2^*(D) + \bar{L}_2^*(G' - I' - D)] + P(G' - I'),$$

$$\bar{\mathcal{L}}_2^*(G') = \min_{D \subseteq G'} [\bar{\mathcal{L}}_2^*(D) + \bar{\mathcal{L}}_2^*(G' - D)] + P(G'),$$

if $G' \neq I'; \bar{L}_2(I') = 0, \bar{\mathcal{L}}(I') = 0.$

It is not necessary to search over all possible induced subgraphs. Thus, for the example of RI codes, the search may be restricted to dominating partitions of G'.

Fast Approximate Algorithms: Fast approximate algorithms to design codes for large graphs may be of interest in practice. We derive two approximate algorithms tailored towards good average-case and worst-case performance respectively. While these algorithms work for arbitrary distributions, we sketch below their performance guarantees for the special, albeit important, case of the uniform distribution. Let a graph G be the input, and let n and c be respectively the size of the vertex set and the minimum cardinality of a coloring of G. We derive a simple algorithm of complexity $O(n \log n)$ to obtain, for almost all graphs G, a UI code of rate within 2 bits of the optimal UI coding rate. We show that, almost surely, the rate of this code is also within $2 + \log e + \log \log n$ bits of the rate of the optimal RI code for G. We then derive a polynomial-time algorithm guaranteed (in the worst-case) to produce a UI code of rate below $\log n - 2 \log \log n + 2 \log \log c$ bits. (Note that this is lower than the entropy of the uniform distribution.)

References

[1] H. S. Witsenhausen, "The zero-error side information problem and chromatic numbers," *IEEE Transactions on Information Theory*, vol.IT-22, (no.5), pp.592-3, Sept. 1976.

[2] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Transactions on Information Theory*, vol.IT-42, (no.5), pp.1329-39, Sept. 1996.