Worst case bounds on Universal Variable-to-Fixed Length Source Codes [1]

K. Visweswariah
Dept. of Electrical Engg.
Princeton Univ., NJ 08544
karthik@ee.princeton.edu

S. R. Kulkarni
Dept. of Electrical Engg.
Princeton Univ., NJ 08544
kulkarni@ee.princeton.edu

S. Verdú
Dept. of Electrical Engg.
Princeton Univ., NJ 08544
verdu@ee.princeton.edu

Abstract — **In [6] universal variable-to-fixed length codes were considered which were shown to be asymptotically optimal. Here we consider their performance on an individual sequence x_1^n. We also look at the performance on individual sequences of the scheme introduced in [1]. We bound the performance in terms of the best code length achievable for the string from a class of coders.**

I. Introduction

We consider the compression of a source that takes values in a finite set A. A variable-to-fixed length code is given by a complete, prefix free dictionary $\mathcal{D}$. The source string is parsed using the dictionary and each phrase is coded using $\log|\mathcal{D}|$ bits. Since the output is of a fixed length the code achieves compression if the input phrase lengths are large.

Most work in the area of variable-to-fixed length coding looks at the expected length of a phrase in the dictionary as a measure of performance. This is justified because for Markov sources for long input strings the compression converges to $\log|\mathcal{D}|/E(L)$. Thus a goal while designing the dictionary is to try and maximize $E(L)$ for a fixed dictionary size. Tunstall [5] gave an algorithm to maximize $E(L)$ for a fixed $|\mathcal{D}|$ for an i.i.d source. Generalization of Tunstall's method to Markov sources were given in [2], [3]. Tjalkens and Willems [4] gave asymptotically optimal universal codes for binary i.i.d sources. Asymptotically optimal universal codes for Markov sources were given in [6].

II. Problem Formulation

Here we will consider the performance of codes given in [6] and [1] on individual sequences. We will consider a sequence $x_1^n \in A^n$ and bound the performance of the codes on this sequence in terms of the best possible code length form a class of codes. Our comparison class will be the class of Markov processes with memory l and output alphabet A. The set of states is $S=\{0,1\}^l$. Let $P_\theta(x_1^n)$ denote the probability of x_1^n when the parameter vector (which gives the probabilities in each state) is θ.

We will allow the encoder and the decoder to choose the dictionary to be used depending on what state the source is in. An encoder $\mathcal{D}$ is a collection of 2^l dictionaries. A dictionary $\mathcal{D}_s$ is used to parse the source output if the source is in state s. Consider a collection of dictionaries $\mathcal{D}_s$ with a maximum of M segments in a dictionary. The compression is $R(x_1^n)=(m(x_1^n)\log M)/n$ where $m(x_1^n)$ is the number of phrases that x_1^n is parsed into by $\mathcal{D}$. We will bound $R(x_1^n)$ in terms of the best achievable compression performance from our class of codes. In [6] a method to construct trees was given which was shown to achieve asymptotically optimal performance. There the performance criterion was $\log M/E(L)$ where $E(L)$ is the expected phrase length.

III. Results

Theorem 1 *There exist codes of size M such that for any $\delta \geq 0$ such that for any sequence $x_1^n \in A^n$ and initial state s_0*

$$\frac{R(x_1^n)}{n} \leq \left(1+\frac{C(l,|A|,\delta)\log\log M}{\log M}\right)\min_\theta \frac{1}{n}\log\frac{1}{P_\theta(x_1^n)} + \frac{1+\delta}{\sqrt{\log M}}$$

for all $M \geq M(\delta)$.

The proof of the theorem above is constructive and the method of construction of the code is as in [6].

Our next result is an upper-bound on the performance of the coding method described in [1], which allows the use of a variable-fixed length code along with an estimator of probabilities. We will use as an estimator the context tree weighting method [7]. Let us denote the compression achieved by the scheme as $R_{\mathrm{NF}}(x_1^n)$. Here our comparison class will be the class of sources with memory smaller than l. As in [7] let $P(x_1^n|\theta,\mathcal{S})$ denote the probability of x_1^n when the source model is $\mathcal{S}$ and parameter vector is θ.

Theorem 2 *For sufficiently large M we have*

$$\frac{R_{\mathrm{NF}}(x_1^n)}{n} \leq \frac{1}{n}\min_{\theta,\mathcal{S}}\log\frac{1}{P(x_1^n|\theta,\mathcal{S})} + c(1+\delta)\frac{\log\log M}{\log M}$$

where $c=(\log A + 2^{l-1}+1)(2^{l-1}+1)$.

References

[1] B. Nissenbaum, and M. Feder , "An adaptive variable-to-fix lossles coding scheme," *IEEE International Symposium on Information Theory*, pp. 193, Trondheim, Norway, 1994.

[2] S. A. Savari, and R. G. Gallager " Generalized Tunstall codes for sources with memory," *IEEE Trans. Information Theory*, vol. IT-43, pp. 658-668, March. 1997.

[3] T. J. Tjalkens, and F. M. J. Willems "Variable to fixed-length codes for Markov sources," *IEEE Trans. Information Theory*, vol. IT-33, pp. 246-257, March. 1987.

[4] T. J. Tjalkens, and F. M. J. Willems "A universal variable-to-fixed length source code based on Lawrence's Algorithm," *IEEE Trans. Information Theory*, vol. 38, pp. 247-253, March. 1992.

[5] B.P. Tunstall, "Synthesis of Noiseless compression codes," Ph.D. dissert., Georgia Inst. of Technology., Atlanta, GA, Sept. 1967.

[6] K. Viswewswariah, and S. Kulkarni , "Universal variable-to-fixed length source codes for Markov sources," *IEEE International Symposium on Information Theory*, Cambridge, MA, 1998.

[7] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens, "The context-tree weighting method: basic properties," *IEEE Trans. Information Theory*, vol. 41, pp. 653-664, May. 1995.

[1] This work was supported in part by the National Science Foundation under Grants NYI award IRI-9457645 and NCR 9523805