Minimum Conditional Entropy Context Quantization

Xiaolin Wu[1] Philip A. Chou[2] Xiaohui Xue[3]

Abstract — **We consider the problem of finding the quantizer Q that quantizes the K-dimensional causal context $C_i = (X_{i-t_1}, X_{i-t_2}, \ldots, X_{i-t_K})$ of a source symbol X_i into one of M conditioning states such that the conditional entropy $H(X_i|Q(C_i))$ is minimized. The resulting minimum conditional entropy context quantizer can be used for sequential coding of the sequence** $X_0, X_1, X_2, \ldots$.

A key problem in sequential source coding of a discrete random sequence $X_0, X_1, X_2, \cdots$ is modeling the underlying conditional distribution of the source $P(X_i|X^{i-1})$, Because of model estimation considerations, it is not possible to directly use all of X^{i-1} as the model's context. Many practical source coders choose *a priori* a model with fixed complexity, based on domain knowledge such as correlation structure and typical data length, and estimate only the model parameters. To avoid context dilution problem, we quantize the modeling context into a relatively small number of conditioning states, and estimate $P(X_i|Q(C_i))$ instead, where Q is a context quantizer. This approach has produced some of the best performing signal compression algorithms such as CALIC and JPEG 2000, despite the fact that they are not strictly universal. A pivotal issue for these source coders, which impacts their rate-distortion performance, is the design of the context quantizer Q. The problem is one of optimal vector quantization design with respect to the Kullback-Leibler distance.

Let Y be a discrete random variable, and let C be a jointly distributed random vector, possibly real. Given a positive integer M, we wish to find the quantizer $Q : C \rightarrow \{1, 2, \ldots, M\}$ such that $H(Y|Q(C))$ is minimized. Clearly, $H(Y|Q(C)) \geq H(Y|C)$ by the convexity of H. However, we wish to make $H(Y|Q(C))$ as close to $H(Y|C)$ as possible. Equivalently, we wish to minimize the non-negative "distortion" of Q

$$\begin{aligned} D(Q) &= H(Y|Q(C)) - H(Y|C) \\ &= \int dP(\mathbf{c}) D(P_{Y|C=\mathbf{c}} || P_{Y|Q(C)=Q(\mathbf{c})}), \qquad (1) \end{aligned}$$

which is the average, over all context vectors $\mathbf{c}$, of the Kullback-Leibler distances between the probability mass functions (pmfs) $P_{Y|C}(\cdot|\mathbf{c})$ and their "reproduction" pmfs $P_{Y|Q(C)}(\cdot|Q(\mathbf{c}))$.

Let $\beta_m(y) = P_{Y|Q(C)}(y|m)$ denote the mth reproduction pmf. Then an optimal Q must map almost all context vectors $\mathbf{c}$ to the conditioning state m that minimizes the Kullback-Leibler distance $D(P_{Y|C=\mathbf{c}}||\beta_m)$, i.e.,

$$Q(\mathbf{c}) = \arg\min_m D(P_{Y|C=\mathbf{c}}||\beta_m). \qquad (2)$$

The quantization regions $A_m = \{\mathbf{c} : Q(\mathbf{c}) = m\}$, $m = 1, \ldots, M$, of a minimum conditional entropy context quantizer are generally quite complex in shape, and may not even be convex or connected. However, their associated sets of pmfs $B_m = \{P_{Y|C}(\cdot|\mathbf{c}) : \mathbf{c} \in A_m\}$ are simple convex sets in the probability simplex for Y, owing to the above necessary condition for optimal Q. Let $\beta_m(y) = P(y|C \in A_m)$ be the conditional distribution of Y given.$C \in A_m$. Then by (2), for each $\mathbf{c} \in B_m$, the Kullback-Leibler distance from $P_{Y|C}(y|\mathbf{c})$ to $\beta_m(y)$ must be less than (or equal to) the Kullback-Leibler distance to $\beta_{m'}(y)$, $m' \neq m$. Hence

$$\sum_y P(y|\mathbf{c}) \log \frac{1}{\beta_m(y)} \leq \sum_y P(y|\mathbf{c}) \log \frac{1}{\beta_{m'}(y)}, \qquad (3)$$

for all $m' \neq m$. In other words, if $\mathbf{c} \in B_m$, then $P(y|\mathbf{c})$ lies in an intersection of halfspaces.

If Y is a binary random variable, then its probability simplex is one-dimensional. In this case, the quantization regions B_m are simple intervals. If the random variable Z is defined as $P_{Y|C}(1|C)$ (the posterior probability that $Y = 1$ as a function of C), then the conditional entropy $H(Y|Q(C))$ of the optimal context quantizer can be expressed

$$H(Y|Q(C)) = \sum_{m=1}^{K} P\{Z \in [q_{m-1}, q_m)\} H(Y|Z \in [q_{m-1}, q_m)) \qquad (4)$$

for some set of thresholds $\{q_m\}$. Therefore, the optimal context quantizer can be found by searching over $\{q_m\}$. This is a scalar quantization problem, which can be solved exactly using dynamic programming, regardless of the dimensionality of the context space. Once the scalar problem is solved, the optimal context quantizer cells A_m are given by

$$A_m = \{\mathbf{c} : P_{Y|C}(1|\mathbf{c}) \in [q_{m-1}, q_m)\}. \qquad (5)$$

In particular, the boundaries between these cells are determined by those vectors $\mathbf{c}$ for which the posterior probability $P_{Y|C}(1|\mathbf{c})$ is a constant: For example, $P_{Y|C}(1|\mathbf{c}) = q_m$ for $\mathbf{c}$ along the boundary between A_m and A_{m+1}. Equivalently, A_m can be expressed in terms of the likelihood ratio

$$L(\mathbf{c}) = \frac{P_{C|Y}(\mathbf{c}|1)}{P_{C|Y}(\mathbf{c}|0)} = \frac{P_Y(0)}{P_Y(1)} \frac{P_{Y|C}(1|\mathbf{c})}{1 - P_{Y|C}(1|\mathbf{c})}. \qquad (6)$$

If both $P_{C|Y}(\mathbf{c}|0)$ and $P_{C|Y}(\mathbf{c}|1)$ are d-dimensional Gaussians, then optimal context quantizer cells are bounded by d-dimensional quadratic surfaces.

The significance of this research is in that it offers a constructive means of designing optimal source codes for minimum code length via high-order context modeling. The problem of controlling model cost in high-order context modeling is addressed by designing optimal context quantizer, which collapses high-order contexts into any given number of coding states in a way to minimize the actual code length. Once the context quantizer Q is designed, on-line estimation of $P(\cdot|Q(C))$ by count statistics and adaptive entropy coding can be done very efficiently, much faster than by context tree methods. We observe that our techniques often outperform the universal source codes of proven optimality by appreciable margins on real data in image, video, and audio compression.

[1] Dept. of Computer Science, Univ. of Western Ontario, London, Ontario, Canada N6A 5B7, wu@csd.uwo.ca.

[2] Microsoft Research, One Microsoft Way, Redmond, WA 98005, USA, pachou@microsoft.com.

[3] Dept. of Computer Science, Harbin Institute of Technology, Harbin, China, xue@csd.uwo.ca.

0-7803-5857-0/00/$10.00 ©2000 IEEE.