LETTER *Special Section on Information Theory and Its Applications*

Conditional Lempel-Ziv Complexity and Its Application to Source Coding Theorem with Side Information

Tomohiko UYEMATSU[†a)], ***Regular Member and*** **Shigeaki KUZUOKA**[†b)], ***Student Member***

SUMMARY This paper proposes the conditional LZ complexity and analyzes its property. Especially, we show an inequality corresponding to Ziv's inequality concerning a distinct parsing of a pair of sequences. Further, as a byproduct of the result, we show a simple proof of the asymptotical optimality of Ziv's universal source coding algorithm with side information.

key words: *conditional LZ complexity, conditional Ziv's inequality, universal source coding*

1. Introduction

The Lempel-Ziv78 (LZ78) code proposed by Ziv and Lempel [1] is one of the most popular lossless source code. Further, LZ78 code has various applications in the areas of information theory such as random number generation, hypothesis testing, etc. [2]. LZ78 code is based on the incremental parsing. Thus, trying to better understand the parsing of the string is an important problem in information theory. Especially, Ziv's inequality [3, Lemma 12.10.3] is known as a useful inequality that provides a connection between partitions of the string and probability distribution. On the other hand, LZ78 code can also be applied to the sequence accompanied by side information. For example, Ziv [4] considered the incremental parsing of the joint string consists of the sequence of ordered pairs of input and side information symbols, and proposed the source coding algorithm with side information, which was proved to be asymptotically optimal [5]. Moreover, Merhav [6] gave a lower bound to the compression ratio of source coding with side information by using the incremental parsing of the joint string. In this paper, we consider distinct parsing of the joint string, and propose the concept of the conditional Lempel-Ziv (LZ) complexity. Then, we show an inequality corresponding to Ziv's inequality and analyze the property of the conditional LZ complexity. Further, by using the result, we show an elementary proof of the asymptotical optimality of Ziv's universal source coding algorithm with side information [5].

Manuscript received January 9, 2003.
Manuscript revised April 18, 2003.
Final manuscript received June 6, 2003.

[†]The authors are with Dept. of Communications and Integrated Systems, Tokyo Institute of Technology, Tokyo, 152-8522 Japan.
a) E-mail: uyematsu@ieee.org
b) E-mail: kuzuoka@it.ss.titech.ac.jp

2. Conditional LZ Complexity and Its Properties

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite alphabets. For given $\boldsymbol{x} = x_1x_2\cdots x_n \in \mathcal{X}^n$ and $\boldsymbol{y} = y_1y_2\cdots y_n \in \mathcal{Y}^n$, the string $(\boldsymbol{xy}) = (x_1y_1)(x_2y_2)\cdots(x_ny_n) \in (\mathcal{X}\times\mathcal{Y})^n$ of pairs of symbols is called a joint string.

Suppose that the joint string $(\boldsymbol{xy})$ is parsed into $c = c(\boldsymbol{x},\boldsymbol{y})$ distinct words as follows:

$$(\boldsymbol{xy}) = (xy)_{n_0+1}^{n_1}(xy)_{n_1+1}^{n_2}\cdots(xy)_{n_{c-1}+1}^{n_c}, \tag{1}$$

where $(xy)_i^j$ denotes a substring $(x_iy_i)(x_{i+1}y_{i+1})\cdots(x_jy_j)$ of the string $(\boldsymbol{xy})$ and $n_0 = 0, n_c = n$. We denote by $c(\boldsymbol{y})$, the number of distinct phrases in the parsing of $\boldsymbol{y}$ induced by the parsing (1), and by $y(l)$, the lth distinct phrase in the induced parsing on $\boldsymbol{y}$. Set $c_l(\boldsymbol{x}|\boldsymbol{y})$ to be the number of distinct x phrases that appear jointly with $y(l)$. Then, the conditional LZ complexity of $\boldsymbol{x}$ with side information $\boldsymbol{y}$ induced by the parsing (1) is defined by

$$\frac{1}{n}\sum_{l=1}^{c(\boldsymbol{y})} c_l(\boldsymbol{x}|\boldsymbol{y})\log c_l(\boldsymbol{x}|\boldsymbol{y}). \tag{2}$$

The conditional LZ complexity satisfies the next theorem that is corresponding to [3, Theorem 12.10.1].

Theorem 1: Let $(\mathbf{X},\mathbf{Y})$ be a stationary ergodic process. Then, for any distinct parsing of the joint string,

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{l=1}^{c(Y_1^n)} c_l(X_1^n|Y_1^n)\log c_l(X_1^n|Y_1^n) \le H(\mathbf{X}|\mathbf{Y}) \quad \text{a.s.} \tag{3}$$

where X_1^n and Y_1^n are sequences of random variables generated by $\mathbf{X}$ and $\mathbf{Y}$ respectively, and $H(\mathbf{X}|\mathbf{Y})$ is the conditional entropy rate of the process $(\mathbf{X},\mathbf{Y})$.

Remark: Merhav [6] showed another characterization of the conditional LZ complexity. When we compress $\boldsymbol{x}$ in the presence of $\boldsymbol{y}$ as side information at both encoder and decoder, for any K-state lossless encoder E and for every $(\boldsymbol{xy})$,

$$L_E(\boldsymbol{x}|\boldsymbol{y}) \ge \frac{1}{n}\sum_{l=1}^{c(\boldsymbol{y})}[c_l(\boldsymbol{x}|\boldsymbol{y}) + K^2]\log\frac{c_l(\boldsymbol{x}|\boldsymbol{y})}{4K^2}, \tag{4}$$

where $L_E(\boldsymbol{x}|\boldsymbol{y})$ is the code length associated with the encoder E. Notice that for any fixed K, right hand side of (4) approaches conditional LZ complexity as n tends to infinity.

Before proceeding to the proof of the theorem, we introduce some notations and lemmas. For a fixed integer k and arbitrarily fixed $s_1 \in \mathcal{S}_k \stackrel{\triangle}{=} (\mathcal{X}\times\mathcal{Y})^k$, let $\tilde{Q}_k((\boldsymbol{xy})|s_1)$ be an arbitrary kth order Markov distribution on $(\mathcal{X}\times\mathcal{Y})^n$. Suppose that the joint string $(\boldsymbol{xy})$ is parsed as (1) and define $s_i = (xy)_{i-k}^{i-1}$. Namely, s_i is the k symbols of $(\boldsymbol{xy})$ preceding $(xy)_i$. Let $c_{l,s}$ be the number of phrases $x_{n_{i-1}+1}^{n_i}(1 \le i \le c)$ which satisfy $y_{n_{i-1}+1}^{n_i} = y(l)$ and $s_{n_{i-1}+1} = s$, where $l = 1, \cdots, c(\boldsymbol{y})$ and $s \in \mathcal{S}_k$. We now describe the conditional Ziv's inequality.

Lemma 1 (Conditional Ziv's inequality): For any stationary kth order Markov distribution $\tilde{Q}_k$ and any joint string $(\boldsymbol{xy})$,

$$\log \tilde{Q}_k(\boldsymbol{x}|\boldsymbol{y}, s_1) \le -\sum_{l=1}^{c(\boldsymbol{y})}\sum_{s\in\mathcal{S}_k} c_{l,s}\log c_{l,s}, \tag{5}$$

where the initial state $s_1 = (xy)_{-k+1}^0$ is arbitrarily fixed.

Proof: Let us define the set $I(l,s)$ as

$$I(l,s) \stackrel{\triangle}{=} \{i : y_{n_{i-1}+1}^{n_i} = y(l), s_{n_{i-1}+1} = s\}.$$

Since $x_{n_{i-1}+1}^{n_i} \ne x_{n_{j-1}+1}^{n_j}(i \ne j)$, for any l and s,

$$\sum_{i\in I(l,s)} \tilde{Q}_k(x_{n_{i-1}+1}^{n_i}|y_{n_{i-1}+1}^{n_i}, s_{n_{i-1}+1}) \le 1.$$

Thus, we have

$$\begin{aligned}
&\log \tilde{Q}_k(\boldsymbol{x}|\boldsymbol{y}, s_1)\\
&= \sum_{l=1}^{c(\boldsymbol{y})}\sum_{s\in\mathcal{S}_k}\sum_{i\in I(l,s)} \log \tilde{Q}_k(x_{n_{i-1}+1}^{n_i}|y_{n_{i-1}+1}^{n_i}, s_{n_{i-1}+1})\\
&= \sum_{l,s} c_{l,s} \sum_{i\in I(l,s)} \frac{\log \tilde{Q}_k(x_{n_{i-1}+1}^{n_i}|y_{n_{i-1}+1}^{n_i}, s_{n_{i-1}+1})}{c_{l,s}}\\
&\le \sum_{l,s} c_{l,s}\log\left(\sum_{i\in I(l,s)} \frac{1}{c_{l,s}} \tilde{Q}_k(x_{n_{i-1}+1}^{n_i}|y_{n_{i-1}+1}^{n_i}, s_{n_{i-1}+1})\right)\\
&\le -\sum_{l,s} c_{l,s}\log c_{l,s},
\end{aligned}$$

where the first inequality follows from Jensen's inequality. □

Proof of Theorem 1: Let Q be a joint distribution that characterizes the process $(\mathbf{X}, \mathbf{Y})$ and $\tilde{Q}_k$ be the kth order Markov approximation of Q defined as

$$\tilde{Q}_k((xy)_{-k+1}^n) \stackrel{\triangle}{=} Q((xy)_{-k+1}^0)\prod_{i=1}^n Q((xy)_i|(xy)_{i-k}^{i-1}). \tag{6}$$

Let $c_l = c_l(\boldsymbol{x}|\boldsymbol{y})$ and $c = \sum_{l=1}^{c(\boldsymbol{y})} c_l$. Then using the fact $\sum_{s\in\mathcal{S}_k} c_{l,s} = c_l$, and applying Lemma 1 to the conditional distribution $\tilde{Q}_k(\boldsymbol{x}|\boldsymbol{y}, s_1)$ which is deduced from (6), we have

$$\begin{aligned}
&\log \tilde{Q}_k(\boldsymbol{x}|\boldsymbol{y}, s_1)\\
&\le -\sum_{l=1}^{c(\boldsymbol{y})}\sum_{s\in\mathcal{S}_k} c_{l,s}\log c_{l,s}\\
&= -\sum_{l=1}^{c(\boldsymbol{y})} c_l \log c_l - c\sum_{l=1}^{c(\boldsymbol{y})}\sum_{s\in\mathcal{S}_k}\frac{c_{l,s}}{c}\log\frac{c_{l,s}}{c_l}\\
&= -\sum_{l=1}^{c(\boldsymbol{y})} c_l \log c_l + cH(S|L)\\
&\le -\sum_{l=1}^{c(\boldsymbol{y})} c_l \log c_l + ck\log|\mathcal{X}||\mathcal{Y}|,
\end{aligned} \tag{7}$$

where L and S are random variables defined by

$$\Pr\{L = l, S = s\} \stackrel{\triangle}{=} \frac{c_{l,s}}{c}.$$

In a similar way as [3, Lemma 12.10.1], it can be proved that for any distinct parsing of $(\boldsymbol{xy})$, $c/n \to 0$ $(n \to \infty)$. So, from (7) we have

$$\frac{1}{n}\sum_{l=1}^{c(\boldsymbol{y})} c_l \log c_l \le -\frac{1}{n}\log \tilde{Q}_k(\boldsymbol{x}|\boldsymbol{y}, s_1) + \delta_k(n), \tag{8}$$

where $\delta_k(n) \to 0$ $(n \to \infty)$. Therefore,

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n}\sum_{l=1}^{c(Y_1^n)} c_l(X_1^n|Y_1^n)\log c_l(X_1^n|Y_1^n)\\
&\le \limsup_{n\to\infty} -\frac{1}{n}\log \tilde{Q}_k(X_1^n|Y_1^n, (XY)_{-k+1}^0)\\
&\stackrel{(a)}{=} \limsup_{n\to\infty} -\frac{1}{n}\sum_{i=1}^n \log \tilde{Q}_k(X_i|Y_i, (XY)_{i-k}^{i-1})\\
&= \limsup_{n\to\infty} -\frac{1}{n}\sum_{i=1}^n \log \tilde{Q}_k(X_{i+k}|Y_{i+k}, (XY)_i^{i+k-1})\\
&\stackrel{(b)}{=} E_{\tilde{Q}_k}[-\log \tilde{Q}_k(X_1|Y_1, (XY)_{-k+1}^0)]\\
&= H(X_1|Y_1, (XY)_{-k+1}^0)\\
&\to H(\mathbf{X}|\mathbf{Y}) \quad \text{as } k \to \infty,
\end{aligned}$$

with probability one. Equality (a) follows from that $\tilde{Q}_k$ is kth order Markov distribution and equality (b) follows from the ergodic theorem. □

3. An Application to Source Coding Theorem with Side Information

In this section, as an application of Theorem 1, we prove the asymptotical optimality of Ziv's universal source

coding algorithm with side information [5].

First, we describe Ziv's coding algorithm. Let $\boldsymbol{x}$ be a sequence to be compressed and $\boldsymbol{y}$ be a side information that can be available at both encoder and decoder.

Ziv's algorithm[4], [5]:

1. Let $(\boldsymbol{xy}) = (xy)_{n_0+1}^{n_1}(xy)_{n_1+1}^{n_2}\cdots(xy)_{n_{p-1}+1}^{n_p}$ be the incremental parsing [1] of $(\boldsymbol{xy})$, where $n_0 = 0, n_p = n$. Set $i = 0$.
2. Encode the length $(n_{i+1} - n_i)$ of the substring $(xy)_{n_i+1}^{n_{i+1}}$ by using Elias's ω^*code [7].
3. Find the integer $j(< i)$ such that $(xy)_{n_j+1}^{n_{j+1}} = (xy)_{n_i+1}^{n_{i+1}-1}$.
4. Let q be the number of earlier y phrases which are identical with $y_{n_i+1}^{n_{i+1}-1}$. Then, encode the serial number of $(xy)_{n_j+1}^{n_{j+1}}$ using $\lceil \log q \rceil$ bits.
5. Encode $x_{n_{i+1}}$ using $\lceil \log |\mathcal{X}| \rceil$ bits.
6. If $i < p-1$, then set $i = i+1$ and return to Step 2. If $i = p-1$, then stop encoding.

The next theorem shows that Ziv's algorithm is asymptotically optimal.

Theorem 2 (Uyematsu and Maeda [5]): Let $(\mathbf{X}, \mathbf{Y})$ be a stationary ergodic process, and $L(\boldsymbol{x}|\boldsymbol{y})$ be the code length of the Ziv's algorithm for a joint string $(\boldsymbol{xy})$. Then

$$\lim_{n\to\infty} \frac{L(X_1^n|Y_1^n)}{n} = H(\mathbf{X}|\mathbf{Y}) \quad \text{a.s.}$$

Proof: The converse part of the theorem can be immediately obtained by [8, Theorem 4.2.1]. Hence, we only prove the direct part of the theorem, i.e.

$$\limsup_{n\to\infty} \frac{L(X_1^n|Y_1^n)}{n} \le H(\mathbf{X}|\mathbf{Y}) \quad \text{a.s.} \tag{9}$$

As shown in [5, lemma 2], the code length $L(\boldsymbol{x}|\boldsymbol{y})$ satisfies

$$L(\boldsymbol{x}|\boldsymbol{y}) \le \sum_{l=1}^{c(\boldsymbol{y})} c_l(\boldsymbol{x}|\boldsymbol{y})\{\log c_l(\boldsymbol{x}|\boldsymbol{y}) + \log \ell(y(l)) + 2\log\log \ell(y(l)) + \log|\mathcal{X}| + \log|\mathcal{Y}| + 9\} \tag{10}$$

Further, Ziv [4] showed that for any joint string $(\boldsymbol{xy}) \in (\mathcal{X}\times\mathcal{Y})^n$,

$$\frac{1}{n}\sum_{l=1}^{c(\boldsymbol{y})} c_l(\boldsymbol{x}|\boldsymbol{y}) \log \ell(y(l)) \le O\left(\frac{\log\log n}{\log n}\right). \tag{11}$$

According to (10),(11) and Theorem 1, we have

$$\begin{aligned}&\limsup_{n\to\infty} \frac{L(X_1^n|Y_1^n)}{n}\\ &\le \limsup_{n\to\infty} \frac{1}{n}\sum_{l=1}^{c(Y_1^n)} c_l(X_1^n|Y_1^n)\log c_l(X_1^n|Y_1^n)\\ &\le H(\mathbf{X}|\mathbf{Y}),\end{aligned}$$

with probability one. Thus, (9) is proved. □

4. Conclusion

In this paper, we proposed the conditional LZ complexity and analyzed its property. Further, by using the result, we provided an elementary proof of the source coding theorem with side information. As shown in this paper, we believe that the conditional LZ complexity is a keystone of analyzing many applications of LZ78 with side information.

References

[1] J. Ziv and A. Lempel, "Compression of individual sequences by variable rate coding," IEEE Trans. Inf. Theory, vol.IT-24, no.5, pp.530–536, Sept. 1978.
[2] T. Uyematsu, "Lempel-Ziv coding as a tool for information theory," IEICE Trans. Fundamentals (Japanese Edition), vol.J84-A, no.6, pp.681–690, June 2001.
[3] T.M. Cover and J.A. Thomas, Elements of Information Theory, Wiley, New York, 1991.
[4] J. Ziv, "Universal decoding for finite-state channels," IEEE Trans. Inf. Theory, vol.IT-31, no.4, pp.453–460, July 1985.
[5] T. Uyematsu and K. Maeda, "Asymptotical optimality for universal data compression algorithm with side information based on incremental parsing," IEICE Trans. Fundamentals (Japanese Edition), vol.J85-A, no.1, pp.95–102, Jan. 2002.
[6] N. Merhav, "Universal detection of messages via finite-state channels," IEEE Trans. Inf. Theory, vol.46, no.6, pp.2242–2246, Sept. 2000.
[7] P. Elias, "Universal codeword sets and representation of the integers," IEEE Trans. Inf. Theory, vol.21, no.2, pp.194–203, March 1975.
[8] J. Muramatsu, Universal data compression algorithms for stationary ergodic sources based on the complexity of sequences, Ph.D. Thesis, Nagoya University, Nagoya, Japan, 1998.