ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 5, No. 3, 2010, *pp*. 193-198

An Improved Bit-level Arithmetic Coding Algorithm

Jianjun Zhang [1, +] and Xingfang Ni [2]

[1]Dept. of Math, Shanghai University, Shanghai 200444, P.R. China
[2]Shanghai Jiao Tong University, Shanghai 200030, P.R. China

(*Received March 16, 2009, accepted October 22, 2009*)

Abstract: Arithmetic coding is the most powerful lossless data compression technique that has attracted much attention in recent years. This paper presents a new implementation of bit-level arithmetic coding using integer additions and shifts. The algorithm has less computational complexity and more flexibility, and thus is very suitable for hardware design. We show that degradation of the proposed algorithm is bounded by 0.2075.

Keywords: Arithmetic coding, multiplication-free.

1. Introduction

Arithmetic coding is the most powerful lossless data compression technique that has attracted much attention in recent years [2-10]. It provides more flexibility and better efficiency than the celebrated Huffman coding. Arithmetic coding completely bypasses the traditional coding paradigm: replace an input symbol with a specific code. Instead, it represents a stream of input symbols by a binary number in the interval [0,1). This approach totally relaxes the constraint upon Huffman coding: each symbol has to be encoded by an integral number of bits and by at least one bit. Thus its coding results are closer to Shannon's entropy bound [1].

The basic arithmetic coding requires infinite precision operations that are difficult to implement on a fixed precision computer. Therefore, it takes about twenty years to implement a practical arithmetic coding program on a fixed precision computer due to Rissanen and Pasco [5][6]. They proposed a LIFO-form (last-in-first-out) and a FIFO-form (first-in-first-out) of arithmetic coding respectively. These illustrate that arithmetic coding can be implemented on a fixed length computer. Combining the advantage of the above two schemes, Rubin [7] proposed a general implementation of arithmetic coding by using of fixed precision registers. A good tutorial can be found in Witten, Neal and Cleary [8], in which an interesting carry-over technique to avoid bit propagation was introduced.

One drawback of arithmetic coding is its slow speed, because arithmetic coding requires multiplications. Some research have been done to avoid multiplications and to improve the efficiency. Langdon and Rissanen[4] proposed a modified scheme for encoding a binary string by using of shift-and-add. Rissanen and Mohiuddin[6] proposed a multiplication-free algorithm for encoding general string. The method was further developed by Lei[5]. Howard and Vitter[2] described an efficient bit-level implementation that uses table lookup as a fast alterative to arithmetic operations.

To further improve the implementation efficiency of arithmetic coding, we present a new bit-level arithmetic coding by using integer additions and shifts. The algorithm has less computational complexity and more flexibility, and thus is very suitable for hardware and software design.

The paper is organized as follows. In section 2, we briefly review the bit-level arithmetic coding. In section 3, we propose an improved multiplication-free arithmetic coding algorithm which has less computational complexity. In section 4, we analyze the efficiency of the algorithm and show that the degradation of the proposed algorithm is bounded by 0.2075. Finally, computer simulation is given.

2. Review of bit-level arithmetic coding

Let $p(0|s)$ be the probability of 0 according to a given model, where s denotes the previous string. If the encoding interval for string s is $[C(s), C(s)+A(s))$, then the bit-level arithmetic coding algorithm

updates $A(s)$ and $C(s)$ according to the following rule:

$$\begin{aligned} C(s0) &= C(s) & A(s0) &= A(s)\,p(0\,|\,s) \\ C(s1) &= C(s) + A(s0) & A(s1) &= A(s) - A(s0) \end{aligned} \tag{2.1}$$

In practice, $C(s)$ and $A(s)$ are represented by finite bits, and probability $p(0\,|\,s)$ is estimated by the zero occurrence frequency. That is $p(0\,|\,s) \approx n(0\,|\,s)/n(s)$, $p(1\,|\,s) \approx n(1\,|\,s)/n(s)$ where $n(0\,|\,s)$ and $n(1\,|\,s)$ represent the count number of 0 and 1 in string s according to a given model, and $n(s) = n(0\,|\,s) + n(1\,|\,s)$. It is obvious that, arithmetic coding requires multiplications even if we only use integer number. This is expensive and usually has slow speed in both hardware and software implementation. Therefore, in practical, we should avoid to use formula (2.1).

Two typical multiplication-free bit-level arithmetic coding are LR algorithm (London & Rissanen) [4] and RM algorithm (Rissssanen & Mohiouddin) [6].

LR algorithm uses $2^{-k(s)}$ as an approximation to probability $p(0\,|\,s)$ to eliminate the multiplications in (2.1), where $k(s)$ is some integer. As the encoding proceeds, $A(s)$ becomes smaller and smaller. In order to keep as many significant bits as possible, it multiplies $2^{L(s)}$ to $A(s)$ such that $2^{L(s)}A(s)$ lies in [1, 2), where $L(s)$ is some nonnegative number adjusted according to magnitude of $A(s)$. LR algorithm then updates $A(s)$ and $C(s)$ as follows:

$$\begin{aligned} C(s0) &= C(s) & A(s0) &= 2^{-L(s)-k(s)} \\ C(s1) &= C(s) + A(s0) & A(s1) &= A(s) - A(s0) \end{aligned} \tag{2.2}$$

RM algorithm is a simplified form of multiplication-free multi-alphabet arithmetic coding. The mechanism to avoid multiplications is choosing $L(s)$ and $k(s)$ such that $2^{L(s)}A(s)$ and $2^{-k(s)}n(s)$ lie inside [0.75, 1.5) simultaneously. RM algorithm then updates $A(s)$ and $C(s)$ as follows:

$$\begin{aligned} C(s0) &= C(s) & A(s0) &= 2^{-L(s)-k(s)}\,n(0\,|\,s) \\ C(s1) &= C(s) + A(s0) & A(s1) &= A(s) - A(s0) \end{aligned} \tag{2.3}$$

It is obvious that, the above algorithm can be implemented by shifts and additions. However, the operation of choosing $2^{L(s)}$ such that $2^{L(s)}A(s)$ lies in [1, 2], and choosing $L(s)$ and $k(s)$ such that $2^{L(s)}A(s)$ and $2^{-k(s)}n(s)$ lie inside [0.75, 1.5] simultaneously is somewhat time-consuming. On the other hand, in practical implementation, the inherent carry-over problem need to be considered. A wide used technique to avoid carry-over is bit-stuffing proposed by Rissanen [6], but this technique reduces the encoding efficiency in some extent. In [8], an algorithm to avoid bit-stuffing is presented, but the technique in the algorithm is very complicated. In [9], a very simple method to avoid carry-over was presented.

3. An improved bit-level arithmetic coding

Zhao et al in [9] used the similar approximation method as LR and RM algorithm. They directly choose integer number $k(s)$ such that $2^{-k(s)}$ approximates $p(0\,|\,s)$. To deal with the carry-over problem, they presented a very simple and efficient technique.

In the subsequent discussion, we use a similar method to deal with the carry-over problem. To avoid integer multiplication, we take a method that is easy to manipulate. The method combines the encoding and the carry-over technique together. The method has less computational complexity and accordingly runs rapidly.

Improved arithmetic coding algorithm:

1. Initialize C and A;
2. Compute $p0$ and $p1$ according to a given statistical model;
3. Input the next symbol x. If end of file, then output the content of C and terminate;

4. If $p0 > p1$, exchange the contents of $p0$ and $p1$ and let $x =!x$;

5. If $A < (p0 + p1)$, repeat:

   a) Output the most significant bit of C;

   b) $C <<= 1, A <<= 1$;

   c) If $A >\sim C$, then $A =\sim C + 1$;

6. If $x = 0, A = p0$; else $A = A - p0, C = C + p0$.

In the above algorithm, A and C can be given arbitrarily. Since the algorithm outputs the results bit by bit, we need to pack the bits into byte or word. In the end of encoding, we must output the content of C. The decoding algorithm is similar to the encoding algorithm. We only need an extra decoding variable V of the same bits as C.

Improved arithmetic decoding algorithm :

1. Initialize C and A;

2. Compute $p0$ and $p1$ according to the given statistical model;

3. Set $x = 0$;

4. If $p0 > p1$, exchange the contents of $p0$ and $p1$, let $x =!x$;

5. If $A < (p0 + p1)$, repeat:

   a) $C <<= 1, A <<= 1, V <<= 1$;

   b) V |= next bit to be decoded;

   c) If $A >\sim C,$ then $A =\sim C + 1$;

6. If $V - C < p0, A = p0$, else $A = A - p0, C = C + p0, x =!x$.

In the decoding algorithm, the parameters must be consistent with those in the encoding algorithm, and $V = 0$. Otherwise, the decoding algorithm will not work correctly. The length of the data to be encoded must be preserved, so that the decoding algorithm can terminate properly.

The $p0$ and $p1$ in the improved arithmetic coding algorithm need to be computed by using of a statistical model. To this end, we introduce two array $P0$ and $P1$ of K bits, and assign variable W of l bits, which actually is a sliding window on input data. Then the statistical model is as follows.

Statistical model:

1. Input: x (0 or 1);

2. $p0 = P0[W], p1 = P1[W]$;

3. If $x = 0$, then $P0[W]$++, else $P1[W]$++;

4. W shift left 1 bit, W |= x;

5. Output $p0$ and $p1$.

To avoid the zero probability, the initial value of $p0$ and $p1$ should not set to zero, but the initial value of W could be any value. Using the improved arithmetic algorithm and the given statistical model, we can carry out the data compression efficiently.

4. Efficiency analysis

Following Lei [5], we consider the degradation of the improved arithmetic coding, and compare it with RM method. It is easy to see that the RM method is using the following modified probability model $p(0 \mid s) = \dfrac{2^{-L(S)-k(s)}\mathrm{n}(0|s)}{A}$, $p(1 \mid s) = 1 - p(0 \mid s)$, instead of the ideal model $n(0 \mid s)/n(s)$ and $1 - n(0 \mid s)/n(s)$, while the improved arithmetic coding uses the following modified probability model $p(0 \mid s) = \dfrac{n(0 \mid s)}{A}$, $p(1 \mid s) = 1 - p(0 \mid s)$, instead of the ideal model $n(0 \mid s)/n(s)$ and $1 - n(0 \mid s)/n(s)$.

The ideal average coding length l is known to be

$$l = -\frac{n(0|s)}{n(s)}\log_2\frac{n(0|s)}{n(s)} - (1-\frac{n(0|s)}{n(s)})\log_2(1-\frac{n(0|s)}{n(s)}).$$

The actual average coding length l_R of the RM coding algorithm is

$$l_R = -\frac{n(0|s)}{n(s)}\log_2\frac{2^{-L(s)-k(s)}n(0|s)}{A} - (1-\frac{n(0|s)}{n(s)})\log_2(1-\frac{2^{-L(s)-k(s)}n(0|s)}{A}),$$

and the actual average coding length l_I of the improved arithmetic coding algorithm is

$$l_I = -\frac{n(0|s)}{n(s)}\log_2\frac{n(0|s)}{A} - (1-\frac{n(0|s)}{n(s)})\log_2(1-\frac{n(0|s)}{A}).$$

Therefore the degradation of RM coding and the improved arithmetic coding is respectively:

$$\delta_R = l_R - l = \frac{n(0|s)}{n(s)}\log_2\frac{\alpha(s)}{n(s)} + (1-\frac{n(0|s)}{n(s)})\log_2((1-\frac{n(0|s)}{n(s)})\frac{\alpha(s)}{n(s)}\Big/(\frac{\alpha(s)}{n(s)}-\frac{n(0|s)}{n(s)}))$$

and

$$\delta_I = l_I - l = \frac{n(0|s)}{n(s)}\log_2\frac{A}{n(s)} + (1-\frac{n(0|s)}{n(s)})\log_2((1-\frac{n(0|s)}{n(s)})\frac{A}{n(s)}\Big/(\frac{A}{n(s)}-\frac{n(0|s)}{n(s)})),$$

where $\alpha(s) = 2^{L(s)+k(s)}A$. The range of $\frac{\alpha(s)}{n(s)}$ is [0.5, 2], and the range of $\frac{A}{n(s)}$ is [1, 2].

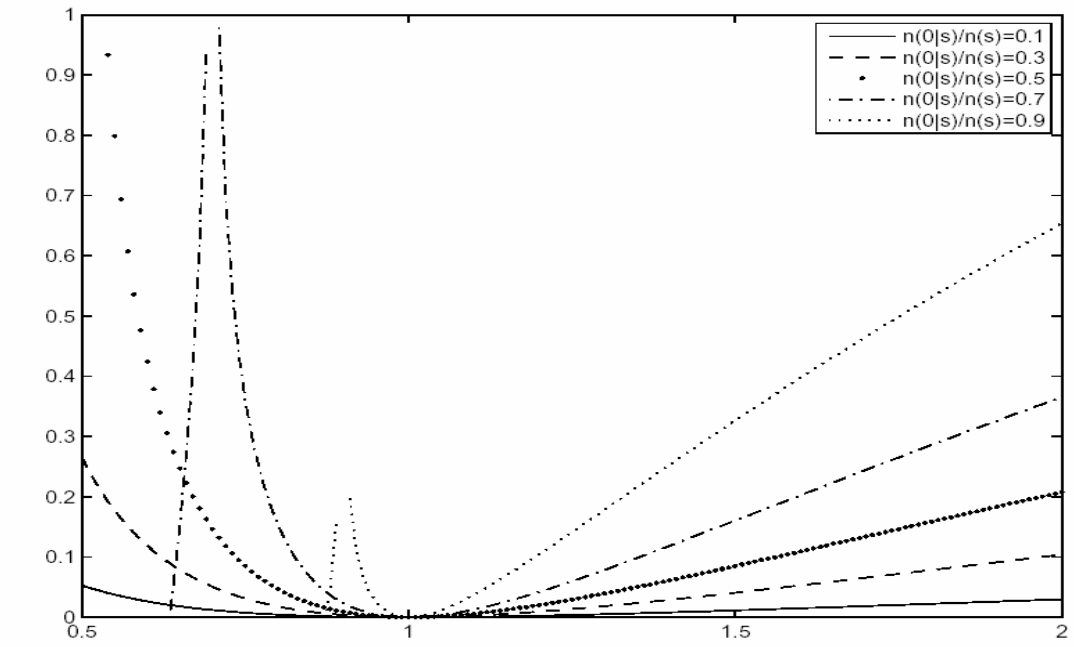

Fig.1 Degradation of RM scheme

Fig. 1 gives plots of δ_R versus $\frac{\alpha(s)}{n(s)}$ for different $\frac{n(0|s)}{n(s)}$, and Fig. 2 gives plots of δ_I versus $\frac{A}{n(s)}$ for different $\frac{n(0|s)}{n(s)}$. Since $\frac{n(0|s)}{n(s)} \le 0.5$ in the improved arithmetic coding, from the figures, we see that δ_I is always less than 0.2075, while δ_R may be very large in some worse cases.

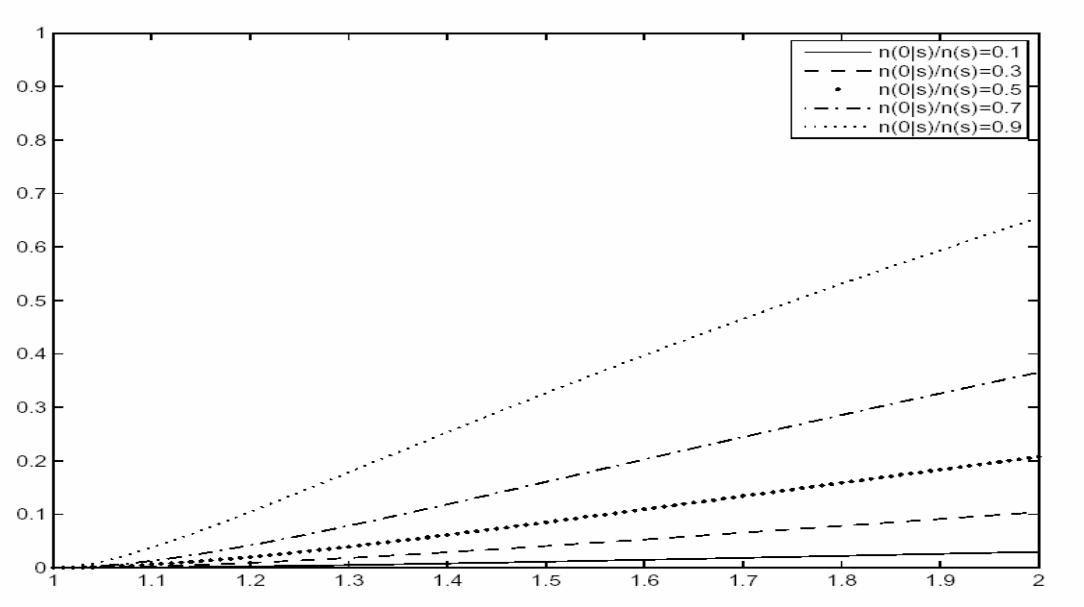

Fig.2 Degradation of improved arithmetic scheme

5. Computer simulation

In this section, the basic bit-level arithmetic coding algorithm and our proposed algorithm were tested. We arbitrarily select four types of test data file: C-language source file(file1), Chinese word document(file2), .exe file(file3) and Lena.tif file(file4). The sizes of these files are 11969, 976896, 121856 and 262330 bytes respectively.

Test results are listed in table 1, where we list the encoding rate which is defined as Er=8×length of encoded file(bytes)/ length of original file(bytes), and IA and BA represent the proposed algorithm and basic arithmetic coding algorithm respectively.

Table 1 Encoding rate for different files

	file1	file2	file3	file4
IA	4.206	2.414	3.361	5.681
BA	4.342	2.350	3.282	6.017

For comparison, we also list the test results in [9] in table 2, where five types of file: C-language sources file(file5), Chinese word document(file6), grey image file(file7), .exe file (file8) and hybrid data file(file9) were tested. The sizes of these files are 27620, 72596, 262330, 54645 and 417192 bytes respectively. In table 2, BA, ZA, RA and LA represent the basic arithmetic coding algorithm, Zhao's algorithm [9], RM algorithm and LR algorithm respectively.

Table 2 Encoding rate for different files

	file5	file6	file7	file8	file9
BA	3.001	3.464	5.433	5.944	5.377
ZA	3.074	3.507	5.504	5.972	5.443
RA	3.198	3.595	5.574	6.053	5.513
LA	3.237	3.692	5.941	6.438	5.945

From table 1 and table 2, we see that Zhao's algorithm, RM algorithm and LR algorithm always have less compression rate than the basic arithmetic coding, while our proposed algorithm sometimes has higher compression rate than the basic arithmetic coding. This may be due to inaccuracy probability estimation of the practical modal. In addition, our proposed algorithm has the least complexity and the highest speed in the above mentioned algorithm.

6. Conclusion

In this paper, we proposed a new bit-level arithmetic coding algorithm. It has the least complexity and the highest speed in Zhao's algorithm, RM algorithm, LR algorithm and the basic arithmetic coding algorithm. Sometimes it has higher compression rate than basic arithmetic encoding algorithm. Therefore, it provides an excellent compromise between good performance and low complexity.

7. Acknowledgements

The authors would like to thank the associate editor and the reviewers for helpful comments.

8. References

[1] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal.* 1948, 27: 379-423.

[2] P. G. Howard, and J. S. Vitter. Arithmetic coding for data compression. *Proc. IEEE.* 1994, 82: 857-865.

[3] J. Jiang. Novel design of arithmetic coding for data compression. *IEE Proc. Comput. Digit.* 1995, 142: 419-424.

[4] G. G. Langdon, and J. J. Rissanen. A simple general binary source code. *IEEE Trans. Inf. Theory.* 1982, IT-28: 800-803.

[5] Shaw-Min Lei. Efficient multiplication-free arithmetic codes. *IEEE Trans. Commun*. 1995, 43: 2950-2958.

[6] J. J. Rissanen, and K. M. Mohiuddin. A multiplication-free multialphabet arithmetic code. *IEEE Trans. Commun.* 1989, 37: 93-98.

[7] F. Rubin. Arithmetic stream coding using fixed precision registers. *IEEE Trans. Inf. Theory.* 1979, IT-25: 672-675.

[8] I. H. Witten, R. M. Neal, and J. G. Cleary. Arithmetic coding for data compression. *Commun. ACM.* 1987, 30: 520-540.

[9] F. Zhao, F. Jiang and L. Shen. Multiplication-free bit-level arithmetic coding and its applications. *J. of China Institute of Commun*. 1998, 19: 82-87.

[10] C. Boyd, J. G. Cleary, S. A. Irvine, I. Rinsma-Melchert, and I. H. Witten. Integrating Error Detection into Arithmetic Coding. *IEEE Trans. on Commun*. 1997, 45: 1-3.