Empirical Evaluation of the Effect of the Symbol Distribution on the Performance of ANS

Danny Dubé
Danny.Dube@ift.ulaval.ca
/
Departement of Computer Science and Software Eng.
Université Laval, Canada

Hidetoshi Yokoo
Yokoo@gunma-u.ac.jp
/
Division of Electronics and Informatics
Gunma University, Japan

Abstract

We consider a variety of construction methods of the symbol distribution used by ABS/ANS and measure the resulting coding efficiency. The evaluation of the efficiency is performed empirically. We evaluate the effect of different, mostly orthogonal construction strategies and problem settings on the efficiency. In particular, we compare: the uniform symbol distribution (as in uABS) versus the range one (as in rANS); the use of a correction on the state probabilities versus none; a binary input alphabet versus larger ones; a small set of states versus a large one. Specific probability distributions are used for each of the considered alphabets.

Basic definitions for ABS/ANS

Source alphabet:

$$\mathcal{A}, \qquad \text{ABS when } |\mathcal{A}| = 2, \qquad \text{ANS when } |\mathcal{A}| > 2$$

Symbol distribution:

$$\overline{s} \;:\; \mathbb{N} \to \mathcal{A} \qquad \text{// or infinite string on } \mathcal{A}$$

$$\overline{s} \overset{\text{e.g.}}{=} \boxed{\texttt{a}}_{0}\boxed{\texttt{b}}_{1}\boxed{\texttt{a}}_{2}\boxed{\texttt{c}}_{3}\boxed{\texttt{a}}_{4}\boxed{\texttt{b}}_{5}\boxed{\texttt{a}}_{6}\boxed{\texttt{a}}_{7}\boxed{\texttt{b}}_{8}\boxed{\texttt{a}}_{9}\boxed{\texttt{a}}_{10}\boxed{\texttt{b}}_{11}\boxed{\texttt{c}}_{12}\boxed{\texttt{a}}_{13}\boxed{\texttt{a}}_{14}\boxed{\texttt{b}}_{15}\boxed{\texttt{a}}_{16}\boxed{\texttt{a}}_{17}\boxed{\texttt{b}}_{18}\boxed{\texttt{a}}_{19}\boxed{\texttt{c}}_{20}\boxed{\texttt{a}}_{21}\boxed{\texttt{b}}_{22}\boxed{\texttt{a}}_{23}\boxed{\texttt{a}}_{24}\ldots$$

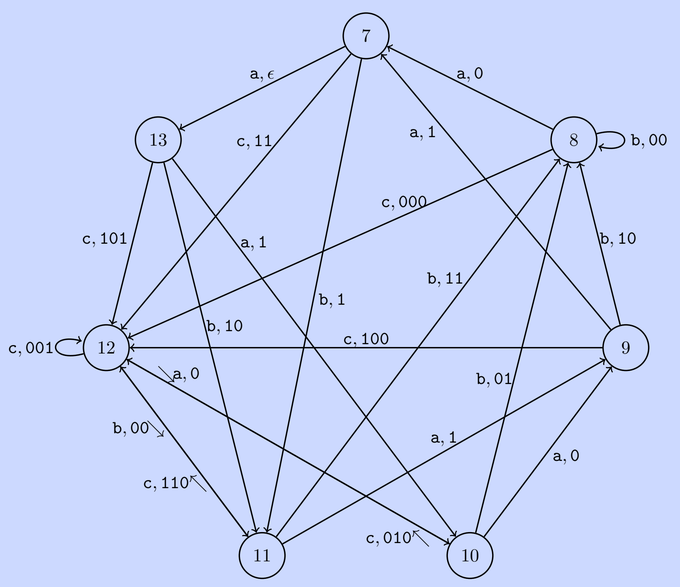
Encoding function:

$$C \;:\; \mathcal{A} \times \mathbb{N} \to \mathbb{N}$$
$$C(s, x) = \text{select}_s(x, \overline{s}) \qquad \text{// counting from 0}$$

Decoding function:

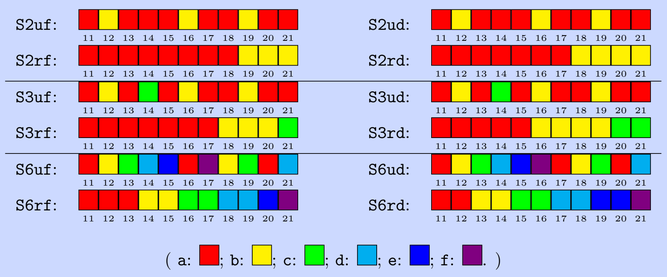
$$D \;:\; \mathbb{N} \to \mathcal{A} \times \mathbb{N}$$
$$D(x) = \textbf{let } s = \overline{s}(x) \textbf{ in } (s, \text{rank}_s(x, \overline{s})) \qquad \text{// counting from 0}$$

Encoding automaton

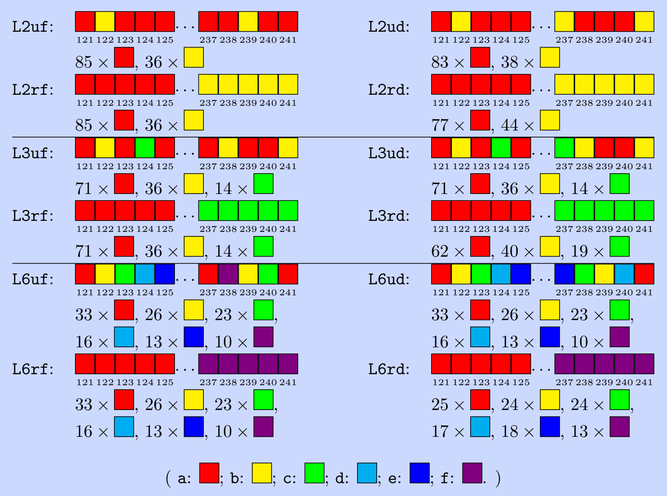


The first example encoder under the form of an automaton.

Symbol distributions in the experiments with $|I| = 11$



Symbol distributions in the experiments with $|I| = 121$



The 41st Symposium on Information Theory and its Applications (SITA2018)
Iwaki, Fukushima, Japan, Dec. 18–21, 2018

Definitions for the stream variant of ABS/ANS

Selected interval of possibilities for the current state:

$$I = \{start, start + 1, \ldots, end - 1, end\}$$

Pre-images for the encoding of each $s \in \mathcal{A}$:

$$I_s = \{x \mid C(s, x) \in I\}$$

Let b be the size of the output alphabet.
Interval I and all intervals I_s must be *b-absorbing*; i.e. there must exist l and l_s, for all $s \in \mathcal{A}$, such that:

$$I = \{l, l+1, \ldots, b \cdot l - 1\} \quad \text{and} \quad I_s = \{l_s, l_s + 1, \ldots, b \cdot l_s - 1\}$$

We consider only the case $b = 2$; so,

$$I = \{l, l+1, \ldots, 2l - 1\} \quad \text{and} \quad I_s = \{l_s, l_s + 1, \ldots, 2l_s - 1\}$$

and

$$|I| = l \quad \text{and} \quad |I_s| = l_s$$

The fact that, for instance, I is 2-absorbing means that for any $n \in \mathbb{N}$, $n \geq 1$, there exists a unique $k \in \mathbb{Z}$ such that:

$$\lfloor 2^k n \rfloor \in I$$

Encoding function for stream variant:

$$\vec{C} \;:\; \mathcal{A} \times I \to I$$
$$\vec{C}(s, x) = C\left(s, \vec{C}_{\text{aux}}(s, x)\right)$$
$$\vec{C}_{\text{aux}} \;:\; \mathcal{A} \times \mathbb{N} \to \mathbb{N} \qquad \text{// with output in } I_s$$
$$\vec{C}_{\text{aux}}(s, x) = \begin{cases} x, & \text{if } x \in I_s \\ \textbf{emit\_bit}(x \bmod 2);\ \vec{C}_{\text{aux}}\left(s, \lfloor \frac{x}{2} \rfloor\right), & \text{otherwise} \end{cases}$$

Decoding function for stream variant:

$$\vec{D} \;:\; I \to \mathcal{A} \times I$$
$$\vec{D}(x) = \textbf{let } (s, x') = D(x) \textbf{ in } \left(s, \vec{D}_{\text{aux}}(x')\right)$$
$$\vec{D}_{\text{aux}} \;:\; \mathbb{N} \to I$$
$$\vec{D}_{\text{aux}}(x) = \begin{cases} x, & \text{if } x \in I \\ \vec{D}_{\text{aux}}(2x + \textbf{receive\_bit}()), & \text{otherwise} \end{cases}$$

Probability distributions of the sources

First source: $\mathcal{A}_2$
a: $7/10$ b: $3/10$
Entropy: 0.8812908992306926

Second source: $\mathcal{A}_3$
a: $10/17$ b: $5/17$ c: $2/17$
Entropy: 1.3328204045850196

Third source: $\mathcal{A}_6$
a: $10/37$ b: $8/37$ c: $7/37$ d: $5/37$ e: $4/37$ f: $3/37$
Entropy: 2.4733643553620004

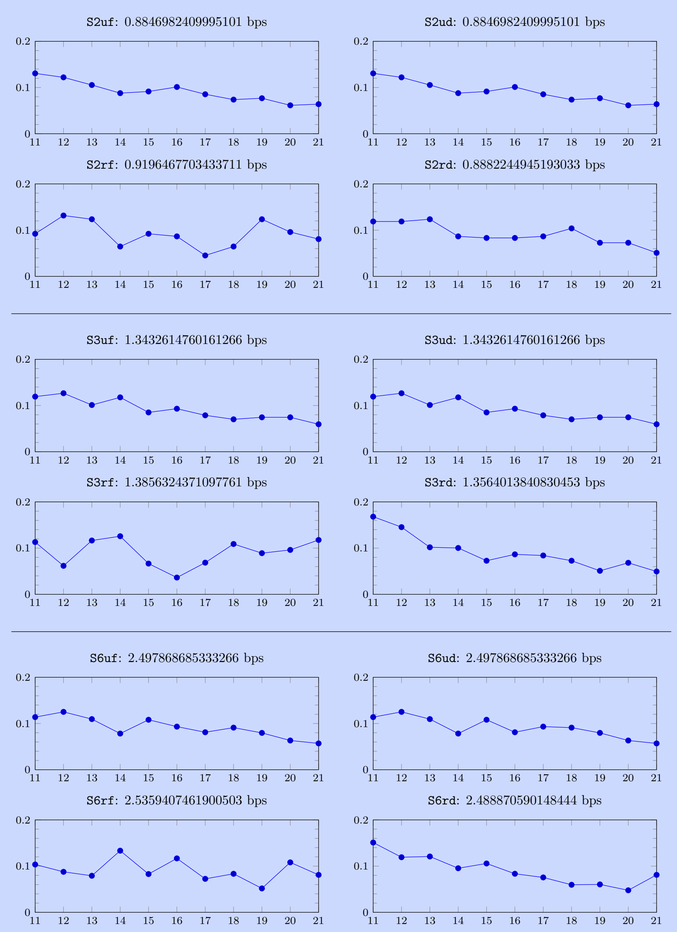
Code names of the experiments

S2uf

1. Size of I; may be:
   - S (for "small"), when $|I| = 11$;
   - L (for "large"), when $|I| = 121$.
2. Size of $\mathcal{A}$; may be:
   - 2, when using $\mathcal{A}_2$;
   - 3, when using $\mathcal{A}_3$;
   - 6, when using $\mathcal{A}_6$.
3. Variant of ANS used; may be:
   - u, when using uANS; cycle σ is:
     $$\textsc{Uniform\_Cycle}(I, W);$$
   - r, when using rANS; cycle σ is:
     $$\textsc{Range\_Cycle}(I, W);$$

   in both cases, $\overline{s}$ is σ^∞.
4. Weights for the states of I; may be:
   - f (for "flat"), when every state has weight 1; choosing W to be W_{flat};
   - d (for "decay"), when every state has weight following the $1/x$ curve; choosing W to be W_{decay}.

Average codeword lengths and stationary probability distributions when $|I| = 11$



Example of an ANS stream encoder

$$\overline{s} = \boxed{\texttt{a}}_{0}\boxed{\texttt{b}}_{1}\boxed{\texttt{a}}_{2}\boxed{\texttt{c}}_{3}\boxed{\texttt{a}}_{4}\boxed{\texttt{b}}_{5}\boxed{\texttt{a}}_{6}\boxed{\texttt{a}}_{7}\boxed{\texttt{b}}_{8}\boxed{\texttt{a}}_{9}\boxed{\texttt{a}}_{10}\boxed{\texttt{b}}_{11}\boxed{\texttt{c}}_{12}\boxed{\texttt{a}}_{13}\ldots$$

$I = \{7, 8, \ldots, 13\}$ $\quad I_{\texttt{a}} = \{4, 5, 6, 7\}$ $\quad I_{\texttt{b}} = \{2, 3\}$ $\quad I_{\texttt{c}} = \{1\}$

$s \backslash x$	7	8	9	10	11	12	13
a	$\frac{\epsilon}{\frac{7}{13}}$	$\frac{0}{\frac{4}{7}}$	$\frac{1}{\frac{4}{7}}$	$\frac{0}{\frac{5}{9}}$	$\frac{1}{\frac{5}{9}}$	$\frac{0}{\frac{6}{10}}$	$\frac{1}{\frac{6}{10}}$
b	$\frac{1}{\frac{3}{11}}$	$\frac{00}{\frac{2}{8}}$	$\frac{10}{\frac{2}{8}}$	$\frac{01}{\frac{2}{8}}$	$\frac{11}{\frac{2}{8}}$	$\frac{00}{\frac{3}{11}}$	$\frac{10}{\frac{3}{11}}$
c	$\frac{11}{\frac{1}{12}}$	$\frac{000}{\frac{1}{12}}$	$\frac{100}{\frac{1}{12}}$	$\frac{010}{\frac{1}{12}}$	$\frac{110}{\frac{1}{12}}$	$\frac{001}{\frac{1}{12}}$	$\frac{101}{\frac{1}{12}}$

Legend:

Emitted bits
$\vec{C}_{\text{aux}}(s, x)$
$\vec{C}(s, x)$

Another example of an ANS stream encoder

$$\overline{s} = \boxed{\texttt{a}}_{0}\boxed{\texttt{b}}_{1}\boxed{\texttt{a}}_{2}\boxed{\texttt{c}}_{3}\boxed{\texttt{a}}_{4}\boxed{\texttt{b}}_{5}\boxed{\texttt{a}}_{6}\boxed{\texttt{a}}_{7}\boxed{\texttt{b}}_{8}\boxed{\texttt{a}}_{9}\boxed{\texttt{a}}_{10}\boxed{\texttt{b}}_{11}\boxed{\texttt{c}}_{12}\boxed{\texttt{a}}_{13}\boxed{\texttt{a}}_{14}\boxed{\texttt{b}}_{15}\boxed{\texttt{a}}_{16}\boxed{\texttt{a}}_{17}\boxed{\texttt{b}}_{18}\boxed{\texttt{a}}_{19}\ldots$$

$I = \{10, 11, \ldots, 19\}$ $\quad I_{\texttt{a}} = \{6, 7, \ldots, 11\}$ $\quad I_{\texttt{b}} = \{3, 4, 5\}$ $\quad I_{\texttt{c}} = \{1\}$

$s \backslash x$	10	11	12	13	14	15	16	17	18	19
a	$\frac{\epsilon}{\frac{10}{17}}$	$\frac{\epsilon}{\frac{11}{19}}$	$\frac{0}{\frac{6}{10}}$	$\frac{1}{\frac{6}{10}}$	$\frac{0}{\frac{7}{13}}$	$\frac{1}{\frac{7}{13}}$	$\frac{0}{\frac{8}{14}}$	$\frac{1}{\frac{8}{14}}$	$\frac{0}{\frac{9}{16}}$	$\frac{1}{\frac{9}{16}}$
b	$\frac{0}{\frac{5}{18}}$	$\frac{1}{\frac{5}{18}}$	$\frac{00}{\frac{3}{11}}$	$\frac{10}{\frac{3}{11}}$	$\frac{01}{\frac{3}{11}}$	$\frac{11}{\frac{3}{11}}$	$\frac{00}{\frac{4}{15}}$	$\frac{10}{\frac{4}{15}}$	$\frac{01}{\frac{4}{15}}$	$\frac{11}{\frac{4}{15}}$
c	$\frac{010}{\frac{1}{12}}$	$\frac{110}{\frac{1}{12}}$	$\frac{001}{\frac{1}{12}}$	$\frac{101}{\frac{1}{12}}$	$\frac{011}{\frac{1}{12}}$	$\frac{111}{\frac{1}{12}}$	$\frac{0000}{\frac{1}{12}}$	$\frac{1000}{\frac{1}{12}}$	$\frac{0100}{\frac{1}{12}}$	$\frac{1100}{\frac{1}{12}}$

Construction of the Symbol Distributions

Algorithm 1 $W_{\text{flat}}(I)$

```
let l = |I|
let w⃗ ∈ {1}^l                    /* Cartesian product of sets */
return w⃗
```

Algorithm 2 $W_{\text{decay}}(I)$

```
let l = |I|                       /* I = {l, l+1, ..., 2l−1} */
return (1/l, 1/(l+1), ..., 1/(2l−1)) · l / Σ_{x=l}^{2l−1} 1/x
```

Algorithm 3 $\textsc{Uniform\_Cycle}(I, W)$

```
let k = |A|                       /* A = {a_0, a_1, ..., a_{k−1}} */
let p⃗ = (p(a_0), p(a_1), ..., p(a_{k−1}))
let l = |I|
let w⃗ = W(I)                      /* step weights */
let b⃗ ∈ {0}^k                     /* budgets */
let σ = ϵ
for x ← l to 2l − 1 do
    b⃗ ← b⃗ + w⃗[x] · p⃗
    let s = arg max_{s′∈A} b⃗[s′]
    σ ← σ · s
    b⃗[s] ← b⃗[s] − w⃗[x]
end for
return σ
```

Algorithm 4 $\textsc{Range\_Cycle}(I, W)$

```
assume p(a_0) ≥ p(a_1) ≥ ... ≥ p(a_{|A|−1})
let l = |I|
let w⃗ = W(I)                      /* step weights */
let σ = ϵ
for i ← |A| − 1 downto 1 do
    let m = arg min_{0<n<l} | Σ_{j=l−n}^{l−1} w⃗[j] / Σ_{j=0}^{l−1} w⃗[j]  −  p(a_i) / Σ_{j=0}^{i} p(a_j) |
    σ ← (a_i)^m · σ
    l ← l − m
end for
σ ← (a_0)^l · σ
return σ
```

Average codeword lengths per symbol in the experiments with $|I| = 121$

L2uf: 0.8813256469832618 bps	L2ud: 0.8817206984224625 bps
L2rf: 0.9038706726344816 bps	L2rd: 0.8882323216125733 bps
L3uf: 1.3328818091895864 bps	L3ud: 1.3328839957721896 bps
L3rf: 1.3619130277699671 bps	L3rd: 1.3415294833690325 bps
L6uf: 2.4735350806225553 bps	L6ud: 2.4735510002785075 bps
L6rf: 2.5202290464824446 bps	L6rd: 2.496463797855398 bps

Conclusions

- We presented a way to build pseudo-uANS symbol distributions.
- We proposed using varying weights for the auction on states according to the $1/x$ curve.
- We empirically demonstrated that using a larger I generally helps.
- uANS is generally superior to rANS.
- Using the weights modelled after the $1/x$ curve generally helps.