Context-Tree Weighting and Maximizing: Processing Betas

Frans Willems, Tjalling Tjalkens, and Tanya Ignatenko,
Eindhoven University of Technology,
Eindhoven, The Netherlands
VI. Binary Tree Sources (Example)

\[ P_a(X_t = 1 | \cdots, X_{t-1} = 1) = 0.1 \]
\[ P_a(X_t = 1 | \cdots, X_{t-2} = 1, X_{t-1} = 0) = 0.3 \]
\[ P_a(X_t = 1 | \cdots, X_{t-2} = 0, X_{t-1} = 0) = 0.5 \]
VII. Context-Tree Weighting

A context tree is a tree-like data-structure with depth $D$. Node $s$ contains the sequence of source symbols that have occurred following context $s$. 
Context-tree *splits up* sequences in subsequences.
Recursive weighting (WST 1995) yields the coding probability:

\[
P_s^w \triangleq P_e(a_s, b_s) \text{ for } s \text{ at level } D,
\]
\[
P_s^w \triangleq P_e(a_s, b_s) + \frac{P_w^0s \cdot P_w^1s}{2} \text{ for } s \text{ elsewhere.}
\]

for the subsequence that corresponds to node \( s \).

In the root \( \lambda \) of the context-tree the coding probability \( P_w^\lambda \) for the entire source sequence \( x_T^1 \).

**Total individual redundancy:**

\[
\rho(x_T^1) < \Gamma_D(S) + \left( \frac{|S|}{2} \log_2 \frac{T}{|S|} + |S| \right) + 2 \text{ bits,}
\]

where

\[
\Gamma_D(S) \triangleq 2|S| - 1 - |\{s \in S, \text{depth}(s) = D\}|.
\]

Asymptotically optimal (achieves Rissanen’s lower bound).
IX. Betas: Introduction

Consider an internal node $s$ in the context tree $T_D$ and the corresponding conditional weighted probability $P_w^s(X_t = 1|x_{1}^{t-1})$. Assuming that $0s$ (and not $1s$) is a suffix of the context $x_1^D, x_1^{t-1}$ of $x_t$, we obtain for this probability that

$$P_w^s(X_t = 1|x_1^{t-1}) = \frac{P_e(x_1^{t-1}, X_t = 1) + P_0w(x_1^{t-1}, X_t = 1)P_1w(x_1^{t-1})}{P_e(x_1^{t-1}) + P_0w(x_1^{t-1})P_1w(x_1^{t-1})}$$

$$= \frac{\beta s(x_1^{t-1})P_e(X_t = 1|x_1^{t-1}) + P_0w(X_t = 1|x_1^{t-1})}{\beta s(x_1^{t-1}) + 1}$$

(1)

where

$$s(x_1^{t-1}) \triangleq \frac{P_e(x_1^{t-1})}{P_0w(x_1^{t-1})P_1w(x_1^{t-1})}.$$  

(2)

If we start in the context-leaf and work our way down to the root, we finally find $P_w^\lambda(X_t = 1|x_1^{t-1})$. 
Implementation

Assume that in node $s$ the counts $a_s(x_{t-1}^1)$ and $b_s(x_{t-1}^t)$ are stored, as well as $\beta_s(x_{t-1}^t)$. We then get the following sequence of operations:

1. Node 0 sends cond. wei. probability $P_{w}^{0s}(X_t = 1|x_{t-1}^t)$ to node $s$. 
2. Cond. est. probability $P_{e}^{s}(X_t = 1|x_{t-1}^t)$ is determined as follows:
   \[
   P_{e}^{s}(X_t = 1|x_{t-1}^t) = \frac{b_s(x_{t-1}^t) + 1/2}{a_s(x_{t-1}^t) + b_s(x_{t-1}^t) + 1}.
   \]  
   (3)
3. Now $P_{w}^{s}(X_t = 1|x_{t-1}^t)$ can be computed as in (1).
4. The ratio $\beta_s(\cdot)$ is then updated with symbol $x_t$ as follows:
   \[
   \beta_s(x_{t-1}^t, x_t) = \beta_s(x_{t-1}^t) \cdot \frac{P_{e}^{s}(X_t = x_t|x_{t-1}^t)}{P_{w}^{0s}(X_t = x_t|x_{t-1}^t)}.
   \]  
   (4)
5. Finally, depending on the value $x_t$, either count $a_s(x_{t-1}^t)$ or $b_s(x_{t-1}^t)$ is incremented.
XV. Conclusion

- Betas simplify the implementation.

- Based on betas we can compute:
  - A posteriori probabilities,
  - MAP tree-model,
  - \( P_{w}^{\lambda}(X_t = 1|x_{1}^{t-1}) \) as convex combination of cond. estim. probabilities along context path,
  - difference between CTW and CTM codeword lengths.

- Similar results hold for weightings other than \((1/2, 1/2)\).