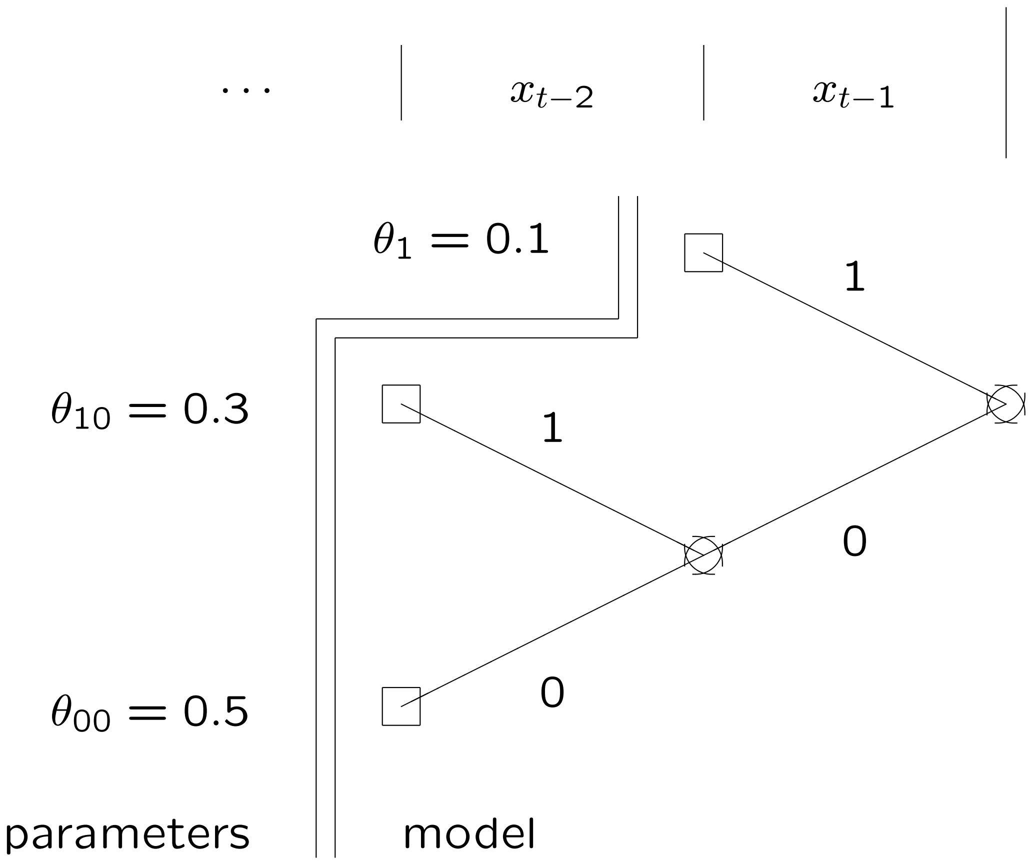
Context-Tree Weighting and Maximizing: Processing Betas

Frans Willems, Tjalling Tjalkens, and Tanya Ignatenko,
Eindhoven University of Technology,
Eindhoven, The Netherlands

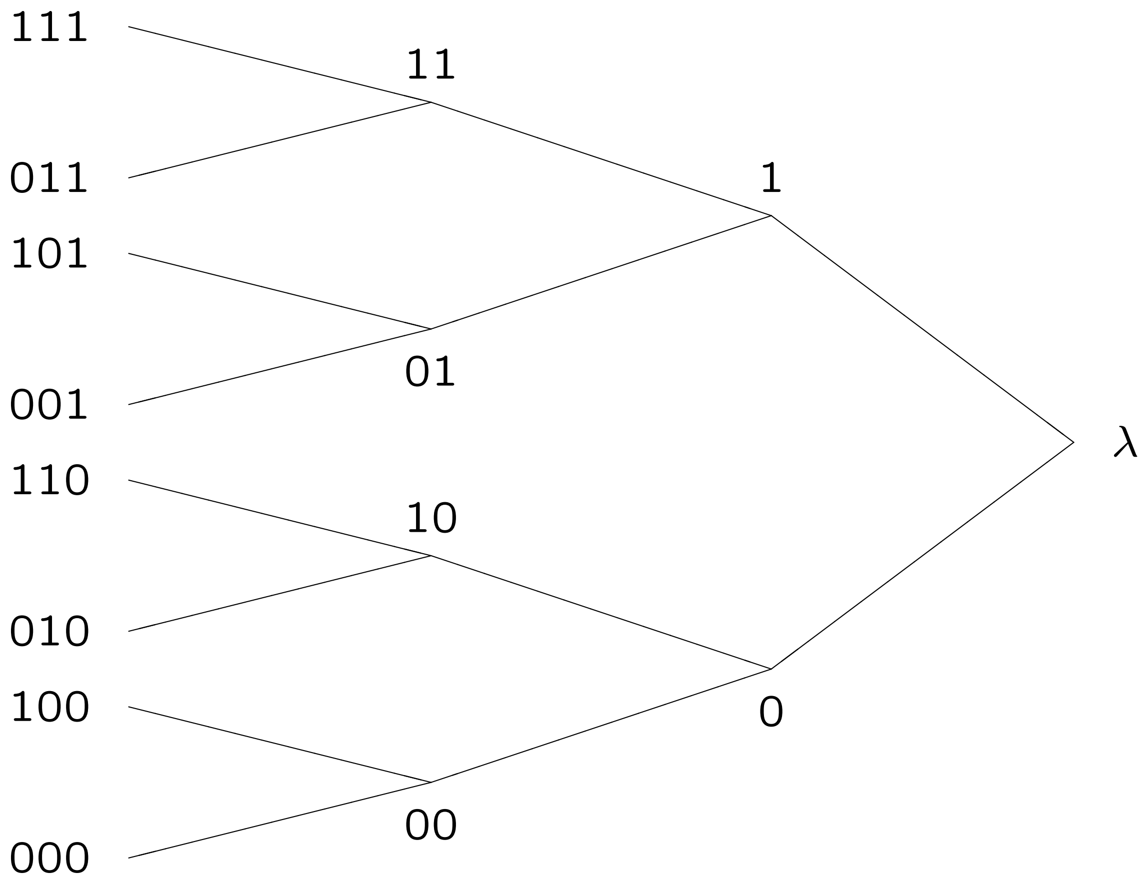
VI. Binary Tree Sources (Example)



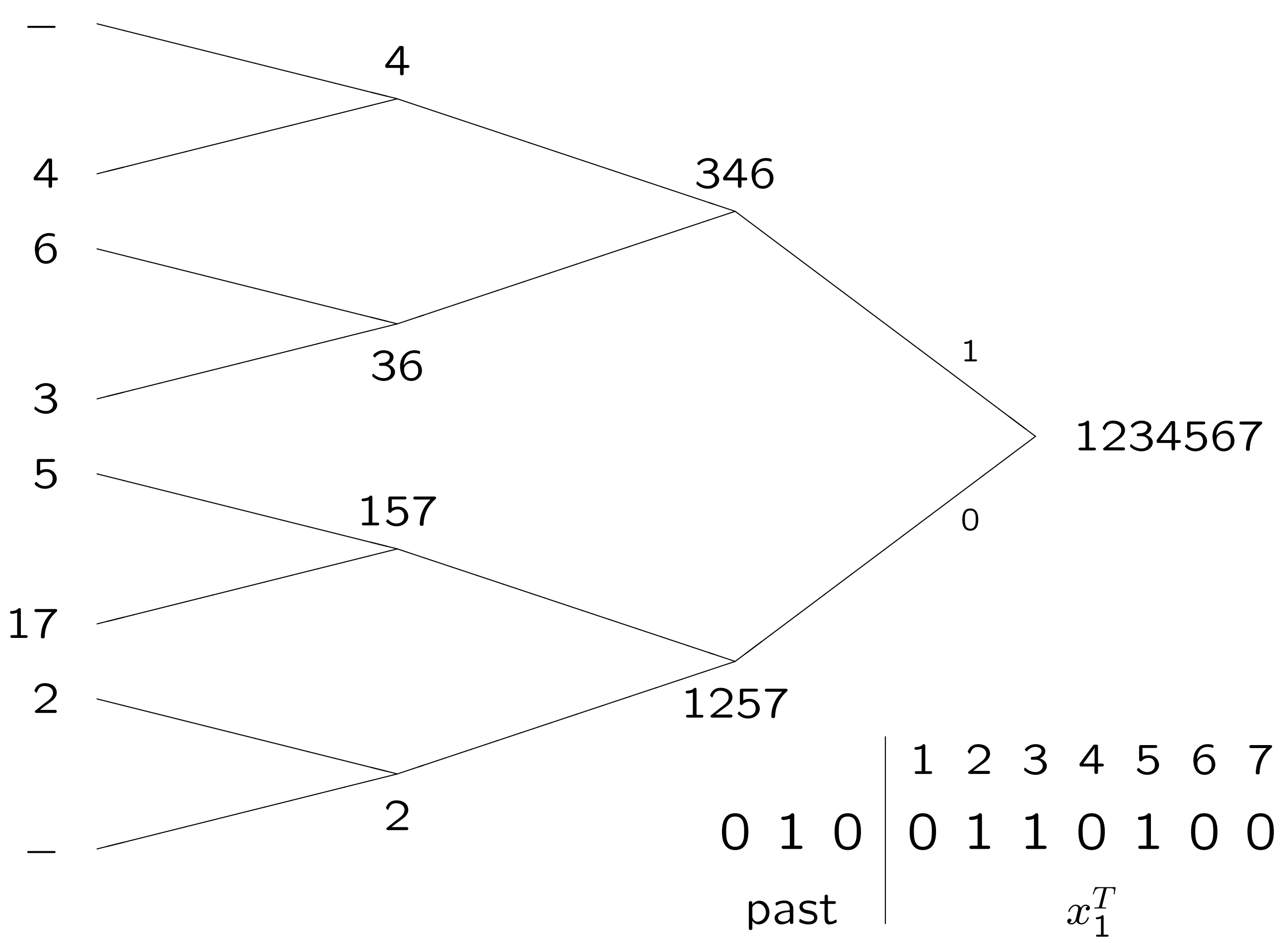
$$
\begin{aligned}
P_a(X_t = 1|\cdots, X_{t-1} = 1) &= 0.1 \\
P_a(X_t = 1|\cdots, X_{t-2} = 1, X_{t-1} = 0) &= 0.3 \\
P_a(X_t = 1|\cdots, X_{t-2} = 0, X_{t-1} = 0) &= 0.5
\end{aligned}
$$

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:

$$
\begin{aligned}
P_w^s &\triangleq P_e(a_s, b_s) \text{ for } s \text{ at level } D, \\
P_w^s &\triangleq \frac{P_e(a_s, b_s) + P_w^{0s} \cdot P_w^{1s}}{2} \text{ for } s \text{ elsewhere.}
\end{aligned}
$$

for the subsequence that corresponds to node s.

In the *root* λ of the context-tree the coding probability P_w^λ for the entire source sequence x_1^T.

Total individual redundancy:

$$\rho(x_1^T) < \Gamma_D(\mathcal{S}) + \left(\frac{|\mathcal{S}|}{2} \log_2 \frac{T}{|\mathcal{S}|} + |\mathcal{S}| \right) + 2 \text{ bits},$$

where

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

Asymptotically optimal (achieves Rissanen's lower bound).

IX. Betas: Introduction

Consider an internal node s in the context tree $\mathcal{T}_{\mathcal{D}}$ 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}^0, x_1^{t-1} of x_t, we obtain for this probability that

$$P_w^s(X_t = 1|x_1^{t-1}) = \frac{P_e^s(x_1^{t-1}, X_t = 1) + P_w^{0s}(x_1^{t-1}, X_t = 1)P_w^{1s}(x_1^{t-1})}{P_e^s(x_1^{t-1}) + P_w^{0s}(x_1^{t-1})P_w^{1s}(x_1^{t-1})}$$
$$= \frac{\beta^s(x_1^{t-1})P_e^s(X_t = 1|x_1^{t-1}) + P_w^{0s}(X_t = 1|x_1^{t-1})}{\beta^s(x_1^{t-1}) + 1} \quad (1)$$

where

$$\text{[redacted]}^s(x_1^{t-1}) \triangleq \frac{P_e^s(x_1^{t-1})}{P_w^{0s}(x_1^{t-1})P_w^{1s}(x_1^{t-1})}. \quad (2)$$

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 $a_s(x_1^{t-1})$ and $b_s(x_1^{t-1})$ are stored, as well as $\beta^s(x_1^{t-1})$. We then get the following sequence of operations:

1. Node $0s$ delivers cond. wei. probability $P_w^{0s}(X_t = 1|x_1^{t-1})$ to node s.

2. Cond. est. probability $P_e^s(X_t = 1|x_1^{t-1})$ is determined as follows:

$$P_e^s(X_t = 1|x_1^{t-1}) = \frac{b_s(x_1^{t-1}) + 1/2}{a_s(x_1^{t-1}) + b_s(x_1^{t-1}) + 1}. \tag{3}$$

3. Now $P_w^s(X_t = 1|x_1^{t-1})$ can be computed as in (1).

4. The ratio $\beta^s(\cdot)$ is then updated with symbol x_t as follows:

$$\beta^s(x_1^{t-1}, x_t) = \beta^s(x_1^{t-1}) \cdot \frac{P_e^s(X_t = x_t|x_1^{t-1})}{P_w^{0s}(X_t = x_t|x_1^{t-1})}. \tag{4}$$

5. Finally, depending on the value x_t, either count $a_s(x_1^{t-1})$ or $b_s(x_1^{t-1})$ 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 $(1/2, 1/2)$.