Exact Prediction and Universal Coding for Trees

Flemming Topsøe
Department of Mathematics,
University of Copenhagen,
Denmark
e-mail: topsoe@math.ku.dk

Abstract — **No closed formula for the optimal predictor (estimator) exists for the model of order preserving distributions on a finite tree. We present an algorithm of low complexity for a large class of trees.**

Let $T = (T, \leq)$ be a finite rooted tree and consider the model $\mathcal{P}$ of all distributions over the nodes of T which respect the ordering ($P \in \mathcal{P}$ means $P(a) \leq P(b)$ whenever $a \leq b$ – hence the root has minimal, the leaves maximal probability). For the unique *optimal predictor* $P^* \in \mathcal{P}$, $\sup_{P \in \mathcal{P}} D(P\|P^*)$ is minimal among all distributions in $\mathcal{P}$. It is our goal to determine P^* exactly.

Presently, results in this direction (exact rather than numerical or asymptotic results) are somewhat sporadic. The starting point of exact results of the type here considered is Ryabko [1]. The author, jointly with Peter Harremoës, has pointed to exact results for Bernoulli sources, cf. [3]. In [2] numerical methods are indicated, but based on the same theoretical reasoning as here (and with some associated exact results, not stated there). As a distant goal we mention the possibility to base asymptotically optimal prediction for, say Bernoulli sources on exact prediction results.

Our concern here is a case-study: No side information, models defined via a tree structure on the basic alphabet, taken finite. Even so, the problem is complicated and to further simplify, we limit the discussion: For a sequence $\mathbf{k} = (k_1, \ldots, k_n)$ of natural numbers, $T = T[\mathbf{k}]$ denotes the rooted tree with k_1 branches emerging from the root, with k_2 branches emerging from each node in level 1 etc. until k_n nodes emerging from each node in level $n-1$. Trees of this type have *uniform branching* and $\mathbf{k}$ is the *branching pattern.*

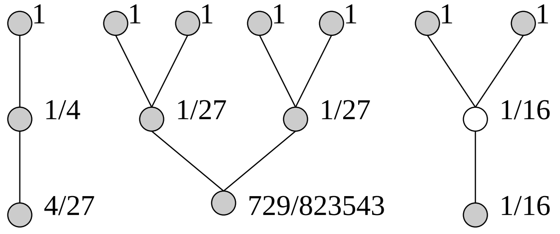

Fig. 1: Trees with optimal weights and marking of active anchors

The figure depicts the trees $T[1,1], T[2,2]$ and $T[1,2]$. For each of them, the optimal predictor is indicated by showing the weights assigned to each node. The last example is a bit surprising in that the root and its immediate successor are assigned the same weight, hence the same probability. The marked nodes show the *spectrum* $\sigma(T)$ of the trees. $\sigma(T)$ consists of the *active anchors*, nodes a for which the uniform distribution over $\{b | b \geq a\}$ contributes to the optimal predictor (i.e., the uniform distribution has positive weight in the unique convex combination representing P^*). For the two first examples, $\sigma(T) = T$ but for the last one, there is an inactive node. This phenomenon accounts for the contra-intuitive feature that nodes in different levels may have the same weight.

The key difficulty is the determination of $\sigma(T)$. Once known, it is easy to calculate P^* exactly. If $k_1 \geq \cdots \geq k_n$, then $\sigma(T) = T$, and the tree is *balanced.* The case $k_1 = \cdots = k_n = 1$ then essentially is Ryabko's result, [1].

For the general tree $T = T[\mathbf{k}]$ we base the analysis on the quantities $T_{i,j}, N_i, [i,j], \sigma_i, \alpha_i$ and ν defined as follows: $T_{i,j} = \Pi_{i+1}^{j} k_t$ (with special cases given by $T(i,i) = 1$ and $T(i, n+1) = 0$), $N_i = \sum_i^n T(i,j)$,

$$[i,j] = \frac{\bar{N}_i - T_{i,j+1}\bar{N}_{j+1}}{N_i - T_{i,j+1}N_{j+1}} \text{ for } 0 \leq i \leq j \leq n$$

(with $\bar{a} = a \ln a$), $\sigma_n = n$,

$$\sigma_i = \min\{j \geq i | [i,j] \geq [j+1, \sigma_{j+1}]\} \text{ for } i = n-1, \cdots, 0$$

and, finally, ν and α's are defined by $\alpha_0 = 0$, $\alpha_{i+1} = \sigma_{\alpha_i}$ for $0 \leq i < \nu$ and $\alpha_\nu = n$.

Theorem. *The spectrum $\sigma(T)$ consists of all nodes which lie in one of the levels $\alpha_0, \cdots, \alpha_\nu$. For $0 \leq i \leq \nu$, the optimal predictor assigns the probability*

$$\frac{e^{-[\alpha_i, \sigma_{\alpha_i}]}}{Z},$$

to all nodes in the levels $\alpha_i, \cdots, \sigma_{\alpha_i}$. Here, Z is a normalization factor.

Special structure guarantees that the algorithm is of low complexity. In fact, the maximal number of comparisons necessary (to determine the σ's) is $2n - 3$ for any branching pattern of length n. The author looks forward to present orally facts about the algorithm which, more understandable than by a series of definitions and formulas as here, will reveal the true nature of the algorithm. As we shall see, this depends on a set of combinatorial identities.

Acknowledgments

Discussions with Peter Harremoës and Boris Ryabko are gratefully acknowledged.

References

[1] B. Ya. Ryabko, “Coding of a source with unknown but ordered probabilities,” *Probl. Peredachi Inf.*, vol. 14, pp. 46–52, 1978.

[2] B. Ya. Ryabko and F. Topsøe, “Universal coding for sources with partially ordered probabilities,” these proceedings, p.xxx.

[3] F. Topsøe, “Instances of exact prediction and a new type of inequalities obtained by anchoring,” in Proc. of the 1999 IEEE Information Theory and Communications Workshop, (Kruger National Park, South Africa), p. 99, 1999.

[1] Research supported by a grant from the Danish Natural Science Research Council.