Universal coding for sources with partially ordered probabilities

Boris Ya. Ryabko
Siberian State University of Telecommunication and Computer Science, Novosibirsk, Russia
e-mail: ryabko@neic.nsk.su

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

Abstract — **We suggest an algorithm for the calculation of the optimal universal code for a source which generates letters with unknown probabilities but it is known that they are ordered according to a given partial order. This problem is well known in biology and other applications of Information Theory, see[1]. The algorithm is based on the relation between the redundancy of universal codes and channel capacity.**

I. Introduction

Let A be a set of objects. The binary tree L is a *retrieval tree* for A if, for every $a \in A$ there exists a leaf in L corresponding to a. Each node of the tree corresponds to an attribute. In order to use the tree for identification of an object, one should first check the attribute which corresponds to the root of the tree. If the object possesses that attribute, one should go to the left and if not, one should go to the right of the root. Moving in this way from one node to another, one finally reaches a leaf which is labelled with the name of the object. Let $L(a)$ be the length of the path from the root to the leaf. The smaller the average number of $L(\cdot)$, the shorter the time of identification. If it is known that $a \in A$ has probability $p(a)$, then the average cost of the tree L is defined as $c(L,p) = \sum_{a \in A} p(a)L(a)$. It is well known that the minimum of the average cost over the set of all trees is close to the *Shannon entropy* $h(p) = -\sum_{a\in A} p(a)\log p(a)$. The difference $r(L,p) = c(L,p) - h(p)$ is the *redundancy* of the tree L.

We address the problem of construction of the tree when the probability distribution is not known exactly. For instance, the probability to meet a biological object is never known exactly because it depends on the year, the month, the place, etc. On the other hand, there exists information about frequencies of occurrence of species. Thus, it is known that some species are observed very rarely (and a few examples are saved at zoological or botanical museums), and other species are observed quite often, but never in big numbers, and, at last, there exist a small number of species that are observed everywhere and quite often. Such information can be naturally presented as a partial order π on the set of probabilities of species occurrence. Let the set P_π contain all probability distributions $p(.)$ on A which correspond to the order π. (It means that if $a_i \le a_j$ according to the order π then $p(a_i) \le p(a_j)$ for every $p \in P_\pi$). For every retrieval tree L on A and every partial order π we define the *redundancy* of L for the model P_π by: $R(L, P_\pi) = \sup_{p \in P_\pi} r(L,p)$ We will say that a tree α is *optimal* if $R(\alpha, P_\pi)$ is minimal (over all trees).

II. The algorithm

In this report we suggest an algorithm for the construction of a (near–) optimal tree corresponding to an arbitrary order.

[1]Both authors have been supported by a grant from NATO Science Affairs Division.

Let us give some definitions. For two distributions p and q the *Kullback-Leibler divergence* is defined by : $r^*(p,q) = \sum_{a \in A} p(a)\log(p(a)/q(a))$. For a family of distributions P and a distribution q we define $R^*(P,q) = \sup\{r^*(p,q); p \in P\}$ and let $R^*(P) = \inf R^*(P,q)$ where inf is taken over all distributions q on A. Let P be an arbitrary family of probability distributions on A. Let $M(P)$ be the set of all probability distributions on the family P and by definition, $\mu(p)$ means the probability of a distribution $p \in P$. The *output distribution* $\gamma_\mu(.)$, the *mutual information* $I(\mu, P)$ and the *information rate* of a channel (or the *channel capacity*) $c(P)$ are defined by: $\gamma_\mu(a) = \sum_{p \in P} \mu(p)p(a), a \in A$, $I(\mu,P) = \sum_{p\in P} \mu(p) r^*(p, \gamma_\mu)$, $c(P) = \sup\{I(\mu,P); p \in P)\}$, see [2]. It is known that for every family of distributions P on A, $R^*(P) = c(P)$ and, if $c(P) = I(\nu, P)$ then $R^*(P) = R^*(P, \gamma_\nu)$ (this result was obtained by Gallager [3], but his paper was not published. Then, the result was independently found and published in [4]; see also the note[5] and editors comment after the note).

For an arbitrary partial order π on A let C_π contain all uniform distributions in P_π over connected subsets of A (*connected* subsets are defined in the natural way for partial orders and coincides with *pathwise connected* subsets).

Theorem 1. *Let A be a finite set and π a partial ordering on A. Then the maximal redundancy of any code for A is no less than the information rate $c(C_\pi)$ of the set C_π. On the other hand, there exists such a code L_π that its maximal redundancy is no more than $c(C_\pi) + 1$.*

Remark. It is important to note that C_π (in contrast to P_π) is finite. This is why a numerical algorithm can be used to find ν with $c(C_\pi) = I(\nu, C_\pi)$. Once ν is found, the desired Shannon code L_π can be constructed in the usual way from the distribution q_π given by $q_\pi(a) = \sum_{p\in P} \nu(p)p(a), a \in A$. Then $L_\pi(a) \le -\log q_\pi(a) + 1$ for $a \in A$.

Acknowledgments

We acknowledge contributions by Alexei A. Fedotov and by Peter Harremoës.

References

[1] R. Ahlswede, J. Wegener. Such probleme. *B. G. Teubner*, 1979.

[2] I. Csiszar and J. Körner. Information Theory. Coding Theorems for Discrete Memoryless Systems. *Akademiai Kiado*, Budapest, 1981.

[3] R.G. Gallager. Source coding with side information and universal coding. *Unpublished manuscript.*

[4] B. Ryabko. Encoding of a source with unknown but ordered probabilities. *Problems of Information Transmission*, vol.15, no.2, 1979, pp.71-77.

[5] B. Ryabko. Comments on "A source matching approach to finding minimax codes". *IEEE Trans. Information Theory*, vol.27, no. 6, 1981, pp.780-781.