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 of calculations 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, mineralogy and other applications of Information Theory, see [1]. This algorithm is based on the relation between the redundancy of universal codes and channel capacity.**

I. Introduction

Let A be a set of objects. By definition, the binary tree L is a *retrieval tree* for the set A if for every a from A there exists in L a leaf corresponding to a. Each node of the tree corresponds to an attribute. If somebody wants to use the tree for identification of an object he should do it as follows: first, check the attribute which corresponds to the root of the tree. If the object possesses that attribute, go to the left of the root; if it does not, go to the right. Moving in this way from one node to another, one finally reaches a leaf which is labelled with the name of the object. Let us denote by $L(a)$ the length of the path from the root to the leaf. Of course, the smaller the average number of $L(.)$, the smaller the time of identification and the better the tree. More exactly, if it is known that an object $a \in A$ has probability $p(a)$, then, it is natural to define the average cost of the tree L by: $c(L,p) = \sum_{a\in A} p(a)L(a)$. It is well known that the minimum of the average cost $c(L,p)$ over the set of all trees is close to the Shannon entropy $h(p) = -\sum_{a\in A} p(a)\log p(a)$ and the difference $r(L,p) = c(L,p) - h(p)$ is defined to be 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 all i, j). 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 an optimal tree corresponding to an arbitrary order.

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 channel capacity) $c(P)$ are defined as follows: $\gamma_\mu(a) = \sum_{p\in P} \mu(p)p(a), a \in A$, $I(\mu,P) = \sum_{p\in P} \mu(p) \sum_{a\in A} p(a)\log(p(a)/(\sum_{p\in P} \mu(p)p(a))$, $C(P) = \sup\{I(\mu,P); p \in P)\}$,see [2]. It is also known in Information Theory 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 a note[5] and editor's comment after the note).

For an arbitrary partial order π on A let C_π contain all distributions on A that can be presented as permutations of coordinates of one of the vectors $(1,0,\dots,0)$, $(1/2,1/2,0,\dots,0)$, $\dots$, $(1/n,1/n,\dots,1/n)$ and each distribution should correspond to the order π.

Theorem 1. *Let A be a finite set and π be an partial order 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 redundancy is no more than $c(C_\pi)+1$.*

Remark To make such a code, take the distribution ν for which $c(C_\pi) = I(\nu, C_\pi)$ and develop for the distribution $q_\pi(a) = \sum_{p\in P} \nu(p)p(a), a \in A$, the Shannon code L_π for which $L_\pi \le \log q_\pi(a) + 1$.

It is important to note that C_π (in contrast to P_π) is finite, that is why a numerical algorithm can be used to find ν.

References

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

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

[3] Gallager R.G. 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*, v.15, n.2,1979, pp.71-77.

[5] B.Ryabko.Comments on "A source matching approach to finding minimax codes".*IEEE Trans. Information Theory*, v.27,n 6, 1981, pp.780-781.