INTERVALS AS ULTRAMETRIC APPROXIMATIONS ACCORDING TO THE SUPREMUM NORM

Bernard Fichet

Abstract: *Given two distances d and d' defined on a finite set I, with $d \leq d'$, we characterise the set of all ultrametrics lying between d and d' (if any), showing they are the union of finitely many intervals with a common top endpoint, the subdominant ultrametric of d'. We also provide an algorithm to compute the bottom of any interval, given by any upperminimal ultrametric of d less than d'. A series of approximations according to the supremum norm derive from this.*

Keywords: *ultrametric, subdominant ultrametric, upperminimal ultrametrics, approximation, supremum norm*

MSC*: G3: Statistics*

Given a dissimilarity (or a distance) d defined on a finite set I, we here propose an algorithm for computing an upperminimal ultrametric d^* of d bounded by a fixed dissimilarity d' (if any), so yielding a solution of the sandwich problem : find an ultrametric lying between d and d' (we recall that an ultrametric u on I obeys the ultrametric inequality: $u(i,j) \leq max[u(i,k), u(j,k)]$ for all i, j, k in I).

As observed by Chepoi and Fichet (2000), the latter problem admits a very simple characterisation via a subdominant approximation. Namely, the subdominant ultrametric d'_* of d' is (the greatest) solution, provided that it is greater than d (a contrario, there is no solution). Moreover, it is well known that there are efficient algorithms for computing a subdominant, such as the one defined by the single linkage procedure.

Similar results hold with upperminimal ultrametric approximations of d: the sandwich problem has a solution if and only if there is an upperminimal ultrametric of d which is less than d'. We recall that, as opposed to the subdominant which is the best lower approximation, there are finitely many best upper approximations, the upperminimal ultrametrics of d. The complete linkage algorithm builds a particular approximation (several in case of ties), but it was not before the eighties that some dividing procedures compute any of them (Van Cutsem (1984), Leclerc (1986)).

Our algorithm extends the latter ones, in order to satisfy the boundary constraint. It is recursive and (formally) depicts all the solutions (if any). It involves a so-called d-admissibility condition of a given partition of I, and a threshold graph G associated with d'. At a given step, we know whether there is no solution, or there is a possible one related to a new minimal partition less than the one given by the connected components of G.

Thus, the set of solutions of the sandwich problem appears as the finite union of intervals of the type $[d^*, d'_*]$, with common top endpoint d'_*. A series of derivative problems may be solved by this way, such as :

· Characterising all ultrametrics at a least supremum norm of d and d', $d \leq d'$, so recovering the results of (Chepoi-Fichet, 2000), and (Farach et al., 1995).

· Given d, find the minimum $\lambda \geq 1$, such that there is an ultrametric lying between $(1/\lambda)\, d$ and λd.

· Given a partial dissimilarity d, characterising all ultrametric extensions of d.

Authors' Information

Bernard Fichet *- Aix-Marseille University. Laboratoire d'Informatique Fondamentale. bernard.fichet@lif.univ-mrs.fr*