PriorID: Model and software for the construction of rankings to prioritize the search of unlocated people as support in the identification process.

Short description:

The model extracts knowledge provided by the already-solved cases of a closed and bounded particular event to be used in new searches of the same event, by constructing probabilistic priorities (rankings) within a Bayesian inference frame. The methodology includes cross-validation of the results using reserved information, and is implemented using software based on R.

Methodology

This software implements the methodology developed to construct prioritized rankings of possible victims that could correspond to each one of the non-identified skeletal remains of a certain event involving deaths, taking advantage of the information extracted from a fraction of already solved-cases within the same event. The objective of this software is to assist researchers with the work related to background information by prioritizing cases.

The key consists in defining cells on the basis of a set of two, three, etc, non-genetic variables associated with the massive event (for example, geographical and temporal variables associated to the death of every individual). Each cell is associated to a particular set of values (or a partition value) for each of the non-genetic variables considered. The crucial point is that for a certain massive event, well limited in time, place and total number of dead individuals, there is a set of possible victims and a set of already-solved cases.

The methodology learns from the information of the cells of the already-solved cases in order to update a prior probability to a posterior probability, which is later used to construct the rankings of individuals for each skeletal remains. The assumptions made and the Bayesian statistical details are described in *Statistical Calculations*.

There is a set of functions which enables the addition of other information like anthropological data of the skeletal remains, such as gender and age at the time of death. In these cases, the results include only those possible victims compatible with this information. Another function implements a cross-validation analysis using reserved information to calculate method success and effectiveness.

An open, free, multi-platform and standalone interface for users to implement this methodology in diverse problems and incorporate feedback from forensic researchers is now under construction.

Details of the Functions

rankings(V,S)
This function generates a ranking of individuals from a set of individuals, each one associated with a cell. Moreover a plot of probabilities is generated. Ccells are numbered from 1 to m (see statistical calculations).

V is a matrix with the information of the individuals (ID in column 1, cell in column 2)
S is a vector with the ID of the already-solved cases.

rankings.antro(Va,S,gender_s,age_s)
This function generates a ranking of individuals compatible with the anthropological information of some skeletal remains from a set of individuals, each one associated with a cell, gender and age at the time of death. Moreover a plot of probabilities is generated. Cells are numbered from 1 to m (see statistical calculations).

Va is a matrix with the information of the individuals (ID in column 1, cell in column 2, age at the time of death in column 3, gender in column 3 (F or M))
S is a vector with the ID of the already-solved cases
gender_s is the gender of the skeletal remains ("F" or "M")
age_s is a vector with the information of the estimated age of the skeletal remains at the time of death (a sequence of two values, the lower and upper limits)

rankings.cv(V,S,rounds)
This function implements a cross-validation analysis by selecting 75 % of the already-solved cases as the learn sample and the remaining 25 % of the reserved data as the reserved data set for the evaluation of the results. Two magnitudes are calculated: *Discriminating Power* (success), which is the ratio between the size of the reserved cases which improve their scores respect to the initial instance of knowledge and the total number of cases which improve their scores) and *Efficacy Rate (*effectiveness), which is the ratio between the reserved cases which obtain odds posterior greater than odds priors and the total number of cases which obtain odds posterior greater than odds priors. Several independent realizations of this cross validation are implemented. Parameter "rounds" is the number of realizations. Standard deviations of the *Discriminating Power* and *Efficacy Rate* are calculated from these realizations.
rounds is the number (integer number) of realizations of cross validation performed.

rankings.cv.random(V,S,rounds)
This function generates the same results as rankings.cv and also shows the values of *Discriminating Power* (success) and *Efficacy Rate* (effectiveness) (and their standard deviations) that are obtained by choosing the same ODDS distribution that is generated with the methodology, but randomly assigning those values among the individuals.