

PriorID: Modelo y software para la construcción de rankings para priorizar la búsqueda de personas no localizadas como apoyo en el proceso de identificación.

Breve descripción:

El modelo extrae el conocimiento brindado por los casos resueltos de un evento particular cerrado y acotado para que sea usado en nuevas búsquedas del mismo evento, mediante la construcción de rankings de prioridades, dentro de un marco de inferencia bayesiana. La metodología incluye la validación cruzada de los resultados usando información reservada y se implementa mediante un software basado en R.

Metodología:

Este software implementa la metodología desarrollada para construir rankings de prioridades de posibles víctimas candidatas a corresponderse con cada uno de los esqueletos o restos no identificados de un determinado evento de muerte, para el cual hay una fracción de casos ya resueltos, a partir de los cuales se buscará aprender. El objetivo de este software es ayudar a los investigadores con el trabajo relacionado con la investigación preliminar, priorizando los casos.

La clave consiste en definir casilleros en base a un conjunto de dos, tres, etc. variables no genéticas asociadas con el evento masivo de muerte (por ejemplo, variables geográficas y temporales asociadas a la muerte de cada individuo). Cada casillero está asociado a un conjunto particular de valores (o un valor de partición) de cada una de las variables no genéticas consideradas. El punto crucial es que para un cierto evento masivo, limitado en el tiempo, el lugar y en cuanto al número total de individuos muertos, hay un conjunto de posibles víctimas y un conjunto de casos ya resueltos.

La metodología aprende a partir de la información de los casilleros observados de los casos ya resueltos para actualizar una probabilidad a priori en una probabilidad a posteriori de los casos ya resueltos, que luego se utiliza para construir los rankings de individuos para cada esqueleto buscado. Los supuestos asumidos y los detalles estadísticos bayesianos se describen en *Cálculos Estadísticos*.

Existe un conjunto de funciones que permiten la adición de otra información como la información antropológica de los restos óseos, como el sexo y la edad en el momento de la muerte. En estos casos, los resultados incluyen sólo las posibles víctimas compatibles con esta información. Otra función implementa un análisis de validación cruzada usando información reservada para calcular éxito y eficacia del método.

Detalles de las funciones

rankings(V, S)

Esta función genera un ranking de los individuos a partir de un conjunto de individuos, cada uno asociado con un casillero. Además, se genera una gráfico con las ODDS a posteriori. Los casilleros están numerados del 1 al m (ver cálculos estadísticos).

V es una matriz con la información de los individuos (ID en la columna 1, casillero en la columna 2)

S es un vector con el ID de los casos ya resueltos del evento considerado

rankings.antro(Va,S,gender_s,age_s)

Esta función genera un raning de los individuos compatibles compatibles con la información antropológica de los restos óseos a partir de un conjunto de individuos, cada uno asociado con un casillero,sexo y edad en el momento de la muerte. Además, se genera una gráfico con las ODDS a posteriori. Los casilleros están numerados del 1 al m (ver cálculos estadísticos).

Va es una matriz con la información de los individuos (ID en la columna 1, casillero en la columna 2, edad en el momento de la muerte en la columna 3, sexo en la columna 3 (F o M))

S es un vector con el ID de los casos ya resueltos del evento considerado.

gender_s es el género de los restos óseos ("F" o "M")

age_s es un vector con la información de la edad estimada de los restos óseos al momento de la muerte (una secuencia de dos valores, los límites inferior y superior del rango de edad)

rankings.cv(V,S,rounds)

Esta función implementa un análisis de validación cruzada seleccionando el 75% de los casos ya resueltos del evento (S) como muestra de aprendizaje y el 25% restante como datos reservados para la evaluación de los resultados. Se calculan dos magnitudes: el éxito (la fracción de los casos reservados que obtienen ODDS a posteriori mayores que las ODDS a priori) y la efectividad (la relación entre los casos reservados que obtienen ODDS a posteriori mayores que las ODDS a priori y el número total de casos que obtienen obtienen ODDS a posteriori mayores que las ODDS a priori). Se implementan varias realizaciones independientes de esta validación cruzada. El parámetro "rondas" es el número de realizaciones. La desviación estándar del éxito y la efectividad se calculan a partir de estas realizaciones. Los casilleros están numerados del 1 al m (ver cálculos estadísticos).

rounds es el número total (entero) de realización independientes para implementar la validación cruzada.

rankings.cv.random(V,S,rounds)

Esta función genera los mismos resultados que rankings.cv y además muestra los valores de éxito y efectividad (y sus desviaciones estándar) que se obtienen eligiendo la misma distribución de ODDS que se genera con la metodología, pero distribuyendo esos valores al azar entre los individuos.