

Un método de seriación automática

El objeto de este artículo es la presentación de un método de seriación automática. En principio, éste ha sido concebido para ser aplicado a los materiales recogidos por la Misión Arqueológica Española en el Ecuador, que constan fundamentalmente de unas ciento treinta y cinco colecciones cerámicas de superficie, resultado de la exploración de una zona de la costa de la provincia de Esmeraldas y de los recuperados en dieciocho excavaciones estratigráficas, repartidas entre siete yacimientos seleccionados, concentrándose los trabajos en tres de ellos: Balao, La Propicia y Atacames. Una información más detallada acerca de estos materiales y de los planteamientos generales de la investigación puede encontrarse en las numerosas publicaciones sobre el tema (Alcina, 1977).

En principio, el problema que tratamos de resolver con el método que presentamos es la seriación de las colecciones cerámicas de superficie, para las cuales no disponemos de otros medios posibles de datación.

En arqueología se entiende por seriación una manera particular de ordenación, que coloca las unidades bajo estudio en una serie secuencial de acuerdo a su posición cronológica.

La idea base que nos permite reconstruir un orden cronológico es «el hecho establecido empíricamente, que en el transcurso del tiempo los tipos cerámicos entran y salen en el uso habitual de un pueblo», y que «la forma en que un determinado tipo entra y sale del uso habitual es lenticular» (Robinson, 1951: 293).

Esta definición, que fue una de las primeras, ha sido luego completada y matizada por muchos otros autores. Por ejemplo, el desarrollo lenticular también se observa en las variaciones que ocurren a la

hora de manufacturar artefactos, en la presencia o ausencia de determinado objeto en un estrato o tumba, etc.; por su parte, el término lenticular ha sido sustituido por definiciones matemáticas más precisas, pero hemos seleccionado ésta porque es la de más fácil comprensión intuitivamente.

Los problemas surgen a la hora de ordenar las unidades implicadas de acuerdo a este criterio, y es aquí donde nos vemos obligados a hacer un poco de historia, ya que el método que nos va a ocupar propone un enfoque diferente para la solución de algunos problemas tradicionales.

En el modelo propuesto por Brainerd y Robinson, las unidades a seriar (yacimientos o niveles estratigráficos) representan instantes temporales y vemos que su emplazamiento en el eje temporal es una función de similaridad, de manera que aquellas unidades con fuerte similaridad en la lista cuantitativa o cualitativa de los tipos deben de caer cerca unas de otras en la secuencia temporal (Brainerd, 1951: 304).

Partiendo de este supuesto, elaboran un método de trabajo en la forma siguiente:

1. Dado el porcentaje de cada tipo que aparece en cada yacimiento, definen una medida de similaridad entre cada par de yacimientos.

2. Calculan la medida de similaridad entre cada par de yacimientos. Las medidas así calculadas se disponen en una matriz cuyas filas y columnas son los yacimientos en orden arbitrario.

3. La matriz se reordena, filas y columnas, hasta conseguir que los valores máximos estén concentrados en torno a la diagonal principal. El orden así obtenido se considera el orden cronológico de los yacimientos.

Con estos planteamientos, los problemas relacionados con la seriación se centran en tres: la forma más adecuada de encontrar la secuencia ideal; el criterio para elegir entre varias secuencias alternativas no del todo ideales, cuando la ideal no exista, y en cómo vamos a medir la similaridad, implicando esto últimas decisiones acerca de los rasgos de las unidades que deben de ser tenidos en cuenta y decisiones acerca de las funciones matemáticas que se han de emplear en el cálculo (Cowgill, 1972: 383).

Con referencia al primer punto, la forma de encontrar la secuencia, Robinson utiliza un sistema manual de reordenamiento de filas y columnas bastante laborioso, y que de cualquier forma sólo puede ser emprendido si el número de yacimientos implicado es pequeño. El primer intento de resolver este problema mediante el empleo de un

ordenador lo realizan Ascher y Ascher (1963). El proceso seguido sería básicamente como sigue:

1. Empleando números aleatorios se determina la matriz de partida.
2. Modificadores u operadores hacen una determinada permutación; cada vez que la aplicación de un operador mejora el orden, el cambio es retenido. Cuando después de haber aplicado la lista completa de operadores no conseguimos mejorar el orden, hemos alcanzado un mínimo local.
3. El proceso es repetido numerosas veces con distintas matrices de partida. El mejor de los ordenamientos finales es la solución que devuelve el programa.

Después de Ascher se ensayan otros sistemas de permutaciones para las matrices de similaridad; Kuzara, Mead, y Dixon, en 1966; Hole y Saw, en 1967, y Craitor & Johnson, en 1968, son algunos ejemplos. En los últimos años representa una innovación el paso a permutar directamente la matriz de datos, sin necesidad de convertirla en matriz de similaridad, y se aportan diversas técnicas para realizarlo (Kendall, 1971; Doran, 1971; Goldman, 1972), pero unos y otros continúan realizando las permutaciones mediante un proceso heurístico, que tiene serias limitaciones.

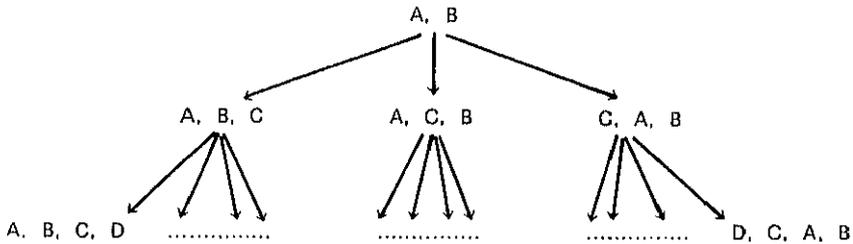
Esta manera heurística de realizar las permutaciones es un intento de obviar el problema que supone el rápido crecimiento de la función factorial, de forma que si para cinco yacimientos el número de órdenes posibles es de 120 y para ocho de 40.320, para once se acercaría a los 40 millones. Al aplicar un método heurístico, el orden que obtenemos no es el mejor de los posibles, pero sí suficientemente bueno de acuerdo a un test, que realiza el programa antes de cada iteración. El propio Kendall se refiere a este asunto en los siguientes términos: «Incluso en este problema matemático, un ataque directo (implicando un escrutinio de todas las posibles permutaciones de la matriz) puede ser a menudo impracticable, y, por tanto, esto conduce a buscar procedimientos más simples "heurísticos", la eficacia de los cuales sería asunto del *crédito* que se les conceda más que de *conocimiento*» (1969: 72; el subrayado es nuestro).

El método propuesto por Bordaz y Bordaz (1970) utiliza otro recurso para solucionar este problema. Emplea la función de Tanimoto, utilizada en Teoría de la Información, para seleccionar la más correcta entre varias recepciones de un mismo mensaje. Se considera que el mensaje que nos quieren transmitir los yacimientos es el instante

temporal que representan. De un subconjunto de yacimientos aproximadamente contemporáneos, y que, por tanto, nos quieren transmitir el mismo mensaje, tomamos solamente aquel que nos lo transmite con más exactitud; de esta forma reducimos el problema de ordenar cien yacimientos a ordenar unos diez que representan los instantes temporales básicos en torno a los cuales posteriormente ordenaremos todos.

El método que nosotros proponemos es el siguiente:

- Para un conjunto de N yacimientos (o niveles estratigráficos) y otro conjunto de tipos cerámicos (o clases), nos proponemos obtener, de forma automática, todas las secuencias de yacimientos que sean lenticulares con respecto al conjunto de tipos dado.
- Todas las secuencias posibles para N yacimientos, las obtenemos mediante un proceso arborescente. Designamos los yacimientos por letras A, B, C, ...



Para realizar este árbol partimos de dos yacimientos y obtenemos un nuevo nivel incorporando otro yacimiento, que colocaremos en todos los lugares posibles, partiendo de los nodos terminales del nivel anterior. Así hasta agotar los N yacimientos. Obtenemos de este modo $N!/2$ secuencias posibles.

Otro árbol similar a éste, pero empezando con B, A, no lo consideraremos, pues las secuencias que nos daría serían las inversas de las ya obtenidas, y dado que el criterio de selección de las secuencias válidas no nos permite distinguir entre una secuencia y su inversa, sería un trabajo inútil.

- Observemos que si una secuencia intermedia, por ejemplo, A, B, C, no es unimodal (lenticular) con respecto a un tipo, ya no lo será en ninguna de las que se formen a partir de ella, y, por tanto, podemos prescindir de éstas en nuestra búsqueda. De esta forma, para un solo tipo cerámico las posibilidades quedan reducidas a 2^n . Si ahora imponemos que la secuencia sea válida para más de un tipo cerámico simultáneamente, los posibles órdenes se ven drásticamente reducidos, siempre que no utilicemos tipos con perfil idéntico.

El proceso descrito en los últimos puntos es realizado automáticamente por el ordenador. La herramienta con la que contamos es la siguiente: dado un conjunto de N yacimientos, de los que conocemos la frecuencia relativa por lo menos de dos tipos (o clases), podemos saber todas las secuencias en las que esos tipos son unimodales.

Ya hemos dicho que nuestro primer objetivo fue la obtención de la cronología relativa; para llegar a ésta tenemos aún que analizar dos problemas: 1.º) En el conjunto de yacimientos propuesto, puede haber uno o varios que pertenezcan a distinto complejo cultural. 2.º) Que el conjunto de tipos propuesto no sea coherente (para nosotros, un conjunto de tipos es coherente, si éstos son unimodales con respecto a la misma variable), o que siendo coherentes no expresen tiempo, sino distribución espacial, función, relaciones sociales, etc., que también pueden presentar un desarrollo lenticular.

El primero lo resolvemos en parte incorporando a nuestro logaritmo las etapas necesarias para que un yacimiento, que no produzca órdenes válidos, sea eliminado de la secuencia.

Para solucionar el segundo efectuamos una búsqueda sistemática de todos los conjuntos de tipos coherentes. Esta la realizará el ordenador, a partir de los datos que nosotros le facilitemos, descendiendo en nuestro nivel de análisis hasta los atributos de los fragmentos cerámicos, formando clases en función de cada uno de los atributos reconocidos. Pudiendo posteriormente combinar éstos, para obtener tipos cerámicos «tradicionales», si esto fuera más conveniente para la investigación. Si lo que tuviésemos fuera una «matriz de ocurrencia», indicándonos sólo la presencia o ausencia del atributo en un fragmento u objeto, ésta podría ser igualmente tratada con este mismo programa.

En el caso de los materiales de Esmeraldas, la posibilidad de utilizar varias clasificaciones de nuestros datos es especialmente importante, dado que el mal estado de conservación en las cerámicas a causa del clima de la región, podría invalidar algunas de éstas, al introducir errores en la clasificación.

Queremos hacer resaltar aquí algunas de las características del programa que empleamos:

1.º Los conjuntos de tipos que empleamos no tienen por qué agotar las colecciones cerámicas (dicho en términos más precisos, el conjunto de tipos no es una partición de nuestro universo cerámico).

2.º Hemos solucionado el problema del crecimiento del factorial, sin renunciar al conocimiento de todos aquellos órdenes que cumplen el criterio deseado. Ya hemos visto cómo el sistema heurístico de los métodos anteriores nos ofrecía un solo orden final, aunque pudiera

haber más; incluso éste podía no ser el mejor, sino simplemente bastante bueno.

3.º Al realizar las permutaciones, fijamos nuestra atención no en las similitudes que presentan los yacimientos entre sí, sino en la búsqueda de aquellos tipos que, creciendo hacia un solo pico de frecuencia de ocurrencia o relativa abundancia, decrecen más tarde hacia un nivel de poca o ninguna ocurrencia, es decir, que evolucionan de la manera a la que normalmente nos referimos como desarrollo lenticular.

Con referencia al segundo de los puntos, pensamos que la recuperación de todos los órdenes es importante, ya que amplía las posibilidades interpretativas de los resultados. El método de seriación como tal ha sufrido duras críticas, que han llevado a algunos autores (Doran, 1975; McNutt, 1973) a exigir para tomar un consideración los resultados, que éstos sean cuestionados e interpretados seriamente y exhaustivamente.

Ciertamente, la seriación cronológica tropieza con la dificultad, ya señalada antes, de que las unidades seriadas pueden diferir por otras razones no relacionadas con el transcurso del tiempo, como las espaciales, funcionales o sociales con desarrollo lenticular; un caso evidente serían los fenómenos de difusión.

Una seriación sólo sería fiable en el caso de que elimináramos de nuestros datos todas aquellas otras variables que no fueran debidas al tiempo. Tratando de paliar esto, la mayoría de los autores le ponen determinadas condiciones a las unidades a seriar, pero aun así no consiguen resolver el conflicto. Estamos de acuerdo con los autores arriba citados en que la interpretación es la solución más adecuada, y para ello es fundamental el conocimiento de todos los órdenes correctos posibles.

De esta forma llegamos a las siguientes conclusiones: si nos proponemos la búsqueda del orden cronológico, utilizaremos tipos, o clases, haciendo con ellos todos los conjuntos que creamos que pueden ser coherentes, ofreciéndoselos como datos al ordenador; éste recuperará las secuencias válidas que se produzcan para cada conjunto de tipos. Obtendremos así secuencias cronológicas en función de los conjuntos de tipos que tengan significación temporal. Pero tendremos además otras secuencias, que expresaran difusión, relaciones estructurales, distribución espacial, etc., en función de los conjuntos de tipos que expresen estas variables.

La labor del arqueólogo consistiría en la interpretación del significado real de cada una de estas secuencias y se facilitaría con la inclusión en éstas de yacimientos ya datados, o de unidades cuya posi-

ción estratigráfica sea conocida, para el caso del orden temporal y con la utilización de todos los datos posibles con referencia a las situaciones espaciales o de dependencia entre unos yacimientos y otros. En el caso de los materiales objeto de nuestro estudio contamos, como ya hemos dicho anteriormente, con cortes estratigráficos que favorecerán la tarea interpretativa, también con una serie de objetos y datos que pueden ayudarnos a separar los distintos complejos culturales, así como con información acerca de dos posibles redes comerciales, una marítima y otra fluvial.

El programa está escrito en FORTRAN IV, y ha sido probado en el IBM 360 del Centro de Cálculo de la Universidad Complutense. Los datos se los suministramos como una matriz cuyas columnas son los yacimientos, y las filas, los tipos. En él podemos distinguir tres niveles:

El programa principal, que maneja los conjuntos de tipos o clases hipotéticamente coherentes.

Subpr. secuencia, que obtiene, para cada conjunto de tipos supuesto coherente, las secuencias válidas y es llamado por el programa principal.

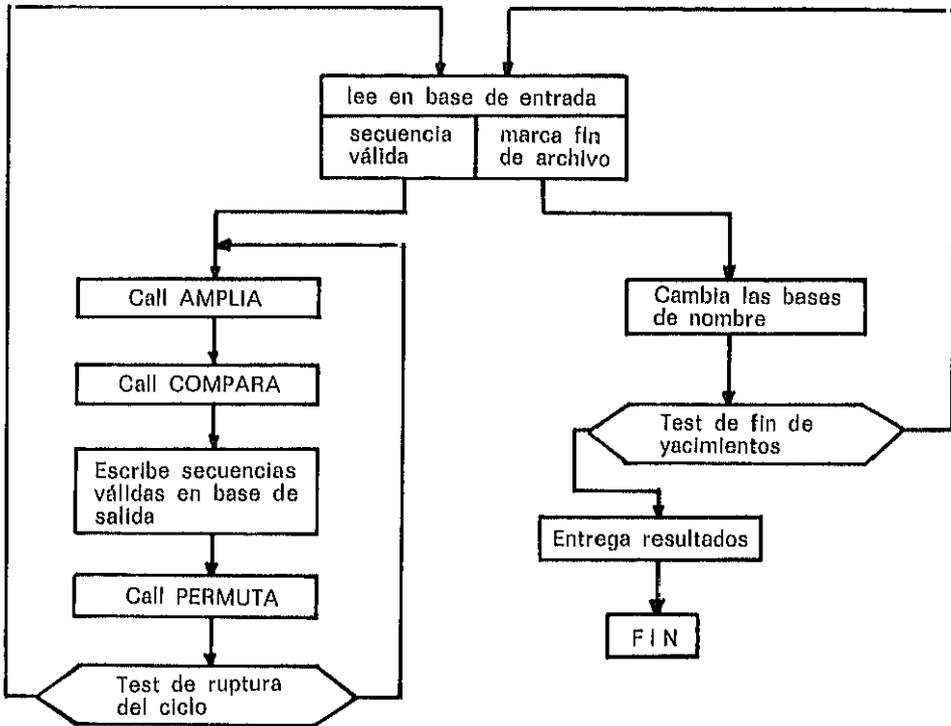
Subpr. amplía, permuta y compara, que realizan las funciones elementales, y son llamados por secuencia.

Para su realización, la dificultad principal que tuvimos que superar fue el no poder conocer de antemano la cantidad de memoria que íbamos a necesitar durante la ejecución.

En el proceso de obtención de las permutaciones de los N yacimientos dados, según el procedimiento descrito anteriormente, cada nuevo nivel del árbol lo obtenemos a partir de las secuencias válidas del nivel anterior; luego debemos tenerlas en memoria para poder operar con ellas, así como las obtenidas a partir de éstas que resulten válidas, y el programador no puede saber cuántas serán éstas, más que *a posteriori*.

A sugerencia de Miguel Guinea, que colaboró con nosotros en la escritura del programa, resolvimos este problema utilizando memoria externa. Tenemos situadas en una cinta, que llamamos Base de entrada, las secuencias válidas de un determinado nivel del árbol; las nuevas secuencias válidas son obtenidas a partir de las anteriores, ampliando con un nuevo yacimiento, permutando éste y comparando con el modelo lenticular, las situamos en otra cinta que llamamos Base de salida. En el paso siguiente repetimos el mismo proceso, pero cambiando los nombres de las bases.

El diagrama de flujo sería:



En el caso de matrices de presencia-ausencia, el test de ruptura de ciclo es el número de permutaciones posibles en cada nivel, y en el caso de frecuencias relativas de tipos puede limitarse a la obtención de dos salidas válidas, siempre que los tipos tengan frecuencia relativa distinta en cada yacimiento.

BIBLIOGRAFIA

ALCINA FRANCO, José:

1977 La arqueología de Esmeraldas (Ecuador): Perspectiva actual. *Antropología de España y América*: 159-174. Ed. Miguel Rivera. Dosbe. Madrid.

ASCHER, Marcia, y ASCHER, Robert:

1963 Chronological Ordering by Computer. *American Anthropologist*. Vol. 65: 1045-1052. Menasha.

- BORDAZ, V., y BORDAZ, J.:
1970 A computer-assisted pattern recognition method of classification and seration applied to archaeological material. *Archéologie et Calculateurs*: 229-244. Ed. Gardin. CNRS. París.
- BRAINERD, George W.:
1951 The place of cronological ordering in archaeological analysis. *American Antiquity*. Vol. XVI; núm. 4; 301-312. Salt Lake City.
- COWGILL, G. L.:
1972 Models, methods and techniques for seriation. *Models in Archaeology*: 381-424. Ed. David L. Clarke. Methuen & Co. Ltd. Londres.
- CRAYTOR, W. B. y JOHNSON, L.:
1968 *Refinements in computerized item seration*. Bulletin núm. 10, Natural History Museum. University of Oregon.
- DORAN, J. E.:
1971 Computer analysis of data from the La Tène cemetery at Münsingen-Rain. *Mathematics in the Archaeological and Historical Sciences*: 422-431. Ed. Hodson, Kendall y Tautu. Edimburg University Press.
- DORAN, J. E., y HODSON, F. R.:
1975 *Mathematics and Computers in Archaeology*. Edimburg University Press.
- GOLDMAN, K.:
1972 Zwei Methoden Chronologischer Gruppierung. *Acta Praehistorica at Archaeologica*. Vol. 3: 1-34. Buenos Aires.
- HOLE, F., y SHAW, M.:
1967 Computer analysis of cronological seriation. *Rice University Studies*. Vol. 53. Houston.
- KENDALL, David G.:
1969 Some problems and methods in statistical archaeology. *World Archaeology*. Vol. 1, núm. 1: 68-76. London.
1971 Seration from abundance matrices. *Mathematics in Archaeological and Historical Sciences*: 215-252. Edimburg University Press.
- KUZARA, Richard; MEAD, G., y DIXON, K.:
1966 Seration of Anthropological Data: A Computer Program for Matrix-Ordering. *American Anthropologist*. Vol. 68: 1443-1455. Menasha.
- McNUTT, Charles:
1973 On the Methodological Validity of Frequency Seration. *American Antiquity*. Vol. 38, núm. 1: 45-60. Salt Lake City.
- ROBINSON, W. S.:
1951 A Method for Cronological Ordering Archaeological Deposits. *American Antiquity*. Vol. XVI, núm. 4: 293-301. Salt Lake City.

Universidad Complutense de Madrid.