

Ejercicio sobre Exploración Gráfica de Datos

Objetivo y Tareas a Realizar

El objetivo de este ejercicio consiste en utilizar los métodos exploratorios vistos en teoría para obtener información y conclusiones sobre un conjunto de datos de dominio.

Para ello, se utilizará un dataset relacionado con el resultado de un *Card Sorting* obtenido a través de una evaluación con distintos usuarios. El Card Sorting es una técnica de investigación y evaluación utilizada en Interacción Persona-Ordenador y Experiencia de usuario que permite la investigación inicial con usuarios y también la evaluación de estructuras de información (por ejemplo, contenidos de una aplicación web o composición de un menú de navegación). La técnica consiste en suministrar al usuario una serie de tarjetas (que representan objetos del dominio) que debe ordenar situándolas dentro de un conjunto de categorías (que pueden venir predefinidas, pueden ser creadas por el propio usuario, o una combinación de ambas cosas a la vez). Cuando esta ordenación se realiza repetidas veces con distintos usuarios, se obtiene información sobre las relaciones entre tarjetas y categorías, lo que permite obtener información sobre el modelo mental del usuario (en fases iniciales de una investigación), o bien una indicación sobre la idoneidad de estructuras ya existentes (por ejemplo, la idoneidad de las opciones dentro de las categorías de un menú de una aplicación software).

El dataset a utilizar es el resultado de un Card Sorting donde se les pidió a 24 usuarios clasificar 40 tarjetas que representaban distintos alimentos. Al tratarse de un Card Sorting *abierto*, fueron los propios usuarios los que crearon las categorías que consideraron convenientes, por lo que se tiene un total de 240 categorías clasificatorias. En concreto, el dataset es un fichero csv que puede obtenerse a través del siguiente enlace:

<http://cardsorting.net/tutorials/25.csv>

El dataset se dispone en forma de matriz (M), de forma que las categorías se sitúan en filas, constituyendo observaciones ($n = 240$), mientras que las tarjetas se distribuyen en columnas, constituyendo variables ($p = 40$). Por tanto, $M[i, j] = N$ implica que N usuarios han clasificado la tarjeta j en la categoría i . Para el objetivo de este ejercicio, se puede prescindir de las columnas *Uniqid*, *Startdate*, *Starttime*, *Endtime*, *QID* y *Comment*.

Supongamos que os ponéis en el rol de un Ingeniero de Usabilidad, o en el del evaluador de Experiencia de Usuario, y queréis explorar visualmente el dataset para obtener la siguiente información:

1. Tipología y rangos de los datos numéricos.
2. Tarjetas más relacionadas entre sí. Entiéndase que, en este ámbito concreto, dos tarjetas se consideran similares si son ordenadas en categorías similares.

Tareas a realizar:

1. Crear un código en R (bien documentado internamente, que permita:
 - a. Leer el dataset desde su origen (a través de la dirección web suministrada).
 - b. Realizar las transformaciones que se consideren convenientes para trabajar de manera efectiva con las categorías y las tarjetas. Se deberá obviar toda la información que no sea de utilidad.
 - c. Representar un histograma, u otro gráfico basado en frecuencias o densidad, para estudiar los datos numéricos que aparecen en el dataset, así como su frecuencia de aparición.
 - d. Crear una matriz de distancia o de similitud de tarjetas. ¿Qué visualización es la más adecuada para esta matriz? Representala convenientemente.
 - e. Representar gráficamente las relaciones entre las tarjetas a través de un grafo, utilizando para ello la librería `qgraph` de R, de forma que las tarjetas más relacionadas se distingan de manera visual.
 - f. Finalmente, ¿cuáles son las tarjetas que están más relacionadas? ¿Tiene esta relación sentido a nivel semántico (en función de los ítems de dominio que representan)?
2. Elaborar una memoria con la solución razonada a cada una de las cuestiones anteriores, incluyendo también las capturas de las figuras con las visualizaciones gráficas correspondientes. La memoria deberá estar identificada en su portada con el nombre de los estudiantes que han trabajado sobre ella.

Entrega

El ejercicio se entregará a través de la tarea correspondiente de Moodle. Para ello, se enviará un fichero .ZIP que contenga la memoria (en formato PDF) y el código fuente en R elaborado. El ejercicio se entregará antes de las 23:55 horas del 13 de diciembre de 2020.

Criterios de Evaluación

El ejercicio se realizará en equipos de dos estudiantes, y puntuará entre 0 y 10 puntos, suponiendo un 25% de la nota de prácticas de la asignatura. Se valorará especialmente la originalidad del aporte del estudiante, la contestación justificada y convincente a cada uno de los apartados propuestos anteriormente, y la calidad y completitud de las representaciones gráficas aportadas.