

Las relaciones de un todo

Ordaz Guillén Fabián

Junio 2024

1. Introducción

A diferencia de las llamadas “ciencias exactas” aquellos ámbitos donde la humanidad es el protagonista han presentado una misma dificultad; la variabilidad del ser humano y su amplio espectro de posibles respuestas aún a situaciones ya conocidas. Por ese motivo se han diseñado varias herramientas a lo largo de la historia que buscan generar grupos que permitan describir, para “bien o para mal” [Dattner(2004)], mediante lo que es y no es el ser humano, responder la pregunta ¿cual es la mejor manera de resumir a un individuo?.

Algunas de estas herramientas son test como el; *Inventario Multifásico de Personalidad de Minnesota (MMPI)*, *16 Factores de Personalidad (16PF)*, *Cuestionario de Personalidad de Eysenck (EPQ)*, *Test de Rorschach* o bien aquel en el que profundizaremos el ***Inventario de Personalidad NEO (NEO PI-R)***.

El NEO PI-R (1980, Paul T. Costa Jr. y Robert R. McCrae) tiene sus fundamentos en la teoría de los cinco grandes (*The Big Five*), que busca medir los cinco principales dominios de la personalidad (Neuroticismo, Extraversión, Apertura a la Experiencia, Amabilidad y Estabilidad emocional) y sus 30 facetas específicas. Es un modelo muy robusto y respetado dentro de la comunidad tanto en fines clínicos preventivos o reactivos, académicos y laborales, valorado por su capacidad para capturar una amplia gama de comportamientos y características de personalidad con un alto grado de validez. Dando una serie de reactivos en los que en una escala del 1 (completamente en desacuerdo) al 5 (muy de acuerdo) se presentan (habitualmente) 48 situaciones, el usuario debe elegir generando una serie de notas que se someten a alphas de evaluación ya preestablecidos para definir una “conclusión”, de recordar que no hay una respuesta correcta o esperada, simplemente es la convicción y naturalidad quien indica que se elige en cada área de la personalidad, así en cada una de ellas.

Ahora bien hemos mencionado ya la duda sobre el modo “más óptimo” de como podemos resumir a un individuo, mediante de una serie de incisos que si bien transparentes, bajo una construcción que nos arrojará información confiable y fundamentados siguen siendo individuales, nos referimos a que tenemos 48 incisos de 5 campos mutuamente excluyentes, es decir, la prueba y sus bases teóricas no están construidas para devolver o tener alguna relación entre áreas y aunque existen cualidades tales como un alto índice de neuroticismo que podría

tener tendencias que también afectan su extraversión o amabilidad, estas últimas se calificarán únicamente con los puntajes de sus escalas así que ¿podrían estas individualidades esconder un patrón de grupo? mejor dicho ¿un alto A siempre venir causado por un bajo B que consecuente a un C moderado?

2. Metodología

Los datos evualuados fueron reunidos durante 2 años (2016-2018) mediante un sitio en línea de personalidad. Es importante mencionar que sus participantes fueron avisados que las respuestas podrían ser grabadas y usadas, así como se les pidió confirmaran su aprobación para el uso ético de los mismos.

La respuesta por cada área de la personalidad se agrupo en columnas abreviadas (EXT = extraversión, EST = estabilidad emocional , AGR = amabilidad , CSN = neuroticismo, OPN = apertura a nuevas experiencias) y enumero con un consecutivo después de cada grupo de siglas, Una de las debilidades mas fuertes de este modelo de interpretación es su larga duración y la facilidad de perder a los participantes, así que buscando que se eligiera la mayor cantidad de sentencias en cada área se les mostró a los usuarios de una por cada área a la vez. También generó una marca en milisegundos del tiempo que paso cada usuario en cada reactivo, las dimensiones del dispositivo donde se tomó la prueba (en pixeles) así como una aproximación de la ciudad, latitud y longitud de origen de información. Quizá para otro tipo de enfoques o resolver interrogantes mas adelante podamos usar todo el conjunto de datos pero para este primer acercamiento solo ahondamos en las 50 preguntas con base en los 5 grandes.

Siguiendo un enfoque de estudio distinto a lo habitual, no se buscaron respuestas dentro de estándares psicológicos, aunque si nos documentamos al respecto para una mejor interpretación, contraste y debate de resultados [Costa and McCrae(1992)], sino una causalidad de A entonces B dentro de una misma área así como entre ellas y sus posibles combinaciones

2.1. Descripción de los datos

Como ya hemos hablado se trata de un conjunto de datos discretos de 50 columnas por 1,015,341 filas, tomando valores del 1 al 5. Para un correcto uso de la información trataremos la misma en búsqueda y eliminación de valores atípicos o vacío que nos impidan seguir con el análisis

Dominio	Preguntas (total)	Respuestas (c/u)	Nulos	No nulos
EXT	10	874,434	0	874,434
EST	10	874,434	0	874,434
AGR	10	874,434	0	874,434
CSN	10	874,434	0	874,434
OPN	10	874,434	0	874,434

Cuadro 1: Resumen de datos tratados a analizar

2.2. Aprendizaje no supervisado

Dentro de los modos que tenemos hoy para analizar patrones dentro de los modelos de aprendizaje no supervisados tenemos el algoritmo FP-Growth (FPG en adelante), este utilizado comúnmente para revisar y encontrar patrones de asociación de eventos, por ejemplo en una lista de supermercado para cuantificar la cantidad de veces que un alimento esta en la misma compra que otro.

Funciona escaneando dos veces la base de datos, la primera nos entrega el número de veces que ocurre un evento en todas las transacciones, entonces, marcamos un mínimo significativo y aquellos que no lleguen a tal frecuencia se descartan y ordenamos de mayor a menor los datos conservados.

Supongamos que tenemos las siguientes transacciones en orden original:

- 1.- [A, B, C]
- 2.- [A, C]
- 3.- [A, B, D]
- 4.- [B, C]
- 5.- [A, C, D]

Y pensando que el umbral mínimo de soporte es 2 tenemos

Frecuencias: A: 4 B: 3 C: 4 D: 2

Todos los artículos cumplen con el umbral mínimo de soporte así que los ordenamos

Orden: A (4), C (4), B (3), D (2)

Transacciones ordenadas por frecuencia

- 1.- [A, C, B]
- 2.- [A, C]
- 3.- [A, B, D]
- 4.- [C, B]
- 5.- [A, C, D]

Para la segunda pasada se empieza la construcción de un árbol, pudiera verse como uno genealógico, con cada evento se crea una rama, si el evento ya existe se agrega un contador, si es la primera vez se agrega un nodo para una nueva rama

Para la primera transacción [A, C, B]:

$$(A : 1) - (C : 1) - (B : 1)$$

Para la segunda transacción [A, C]:

$$(A : 2) - (C : 2) - (B : 1)$$

Para la tercera transacción [A, B, D]:

$$\begin{array}{c} (A : 3) - (C : 2) - (B : 1) \\ \quad \quad \quad \backslash \\ \quad \quad \quad \quad (B : 2) - (D : 1) \end{array}$$

Para la cuarta transacción [C, B]:

$$\begin{array}{c} (A : 3) - (C : 2) - (B : 1) \\ \quad \quad \quad \backslash \\ \quad \quad \quad \quad (B : 2) - (D : 1) \\ (C : 1) - (B : 1) \end{array}$$

Para la quinta transacción [A, C, D]:

$$\begin{array}{c} (A : 4) - (C : 3) - (B : 1) \\ \quad \quad \quad \backslash \\ \quad \quad \quad \quad (B : 2) - (D : 1) \\ \quad \quad \quad \quad \quad \quad \backslash \\ \quad \quad \quad \quad \quad \quad \quad (D : 1) \\ (C : 1) - (B : 1) \end{array}$$

Ahora para el análisis necesitamos contemplar los caminos (secuencias de datos desde la raíz hasta cualquier nodo en el árbol FPG) y el soporte (cantidad de transacciones o eventos que contienen esa secuencia exacta de datos). Por ejemplo veamos el evento [A,C], tenemos 3 eventos que tienen dicha secuencia (de los eventos ordenados por frecuencia tendríamos el 1, 2 y 5) y el soporte de [A,C] se define como las 3 transacciones sobre el número total de eventos 3/5 interpretado como el evento [A,C] aparece en el 60 % de todas las transacciones

De este modo, podemos obtener la frecuencia (soporte) con el que los eventos causan o se relacionan con la aparición de otro.

3. Resultados

Luego de aplicar FPG se obtuvo una relación del número de pregunta de cada dominio, presentando los resultados en la Figura 1

No solo podemos ver la relación entre los dominios globales de personalidad, sino también, la frecuencia con la que preguntas y el nivel de agrado que se eligió anteceden a otra pregunta y su propio nivel de agrado, podríamos de este modo y con un apoyo entendido en psicología generar un nuevo diseño de NEO PI-R mas eficiente, con menos preguntas que ayuden a los participantes con una prueba mas corta que evite una perdida de atención o foco.

Referencias

[Dattner(2004)] Ben Dattner. El uso y mal uso de los tests de personalidad. *Suplemento Selección de Personal. New York, 2004.*

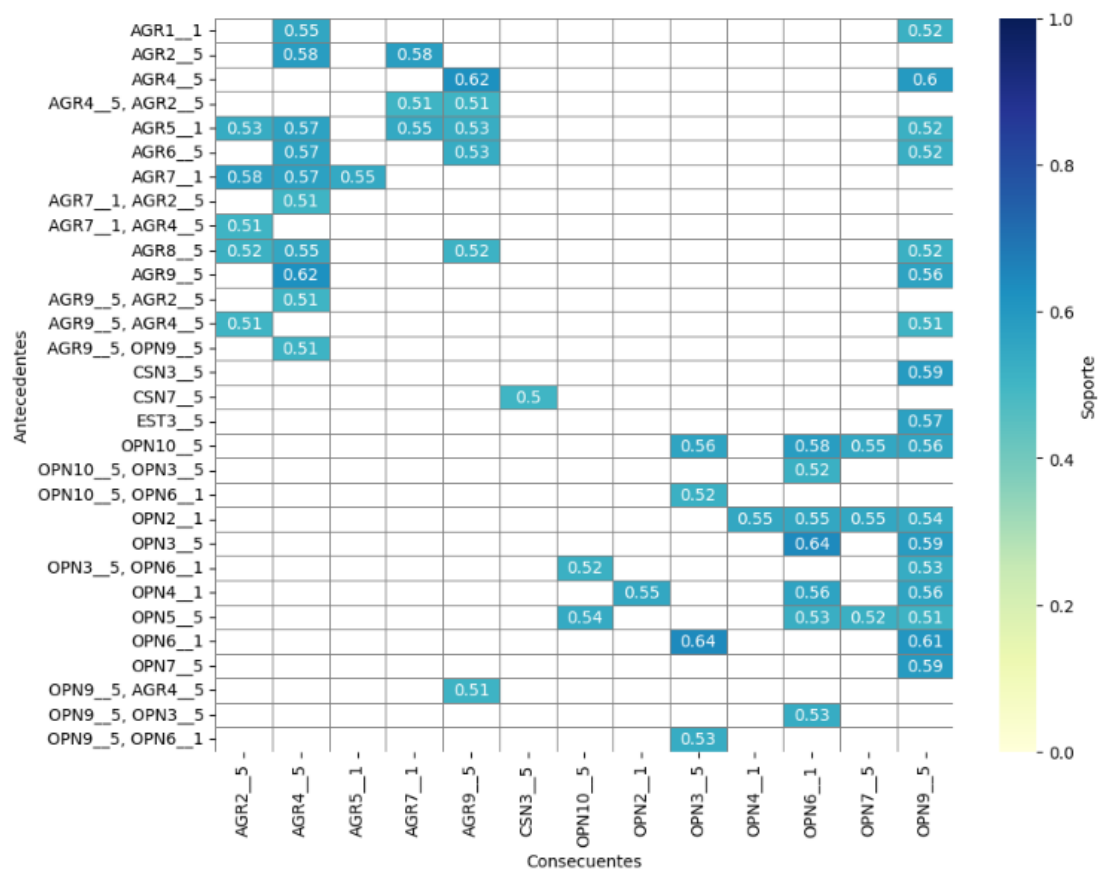


Figura 1: Mapa de calor: Dominios de la personalidad

[Costa and McCrae(1992)] Paul T Costa and Robert R McCrae. *Neo personality inventory-revised (NEO PI-R)*. Psychological Assessment Resources Odessa, FL, 1992.