



El modelo de personalidad humana Big Five y predicción mediante números difusos.

Modelización con Incertidumbre, Información Difusa y Soft Computing

Daniel Gómez Aguado[†]

20 de abril de 2022

[†] Universidad Politécnica de Madrid y Universidad Complutense de Madrid

ÍNDICE

1. Trabajo técnico previo
2. Introducción a la lógica difusa
3. Resultados
4. Conclusiones

INTRODUCCIÓN Y OBJETIVOS

Modelos de personalidad humana: clasifican a individuos según ciertas categorías que buscan englobar (*en medida de lo posible*) una personalidad.



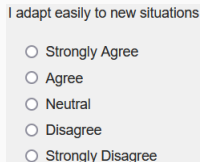
Figura 1: modelo OCEAN - clasificación por 5 características.

- Ejemplos: **Big Five u OCEAN**, o **HEXACO** (considerados de mayor rigor); *Myers-Briggs* (no tan aceptado).

INTRODUCCIÓN Y OBJETIVOS (II)

Aplicación del modelo de personalidad: **cuestionarios**.

- Un usuario responde a un serie de preguntas.
- Cada pregunta puntúa hacia ciertas categorías.



I adapt easily to new situations

- ☐ Strongly Agree
- ☐ Agree
- ☐ Neutral
- ☐ Disagree
- ☐ Strongly Disagree

Figura 2: cinco opciones para responder a la pregunta.

INTRODUCCIÓN Y OBJETIVOS (III)

**¿Son los métodos de categorización
empleados válidos?**

INTRODUCCIÓN Y OBJETIVOS (IV)

Los **modelos predictivos** responderán a la pregunta anterior.

OBJETIVOS DEL PROYECTO

- Generar **un nuevo modelo de predicción** para estos cuestionarios, aprendiendo de datos.
- **Clasificación de individuos:** ¿cómo definimos si puntúa como “alto” o “bajo” en cada categoría?

Se extraen las respuestas a las preguntas de un cuestionario de 300 preguntas (test IPIP-NEO 300) para **95450 usuarios**, con las puntuaciones calculadas para cada categoría.

Trabajo técnico previo

MODELO BIG FIVE

El cuestionario estudiado fue diseñado para el **modelo Big Five u OCEAN**, cuyas siglas vienen dadas por sus cinco categorías de puntuación.

- ***Openness*** (apertura a la experiencia).
- ***Conscientiousness*** (escrupulosidad).
- ***Extraversion*** (extroversión).
- ***Agreeableness*** (amabilidad).
- ***Neuroticism*** (neuroticismo).

Cada una de estas categorías cuenta con seis subdivisiones que aportan información más concreta al modelo.

MODELO BIG FIVE (II)

Agreeableness

score: 82 - high

Agreeableness reflects individual differences in concern with cooperation and social harmony. Agreeable individuals value getting along with others.

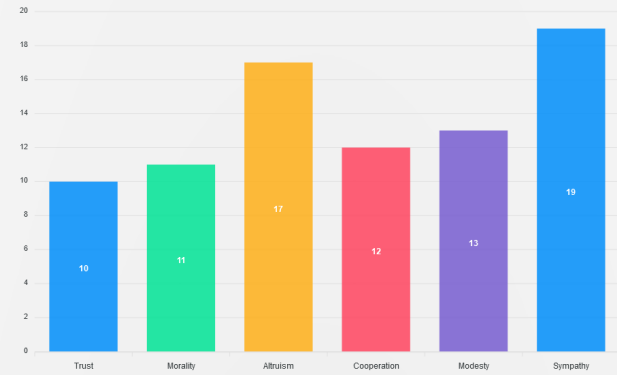


Figura 3: categoría *Agreeableness* desglosada.

SELECCIÓN DE CLASES

Cada usuario tiene una **puntuación continua por categoría**. La primera clasificación realizada, **binaria**, tomó el promedio entre todos los usuarios para cada categoría y, según se estuviera por encima o por debajo, se asignaba una clase en la misma.

EJEMPLOS

- El promedio de puntuación en *Conscientiousness* (*C* abreviado) es de 0.65.
- Un usuario posee 0.77 de *C*: clasifica en este ámbito como **C+**.
- Este mismo usuario obtiene: **O-**, **C+**, **E+**, **A-**, **N+**.
Su clase es: **O-C+E+A-N+**.

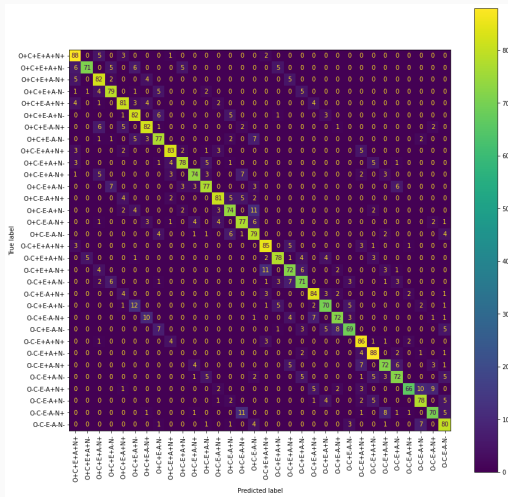
TÉCNICAS DE MODELADO

SVM o máquina de soporte vectorial: se busca predecir la clase de cada usuario según las respuestas dadas al test.

Realizamos una comparativa *one vs all* para categorizar.

Árbol de decisión: posee una gran interpretabilidad y nos permitiría encontrar qué preguntas del test son más relevantes en la obtención de un resultado u otro para cada categoría.

RESULTADOS EN LÓGICA BINARIA: MATRIZ DE CONFUSIÓN



Forma de los datasets:

X_train (7347, 300), y_train (7347,)
X_train (3149, 300), y_train (3149,)

Mejor configuración de parámetros:

pca_n_components: 43
svc_kernel: rbf

Exactitud (val): 0.7426

Exactitud (test): 0.7869

RESULTADOS EN LÓGICA BINARIA: ÁRBOL DE DECISIÓN

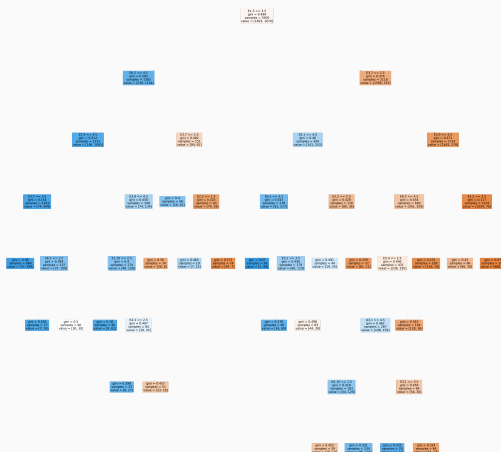


Figura 4: árbol de decisión para la categoría *Extraversión*.

Los resultados presentados son bastante inocentes.

Lógica difusa

SOBRE EL MODELADO PREVIO

El modelo planteado predice muy bien si cada usuario se detecta por encima o por debajo de la media, pero le falta **aportar más información o complejidad**:

- Nos interesa información del tipo: “este usuario tiene un valor **alto** de *Neuroticism*” o “se encuentra en el promedio”.
- Buscamos categorizar con clases (discretas) a su vez, pero tal vez éstas no sean del todo certeras (**grado de pertenencia**).

La lógica difusa resuelve ambos problemas.

NÚMEROS BORROSOS

Noción de número borroso: se define como un subconjunto borroso A de los números reales \mathcal{R} tal que:

- Es convexo: para todo $x, y \in \mathcal{R}$ y $\lambda \in [0, 1]$ se cumple $A(\lambda x + (1 - \lambda)y) \geq \min(A(x), A(y))$.
- Tiene un único valor modal $x \in \mathcal{R} \mid A(x) = 1$.
- Soporte acotado $\text{sop}(A) = \{x \in \mathcal{R} \mid A(x) > 0\}$.
- Función de pertenencia $\mu_A : \mathcal{R} \rightarrow [0, 1]$ continua.

NÚMEROS BORROSOS: APLICACIÓN

Las puntuaciones por categoría de cada usuario van de 0 a 1.

Valores que utilizaremos para definir nuestros números borrosos:

- **Promedio** de puntuaciones por categoría en todo el DataBase (del que ya disponemos por el caso previo).
- **Mínimos y máximos** en cada puntuación.
- **Valores intermedios** entre promedio y mínimos/máximos.
Incluiremos 2: uno entre mínimo y promedio, y otro entre promedio y máximo.

NÚMEROS BORROSOS: TRIANGULAR

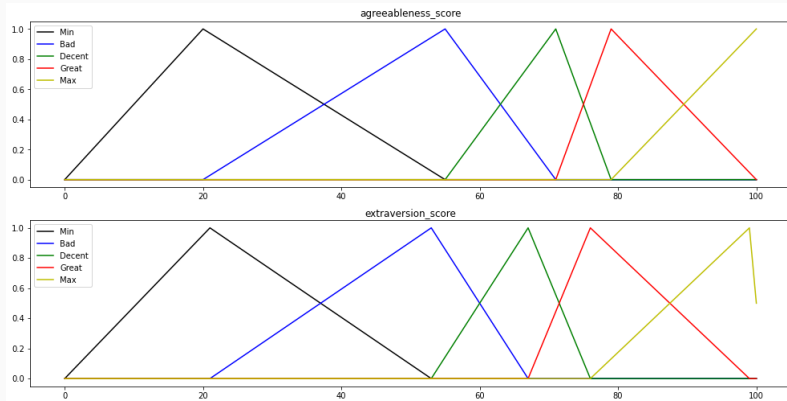


Figura 5: numeros borrosos triangulares para dos de las cinco categorías.

NÚMEROS BORROSOS: GAUSS

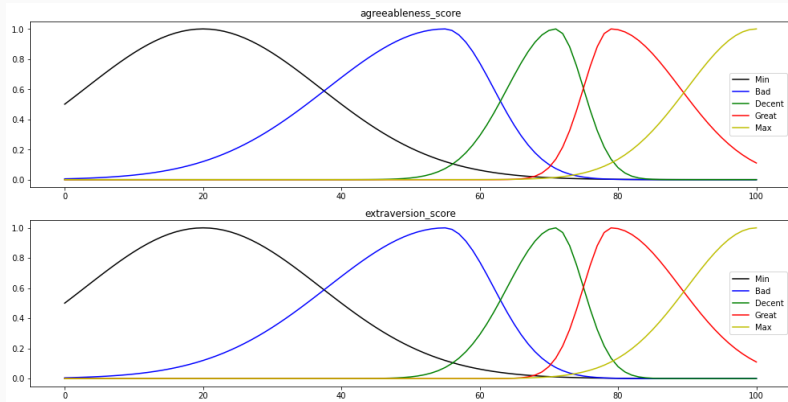


Figura 6: numeros borrosos gaussianos para dos de las cinco categorías.

NÚMEROS BORROSOS: GAUSS + SIGMOIDAL

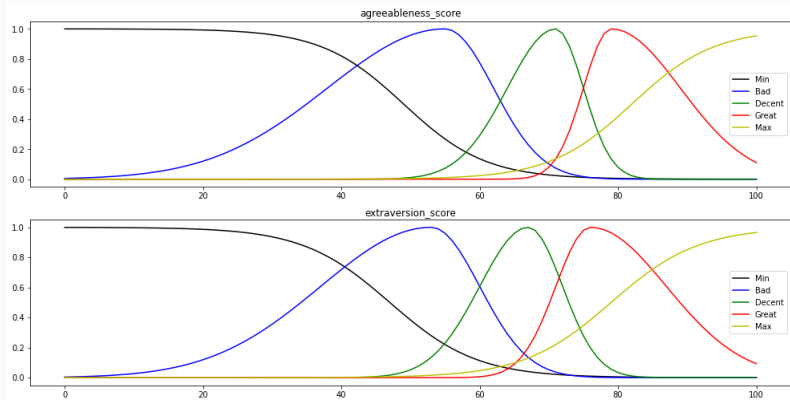


Figura 7: numeros borrosos (gauss + sigmoidal) para para dos de las cinco categorías.

GRADO DE PERTENENCIA POR CATEGORÍA

El grado de pertenencia para un x dado es el valor $A(x)$, siendo A el número borroso.

- Por valor continuo en cada categoría, un usuario tendrá un **grado de pertenencia** a cada una de sus cinco posibles puntuaciones.
- Para clasificar en O , tenemos las opciones: $[O-, o-, \dots, o+, O+]$. Buscamos **a cuál corresponde el grado de pertenencia máximo**.
- **Ejemplo:** para el número borroso de *Openness* u O , un usuario obtiene $[0,17, 0,76, 0,45, 0,00, 0,00]$. En este caso, la clasificación es **$o-$** con $A_{o-}(x) = 0,76$.

GRADO DE PERTENENCIA GLOBAL

Para comparar este nuevo modelado con el anterior, componemos las clases con la misma idea que anteriormente.

Ahora hay que tener en cuenta los grados de pertenencia.

En el caso de una clase compuesta, este grado será tomado como el **mínimo** entre las cinco categorías involucradas.

- **Ejemplo previo:** $A_{o-}(x) = 0,76$. Se le encuentra también $A_{c-}(x) = 0,84$, $A_{E+}(x) = 0,70$, $A_{a+}(x) = 0,77$, $A_{-}(x) = 0,74$. El grado de pertenencia $A_{o-c-E+a+-}(x) = 0,70$.

GRADO DE PERTENENCIA GLOBAL: REGLAS

Sean A_O , A_C , A_E , A_A , A_N los grados de pertenencia para cada categoría. Para una clase cualquiera:

if A_O is low **AND** A_C is low **AND** A_E is very high **AND** A_A is high **AND** A_N is medium \rightarrow clase **o-c-E+a+...**

El código, tras buscar la clasificación de máxima pertenencia en cada categoría por separado, compone la clase correspondiente junto al nuevo grado de pertenencia (el mínimo entre los cinco).

Resultados

CONSIDERACIONES INICIALES

- Ahora disponemos de grados de pertenencia para las clases, por lo que **puede ser interesante filtrar según una cota mínima.**
- En este DataBase de 95450 usuarios, podemos encontrarnos con $5^5 = 3125$ **clases disponibles.**
Dispondremos de un **método de filtrado de clases** tal que si una de éstas no aparece un número mínimo de veces, no se considera.
- Quitaremos las clases con resultados neutros para cada categoría “__”.

PARÁMETROS ESCOGIDOS

- **Filtrado de clases:** se requiere que tengan una ocupación mínima del 0,05 % ($954500,0005 = 47$).
- **Cotas mínima de grado de pertenencia:** se han computado resultados para 0,5 (baja), 0,65 (media) y 0,75 (alta).
- **Números borrosos:** mejores resultados para el caso gauss+sigmoidal.
- **División conjuntos de entrenamiento y test:** 70 % – 30 %.
- Realizamos **bajomuestreo**.

CASO: GAUSS + SIGMOIDAL, GRADO DE PERTENENCIA ≥ 0.5

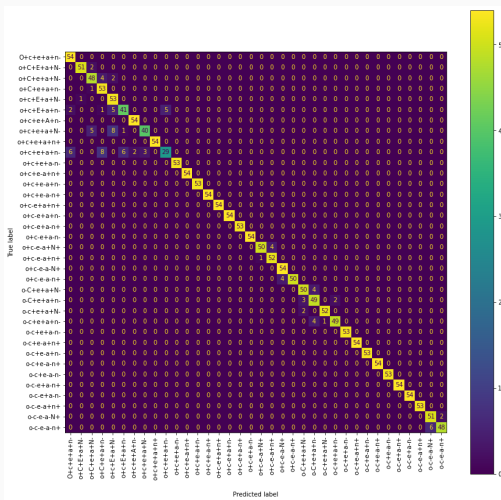


Figura 8: matriz de confusión. Pertenencia 0.5.

CASO: GAUSS + SIGMOIDAL, GRADO DE PERTENENCIA ≥ 0.65

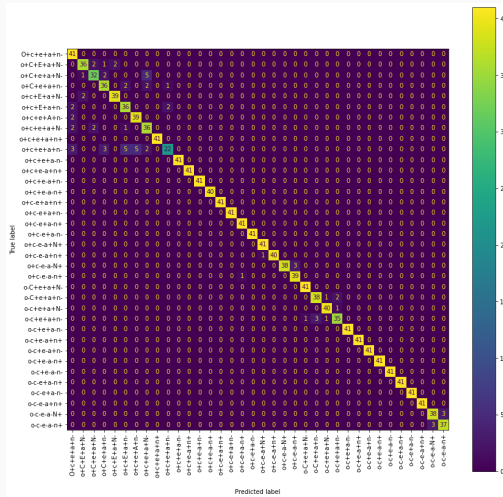


Figura 9: matriz de confusión. Pertenencia 0.65.

CASO: GAUSS + SIGMOIDAL, GRADO DE PERTENENCIA ≥ 0.75

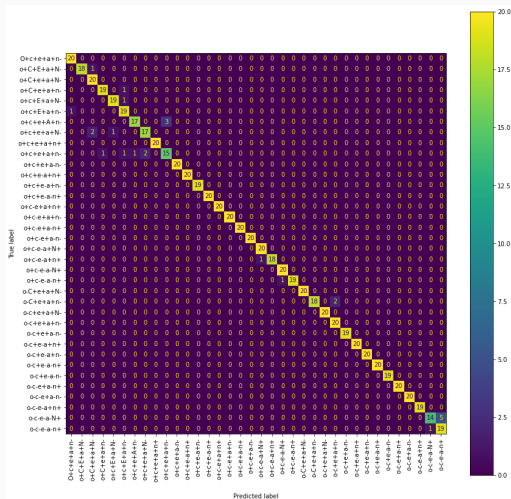


Figura 10: matriz de confusión. Pertenencia 0.75.

CASO: GAUSS + SIGMOIDAL, DATOS AUXILIARES

Grado de pertenencia: **0.5 (bajo).**

Dadas las condiciones, actualmente disponemos de 36 clases
Forma de los datasets:
X_train (4510, 300), y_train (4510,)
X_train (1934, 300), y_train (1934,)

Mejor configuración de parámetros:
pca_n_components: 110
svc_kernel: rbf
Exactitud (val): 0.9379
Exactitud (test): 0.9509

Grado de pertenencia: **0.65 (medio).**

Dadas las condiciones, actualmente disponemos de 36 clases
Forma de los datasets:
X_train (3427, 300), y_train (3427,)
X_train (1469, 300), y_train (1469,)

Mejor configuración de parámetros:
pca_n_components: 110
svc_kernel: rbf
Exactitud (val): 0.9300
Exactitud (test): 0.9530

Grado de pertenencia: **0.75 (alto).**

Dadas las condiciones, actualmente disponemos de 36 clases
Forma de los datasets:
X_train (1663, 300), y_train (1663,)
X_train (713, 300), y_train (713,)

Mejor configuración de parámetros:
pca_n_components: 11
svc_kernel: rbf
Exactitud (val): 0.9609
Exactitud (test): 0.9649

Conclusiones

CONCLUSIONES

- Se ha logrado **identificar usuarios por clases o tipos de personalidad** mediante modelos de aprendizaje estadístico.
- El empleo de números borrosos ha mejorado la precisión del modelo de un 0.7426 en validación y 0.7869 en test, a **0.9379 en validación y 0.9509 en test** (caso de grado de pertenencia mínimo de 0.5, el más global).
- Categorizar personalidades humanas es un problema muy complejo, y muchas veces **los resultados de un cuestionario no aportan información de valor (grado de pertenencia bajo)**.

FIN