Análisis de Brechas Explicable y Mínimo Esfuerzo Psicológico para Alcanzar Consenso en Toma de Decisiones

Cristina Zuheros[1] and Francisco Herrera[1,2]

[1] Departamento de Ciencias de la Computación e Inteligencia Artificial
[2] Instituto Andaluz de Data Science and Computational Intelligence (DaSCI)
Universidad de Granada, España
czuheros@ugr.es, herrera@decsai.ugr.es

Abstract. En ese artículo se propone *MPEGAS*, un *framework* para fomentar el consenso en la toma de decisiones multi-criterio lingüística. El *framework* unifica evaluaciones con formatos heterogéneos, identifica discrepancias entre expertos y genera peticiones de cambio en lenguaje natural. Un caso de estudio muestra cómo MPEGAS mejora progresivamente el consenso grupal minimizando el esfuerzo de los expertos.

Keywords: Toma de Decisiones en Grupo · Proceso de Alcance del Consenso · IA Explicable · Semánticas Individuales Personalizadas

1 Introducción

Los procesos de toma de decisiones en grupo (*Group Decision Making*, GDM) en entornos multi-criterio requieren mecanismos capaces de alinear las posturas de los expertos y alcanzar un consenso representativo [1]. Cuando los expertos expresan sus evaluaciones mediante información lingüística, la interpretación puede variar según su nivel de experiencia e incertidumbre y su rol profesional. Aunque los modelos basados en semánticas individuales personalizadas (PIS) permiten unificar información heterogénea [3], todavía persisten limitaciones en el proceso de consenso. Este artículo aborda los siguientes desafíos clave en GDM:

1. Heterogeneidad cognitiva y profesional: los expertos pueden emplear distintos formatos de evaluación lingüística y asignar diferente importancia a los criterios según su perfil, lo que dificulta la agregación de sus evaluaciones.
2. Falta de explicabilidad natural: las discrepancias entre evaluaciones individuales y colectivas deben traducirse en argumentos comprensibles que justifiquen las recomendaciones de cambio y fomenten el consenso [4].
3. Resistencia psicológica al cambio: las recomendaciones de ajuste pueden ser rechazadas si afectan a criterios considerados relevantes por el experto o si el experto desconoce el efecto de sus evaluaciones frente al grupo.

Este artículo aborda los desafíos indicados proponiendo el marco de trabajo o *framework MPEGAS* (*Minimized Psychological Effort via Gap-Analysis Semantic Group Decision Making*). Dicho *framework* se presenta en la Sección 2 y se aplica a un problema de GDM multi-criterio lingüístico en la Sección 3.

2 Metodología propuesta

MPEGAS es un *framework* diseñado para fomentar el consenso en GDM multicriterio lingüístico, que identifica las discrepancias más relevantes entre expertos y genera peticiones de ajuste explicables que fomentan el acuerdo con el mínimo esfuerzo psicológico posible. Está compuesto por tres fases fundamentales: 1) módulo unificador de evaluaciones lingüísticas heterogéneas y pesos de criterios subjetivos, 2) motor analítico de explicabilidad basado en análisis de brechas, 3) mecanismo de realimentación orientado al mínimo esfuerzo psicológico.

2.1 Módulo Unificador de Evaluaciones y Criterios Subjetivos

Esta fase tiene por objetivo resolver dos principales fuentes de heterogeneidad: la variabilidad en los formatos de representación de las evaluaciones lingüísticas por parte de los expertos (derivada de su nivel de experiencia e incertidumbre) y la discrepancia en la importancia que cada experto otorga a los criterios de evaluación (derivada de su perfil o rol profesional).

Sea un conjunto de expertos $E = \{e_1, \ldots, e_m\}$, alternativas $X = \{x_1, \ldots, x_n\}$, criterios $C = \{c_1, \ldots, c_q\}$ y un conjunto de términos lingüísticos $S = \{s_0, \ldots, s_g\}$ para la evaluación. Se permite evaluar mediante etiquetas lingüísticas simples, de intervalo o distribuidas. Cada e_k establece un vector de pesos sobre los criterios, $W^k = (w_{k1}, \ldots, w_{kq})$ tal que $\sum_{j=1}^{q} w_{kj} = 1$, que captura el sesgo de su rol (*p. ej.*, un perfil financiero priorizará criterios de costes frente a los de seguridad).

Las evaluaciones heterogéneas se unifican basándose en el modelo PIS [3]. Mediante una función de escala numérica adaptada a cada formato de evaluación, se construye para cada experto e_k una matriz $P^k = \left(p_{ij}^k\right)_{n\times q}$, donde $p_{ij}^k \in [0,1]$ representa la evaluación numérica de la alternativa x_i bajo el criterio c_j. Se calcula la matriz colectiva $P^c = (p_{ij}^c)_{n\times q}$ mediante la media ponderada de las preferencias individuales y el peso de cada experto α_k siendo $\sum_{k=1}^{m} \alpha_k = 1$.

Se modela el esfuerzo psicológico que supone el cambio en la evaluación de un criterio c_j para un experto e_k definiendo el *índice de resistencia al cambio* $IRC = (irc_{kj})_{m\times q}$. Inspirados por la Teoría de las Perspectivas [2], el *framework* reconoce que la resistencia psicológica es asimétrica: un experto tiene mayor aversión a perder influencia sobre un criterio frente a ganarla. Este comportamiento se modela mediante una función por tramos en base a W^k y los pesos colectivos. Finalmente, este índice se proyecta mediante una función exponencial en un matriz de *disposición al cambio*, $DC = (dc_{kj})_{m\times q}$, donde valores cercanos a 1 identifican los criterios de menor coste psicológico de cambio para cada experto.

2.2 Motor Analítico de Explicabilidad

Esta fase tiene por objetivo evaluar la cohesión del grupo y diagnosticar la raíz fundamental de las discrepancias mediante el análisis de brechas.

Se construye una matriz de *brechas individuales* $B^k = \left(b_{ij}^k\right)_{n\times q}$ siendo $b_{ij}^k = p_{ij}^k - p_{ij}^c$, para cada experto, donde el valor $|b_{ij}^k| \in [0,1]$ mide el desacuerdo y

su signo determina la dirección de la desviación. Seguidamente, se establece el *índice de consenso del grupo* $CI^G \in [0,1]$ mediante la media ponderada, del complemento de la distancia media de las brechas de los expertos, en base a los pesos de los expertos. Si el consenso del grupo no alcanza un umbral establecido ($CI^G < \gamma \in [0,1]$) es necesario modificar alguna evaluación de los expertos. En ese caso, el *framework* identifica automáticamente los cambios más viables. Para ello, genera la matriz de *oportunidad de consenso* para cada experto, $OC^k = \left(oc_{ij}^k\right)_{n\times q}$, siendo $oc_{ij}^k = |b_{ij}^k| \cdot dc_{kj}$. Este índice prioriza aquellos elementos de evaluación con amplio margen de mejora y con alta disposición al cambio.

2.3 Mecanismo de Realimentación

Esta fase tiene por objetivo generar recomendaciones de cambio personalizadas y argumentadas en lenguaje natural cuando el consenso colectivo es insuficiente.

El *framework* identifica la tripleta experto-alternativa-criterio que maximiza el consenso a la vez que requiere el menor esfuerzo psicológico de cambio mediante $(e_{k^*}, x_{i^*}, c_{j^*}) = \arg\max_{(k,i,j)} \left(oc_{ij}^k\right)$. A continuación, se traducen los factores analíticos de cambio en una propuesta lingüística fácilmente comprensible para el experto seleccionado. Aprovechando la retranslación en el modelo PIS, el *framework* calcula la etiqueta óptima $s^* \in S$ que minimiza la distancia semántica con el valor colectivo $p_{i^*j^*}^c$. Finalmente, esta etiqueta se integra en un mensaje explicativo mediante reglas de inferencia que evalúan el signo de la brecha y el índice de resistencia al cambio del experto.

Este enfoque garantiza que los expertos modifiquen voluntariamente evaluaciones que consideran menos significativas, reduciendo sus resistencias psicológicas al cambio y acelerando la convergencia del proceso.

3 Caso de estudio

Para ilustrar el potencial del *framework* MPEGAS, se resuelve un problema GDM con tres expertos, tres alternativas, tres criterios y el conjunto de términos lingüísticos $S = \{s_0 = \text{Muy Bajo}, s_1 = \text{Bajo}, s_2 = \text{Medio}, s_3 = \text{Alto}, s_4 = \text{Muy Alto}\}$. Las evaluaciones de los expertos se muestran en la Figura 1. Además, los expertos establecen $W_1 = (0.40, 0.10, 0.50)$, $W_2 = (0.30, 0.30, 0.40)$ y $W_3 = (0.30, 0.50, 0.20)$. El umbral de consenso necesario establecido es $\gamma = 0.93$.

e_1	c_1	c_2	c_3
x_1	s_3	s_2	s_1
x_2	s_2	s_0	s_2
x_3	s_2	s_3	s_2

e_2	c_1	c_2	c_3
x_1	$[s_2, s_4]$	$[s_2, s_2]$	$[s_1, s_1]$
x_2	$[s_3, s_3]$	$[s_1, s_3]$	$[s_2, s_2]$
x_3	$[s_2, s_2]$	$[s_2, s_2]$	$[s_1, s_3]$

e_3	c_1	c_2	c_3
x_1	$(s_2, 0.2), (s_3, 0.8)$	$(s_1, 0.2), (s_2, 0.8)$	$(s_2, 0.2), (s_3, 0.8)$
x_2	$(s_3, 0.7), (s_4, 0.3)$	$(s_2, 0.9), (s_3, 0.1)$	$(s_2, 0.8), (s_3, 0.2)$
x_3	$(s_1, 0.2), (s_2, 0.8)$	$(s_2, 1.0)$	$(s_2, 0.7), (s_3, 0.3)$

Fig. 1. Matrices de evaluación proporcionadas por los expertos.

El framework genera tres rondas iterativas para alcanzar el consenso deseado. En la primera, $CI_{\mathrm{G}} = 0.912$ y se identifica la máxima oportunidad de consenso en el experto e_1. En la segunda ronda, $CI_{\mathrm{G}} = 0.929$ y se identifica la oportunidad de consenso en e_3. MPEGAS emite peticiones explicativas de cambio como se muestra en la Figura 2. Tras aceptar ambos expertos el cambio, en la tercera ronda se alcanza el consenso deseado ya que $CI_{\mathrm{G}} = 0.943$. Se da por concluido el proceso y se genera el *ranking* final de las alternativas $X_2 > X_3 > X_1$.

Ronda 1: Retroalimentación para e_1

Estimado experto e_1*, le sugerimos aumentar su evaluación para la alternativa* x_2 *bajo el criterio* c_2*, aproximándose a* s_1 *(Bajo). Tu evaluación actual es* s_0 *(Muy Bajo).*

Motivo: Se ha realizado un balance entre discrepancia de los expertos y menor impacto en su perfil. Usted es el candidato óptimo:

1) Discrepancia: su desviación respecto a la tendencia colectiva es 0.34, frente a 0.17 en media del resto de expertos.
2) Impacto personal: para su perfil, el criterio c_2 tiene peso 0.10, frente a 0.40 de relevancia media para el resto.

Al fusionar ambos factores, modificar su postura es la acción que menos esfuerzo psicológico genera en todo el grupo y la que más favorece el proceso de decisión.

Ventajas del cambio:

– Consenso grupal actual: 0.912.
– Incremento del consenso: 1.216%.

Ronda 2: Retroalimentación para e_3

Estimado experto e_3*, le sugerimos disminuir su evaluación para la alternativa* x_1 *bajo el criterio* c_3*, aproximándose a* s_2 *(Medio). Tu evaluación actual es* $(s_2, 0.2)$ *y* $(s_3, 0.8)$.

Motivo: Se ha realizado un balance entre discrepancia de los expertos y menor impacto en su perfil. Usted es el candidato óptimo:

1) Discrepancia: su desviación respecto a la tendencia colectiva es 0.30, frente a 0.15 en media del resto de expertos.
2) Impacto personal: para su perfil, el criterio c_3 tiene peso 0.20, frente a 0.45 de relevancia media para el resto.

Al fusionar ambos factores, modificar su postura es la acción que menos esfuerzo psicológico genera en todo el grupo y la que más favorece el proceso de decisión.

Ventajas del cambio:

– Consenso grupal actual: 0.929.
– Incremento del consenso: 1.081%.

Fig. 2. Mensajes explicativos generados por MPEGAS y enviados a los expertos.

Agradecimientos. Esta investigación es parte del Proyecto TSI-100927-2023-1, financiado por Plan de Recuperación, Transformación y Resiliencia Europeo, de la Unión Europea-Next Generation EU a través del Ministerio para la Transformación Digital y de la Función Pública y está parcialmente financiada por el Ministerio de Ciencia y Tecnología de España en el marco del proyecto PID2023-150070NB-I00, financiado por MICIU/AEI/10.13039/501100011033 y por FEDER, UE. Los autores declaran que no existen conflictos de interés.

Bibliografía

1. Guo, J., Zhang, Z., Wang, Z.: A personalised consensus-reaching process for group decision-making under dynamic trust networks. Group Decision and Negotiation **35**(1), 5 (2026)
2. Kahneman, D., Tversky, A.: Prospect theory: An analysis of decision under risk. In: Handbook of the fundamentals of financial decision making: Part I, pp. 99–127. World Scientific (2013)
3. Li, C.C., Dong, Y., Herrera, F., Herrera-Viedma, E., Martínez, L.: Personalized individual semantics in computing with words for supporting linguistic group decision making. an application on consensus reaching. Information Fusion **33**, 29–40 (2017)
4. Xu, M., Wang, Y.: Explainability increases trust resilience in intelligent agents. British Journal of Psychology **117**(2), 528–547 (2026)