Fuzzy-Transformer Fusion for Sentiment Analysis

Jesús González Gironda[1][0009−0007−6517−4772], Clara Simón de Blas[1][0000−0002−6963−1295], and Daniel Gómez[2,3][0000−0001−9548−5781]

[1] Universidad Rey Juan Carlos, Departamento de Informática y Estadística, Móstoles (España) jesusggironda01@gmail.com; clara.simon@urjc.es
[2] Universidad Complutense, Facultad de Estudios Estadísticos, Madrid (España)
[3] Instituto Universitario de Estadística y Ciencia de los Datos
dagomez@ucm.es

Abstract. El análisis de sentimientos es una disciplina que ha evolucionado rápidamente: desde los modelos originales basados en diccionarios (propios de la minería de textos o text-mining) y los modelos de clasificación tipo RNN y LSTM, hasta llegar a las actuales arquitecturas basadas en Transformers. Estas últimas capturan de manera efectiva relaciones semánticas más complejas. Sin embargo, los algoritmos basados en Transformers aún se enfrentan a numerosos desafíos en la actualidad. Entre ellos, uno de los más preocupantes se relaciona con cómo se aborda la ambigüedad y la variabilidad asociadas a la respuesta del Transformer, así como la manera en la que se modela la incertidumbre lingüística. Esta investigación en curso presenta un marco de trabajo que integra la lógica difusa en la capa de salida de Transformers pre-entrenados para modelar mejor el lenguaje subjetivo y ambiguo. Al permitir una representación gradual del sentimiento, en lugar de una clasificación binaria estricta, el enfoque propuesto mejora la interpretabilidad y la detección de emociones mixtas o poco claras. El resultado es un sistema más robusto y matizado, que preserva la eficiencia computacional original del Transformer.

Keywords: Transformers, Lógica difusa · Análisis del Sentimiento · Procesamiento del lenguaje · Interpretabilidad.

El paradigma digital ha redefinido fundamentalmente la interacción humana, estableciendo la comunicación digital como el principal medio de expresión individual y colectiva. En este contexto, la generación masiva de datos textuales en las plataformas de redes sociales constituye una persistente «huella comunicativa». Este rastro digital ofrece una perspectiva inestimable sobre las preferencias de los usuarios, la percepción de las marcas y el rendimiento de los productos, lo que posiciona al Análisis de Sentimientos (AS) como una herramienta fundamental para descifrar la opinión pública en la era moderna [1][3][4].

En sus inicios, los primeros modelos de análisis de sentimientos procesaban el texto mediante la tokenización de sus palabras y el uso de diccionarios para estimar la probabilidad de que un mensaje tuviera una connotación positiva, negativa o neutra. Esta vectorización, típica de la minería de textos (text-mining),

permitió abordar el problema desde la perspectiva del aprendizaje automático (Machine Learning). Representar un texto de manera vectorial posibilitó el salto hacia algoritmos más complejos, como las redes neuronales recurrentes (RNN) y las arquitecturas de memoria a corto y largo plazo (LSTM) [5].

Estos modelos sentaron las bases para el procesamiento secuencial; sin embargo, la aparición de la arquitectura Transformer y su mecanismo de autoatención marcaron un cambio definitivo en este campo [6]. Al asignar numéricamente una importancia relativa a cada token dentro de una secuencia, los Transformers facilitan la detección de dependencias semánticas complejas y relaciones contextuales de largo alcance que los modelos anteriores tenían dificultades para capturar.

A pesar de su notable capacidad para capturar el contexto, los algoritmos basados en Transformers no están exentos de limitaciones. Uno de los desafíos más críticos radica en la variabilidad y la incertidumbre predictiva intrínseca a estos modelos. Ante un mismo texto de entrada, diferentes arquitecturas preentrenadas (o incluso variaciones en el proceso de inferencia) pueden generar distribuciones de probabilidad divergentes. Esta falta de consenso refleja una vaguedad inherente en la respuesta del modelo, la cual es ignorada por las aproximaciones tradicionales que fuerzan una clasificación categórica basada en el modelo individual de mayor rendimiento.

Para abordar esta problemática, este estudio propone un marco de trabajo que integra la lógica difusa no solo para modelar la subjetividad del lenguaje humano, sino para modelar la incertidumbre predictiva de un conjunto de Transformers interpretando las diferentes salidas de los modelos como grados de pertenencia borrosas. La integración de las predicciones de múltiples modelos permite crear un ensamble robusto que capitaliza la divergencia de los algoritmos para alcanzar un consenso más fiable.

En este contexto, la lógica difusa emerge como un marco complementario ideal. Si bien el aprendizaje profundo destaca en el reconocimiento de patrones en grandes volúmenes de datos, la lógica difusa proporciona una base matemática rigurosa para abordar la incertidumbre y la ambigüedad lingüísticas ([2]). La integración de estos dos paradigmas permite crear modelos que no solo ofrecen un alto rendimiento, sino que también son más interpretables, capaces de representar las emociones como transiciones graduales en lugar de categorías binarias rígidas.

El estudio utiliza un corpus especializado de 8768 tuits relacionados con el conflicto de Gaza. Tras un riguroso proceso de preprocesamiento —que incluyó la eliminación de duplicados y el filtrado por idioma— se conservó un subconjunto final de 4469 tuits en español. Seleccionamos tres Transformers de última generación: DistilBERT, RoBERTa y ModernBERT. Para cada tuit, los modelos generan una distribución de probabilidad mediante softmax.

Para sintetizar la predicción final del sentimiento, proponemos un marco de agregación formal. Sea $\mathcal{M} = \{m_1, m_2, \ldots, m_n\}$ sea el conjunto de n modelos Transformer preentrenados. Para cada tweet t y cada categoría emocional

$k \in \{l_1 = positive, l_2 = negative, l_3 = neutral\}$, obtenemos una puntuación agregada final s_{agg} aplicando un operador de agregación difusa Φ:

$$s_{\text{agg}}(k \mid t) = \Phi\left(p_1(k \mid t), p_2(k \mid t), \ldots, p_n(k \mid t)\right) \tag{1}$$

donde $p_i(k \mid t)$ representa el grado de pertenencia (probabilidad normalizada) asignado por el modelo m_i a la clase k. La clasificación final se determina por la clase que maximiza esta puntuación agregada:

$$\arg\max_{k \in \mathcal{Y}} s_{\text{agg}}(k \mid t) \tag{2}$$

En este estudio, exploramos varios operadores para Φ, incluyendo un operador OWA con orness de rango medio, máximo, mínimo y producto.

La integración de la lógica difusa en los flujos de trabajo basados en Transformers mejora significativamente la robustez del análisis de sentimientos en español. El conjunto OWA navega con éxito por las zonas "neutrales´´ y las emociones mixtas del discurso político complejo. En trabajos futuros se explorará el uso de ANFIS (Sistemas de Inferencia Neurodifusa Adaptativa) para aprender automáticamente reglas de agregación óptimas, reduciendo aún más la brecha entre el rendimiento neuronal y la interpretabilidad lingüística.

Agradecimientos Esta investigación ha sido financiada por el Plan del Gobierno español PID2024- 155289NB-I00 financiado por
MICIU/AEI/10.13039/501100011033 and ERDF/EU.

References

1. Do, H.N., Phan, H.T., Nguyen, N.T.: Multi-modal sentiment analysis using deep learning and fuzzy logic: A comprehensive survey. Applied Soft Computing **167**, 112279 (2024). https://doi.org/10.1016/j.asoc.2024.112279
2. Ji, D., Meng, W., Wang, H.: Emotional analysis of tourism reviews based on long short-term memory and fuzzy control algorithm. International Journal of Fuzzy Systems (2025). https://doi.org/10.1007/s40815-024-01890-1
3. Qin, J., Zeng, M., Wei, X., Pedrycz, W.: Ranking products through online reviews: A novel data-driven method based on interval type-2 fuzzy sets and sentiment analysis. Journal of the Operational Research Society **75**(5), 860–873 (2024). https://doi.org/10.1080/01605682.2023.2215823
4. Ramezani, E.B.: Sentiment analysis applications using deep learning advancements in social networks: A systematic review. Neurocomputing p. 129862 (2025)
5. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2013)
6. Tai, K.S., Socher, R., Manning, C.D.: Improved semantic representations from tree-structured long short-term memory networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL) (2015)