

# Reporte: A survey on Text Classification Algorithms: From Text to Predictions (2nd Part)

Eber David Gaytán Medina

04 Noviembre 2024

Como ya se había mencionado en el primer reporte sobre la clasificación de texto, existen dos tipos fácilmente separables, los métodos “shallow learning” y los métodos “deep learning”. En el caso de los métodos de shallow learning como Support Vector Machines (SVM), k-Nearest Neighbors (k-NN) o Logistic Regression (LR), un factor limitante importante es su dependencia en la extracción de características, lo que requiere un conocimiento significativo del dominio. Esto también implica que no se pueden generalizar a otras tareas fácilmente. Por lo tanto, el siguiente paso es el desarrollo del paradigma deep learning que tiene la habilidad de capturar información significativa sin las desventajas de los clasificadores clásicos.

Dentro de la categoría de deep learning se tienen a los: Multilayer Perceptrons (MLPs), son los diseños más básicos de las redes neuronales; Recurrent Neural Networks (RNNs) que son populares cuando tratas con información secuencial; Convolutional Neural Networks (CNN), normalmente usados en visión por computadora pero recientemente usadas en forma de Temporal Convolutional Networks que tienen la habilidad de capturar información temporal de alto nivel; Modelos basados en aproximación; RNN Encoder-Decoders, capaces de procesar secuencias de datos secuenciales y usados en traducción, generación de voz, etc; Attention Mechanism, es un mecanismo utilizado para que otras arquitecturas se enfoquen en las partes más relevantes; La Transformer Architecture, es un nuevo tipo de arquitectura neuronal que codifica los datos de entrada como características poderosas a través del mecanismo de atención<sup>1</sup>. A partir de la arquitectura de transformers se han generado avances en la construcción de modelos de deep learning, los cuales están utilizando un nuevo estándar para el procesamiento de lenguaje natural que consiste en tomar un modelo de lenguaje natural preentrenado y después especializarlo en las tareas a través de un “ajuste fino”. Sin embargo, a pesar de que los transformers son efectivos para modelar el

---

<sup>1</sup>Kai Han, An Xiao, Enhua Wu, Jianyuan Guo, Chunjing Xu, y Yunhe Wang. “Transformer in Transformer.” En \*Advances in Neural Information Processing Systems\*, editado por M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, y J. Wortman Vaughan, 15908–15919. Vol. 34. Curran Associates, Inc., 2021. Disponible en: [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/854d9fca60b4bd07f9bb215d59ef5561-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/854d9fca60b4bd07f9bb215d59ef5561-Paper.pdf).

lenguaje suelen aprender información redundante. Por esta razón surgen los modelos GPT y BERT, para prevenir las redundancias el GPT solo utiliza decodificadores mientras que el BERT utiliza codificadores haciendo a ambos modelos más ligeros y eficientes. A partir de estos modelos, surgen adecuaciones y mejoras. Algunos son: RoBERTa, mejora de BERT que utiliza un enmascaramiento dinámico y elimina la tarea de predicción de la siguiente oración, lo que mejora el rendimiento general; XLNet, combina las ventajas de BERT y GPT, utilizando un enfoque autoregresivo y un preentrenamiento que considera todas las permutaciones posibles de las palabras en una secuencia; DeBERTa, introduce una atención desencadenada y utiliza un preentrenamiento basado en MLM, mejorando la robustez del modelo; ByT5, una variante de T5 que trabaja con bytes en lugar de tokens, eliminando la necesidad de tokenización y mejorando la capacidad de manejo de texto; ERNIE 3.0, que integra información de grafos de conocimiento en su fase de preentrenamiento, mejorando así su comprensión de entidades y relaciones; FLAN, se centra en la fine-tuning con instrucciones para mejorar la capacidad de generalización a tareas no vistas. También en este artículo se menciona los Graph Neural Networks (GNN) que básicamente son arquitecturas que utilizan grafos para capturar dependencias y relaciones en sus nodos.

Finalmente, como ya había mencionado en el reporte 3, me hubiera gustado que en una recopilación tan extensa como ésta, no solo hubiesen abordado el área de clasificadores de texto. Sin embargo, por la misma extensión, ahora también creo que este artículo podrá dividirse en dos, modelos shallow y deep, para profundizar más en cada uno de ellos. Me gustaría ver en un solo artículo la información condensada, de tal manera que para el momento de evaluar qué modelo utilizar en mis proyectos, pudiera tener todas las herramientas para tomar decisiones.