

Entrenamiento de clasificación de reseñas de películas IMBD basada en una implementación de TensorFlow en el siguiente notebook

Adrian Pineda Sanchez A00834710

Comparación de Accuracy obtenidos:

Nsamp	Maxtokens	Maxtokenlen	Feature	LR	SVM	RF	GBM	BERT (N/A TF IDF o BOW)
1000	50	20	BOW	76.83%	75.66%	75.16%	77.16%	74.83%
1000	100	100	BOW	77.66%	77.33 %	76.33%	75.16%	74.17%
1000	200	200	BOW	81.16%	80.66%	79.83%	76.83%	76.83%
1000	230	200	BOW	79.16%	79.33%	80.16%	77%	79.5%
1000	50	20	TF-IDF	80.33%	79.16%	81.16%	80.16%	/
1000	100	100	TF-IDF	80.83%	79.33%	81.83%	79.16%	/
1000	200	200	TF-IDF	79.16%	81.16%	80.33%	79.16%	/
1000	230	200	TF-IDF	79.33%	81.33%	80.66%	79.66%	/

Comparativa de Resultados entre BERT y Modelos Tradicionales

Tarea de Clasificación Binaria (Reseñas IMDB)

En esta tarea, se evaluaron múltiples modelos de clasificación (Regresión Logística, SVM, RF, GBM y BERT) utilizando distintas configuraciones de características (BOW y TF-IDF). A pesar de que BERT logró resultados competitivos en términos de accuracy, las diferencias en el desempeño fueron mínimas en comparación con los modelos tradicionales. Por ejemplo:

Con la configuración de maxtokens = 230 y maxtokenlen = 200, los modelos SVM y RF alcanzaron precisiones de 81.33% y 80.66%, respectivamente, mientras que BERT obtuvo un 79.5% (en ocasiones podíamos obtener resultados de 80% hasta 81-82% dependiendo de la iteración). Esto indica que los modelos tradicionales, al utilizar representaciones manuales como TF-IDF, fueron capaces de lograr desempeños similares al de BERT para esta tarea binaria.

Ventajas en Tiempos de Ejecución

- Una de las principales ventajas observadas al utilizar embeddings preentrenados como los de BERT fue la drástica reducción en los tiempos de ejecución:
- Modelos tradicionales: El entrenamiento con TF-IDF y clasificadores como RF o SVM podía tomar varias horas debido al procesamiento intensivo que requería la construcción de matrices dispersas y el ajuste de hiperparámetros.

- BERT: Gracias al uso de embeddings preentrenados, se pudieron realizar pruebas en cuestión de minutos, acelerando significativamente el desarrollo y ajuste de los experimentos.

Otras Ventajas del Uso de Embeddings

- Generalización: BERT captura relaciones semánticas profundas del texto, lo que lo hace más robusto frente a datos no vistos.
- Menor preprocesamiento: A diferencia de los métodos tradicionales, no es necesario eliminar stop words ni realizar tokenizaciones complejas.
- Rapidez en experimentos iterativos: La integración de embeddings preentrenados permite realizar pruebas y ajustes de manera ágil, reduciendo los tiempos de ciclo de desarrollo.

Explicación del Desempeño Similar en la Clasificación Binaria (IMDB)

El desempeño similar entre BERT y los modelos tradicionales en esta tarea de clasificación binaria se debe a varios factores específicos del problema y la naturaleza de los datos:

1. Tamaño y Simplicidad del Conjunto de Datos:

- El conjunto de datos IMDB contiene reseñas de películas etiquetadas como positivas o negativas. Este tipo de datos tiende a ser más sencillo de modelar porque las diferencias entre clases suelen estar asociadas a palabras clave (e.g., "excelente" para positivas, "terrible" para negativas).
- Representaciones tradicionales como *TF-IDF* son efectivas para capturar estas diferencias, ya que asignan mayor peso a palabras distintivas.

2. Poca Necesidad de Contexto Complejo:

- Las reseñas de IMDB suelen ser relativamente cortas, y su clasificación depende principalmente de palabras o frases individuales más que de relaciones complejas entre términos.
- Modelos como *TF-IDF* y algoritmos como SVM son suficientemente buenos para este tipo de tarea porque no requieren modelar relaciones profundas en el texto, donde BERT tiene su mayor ventaja.

3. Configuración Limitada de BERT:

- En este experimento, no se realizó **fine-tuning** del modelo BERT en los datos específicos del conjunto de IMDB. Esto significa que BERT solo utilizó sus

embeddings generales preentrenados, lo que puede limitar su capacidad para especializarse en las características del conjunto de datos.

¿Cómo BERT Podría Mejorar?

1. Fine-Tuning en el Conjunto de Datos IMDB:

- Realizar ajuste fino (fine-tuning) permitiría que BERT se adapte específicamente a las características del conjunto de datos IMDB. Esto implicaría entrenar el modelo en las reseñas del conjunto de datos con un enfoque supervisado, ajustando los pesos del modelo preentrenado.

2. Incrementar el Tamaño del Conjunto de Datos:

- BERT es más efectivo cuando se entrena con grandes cantidades de datos, ya que puede aprovechar su capacidad para modelar relaciones complejas en texto. Si el conjunto de datos de reseñas se ampliara, BERT podría superar más fácilmente a los modelos tradicionales.

3. Ajustes en los Hiperparámetros:

- Ajustar parámetros como el número de épocas, la tasa de aprendizaje (*learning rate*) y el tamaño del lote (*batch size*) puede mejorar el rendimiento de BERT en esta tarea.

4. Contexto más Amplio:

- Si las reseñas fueran más largas y necesitaran análisis de contexto más profundo (e.g., sarcasmo, múltiples opiniones en un solo texto), BERT tendría una ventaja significativa sobre modelos tradicionales que no pueden capturar relaciones entre palabras en diferentes partes del texto.

5. Uso de Técnicas de Regularización:

- Aplicar técnicas como *dropout* o *early stopping* durante el entrenamiento puede mejorar la capacidad de generalización de BERT y reducir el sobreajuste.

Tarea de Clasificación Multiclase (AG News)

Por otro lado, en la clasificación multiclase, BERT demostró una ventaja significativa sobre los métodos tradicionales con un 95% cercano de accuracy. Su capacidad para generar

representaciones profundas y modelar relaciones complejas en los textos resultó en un desempeño muy superior en términos de precisión. En este caso, el uso de embeddings preentrenados fue crucial para abordar las diferencias semánticas entre las cuatro clases.

Conclusión

En la clasificación binaria de IMDB, las diferencias en accuracy entre BERT y los modelos tradicionales fueron pequeñas, pero la velocidad y facilidad de experimentación con BERT lo hacen una alternativa muy eficiente. En la clasificación multiclase de AG News, BERT superó ampliamente a los modelos clásicos, demostrando ser una herramienta poderosa para problemas más complejos. Esto refuerza el valor del uso de embeddings preentrenados en escenarios donde el tiempo y la precisión son críticos.

Referencias (entendimiento de BERT)

- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Disponible en: <https://arxiv.org/abs/1810.04805>