

Reporte sobre la Mejora del Modelo de Traducción.

1. Introducción Este documento presenta el proceso de optimización y análisis de un modelo de traducción automática basado en la arquitectura **Transformer**, diseñado para traducir del inglés al español. Se describen los ajustes realizados, la implementación de embeddings preentrenados y los resultados obtenidos mediante métricas de evaluación.

2. Objetivos del Proyecto

- Implementar y hacer funcional el modelo de traducción basado en Transformer.
- Incorporar embeddings preentrenados (GloVe) para mejorar la calidad de las traducciones.
- Evaluar el desempeño del modelo mediante métricas estándar como BLEU y ROUGE.

3. Metodología

3.1. Configuración del Modelo

Se partió del código base de Keras para modelos de traducción con Transformer, adaptándolo para permitir el almacenamiento del modelo y la integración de embeddings preentrenados.

3.2. Uso de Embeddings Preentrenados

Para mejorar la representación semántica de las palabras, se utilizaron los embeddings de **GloVe (Global Vectors for Word Representation)**:

- Se descargaron los vectores glove.6B.100d.txt.
- Se construyó un diccionario de palabras y sus representaciones vectoriales.
- Se reemplazaron las inicializaciones aleatorias del modelo con estos embeddings.

3.3. Ajuste del Proceso de Entrenamiento

- Se realizaron **100 épocas** de entrenamiento.
- Se utilizó el optimizador **Adam** con una tasa de aprendizaje ajustada a 0.0005.
- Se empleó la técnica de Dropout para mitigar el sobreajuste.

3.4. Evaluación del Desempeño

Se emplearon las siguientes métricas para analizar la calidad de las traducciones:

1. **BLEU (Bilingual Evaluation Understudy)**: Compara la traducción generada con una referencia humana basándose en n-gramas.
2. **ROUGE (Recall-Oriented Understudy for Gisting Evaluation)**: Evalúa la similitud entre la traducción y la referencia humana a nivel de secuencias.

Resultados obtenidos:

- **BLEU Score promedio:** $\sim 7.46e-155$ (sugiere dificultades en la generación de n-gramas)
- **ROUGE Score promedio:**
 - ROUGE-1 (unigramas): Precisión = 1.0, Recall = 1.0, F1 = 0.9999
 - ROUGE-2 (bigramas): Precisión = 0.5, Recall = 0.5, F1 = 0.4999
 - ROUGE-L (longest common subsequence): Precisión = 0.75, Recall = 0.75, F1 = 0.75

Estos resultados sugieren que el modelo aún enfrenta dificultades con la generación de secuencias más largas.

4. Conclusiones y Próximos Pasos

- **Avances logrados:** Se completó el entrenamiento del modelo y se integraron embeddings preentrenados.
- **Retos detectados:** Se identificaron áreas de mejora en la precisión de n-gramas más largos.
- **Pasos futuros:**
 - Ajustar hiperparámetros para optimizar aún más el rendimiento.
 - Ampliar el conjunto de pruebas para evaluar mejor la capacidad de generalización del modelo.