

Laboratorio 4

Laboratorio: Mejorando el Análisis de Sentimientos con LSTM y Características Adicionales

Objetivo: Incrementar la precisión en el análisis de sentimientos sobre las críticas de películas utilizando RNNs con unidades LSTM y la incorporación de características (features) adicionales.

Instrucciones:

1. **Importación de Datos:** Utilice el conjunto de datos IMDB proporcionado por Keras. pero esta vez, en lugar de utilizar sólo las 20.000 palabras más frecuentes, utilice las 50.000 palabras más frecuentes.

2. **Pre-procesamiento:**

- Secuencie y rellene las críticas para que todas tengan una longitud uniforme.
- De las críticas, extraiga características (features) adicionales, por ejemplo. la longitud de la crítica, la proporción de palabras positivas/negativas y cualquier otra que considere pueda ser útil.

3. **Modelo:**

- Cree un modelo LSTM que acepte las características (features) adicionales junto con la secuencia de palabras.
- Intente usar una arquitectura más compleja, incorporando más capas LSTM, capas de *Dropout* para la regularización y tal vez alguna capa densamente conectada después de la LSTM. (ver también la referencia al final de este documento)

4. **Entrenamiento y Evaluación:** Entrene su modelo con el conjunto de datos de entrenamiento y evalúe su desempeño con el conjunto de datos de prueba.

5. **Informe:** Escriba un breve informe que describa:

- Las características adicionales seleccionadas y la razón de su elección.
- La arquitectura del modelo y las razones detrás de sus elecciones.
- Los resultados obtenidos y una breve comparación con el modelo simple del ejercicio anterior.

Rúbrica de Evaluación:

1. Importación y Pre-procesamiento (30 puntos)

- Correcta importación del conjunto de datos con 50,000 palabras más frecuentes: 10 puntos.
- Secuenciación y relleno de las críticas: 10 puntos.
- Extracción y adecuada justificación de características adicionales: 10 puntos.

2. Arquitectura del Modelo (30 puntos)

- Adecuado diseño de la arquitectura LSTM: 10 puntos.
- Incorporación efectiva de características adicionales en el modelo: 10 puntos.
- Uso de técnicas adicionales (e.g., Dropout, capas densamente conectadas): 10 puntos.

3. Entrenamiento y Evaluación (20 puntos)

- Correcto entrenamiento del modelo sin errores: 10 puntos.
- Evaluación y comparación adecuada con el modelo del ejercicio anterior: 10 puntos.

4. Informe (20 puntos)

- Clara descripción de características adicionales y su relevancia: 5 puntos.
- Explicación coherente de la arquitectura del modelo: 5 puntos.
- Presentación de resultados y comparativa efectiva: 10 puntos.

Recurso de Datos:

- [IMDB Movie reviews sentiment classification](#)

Referencia:

- https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf