

Machine translation

NMT-Keras

Adrián Vázquez Barrera
Polytechnic University of Valencia
January 2022

Introducción

En esta práctica se pretenden construir sistemas de traducción utilizando pares de frases bilingües. Para este fin, utilizaremos la NMT-Keras para entrenar modelos de traducción basados en redes neuronales recurrentes, LSTM y transformers.

Contexto

Para la realización de esta práctica, utilizaremos un conjunto de datos bilingüe de la tarea EuTrans para Español-Inglés, en el contexto de un turista frente al mostrador de un hotel.

Ejercicio 1

Se solicita estudiar el comportamiento de la red al variar el tamaño de la capa de embeddings tanto del encoder como del decoder.

Como puede observarse en la tabla de resultados, vemos como el mejor valor se encuentra en torno a los 256. Debemos tener en cuenta que estamos ante un corpus bastante pequeño, por lo que es de esperar que un tamaño para capa de embeddings muy alto, no sea capaz de aprender las reglas del idioma.

Embedding size	BLEU
32	87.74%
64	94.00%
128	97.70%
256	98.20%
512	97.10%

Ejercicio 2

Se pide analizar el comportamiento de la red al modificar el tamaño del modelo LSTM. Para ello se ha optado por mantener el tamaño óptimo de embeddings obtenido en ejercicio anterior. Tras obtener los resultados, vemos como los mejores valores se encuentran en un tamaño de LSTM de entre 128 y 256. De nuevo, puede deberse al reducido tamaño del corpus de entrenamiento, ya que un valor mayor puede inducir al sobre-ajuste.

Encoder/decoder size	BLEU
32	97.24%
64	98.47%
128	98.29%
256	97.64%
512	89.69%

Ejercicio 3

En este ejercicio se pide analizar el comportamiento de la red al utilizar varios optimizadores. Para que la comparación sea justa se ha establecido un máximo de 5 epochs.

Optimizador	BLEU
Adam	98.47%
Adagrad	13.62%
Adadelata	1%

A la vista de los resultados vemos como Adam presenta la convergencia más rápida, tanto Adagrad como sobre todo Adadelata presentan un BLEU mucho más bajo en las mismas condiciones.

Ejercicio 4

Al utilizar una arquitectura basada en transformers se observa una disminución del BLEU con respecto a lo visto en redes LSTM, a cambio, cabe destacar el grado de paralelismo que ofrecen los modelos transformer, lo que agiliza el entrenamiento de dichas redes.

Model size	BLEU
32	63.51%
64	87.50%
128	90.25%
256	86.63%