

Descripción automática de imágenes

Miembros

- Jose Arias Moncho
- Victoria Beltrán Domínguez

Descripción

Este proyecto, como su título indica, tiene el objetivo de crear un modelo para la descripción automática de imágenes, tarea que se suele conocer como *Image Captioning*. Para ello, utilizaremos el conjunto de datos de *Flickr*, en concreto *Flickr30k*, que contiene 31783 imágenes (cada una de las cuáles tiene cinco descripciones diferentes) y que se encuentra disponible en [Kaggle](#). Este conjunto de datos deberá ser correctamente preprocesado y separado en conjuntos de entrenamiento, validación y test. Dentro del abanico de técnicas que podríamos abordar para solucionar este problema, vamos a centrarnos en las redes neuronales. En concreto, vamos a trabajar con una arquitectura basada en un codificador y decodificador, que respectivamente nos servirán para obtener una representación vectorial de la imagen concreta y utilizarla para generar una descripción textual. La implementación de esta arquitectura se realizará con Keras y Tensorflow. Para la obtención de la representación de la imagen trabajaremos con distintas redes convolucionales preentrenadas, comparando las métricas resultantes. Por otro lado, para generar la descripción a partir de la representación obtenida con la red convolucional, vamos a experimentar con diferentes arquitecturas (como LSTMs o opciones más complejas). Debido a la alta variabilidad de las medidas de evaluación empleadas en los diferentes trabajos realizados en Kaggle sobre este conjunto de datos, hemos optado por trabajar con el BLEU como medida de evaluación, pues es una medida que sí parecía estar extendida entre la literatura científica dentro de esta tarea.

Lista de objetivos

1. Preparar los datos (preprocesar el texto e imágenes por separado).
2. Definir y generar los conjuntos de datos de entrenamiento, validación y test.
3. Seleccionar de una primera arquitectura.
4. Implementar esta arquitectura.

5. Realizar el entrenamiento y evaluación.
6. Revisar el estado del arte para este conjunto de datos.
7. Volver al tercer paso y seleccionar varias arquitecturas diferentes en función del tiempo.
8. Comparar y analizar los resultados obtenidos con las diferentes arquitecturas y con el estado del arte.

Objetivos a cumplir en la primera revisión: 1-5