Final Proyect Progress Report
Gossip Alert Generation

Bishops

Sofía Negueruela Avellaneda
Enrique Fernández-Baíllo Rodríguez de Tembleque
Jacobo Banús Bertram
Almudena Garrido García-Pita

Aprendizaje Profundo &
Procesamiento de Lenguaje Natural
3º Grado en Ingeniería Matemática e Inteligencia Artificial

Progress Report

The rise of entertainment journalism and celebrity culture has led to an increasing demand for automated analysis of news. Our goal is to develop a model that processes news articles of this nature and extracts meaningful insights, focusing on Named Entity Recognition (NER) and Sentiment Analysis (SA).

At first, we were following the pipeline proposed below:

- Data acquisition. We source datasets containing multimodal input (text and images) for analysis.

- Models. Once we got the data, we are going to build our NER and SA models.

  o For the NER (Name Entity Recognition) model, we implement model to recognize entertainment-related entities.

  o For the SA (Sentiment Analysis) model, we train a sentiment classifier to determine positive, negative, or neutral tones in articles.

- Training & Evaluation: Both models will be trained with a combined loss function and evaluated using accuracy, MAE, and F1-score. These loss functions may not be the ones we end up using, we just want to evaluate the performance of our models using several measurements. In the end, we want the model to be able to generalize and classify correctly new input data, so we will have different data for training and evaluating our models.

- Automatic Summary Generation. After both models show good enough results, our goal is to generate a summary of the key aspects of the information received as input.

- Image Captioning. Finally, we want to integrate an image-text model, which extracts relevant descriptions from images, and integrate it with the rest of the model. To simplify the task, we used a pre-trained model (transformers) to generate captions from images.

However, as we wanted to implement image captioning, we decided to do that first so that the rest of the model could be trained with the image captions and the articles at the same time. Therefore, we downloaded a dataset with multimodal input (image and text). The text also included the NER tags. Before start designing and training our model, we focused on obtaining accurate data: from the images, we generated captions using a pre-trained model. From the text and the captions generated, we obtained SA tags also using a pre-trained model. In the end, we created a Dataset containing the image ID, the captions, the texts, the NER tags for each word of the text and the SA tags.