Methodology

Multi-modal Feature Extraction - Image

- Compare to existing multi-modal fake news detection studies that often directly apply a pre-trained CNN model (e.g., VGG) to obtain the representation of news images
- Use image2sentence for consistency and to increase insights when computing the similarity across modalities.

a red double decker bus driving down a street.

Methodology

Modal-independent Fake News Detection

- To properly represent news textual and visual information in predicting fake news, we aim to correctly map the extracted textual and visual features of news content to their possibilities of being fake, and further to their actual labels.
  - Possibilities can be computed by $M_p(t, v) = 1 \cdot \text{softmax}(W_p(t \oplus v) + b_p)$
    - $1 = [1,0]^T$, $W_p \in \mathbb{R}^{2 \times 2d}$ and $b_p \in \mathbb{R}^2$ are parameters to be trained.
- Cross-entropy-based loss function:
  - $L_p(\theta_t, \theta_v, \theta_p) = -\mathbb{E}_{(a,y)\sim(A,Y)}(y \log M_p(t, v) + (1 - y)\log(1 - M_p(t, v)))$