Methodology

Cross-modal Similarity Extraction

- Then defined the loss function based on cross-entropy as below, which assumes that news articles formed with mismatched textual and visual information are more likely to be fake compared to those with matching textual statements and images, when analyzing from a pure similarity perspective:
  - $L_S(\theta_t, \theta_v) = -\mathbb{E}_{(a,y)\sim(A,Y)}(y \log(1 - M_s(t, v)) + (1 - y)\log M_s(t, v))$

Methodology

Model Integration and Joint Learning

- When detecting fake new, we aim to correctly recognize fake news stories whose falsity is in their textual and/or visual information, or their relationship.
- Final loss function as
  - $L(\theta_t, \theta_v, \theta_p) = \alpha L_p(\theta_t, \theta_v, \theta_p) + \beta L_s(\theta_t, \theta_v)$
    - $L_p(\theta_t, \theta_v, \theta_p) = -\mathbb{E}_{(a,y)\sim(A,Y)}(y \log M_p(t, v) + (1 - y)\log(1 - M_p(t, v)))$
    - $L_S(\theta_t, \theta_v) = -\mathbb{E}_{(a,y)\sim(A,Y)}(y \log(1 - M_s(t, v)) + (1 - y)\log M_s(t, v))$