Methodology

Reinforced Weakly-supervised Fake News Detection Framework

- First, pre-train the annotator using the labeled report data $\{R, Y\}$ and assign weak labels $\hat{Y}^u$ to the unlabeled news set X^u
- The proposed reinforced selector will select high-quality samples $\{X_s, Y_s\} = \left\{ X_s^{(k)}, Y_s^{(k)} \right\}_{k=1}^{K}$ from the weakly labeled dataset $\{X^u, \hat{Y}^u\}$
- Then both selected dataset $\{X_s, Y_s\}$ and original labeled data $\{X, Y\}$ are fed into the fake news detector for training.

Methodology

Reinforced Weakly-supervised Fake News Detection Framework

- The final loss of fake news detection consists of two sub losses:
- $L_n\left(X, Y, X_s, Y_s; \theta_n\right) = \lambda_l \cdot L_n^l\left(X, Y; \theta_n\right) + \lambda_s \cdot L_n^s\left(X_s, Y_s; \theta_n\right)$
- Simply set the values of λ_l and λ_u as 1
  - Loss on a small amount of manually labeled data:
    - $L_n^l\left(X, Y; \theta_n\right) = -\mathbb{E}_{(x,y)\sim(X,Y)}\left[y \log\left(D_n\left(x; \theta_n\right)\right) + (1-y)\log\left(1 - D_n\left(x; \theta_n\right)\right)\right]$
  - Loss on automatically annotated data set
    - $L_n^s\left(X_s, Y_s; \theta_n\right) = -\mathbb{E}_{\left(x_s,y_s\right)\sim\left(X_s, Y_s\right)}\left[y_s \log\left(D_n\left(x_s, \theta_n\right)\right) + \left(1-y_s\right)\log\left(1 - D_n\left(x_s; \theta_n\right)\right)\right]$