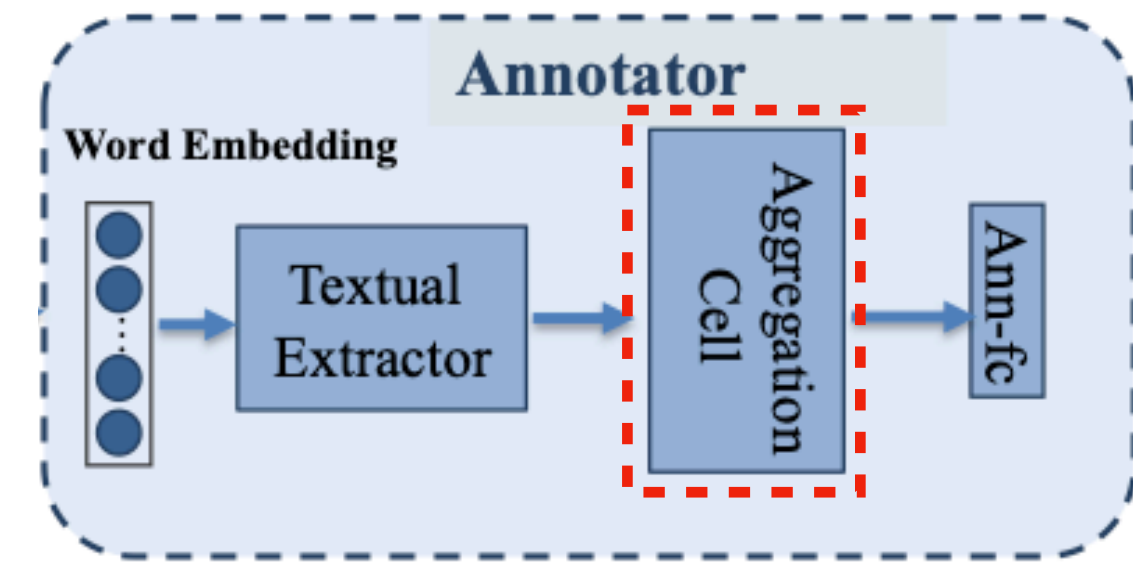


Methodology

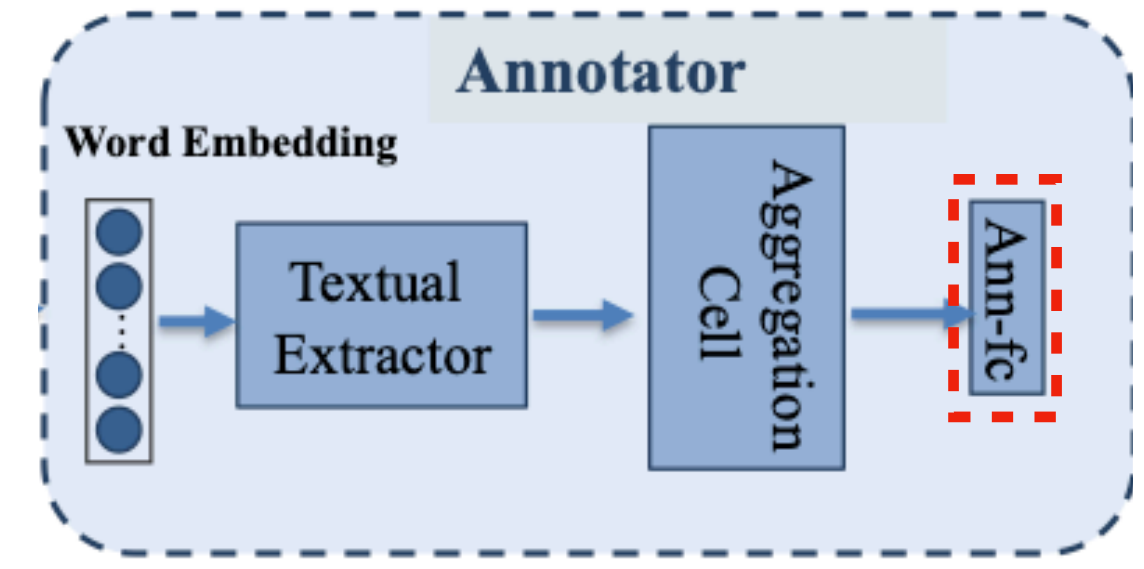
Automatic Annotation based on Reports



- Take the i -th sample as an example, and the j -th report message: $r_j^{(i)}$
- The corresponding report message set is denoted as $R^{(i)} = \{r_1^{(i)}, r_2^{(i)}, \dots, r_{|R^{(i)}|}^{(i)}\}$
 - $|R^{(i)}|$: number of report messages of i -th sample
- $r_j^{(i)} \in R^{(i)}$ is first fed into the textual feature extractor to obtain $\mathbf{h}_j^{(i)}$
- Use the aggregation cell to combine $R^{(i)}$ to learn the hidden feature representation $\mathbf{h}^{(i)}$
- Procedure of aggregation cell: $\mathbf{h}^{(i)} = \text{ReLU} \left(\mathbf{w}_r \cdot \sum_{j=1}^{|R^{(i)}|} \frac{\mathbf{h}_j^{(i)}}{|R^{(i)}|} \right)$, \mathbf{w}_r : weight of the fully-connected layer

Methodology

Automatic Annotation based on Reports



- Feed $\mathbf{h}^{(i)}$ into the fully connected layer, denoted as Ann-fc, to output the corresponding probability of the i -th sample being a fake one
 - $D_r(R^{(i)}; \theta_r)$, θ_r : all parameters of the annotator and corresponding textual feature extractor
- Entire report message dataset $R = \{R^{(1)}, R^{(2)}, \dots, R^{(|R|)}\}$, $|R|$: number of report sets
- Corresponding ground truth labels of news $Y = \{y^{(1)}, y^{(2)}, \dots, y^{(|R|)}\}$
- Loss function for the proposed annotator is defined by cross entropy as follows:

$$L_r(R, Y; \theta_r) = -\frac{1}{|R|} \sum_{i=1}^{|R|} \left[y^{(i)} \log D_r(R^{(i)}; \theta_r) + (1 - y^{(i)}) \log (1 - D_r(R^{(i)}; \theta_r)) \right]$$