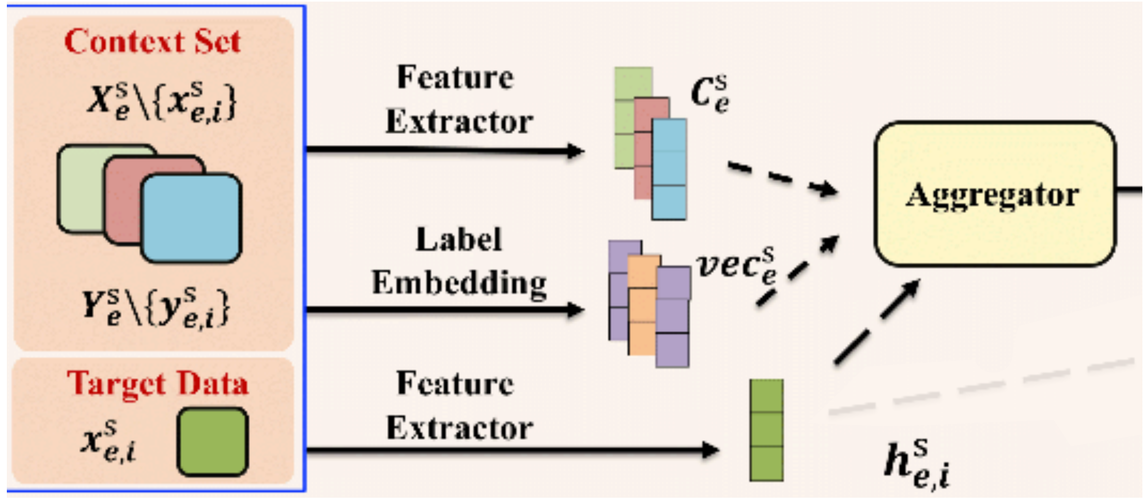
Methodology

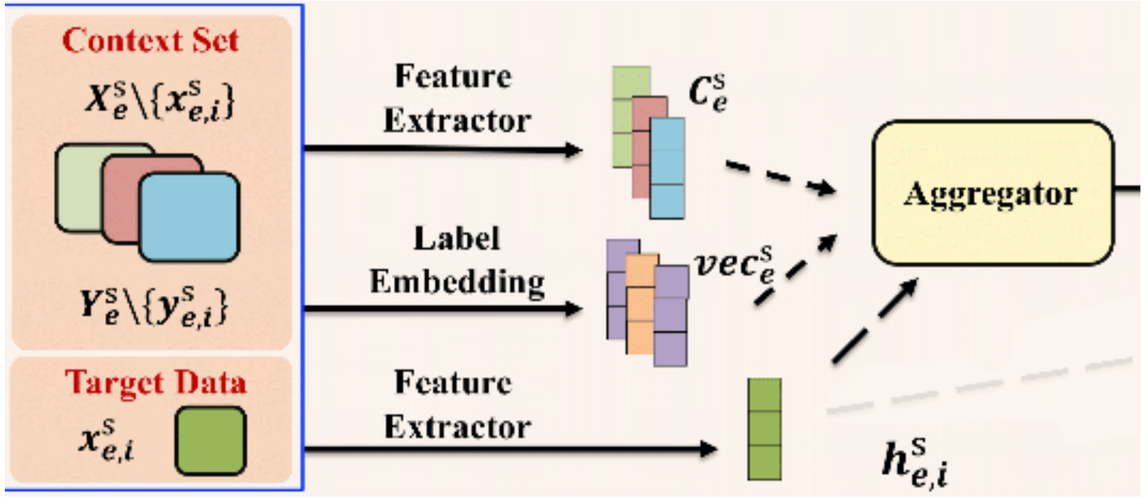
Aggregator: Attention



- Use scaled dot-product attention mechanism
- Mapping a query $\mathbf{Q}$ and a set of key $\mathbf{K}$ - value $\mathbf{V}$ pairs to an output
- $\mathbf{Q}_i = \mathbf{W}_q \mathbf{h}_{e,i}$, $\mathbf{K} = \mathbf{W}_k \mathbf{C}_e$, $\mathbf{V} = \mathbf{W}_v \left(\mathbf{C}_e \oplus \mathbf{vec}_e \right)$
- $\mathrm{a}_i = \mathrm{softmax} \left(\dfrac{\mathbf{Q}_i \, \mathrm{K}^T}{\sqrt{d}} \right)$
- Attention $\left(\mathbf{Q}_i, \mathbf{K}, \mathbf{V} \right) := a_i \mathbf{V}$

Methodology

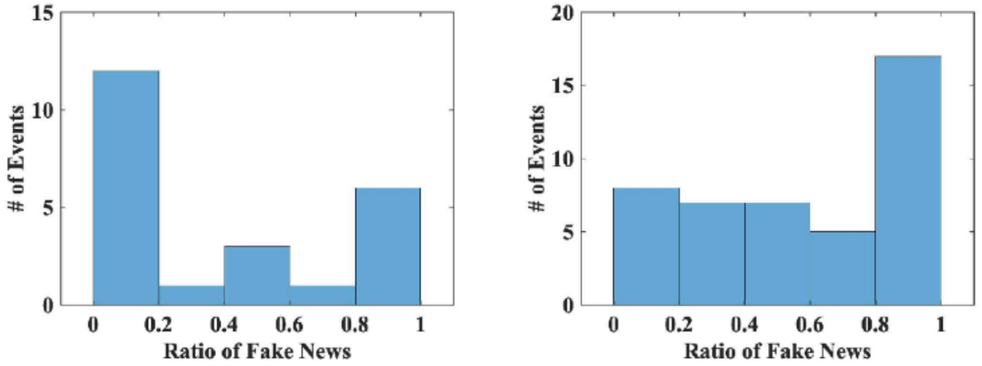
Aggregator: Limitation of Soft-attention



- $$\mathrm{a}_i = \mathrm{softmax}\left(\frac{\mathbf{Q}_i\,\mathrm{K}^T}{\sqrt{d}}\right)$$

- The attention mechanism with soft weight values is categorized into soft-attention.

- However, soft-attention cannot effectively trim irrelevant data especially when have a context set with an imbalanced class distribution as mentioned before.



The number of events with respect to different percentages of fake news.