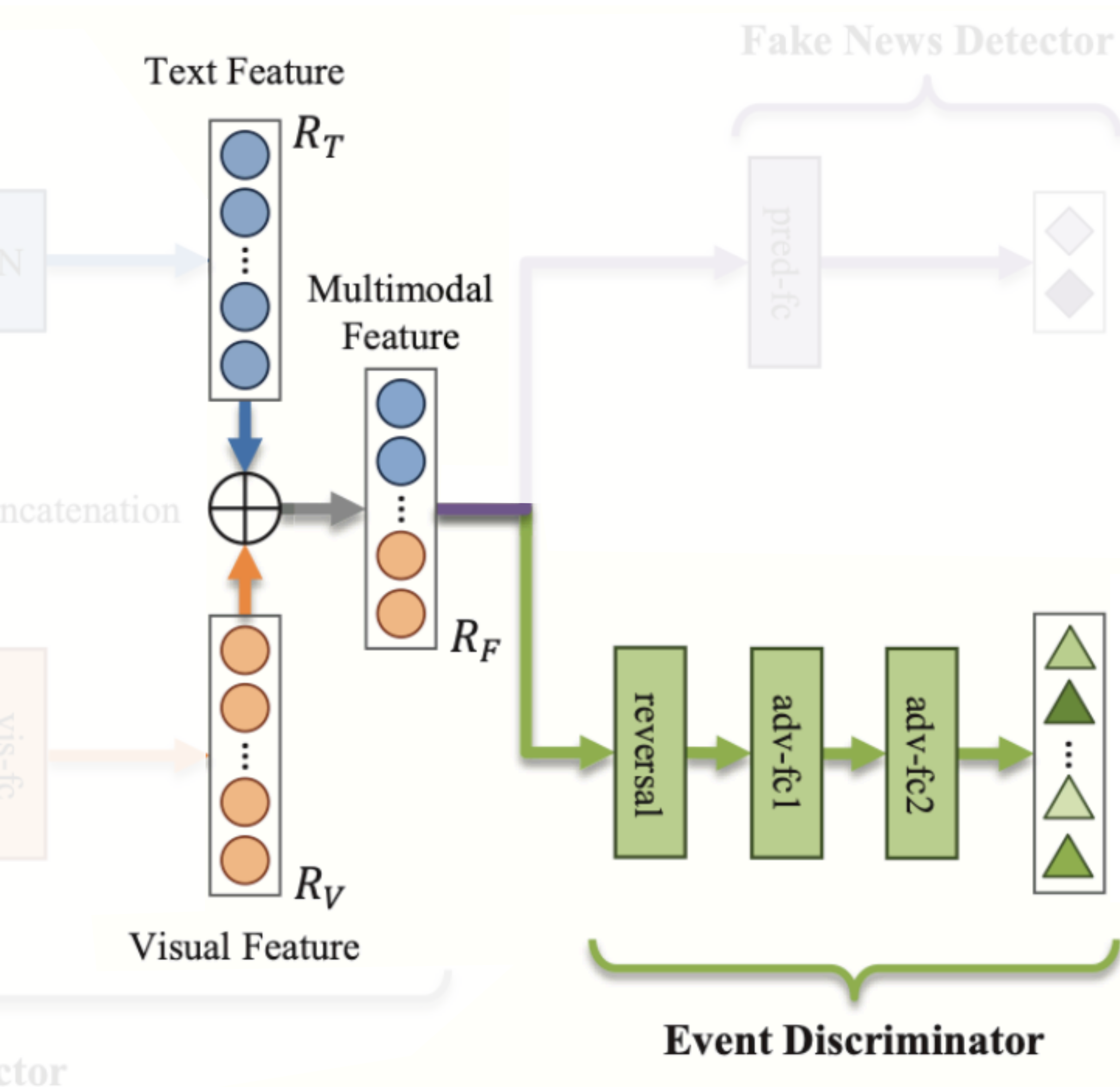


Methodology.....

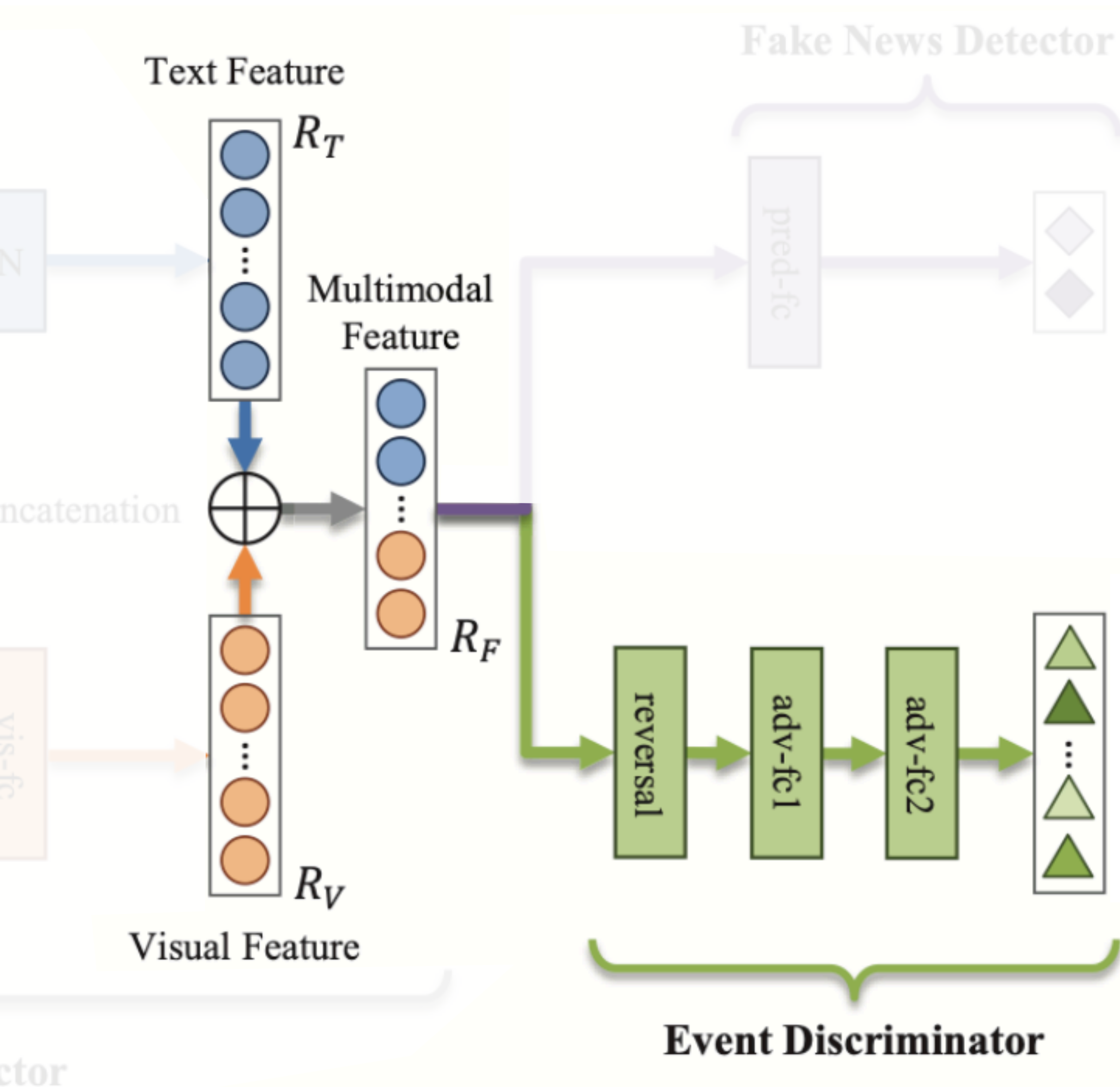
Event Discriminator



- Measure the dissimilarities of the feature representations among events and remove them in order to capture the event invariant feature representation.
- NN which consists of two fully connected layers with corresponding activation functions.
- Aims to correctly classify the post into one of K events based on the multi-modal feature representations.
- Denote as $G_e(R_F; \theta_e)$, θ_e : parameters

Methodology.....

Event Discriminator



- Loss function by cross entropy:

$$L_e(\theta_f, \theta_e) = - \mathbb{E}_{(m,y) \sim (M, Y_e)} \left[\sum_{k=1}^K 1_{[k=y]} \log(G_e(G_f(m; \theta_f); \theta_e)) \right]$$

- Parameters minimizing the loss $L_e(\cdot, \cdot)$:

$$\hat{\theta}_e = \arg \min_{\theta_e} L_e(\theta_f, \theta_e)$$

- Large loss means the events' representations are similar and the learned feature are event-invariant.

- Need to maximize the $L_e(\theta_f, \hat{\theta}_e)$ by seeking the optimal parameters θ_f