

# Comments

## of Similarity-Aware FakeE news detection (SAFE)

- Add across-modal relationship (similarity) to detect fake news detection
- Use image2sentence get image caption
  - Modify cosine similarity equation
- Baseline:
  - Text feature baseline only one of traditional method
  - Multi-modal baseline only one to compared

# Optimization of SAFE

---

## Algorithm 1: SAFE

---

**Input:**  $A = \{(T_j, V_j)\}_{j=1}^m$ ,  $Y = \{y_j\}_{j=1}^m$ ,  $H = \{h_k\}_{k=1}^g$ ,  $\gamma$   
**Output:**  $\theta_p = \{\mathbf{W}_p, \mathbf{b}_p\}$ ,  $\theta_t = \{\mathbf{W}_t, \mathbf{b}_t, \mathbf{w}_t, b_t\}$ ,  $\theta_v = \{\mathbf{W}_v, \mathbf{b}_v, \mathbf{w}_v, b_v\}$

```

1 Randomly initialize  $\mathbf{W}_p, \mathbf{b}_p, \mathbf{W}_t, \mathbf{b}_t, \mathbf{w}_t, b_t, \mathbf{W}_v, \mathbf{b}_v, \mathbf{w}_v, b_v$ ;
2 while not convergence do
3   foreach  $(T_j, V_j)$  do
4     Update  $\theta_p$ :  $\{\mathbf{W}_p, \mathbf{b}_p\} \leftarrow$  Eq. (12);
5     foreach  $h_k$  do
6       Update  $\theta_t$ :  $\{\mathbf{W}_t, \mathbf{b}_t, \mathbf{w}_t, b_t\} \leftarrow$  Eqs. (14-18);
7       Update  $\theta_v$ : similar to updating  $\theta_t$ ;
8     end
9   end
10 end
11 return  $\mathbf{W}_p, \mathbf{b}_p, \mathbf{W}_t, \mathbf{b}_t, \mathbf{w}_t, b_t, \mathbf{W}_v, \mathbf{b}_v, \mathbf{w}_v, b_v$ 

```

---

**Update  $\theta_p$ .** Let  $\gamma$  be the learning rate, the partial derivative of  $\mathcal{L}$  w.r.t.  $\theta_p$  is:

$$\theta_p \leftarrow \theta_p - \gamma \cdot \alpha \frac{\partial \mathcal{L}_p}{\partial \theta_p}. \quad (11)$$

As  $\theta_p = \{\mathbf{W}_p, \mathbf{b}_p\}$ , updating  $\theta_p$  is equivalent to updating both  $\mathbf{W}_p$  and  $\mathbf{b}_p$  in each iteration, which respectively follow the following rules:

$$\mathbf{W}_p \leftarrow \mathbf{W}_p - \gamma \cdot \alpha \Delta \mathbf{y} (\mathbf{t} \oplus \mathbf{v})^\top, \quad \mathbf{b}_p \leftarrow \mathbf{b}_p - \gamma \cdot \alpha \Delta \mathbf{y}, \quad (12)$$

where  $\Delta \mathbf{y} = [\hat{y} - y, y - \hat{y}]^\top$ .

**Update  $\theta_t$ .** The partial derivative of  $\mathcal{L}$  w.r.t.  $\theta_t$  is generally computed by

$$\theta_t \leftarrow \theta_t - \gamma \left( \alpha \frac{\partial \mathcal{L}_p}{\partial \mathcal{M}_t} \frac{\partial \mathcal{M}_t}{\partial \theta_t} + \beta \frac{\partial \mathcal{L}_s}{\partial \mathcal{M}_t} \frac{\partial \mathcal{M}_t}{\partial \theta_t} \right). \quad (13)$$

Let  $\nabla \mathcal{L}_*(\mathbf{t}) = \frac{\partial \mathcal{L}_*}{\partial \mathcal{M}_t}$ ,  $\mathbf{t}_0 = \frac{\mathbf{t}}{\|\mathbf{t}\|}$ ,  $\mathbf{v}_0 = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ , and  $\mathbf{W}_{p,L}$  denote the first  $d$  columns of  $\mathbf{W}_p$ , we can have

$$\nabla \mathcal{L}_p(\mathbf{t}) = \mathbf{W}_{p,L}^\top \Delta \mathbf{y}, \quad (14)$$

$$\nabla \mathcal{L}_s(\mathbf{t}) = \frac{1-y}{2s\|\mathbf{t}\|} ((2s-1)\mathbf{t}_0 - \mathbf{v}_0), \quad (15)$$

based on which the parameters in  $\theta_t$  are respectively updated as follows:

$$\mathbf{W}_t \leftarrow \mathbf{W}_t - \gamma \cdot \mathbf{D}_t \mathbf{B}_t, \quad \mathbf{b}_t \leftarrow \mathbf{b}_t - \gamma \cdot \mathbf{B}_t, \quad (16)$$

$$\mathbf{w}_t \leftarrow \mathbf{w}_t - \gamma \cdot \mathbf{x}_t^{\hat{i}:(\hat{i}+h-1)} \mathbf{W}_t^\top \mathbf{B}_t, \quad b_t \leftarrow b_t - \gamma \cdot \mathbf{W}_t^\top \mathbf{B}_t, \quad (17)$$

where  $\hat{i} = \arg \max_i \{c_t^i\}_{i=1}^{n-h+1}$ ,  $\mathbf{D}_t \in \mathbb{R}^{d \times d}$  is a diagonal matrix with entry value  $\hat{c}_t^{\hat{i}}$ , and

$$\mathbf{B}_t = \alpha \nabla \mathcal{L}_p(\mathbf{t}) + \beta \nabla \mathcal{L}_s(\mathbf{t}). \quad (18)$$

**Update  $\theta_v$ .** It is similar to updating  $\theta_t$ ; we omit details due to space constraints.