MixMatch with GAT

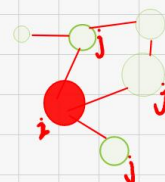## 1. 关键机制

1.1 Attention mechanism

① self-attention 计算 attention 系数

$$c_{ij} = a(W\vec{h_i}, W\vec{h_j})$$

其中, $a : R^{\vec{F'}} \times R^{\vec{F'}} \longrightarrow R$

② multi-head attention

$$h_i' = \big\|_{k=1}^{K} \sigma\big(\sum_{j \in N_i} \alpha_{ij}^k W^k \vec{h_j}\big)$$

1.2 MixMatch 机制

① Data Augmentation

$$\begin{cases} x_b \in X, & X 为 \text{ labeled data} \\ u_b \in U, & U 为 \text{ unlabeled data} \end{cases}$$

$$\hat{x_b} = Augment(x_b)$$

$$\hat{u}_{b,k} = Augment(u_b) \quad k \in (1, \cdots, K), \ K 次 Augmentation.$$

② Label Guessing

$$\bar{q}_b = \frac{1}{K} \sum_{k=1}^{K} model(\hat{u}_{b,k})$$

$$q_b = Sharpen(\bar{q}_b, T)$$

其中 $Sharpen(P, T) = \dfrac{P_i^{\frac{1}{T}}}{\sum_{j=1}^{L} P_j^{\frac{1}{T}}}$, L为标签类别总数

③ Mix Up

集合: $\hat{X} = ((\hat{x}_b, P_b); b \in (1, \cdots B))$

$$\hat{U} = ((\hat{u}_{b,k}, q_b); b \in (1 \cdots B), k \in (1, \cdots K))$$

$$W = shuffle(Concat(\hat{X}, \hat{U})$$

$$X' = MixUp(\hat{X}_i, W_i) \quad i \in (1, \cdots |\hat{X}|)$$

$$U' = MixUp(\hat{U}_i, W_{i+|\hat{X}|}) \quad i \in (1, \cdots |\hat{U}|)$$

④ Loss function

$$L_x = \frac{1}{|X'|} \sum_{x,P \in X'} H(p, model(x)) \quad (3)$$

$$L_u = \frac{1}{L|U'|} \sum_{u,q \in U'} \| q - model(u) \|_2^2 \quad (4)$$

$$L = L_x + \lambda_u L_u \quad (5)$$

$$\lambda u = 0 : 0.005 : 1$$

$$K = 2$$

$$T = 0.5$$