

MixMatch with GAT

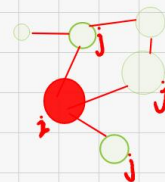
1. 关键机制

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, $a: \mathbb{R}^F \times \mathbb{R}^F \rightarrow \mathbb{R}$



② multi-head attention

$$h'_i = \parallel_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^k W^k \vec{h}_j \right)$$

1.2 Mix Match 机制

① Data Augmentation

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

$$\hat{x}_b = \text{Augment}(x_b)$$

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

② Label Guessing

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

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

其中, $\text{Sharpen}(p, T) = \frac{p_i^{\frac{1}{T}}}{\sum_{j=1}^L p_j^{\frac{1}{T}}}$, L 为标签类别总数

③ Mix Up

$$\text{集合: } \hat{X} = (\hat{x}_b, p_b); b \in (1, \dots, B)$$

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

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

$$X' = \text{Mix Up}(\hat{X}_i, W_i) \quad i \in (1, \dots, |\hat{X}|)$$

$$U' = \text{Mix Up}(\hat{U}_i, W_{i+|\hat{X}|}) \quad i \in (1, \dots, |\hat{U}|)$$

④ Loss function

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

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

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

$$\lambda_u = 0:0.0005:1$$

$$K=2$$

$$T=0.5$$

流程:
(Add unlabeled only)

Add Augmentation

unlabeled only \leftarrow idx-ulb
features
adj

Model

new feature
new-adj
coef
idx-aug

T

Guess Label

(detach)
guesses

Shuffle W \rightarrow W

(U)

Mix Up

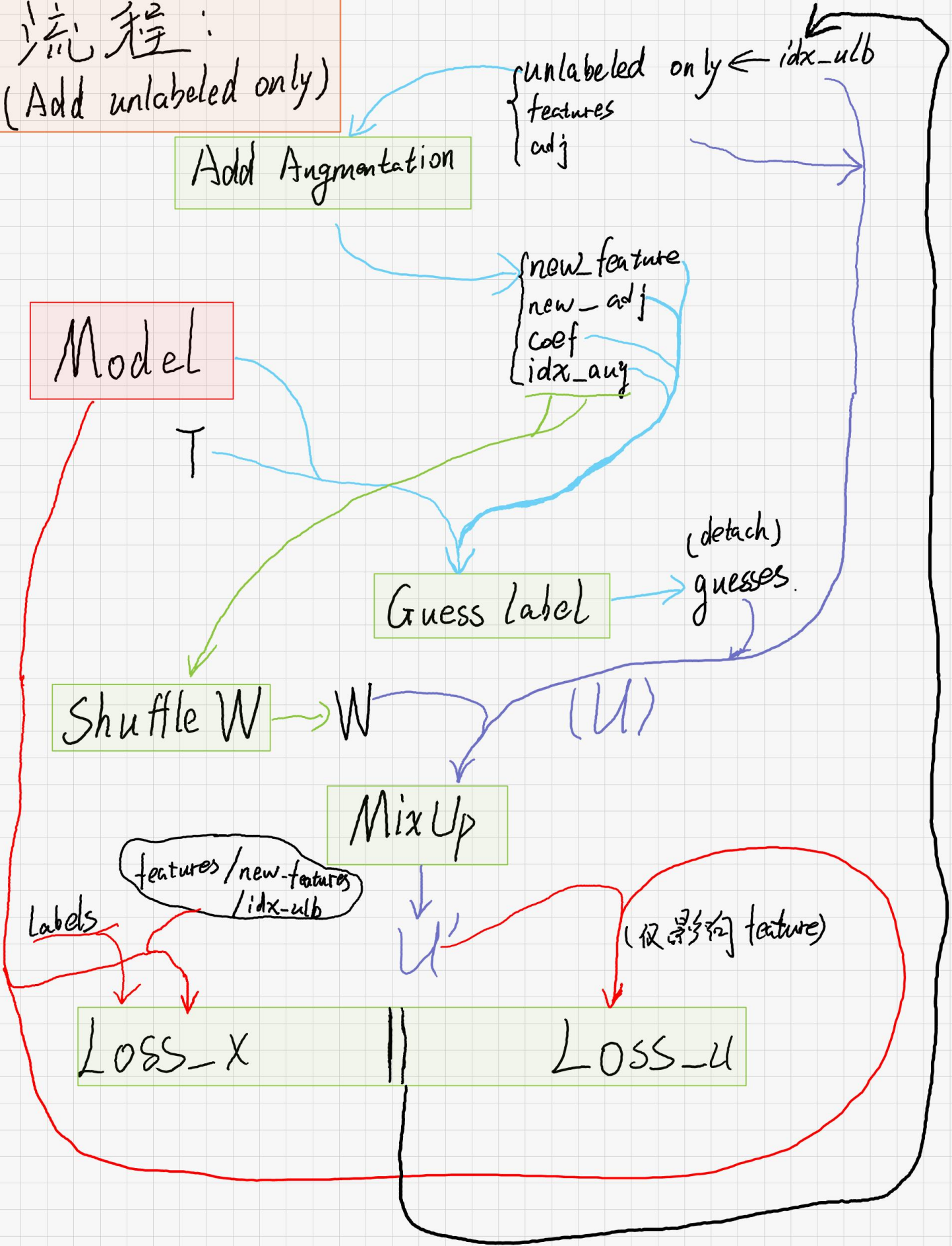
features/new-features
idx-ulb

Labels

(仅影响 feature)

Loss-X

Loss-U



Points.

1. Guess label $\begin{cases} \textcircled{1} \text{ 使用 new-features} \\ \textcircled{2} \text{ 使用 features + idx_wtxcoef 进行混合} \end{cases}$

2. MixUp \rightarrow Shuffle

(a) $\begin{cases} \textcircled{1} \text{ Shuffle (idx_wtx, idx_aug)} \\ \textcircled{2} \text{ Shuffle (idx_aug, idx_aug)} \\ \textcircled{3} \text{ Shuffle (idx_train, idx_aug)} \end{cases}$

(b) MixUp 的系数.

$\textcircled{1} \lambda \sim \text{Beta}(\alpha, \alpha), \lambda' = \max(\lambda, 1-\lambda)$

$$x' = \lambda' x_1 + (1-\lambda') x_2$$

$$p' = \lambda' p_1 + (1-\lambda') p_2$$

$\textcircled{2}$ Mixup 同类的节点.

(c) MixUp 的对象.

MixUp(idx_aug, W)

new-features \leftarrow Mix-features

3. Loss function 的计算

(1). $Loss_X = \frac{1}{|X'|} H(\sim)$ $X' = MixUp(\hat{x}, W)$

$$\hat{x} = \begin{cases} \text{idx-aug.} \\ \text{idx-train.} \end{cases}$$

\therefore $\left\{ \begin{array}{l} \text{features} \leftarrow \text{idx-train.} \\ \text{new features} \left\{ \begin{array}{l} \text{idx-train} \\ \text{idx-train} + \text{idx-aug} \\ \text{idx-aug} \end{array} \right. \end{array} \right.$

X' 来源

idx-train
 $\text{idx-train} + \text{idx-aug} \rightarrow \text{MixUp}$
 $\text{idx-aug} \left\{ \begin{array}{l} \text{unlabel} \end{array} \right.$

$(\text{label} + \text{unlabel}, \text{PP Mix}(\text{label}, \text{guess}))$

(2). $Loss_U$

$$\lambda_u = \begin{cases} 0.82 \\ 0:0.0005:1 \end{cases}$$

$$| \left[\text{tf.clip_by_value}(\sim, 0, 1) \right]$$

U' 来源

$\left\{ \begin{array}{l} \text{new-features} \left\{ \begin{array}{l} \text{guess} + \text{idx-ulb.} \\ \text{idx } U' \end{array} \right. \end{array} \right.$

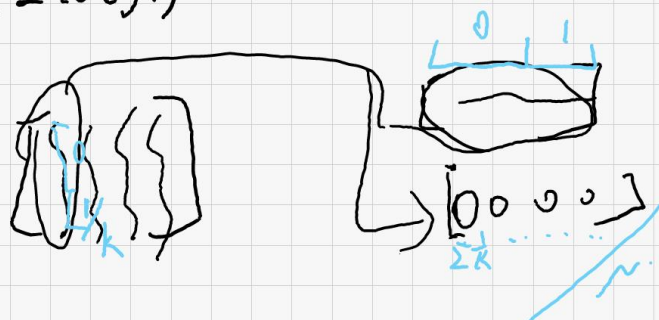
\rightarrow guess 为 aug-ulb 的标签
 idx-ulb 的标签

Function masked_softmax_cross_entropy

tensorflow

log_softmax * mask_resh
(2708,7) (2708)

元素乘 $tf.multiply(a,b)=a*b$



= average_loss

Augment:

Aug = Attention

$$q_b = \text{Aug}(\text{unlabeled_node})$$

训练 $P_{\text{Attention}}$

$$\text{Obj: } \min \| P_{\text{model}}(q_b) - P_{\text{model}}(\text{unlabeled_node}) \|_2^2$$

$$\text{guess_label} = P_{\text{model}}(\text{unlabeled_node})$$

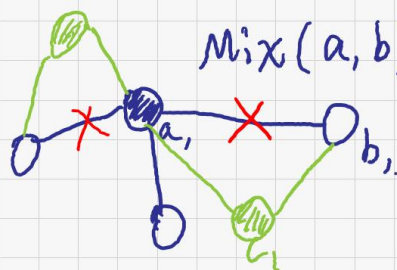
$$\text{Mix_feature}, \text{Mix_label} = \text{Mix}(\text{labeled_node}, \text{unlabeled_node})$$

$$\text{loss_mix} = \| \text{model}(\text{mix_feature}) - \text{Mix_label} \|$$

添加 Augment

{ label node
unlabel node \rightarrow guess label

new node 插入
中间.



Mix (label, feature)

new node			
a_i	b_i	c_i	
a_i	00	10	1
b_i	10	00	1
c_i	1	1	0 0 0 \dots

① Sum 插入节点总数.

$$n = \text{sum}(\text{adj}) / 2$$

$$\text{labels} = \text{np.concatenate}([\text{train_label}, \text{guess}], 0)$$

② 计算新 feature 与 adj, 新 label

$$\text{new_adj} = \text{zeros}(N+n, N+n)$$

$$\text{new_features} = \text{np.concatenate}([\text{features}, \text{zeros}(n, \text{ft_size})], 0)$$

$$\text{new_labels} = \text{np.concatenate}([\text{labels}, \text{zeros}(n, \text{nb_class})], 0)$$

$$\text{idx_new_node} = N$$

for i in range(N):

for j in range(i+1, N):

if adj[i, j] == 1:

$$\text{new_adj}[i, \text{idx_new_node}] = \text{new_adj}[j, \text{idx_new_node}] = [\text{idx_new_node}, i] = [\text{idx_new_node}, j] = 1$$

$$\text{new_feature}[\text{idx_new_node}] = \frac{1}{2} \text{features}[i, :] + \frac{1}{2} \text{features}[j, :]$$

$$\text{new_labels}[\text{idx_new_node}] = \frac{1}{2} \text{labels}[i, :] + \frac{1}{2} \text{labels}[j, :]$$

$$\text{idx_new_node} += 1$$

assert idx_new_node == N+n, 'error'