InstDis

$$P(i|\mathbf{v}) = \frac{\exp(\mathbf{v}_i^T \mathbf{v}/\tau)}{\sum_{j=1}^{n} \exp(\mathbf{v}_j^T \mathbf{v}/\tau)},$$

how to compute probability(vector l_2 norm):

motivation: learn similarity among same features

Memory Bank: low consistence, high negative samples number

Reduce computation: NCE

$$Z \simeq Z_i \simeq n E_j \left[\exp(\mathbf{v}_j^T \mathbf{f}_i/\tau)\right] = \frac{n}{m} \sum_{k=1}^{m} \exp(\mathbf{v}_{j_k}^T \mathbf{f}_i/\tau),$$

$$h(i, \mathbf{v}) := P(D=1|i, \mathbf{v}) = \frac{P(i|\mathbf{v})}{P(i|\mathbf{v}) + mP_n(i)}.$$

r approximated training objective is to minimize the
ve log-posterior distribution of data and noise sampl

$$J_{NCE}(\boldsymbol{\theta}) = -E_{P_d}\left[\log h(i, \mathbf{v})\right] - m \cdot E_{P_n}\left[\log(1 - h(i, \mathbf{v}'))\right].$$

loss

$$w_c = \sum_{i \in \mathcal{N}_k} \alpha_i \cdot 1(c_i = c).$$

classification: knn and voting each class

InvarSpread

motivation

1. memory bank inconsistency
2. same feature close, while different features spread

$$P(i|\hat{\mathbf{x}}_i) = \frac{\exp(\mathbf{f}_i^T \hat{\mathbf{f}}_i/\tau)}{\sum_{k=1}^{m} \exp(\mathbf{f}_k^T \hat{\mathbf{f}}_i/\tau)}.$$

prob: softmax embedding, m is batch size. compare in a batch, higher consistency

$$P_i = P(i|\hat{\mathbf{x}}_i) \prod_{j \neq i} (1 - P(i|\mathbf{x}_j)) \quad (5)$$

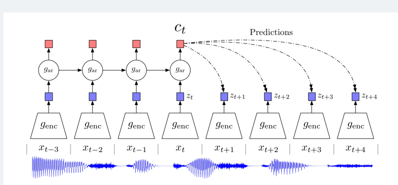
The negative log likelihood is given by

$$J_i = -\log P(i|\hat{\mathbf{x}}_i) - \sum_{j \neq i} \log(1 - P(i|\mathbf{x}_j)) \quad (6)$$

loss: likehood

why $P(i|\hat{x}_i)$ spread : paper gives an analysis

CPC



structure

$$f_k(x_{t+k}, c_t) \propto \frac{p(x_{t+k}|c_t)}{p(x_{t+k})}$$

learn a function f:

$$\mathcal{L}_{\mathrm{N}} = -\mathop{\mathbb{E}}_{X}\left[\log \frac{f_k(x_{t+k}, c_t)}{\sum_{x_j \in X} f_k(x_j, c_t)}\right]$$

loss:

lower bound of Mutual information

under this loss f achieve desired form

CMC

motivation: multiview to compare (pretext task)

loss

$$h_\theta(\{v_1, v_2\}) = \exp\left(\frac{f_{\theta_1}(v_1) \cdot f_{\theta_2}(v_2)}{\|f_{\theta_1}(v_1)\| \cdot \|f_{\theta_2}(v_2)\|} \cdot \frac{1}{\tau}\right)$$

$$\mathcal{L}_{contrast}^{V_1, V_2} = -\mathop{\mathbb{E}}_{\{v_1^1, v_2^1, \ldots, v_2^{k+1}\}}\left[\log \frac{h_\theta(\{v_1^1, v_2^1\})}{\sum_{j=1}^{k+1} h_\theta(\{v_1^1, v_2^j\})}\right]$$

$$\mathcal{L}(V_1, V_2) = \mathcal{L}_{contrast}^{V_1, V_2} + \mathcal{L}_{contrast}^{V_2, V_1}$$

对称

这个loss也能和互信息相关？