Report 9

Shiyu Yi

May 2019

1 Framework

We divide the selection of triplet into 3 parts, firstly we try to approximate each class's data distribution by Gaussian Mixture Model (GMM), then we consider to construct a k-dimensional vector set for each data point by embedding the each class confidence of this point. Secondly, we will propose a Mixture Set Region Sampling (MSRS) for the positive sample and negative sample in the pair of triplet loss. The notion of Mixture Set is about when doing sampling we will pay most of the attention in the region that confidence of other label is significant compared with the intrinsic label. At last, after select the positive sample and negative sample data, we will construct a combined objective function for anchor sampling, which consist the prior information from distance location and approximate probability function. Hence, we apply a trust-region-like numerical optimization algorithm to get the optimal of this problem and select the anchor of this pair.

1.1 Gaussian Mixture Model for Approximating Each Class Distribution

We modeling the approximation distribution of hidden feature sample by Gaussian Mixture Model:

$$p(\mathcal{H}(x)|\theta, c) = \sum_{k=1}^{K} w_i \mathcal{N}\left(\mathcal{H}(x)|\mu_{ck}, \Sigma_{ck}\right) \tag{1}$$

where $\mathcal{H}(x)$ denotes the hidden state of an input $x \in$ the c -th class, θ denotes the DNN classifier of interest (or its parameters), and μ_{ck} and Σ_{ck} are the mean and covariance matrix of the k -th Gaussian component in the mixture model of the c -th class. After training the DNN to transform raw data to feature vectors, we collect the corresponding hidden states for training a GMM for each class using the EM algorithm.

We will apply

1.2 Optimization for Positive Sampling

Let S_A denote sampled anchor, now we need to add the prior information of S_A to the approximating sampling distribution. Considering distance function $D(x) = \|x - S_A\|_2$

$$J_{Positive}(x) = \omega p(\mathcal{H}(x)|\theta, c) + (1 - \omega)D(x) \tag{2}$$

In general, we can write the objective function above as:

$$J_{Positive}(x) = \psi(x, S_A) \tag{3}$$

As for the constructed positive objective function above, we try to use trust-region-like optimization algorithm to solve it.

$$\min_{\eta \in R^n} m(\eta) = J_{Positive}(x) + \partial J_{Positive}(x)\eta + \frac{1}{2}\eta^T H\eta, \quad \|\eta\| \leq \Delta \tag{4}$$

where H is some symmetric matrix and Δ is the trust-region radius. Clearly, a possible choice for H is the Hessian matrix $H_{i,j} = \partial_i \partial_j f(x)$; classical convergence results guarantee super-linear convergence if the chosen H is "sufficiently close" to the Hessian matrix. The algorithm used to compute an approximate minimizer η of the model within the trust region is termed the inner iteration. Once an η has been returned by the inner iteration, the quality of the model m is assessed by forming the quotient

$$\rho = \frac{f(x) - f(x + \eta)}{m(0) - m(\eta)} \tag{5}$$

Depending on the value of ρ, the new iterate $x + \eta$ is accepted or discarded and the trust-region radius Δ is updated.

Note that the iteration step here will not be executed immediately because the before we will update the information of sampled positive data into our objective function and then the modified objective function can be seen as for negative sampling.

1.3 Optimization for Negative Sampling

Let S_P denotes the sampled positive data