Appendix of Affinity Uncertainty-based Hard Negative Mining in Graph Contrastive Learning

Chaoxi Niu, Guansong Pang, *Member, IEEE*, Ling Chen, *Senior Member, IEEE*

I. Details of Other Uncertainty Estimation Methods

A. Softmax-Response

Given the softmax predicted vector $p = \{p_1, \dots, p_C\}$ (C is the number of classes), the uncertainty estimated based on Softmax-Response [1] is calculated as follows:

$$u = 1 - \max_{c=1,\dots,C} p_c \tag{1}$$

The uncertainty is large when the predicted probabilities are more equally distributed, and it is small when predicted probabilities are concentrated on one specific class.

B. Predictive Entropy

Predictive entropy [2] captures the average amount of information presented in the predicted probability distribution. The uncertainty based on predictive entropy is obtained by:

$$u = -\sum_{c=1}^{C} p_c \log p_c \tag{2}$$

Similar to Softmax-Response, the predictive entropy reaches its maximum value when all predicted classes are equiprobable and its minimum is attained when there exists one class with probability one and the probabilities of all other classes are zero.

C. Distance-Based Uncertainty

For each anchor instance, after obtaining the binary partition labels of the negative instances $\{C_{1|2}^{i,j}\}_{j=1}^{N}$, we calculate the uncertainty of the negative instance whose cluster labels are zero. Specially, the distance d_{ij} between the anchor instance and the negative instance $\widehat{z}_j$ ($C_2^{ij} = 0$) is first calculated and then the uncertainty of $\widehat{z}_j$ is set as the reciprocal of d_{ij} weighted by a parameter β i.e.,

$$u_{ij} = \beta \frac{1}{d_{ij}}. \tag{3}$$

Note that for $\widehat{z}_j$ with $C_2^{ij} = 0$, if $\widehat{z}_j$ is close to the anchor $\widetilde{z}_i$, the uncertainty value u_{ij} is large.

II. Proof of Theorem

Theorem 1. *Let $u_{ij} = \phi_i(\widehat{z}_j; \Theta)$ be the affinity uncertainty-based hardness of a negative instance $\widehat{z}_j$ w.r.t. the anchor instance $\widetilde{z}_i$. When the projection function is an identity function and assumes the positive instance is more similar to the anchor than the negative instances, then minimizing the proposed objective in (3) is equivalent to minimizing a modified triplet loss with an adaptive margin $m_{ij} = \frac{\tau}{2}\log(\alpha u_{ij})$, i.e.,*

$$\ell_{AUGCL}(\widetilde{z}_i, \widehat{z}_i) \propto \quad \frac{1}{2\tau}\sum_{j, j\neq i}^{N}\left(\|\widetilde{z}_i^{\prime} - \widehat{z}_i^{\prime}\| - \|\widetilde{z}_i^{\prime} - \widehat{z}_j^{\prime}\| + m_{ij}\right), \tag{4}$$

where $\widetilde{z}_i^{\prime}$ is the normalized embedding.

Proof.

$$
\begin{aligned}
&\ell_{\text{AUGCL}}(\widetilde{z}_i, \widehat{z}_i) \\
&= -\log \frac{\exp(h(\widetilde{z}_i, \widehat{z}_i)/\tau)}{\exp(h(\widetilde{z}_i, \widehat{z}_i)/\tau) + \sum\limits_{j,j\neq i}^{N} \alpha u_{ij} \exp(h(\widetilde{z}_i, \widehat{z}_j)/\tau)} \\
&= \log \left(1 + \sum_{j,j\neq i}^{N} \alpha u_{ij} \exp\left(\frac{h(\widetilde{z}_i, \widehat{z}_j) - h(\widetilde{z}_i, \widehat{z}_i)}{\tau} \right) \right) \\
&= \log \left(1 + \sum_{j,j\neq i}^{N} \exp\left(\frac{h(\widetilde{z}_i, \widehat{z}_j) - h(\widetilde{z}_i, \widehat{z}_i) + 2m_{ij}}{\tau} \right) \right)
\end{aligned} \tag{5}
$$

where $h(\widetilde{z}_i, \widehat{z}_j) = \frac{\widetilde{z}_i^T \widehat{z}_j}{\|\widetilde{z}_i\| \|\widehat{z}_j\|}$ ($-1 \leqslant h(\widetilde{z}_i, \widehat{z}_j) \leqslant 1$) and $m_{ij} = \frac{\tau}{2}\log(\alpha u_{ij})$. Let $\widetilde{z}_i^{'} = \frac{\widetilde{z}_i}{\|\widetilde{z}_i\|}$, $h(\widetilde{z}_i, \widehat{z}_j)$ can be rewrited as $h(\widetilde{z}_i, \widehat{z}_j) = (\widetilde{z}_i^{'})^T \widehat{z}_j^{'}$. Moreover, we can have $h(\widetilde{z}_i, \widehat{z}_j) = 1 - \frac{1}{2}\|\widetilde{z}_i^{'} - \widehat{z}_j^{'}\|$ as $\|\widetilde{z}_i^{'} - \widehat{z}_j^{'}\| = (\widetilde{z}_i^{'})^T \widetilde{z}_i^{'} + (\widehat{z}_j^{'})^T \widehat{z}_j^{'} - 2(\widetilde{z}_i^{'})^T \widehat{z}_j^{'}$ and $\widetilde{z}_i^{'}$ and $\widehat{z}_j^{'}$ are both normalized. Since the positive instance is more similar than the negative instances to the anchor, the value of $h(\widetilde{z}_i, \widehat{z}_j) - h(\widetilde{z}_i, \widehat{z}_i)$ tends to be -2.

Based on the above analysis, we can apply the Taylor expansion of first order and get the following approximation:

$$
\begin{aligned}
&\ell_{\text{AUGCL}}(\widetilde{z}_i, \widehat{z}_i) \\
&= \log \left(1 + \sum_{j,j\neq i}^{N} \exp\left(\frac{h(\widetilde{z}_i, \widehat{z}_j) - h(\widetilde{z}_i, \widehat{z}_i) + 2m_{ij}}{\tau} \right) \right) \\
&\approx \sum_{j,j\neq i}^{N} \exp\left(\frac{h(\widetilde{z}_i, \widehat{z}_j) - h(\widetilde{z}_i, \widehat{z}_i) + 2m_{ij}}{\tau} \right) \\
&\approx e^{-\frac{2}{\tau}} \left((N-1) + \frac{1}{\tau} \sum_{j,j\neq i}^{N} (h(\widetilde{z}_i, \widehat{z}_j) - h(\widetilde{z}_i, \widehat{z}_i) + 2m_{ij} + 2) \right) \\
&\propto \frac{1}{\tau} \sum_{j,j\neq i}^{N} (h(\widetilde{z}_i, \widehat{z}_j) - h(\widetilde{z}_i, \widehat{z}_i) + 2m_{ij}) \\
&= \frac{1}{2\tau} \sum_{j,j\neq i}^{N} \left(\|\widetilde{z}_i^{'} - \widehat{z}_i^{'}\| - \|\widetilde{z}_i^{'} - \widehat{z}_j^{'}\| + m_{ij} \right)
\end{aligned} \tag{6}
$$

which concludes the proof. Note that the equation in the third row is obtained through the Taylor expansion of logarithm at 0 and the equation in the fourth row is obtained through the Taylor expansion of exponential function at -2. □

References

[1] Y. Geifman and R. El-Yaniv, "Selective classification for deep neural networks," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[2] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *Advances in Neural Information Processing Systems*, vol. 30, 2017.