УДК 004.852

Анализ смещения распределений в контрастивном обучении

Л.С. Троешестова, **Р.В. Исаченко**
Московский физико-технический институт
(национальный исследовательский университет)

В работе исследовано влияние смещения распределений в задаче построения представлений без учителя. Зафиксируем объект $\mathbf{x} \in \mathcal{X}$, и пусть $f : \mathcal{X} \to \mathbb{R}^d$ – искомая функция представления. Будем исследовать сдвиги в следующих распределениях: $p_x^+(\mathbf{x}')$ – вероятность наблюдения $\mathbf{x}'$ как позитивного объекта (объекта того же класса) для $\mathbf{x}$ и $p_x^-(\mathbf{x}')$ – вероятность наблюдения $\mathbf{x}'$ как негативного объекта (объекта другого класса) для $\mathbf{x}$. Также обозначим τ^+ за вероятность принадлежности $\mathbf{x}$ данному классу и $\tau^- = 1 - \tau^+$ – любому другому классу. Итак, "идеальная", или несмещенная, функция потерь *Multi-class N-pair Loss* [1] записывается так:

$$L^N_{\text{Unbiased}}(f) = \mathbb{E}_{\substack{\mathbf{x}\sim p, \mathbf{x}^+ \sim p_x^+, \\ \mathbf{x}_i^- \sim p_x^-}} \left[-\log \frac{e^{f(\mathbf{x})^T f(\mathbf{x}^+)}}{e^{f(\mathbf{x})^T f(\mathbf{x}^+)} + \sum_{i=1}^N e^{f(\mathbf{x})^T f(\mathbf{x}_i^-)}} \right] \tag{1}$$

На практике истинное $p_x^-(\mathbf{x}_i^-)$ недоступно, и стандартное решение – семплирование негативных объектов из всего распределения $p(\mathbf{x})$ (*Contrastive Loss*). При таком подходе с вероятностью τ^+ при семплировании в контрастивном обучении возникают ошибки II рода. Метод устранения такого смещения был предложен в работе [2](*DebiasedNeg Loss*).

Однако при построении качественных представлений не менее важно учитывать ошибки I рода, вызванные шумом в данных и в случае компьютерного зрения – аугментациями. Мы предлагаем функцию потерь $L^N_{\text{DebiasedPos}}(f)$, которая в предположении верности негативных объектов оценивает недоступные компоненты эмпирически, и устраняет смещение в распределении $p_x^+(\mathbf{x}^+)$:

$$L^N_{\text{DebiasedPos}}(f) = \mathbb{E}_{\substack{\mathbf{x}\sim p \\ \{\mathbf{u}_i\}_{i=1}^N \sim p_x^{-N} \\ \mathbf{v}\sim p_x^+}} \left[-\log \frac{P_{\text{emp}}(\mathbf{x}, \{\mathbf{u}_i\}_{i=1}^N, \mathbf{v}) - \tau^- P^-_{\text{emp}}(\mathbf{x}, \{\mathbf{u}_i\}_{i=1}^N)}{P_{\text{emp}}(\mathbf{x}, \{\mathbf{u}_i\}_{i=1}^N, \mathbf{v}) + \left(N\tau^+ - \tau^-\right) P^-_{\text{emp}}(\mathbf{x}, \{\mathbf{u}_i\}_{i=1}^N)} \right], \tag{2}$$

где

$$P_{\text{emp}}(\mathbf{x}, \{\mathbf{u}_i\}_{i=1}^N, \mathbf{v}) = \frac{1}{N+2} \left(\sum_{i=1}^N e^{f(\mathbf{x})^T f(\mathbf{u}_i)} + e^{f(\mathbf{x})^T f(\mathbf{v})} + e^{f(\mathbf{x})^T f(\mathbf{x})} \right); \tag{3}$$

$$P^-_{\text{emp}}(\mathbf{x}, \{\mathbf{u}_i\}_{i=1}^N) = \frac{1}{N} \sum_{i=1}^N e^{f(\mathbf{x})^T f(\mathbf{u}_i)}. \tag{4}$$

Таким образом, $L_{\text{DebiasedPos}}$ эмпирически оценивает $\tilde{L}_{\text{DebiasedPos}}$ – предел несмещенной функции потерь. При конечном N и количестве позитивных объектов $M = 1$, теорема 1 ограничивает ошибку этой оценки сверху с асимптотикой $O(N^{-1/2})$.

Теорема 1. *Для произвольного представления f и произвольного $\delta > 0$ существует достаточно большое N, что*

$$\left| \tilde{L}^N_{DebiasedPos}(f) - L^N_{DebiasedPos}(f) \right| \leq \left[\left(1 + \frac{\tau^-}{\tau^+} + \delta \right) \sqrt{\frac{\pi}{2N}} + \left(1 + \frac{1}{\tau^+} \right) \sqrt{\frac{\pi}{2N+2}} \right] e^{3/2}$$

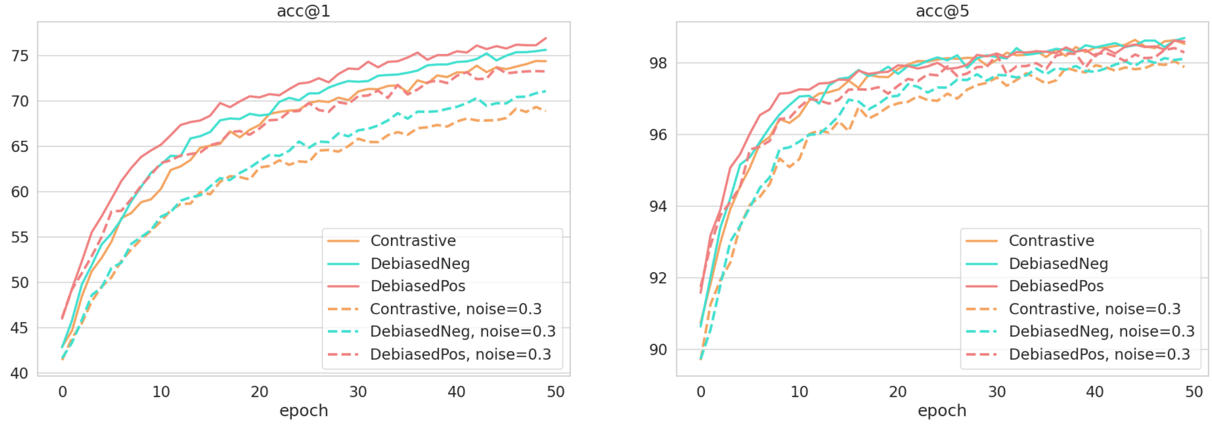

Рис. 1: Точность классификации на датасете CIFAR10 с искусственным зашумлением

В ходе эксперимента были сравнены 3 функции потерь: *Contrastive Loss*, *DebiasedNeg Loss* и *DebiasedPos Loss*. Мы использовали фреймворк SimCLR [3] с ResNet-18 в качестве энкодера, и на обученных представлениях измеряли точность классификации на датасете CIFAR10. Эксперимент показал, что предложенная функция потерь устраняет смещение, вызванное ошибками I рода, и, как следствие, является устойчивой к шумным данным (см. Рис 1). Наконец, было проведено сравнение двух способов агрегации по позитивным объектам. Внешняя агрегация привела к более точным представлениям, а внутренняя – к лучшей производительности.

Литература

Список литературы

[1] K. Sohn, "Improved deep metric learning with multi-class n-pair loss objective," in *Advances in Neural Information Processing Systems* (D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, eds.), vol. 29, Curran Associates, Inc., 2016.

[2] C.-Y. Chuang, J. Robinson, L. Yen-Chen, A. Torralba, and S. Jegelka, "Debiased contrastive learning," 2020.

[3] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," 2020.