

# Анализ смещения распределений при использовании сравнительного подхода в обучении представления данных

Мария Александровна Никитина

Московский физико-технический институт

*Кафедра:* Интеллектуальный анализ данных

*Научный руководитель:* кандидат ф.-м. наук Р. В. Исаченко

2024

# Цель исследования

## Цель

Оценка распределения данных и устранение смещения распределения

## Задача

Нахождение оптимальной лосс-функции, устраняющей смещение, и её применение в алгоритме обучения модели.

## Идея

Выразить распределение положительных элементов через декомпозицию полного распределения:

$$p_x^+(\mathbf{x}') = \frac{p(\mathbf{x}') - \tau^- p_x^-(\mathbf{x}')}{\tau^+},$$

где  $\tau^+$ ,  $\tau^-$  – вероятности нахождения положительных и отрицательных элементов соответственно.

# Contrastive learning

*Contrastive learning* – подход при котором обучение происходит не только по принципу близости, но и по принципу различия.

Положим вектор  $x$  некоторого объекта в качестве основного. Тогда вектор схожего объекта назовём  $x^+$  – позитивный элемент. Он должен быть как можно ближе к  $x$ . А вектор отличного объекта  $x^-$  – как можно дальше, как негативный элемент.



Рис.: Принцип contrastive learning. Взято из [Schroff et al. 2015].

# Список литературы



Ching-Yao Chuang, Joshua Robinson, Lin Yen-Chen, Antonio Torralba, Stefanie Jegelka (2020)

Debiased Contrastive Learning



Yang, Jinyu and Duan, Jiali and Tran, Son and Xu, Yi and Chanda, Sampath and Chen, Liqun and Zeng, Belinda and Chilimbi, Trishul and Huang, Junzhou (2022)

Vision-Language Pre-Training with Triple Contrastive Learning



CTing Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton (2020)

A Simple Framework for Contrastive Learning of Visual Representations



Stanislaw Antol and Aishwarya Agrawal and Jiasen Lu and Margaret Mitchell and Dhruv Batra and C. Lawrence Zitnick and Devi Parikh (2017)

VQA: Visual Question Answering



Florian Schroff, Dmitry Kalenichenko, James Philbin (2015)

FaceNet: A Unified Embedding for Face Recognition and Clustering

# Ложноположительные и ложноотрицательные элементы

Классический лосс, не учитывающий смещения:

$$\mathcal{L}_{N-pair}(f) = -\log \frac{\exp(f(\mathbf{x})^T f(\mathbf{x}_i^+))}{\exp(f(\mathbf{x})^T f(\mathbf{x}_i^+)) + \sum_{i=1}^N \exp(f(\mathbf{x})^T f(\mathbf{x}_i^-))}$$

Решение для устранения смещения ложноотрицательных элементов [Chuang et al., 2020]:

$$L_{\text{Neg}}^{N,M}(f) = \mathbb{E}_{\substack{\mathbf{x} \sim p; \mathbf{x}_+ \sim p_x^+, \\ \{\mathbf{u}_i\}_{i=1}^N \sim p^N, \\ \{\mathbf{v}_j\}_{j=1}^M \sim p_x^+}} \left[ -\log \frac{e^{f(\mathbf{x})^T f(\mathbf{x}^+)}}{e^{f(\mathbf{x})^T f(\mathbf{x}^+)} + N g(\mathbf{x}, \{\mathbf{u}_i\}_{i=1}^N, \{\mathbf{v}_j\}_{j=1}^M)} \right]$$

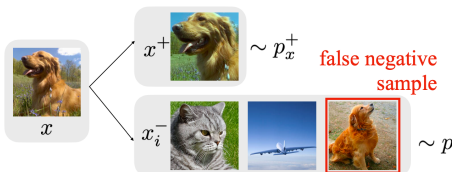


Рис.: Взято из [Chuang et al., 2020].

# DebiasedPos

В отличие от  $L_{\text{Neg}}^{N,M}(f)$  здесь устраняется смещение ложноположительных элементов.

$$\tilde{L}_{\text{Pos}}^N(f) = \mathbb{E}_{\substack{\mathbf{x} \sim p \\ \mathbf{x}^- \sim p_x^-}} \left[ -\log \frac{\mathbb{E}_{\mathbf{x}' \sim p} f(\mathbf{x}, \mathbf{x}') - \tau^- \mathbb{E}_{\mathbf{x}^- \sim p_x^-} f(\mathbf{x}, \mathbf{x}_i^-)}{\mathbb{E}_{\mathbf{x}' \sim p} f(\mathbf{x}, \mathbf{x}') + (N\tau^+ - \tau^-) \mathbb{E}_{\mathbf{x}^- \sim p_x^-} f(\mathbf{x}, \mathbf{x}_i^-)} \right]$$

Финальная оценка:

$$L_{\text{Pos}}^N(f) = \mathbb{E}_{\substack{\mathbf{x} \sim p \\ \{\mathbf{u}_i\}_{i=1}^N \sim p_x^- \\ \mathbf{v} \sim p_x^+}} \left[ -\log \frac{P_{\text{emp}} - \tau^- P_{\text{emp}}^-}{P_{\text{emp}} + (N\tau^+ - \tau^-) P_{\text{emp}}^-} \right].$$

$P_{\text{emp}}, P_{\text{emp}}^-$  — эмпирические оценки матожиданий.

## Lemma

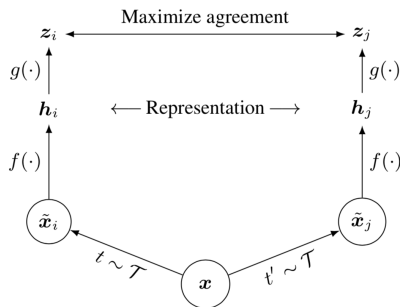
При  $N \rightarrow \infty$ :

$$\mathcal{L}_{N\text{-pair}}^N(f) \rightarrow \tilde{L}_{Pos}^N(f)$$

## Theorem

$\mathcal{L}_{Pos}^N$  максимизирует нижнюю границу взаимной информации.

$$\begin{aligned} I(x, c) &= \sum_{x \in X, c \in C} p(x, c) \log \frac{p(x|c)}{p(x)} \geq \\ &\geq \mathbb{E}_{x \sim p} \log \left( (N+2) \frac{p(x|c)}{p(x)} \right) - \mathcal{L}_{Pos}^N \end{aligned}$$



- ▶  $\mathcal{T}$  — семейство аугментаций
- ▶ Семплируются 2 аугментации  $t, t' \sim \mathcal{T}$ , применяются к каждому объекту.
- ▶ Обучаем сеть-энкодер  $f(\cdot)$  и MLP сеть-проекцию  $g(\cdot)$ , максимизируя соответствие представлений.



# Классификация

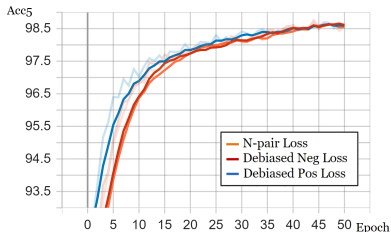
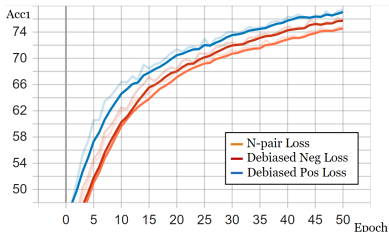
## Цель

Сравнить работу  $L_{\text{Neg}}^N(f)$ ,  $L_{\text{Pos}}^N(f)$  и  $L_{N\text{-pair}}(f)$  на модели SimCLR.

В качестве метрик берётся Accurasy и Top-k-accrasy:

$$\text{Acc}_1 = \frac{TP + TN}{TP + FP + TN + FN}, \quad \text{Acc}_k = \frac{1}{n} \sum_{i=1}^n [y_i \in \hat{y}_i^k]$$

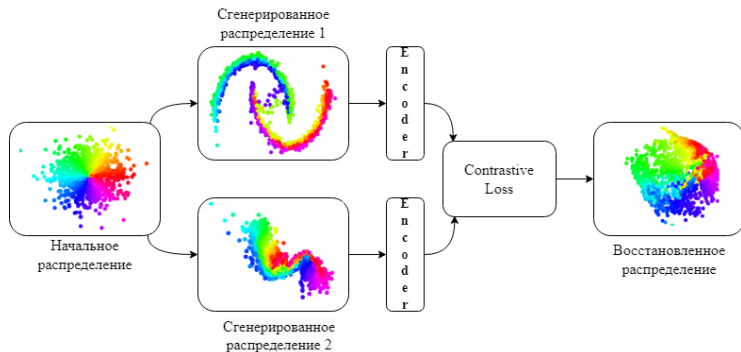
## Результат



# Искусственный эксперимент

## Цель

Проверить способность  $L_{\text{Pos}}(f)$  восстанавливать изначальное распределение.



## Цель

Сравнить работу модели TCL [Yang et. al.] с  $L_{Pos}(f)$  и  $L_{N-pair}(f)$  на задаче Visual Question Answering [Stanislaw Antol et. al.] на датасете MS COCO.

## Метрика

В качестве метрики берётся число попаданий ответа модели в список из 10 ответов, предоставленных авторами датасета:

$$Acc = \frac{1}{n} \sum_{i=1}^n [y_i \in y_i^{10}]$$

# Модели и результаты VQA

## Модель

В качестве бейзлайн модели берётся TCL [Yang et. al.] с  $\mathcal{L}_{N-pair}$  для сравнений эмбеддингов вида изображение-изображение, изображение-текст, текст-текст. Модель состоит из визуального энкодера ViT-B/16 и текстового энкодера BERT-base.

## Результаты VQA для $\mathcal{L}_{N-pair}$ и $\mathcal{L}_{Pos}$

	$\mathcal{L}_{N-pair}$	$\mathcal{L}_{Pos}$
Accuracy	0.67	0.69