

# 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

# Анализ смещения распределений

Исследуется задача восстановления распределения данных при наличии смещения в выборке.

## Проблема

Исходное распределение и способ порождения из него данных неизвестны, функция потерь имеет несколько локальных минимумов, которые не соответствуют истинному восстановлению начального распределения.

## Требуется

Требуется найти оптимальную функцию потерь, устраняющую смещение, и оценить её способность восстанавливать исходное распределение в пространстве представления.

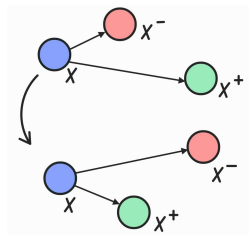
## Решение

Выразить распределение положительных элементов через декомпозицию полного распределения, параметризовав вероятность появления положительных элементов.

# Обучение сравнениями

Обучение сравнениями – подход при котором обучение происходит посредством попарного сравнения элементов друг с другом.

Пусть  $\mathbf{x}$  – вектор основного объекта. Тогда вектор схожего объекта  $\mathbf{x}^+$  – положительный элемент. Вектор отличного объекта  $\mathbf{x}^-$  – негативный элемент.  $f$  – модель. Задача обучения сравнениями:



$$\text{dist}(f(\mathbf{x}), f(\mathbf{x}^+)) \rightarrow \min_f, \quad \text{dist}(f(\mathbf{x}), f(\mathbf{x}^-)) \rightarrow \max_f.$$

Так как решается задача обучения без учителя, появляется смещение ввиду неверного сэмплирования. Например, при чрезмерной аугментации положительной пары или при представлении всего батча в качестве негативных элементов.

$$\text{dist}(f(\mathbf{x}), f(\mathbf{x}_{\text{wrong}}^+)) \nrightarrow \min_f, \quad \text{dist}(f(\mathbf{x}), f(\mathbf{x}_{\text{wrong}}^-)) \nrightarrow \max_f.$$

# Смещение ложноотрицательных элементов

Классическая функция потерь, не учитывающая смещения:

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

Устранение смещения ложноотрицательных элементов<sup>1</sup>:

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

$$\mathcal{L}_{\text{Neg}}^N(f) = \mathbb{E}_{\substack{\mathbf{x} \sim p; \mathbf{x}_+ \sim p_{\mathbf{x}}^+, \\ \{\mathbf{u}_i\}_{i=1}^N \sim p^N, \\ \mathbf{v} \sim p_{\mathbf{x}}^+}} \left[ -\log \frac{e^{f(\mathbf{x})^T f(\mathbf{x}^+)}}{e^{f(\mathbf{x})^T f(\mathbf{x}^+)} + Ng(\mathbf{x}, \{\mathbf{u}_i\}_{i=1}^N, \mathbf{v})} \right],$$

$$g(\mathbf{x}, \{\mathbf{u}_i\}_{i=1}^N, \mathbf{v}) = \frac{1}{\tau^-} \left( \frac{1}{N} \sum_{i=1}^N e^{f(\mathbf{x})^T f(\mathbf{u}_i)} - \tau^+ e^{f(\mathbf{x})^T f(\mathbf{v})} \right).$$

---

<sup>1</sup>Chuang et al., 2020.

# Смещение ложноположительных элементов

Выражение для вероятности получить положительных элемент:

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

## Лемма (Никитина, 2024)

При  $N \rightarrow \infty$  несмещённая функция потерь стремится к функции потерь, учитывающей наличие ложноположительных элементов:

$$\mathcal{L}_{\text{N-pair}}^N(f) = \mathbb{E}_{\substack{\mathbf{x} \sim p \\ \mathbf{x}^+ \sim p_{\mathbf{x}}^+ \\ \{\mathbf{x}_i^-\}_{i=1}^N \sim p_{\mathbf{x}}^{-N}}} \left[ -\log \frac{e^{f(\mathbf{x})^T f(\mathbf{x}^+)}}{e^{f(\mathbf{x})^T f(\mathbf{x}^+)} + \sum_{i=1}^N e^{f(\mathbf{x})^T f(\mathbf{x}_i^-)}} \right] \xrightarrow{N \rightarrow \infty} \\ \xrightarrow{N \rightarrow \infty} \mathbb{E}_{\substack{\mathbf{x} \sim p \\ \mathbf{x}^- \sim p_{\mathbf{x}}^-}} \left[ -\log \frac{R}{R + N \mathbb{E}_{\mathbf{x}^- \sim p_{\mathbf{x}}^-} e^{f(\mathbf{x})^T f(\mathbf{x}^-)}} \right],$$

где

$$R = \frac{1}{\tau^+} (\mathbb{E}_{\mathbf{x}' \sim p} e^{f(\mathbf{x})^T f(\mathbf{x}')} - \tau^- \mathbb{E}_{\mathbf{x}^- \sim p_{\mathbf{x}}^-} e^{f(\mathbf{x})^T f(\mathbf{x}^-)}).$$

# Смещенная функция потерь

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

Финальная оценка:

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

где  $P_{\text{emp}}$ ,  $P_{\text{emp}}^-$  – эмпирические оценки матожиданий:

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

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

# Корректность функции потерь для задачи

## Теорема (Никитина, 2024)

Задача минимизации  $\mathcal{L}_{\text{Pos}}^N$  эквивалентна задаче максимизации совместной информации между положительной парой  $I(\mathbf{x}, \mathbf{c})$ , то есть:

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

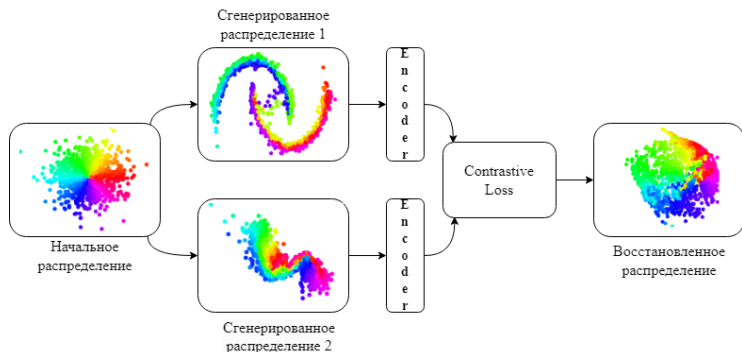
# Эксперимент на искусственных данных

## Цель

Проверить способность функции потерь  $\mathcal{L}_{\text{Pos}}(f)$  восстанавливать начальное распределение.

## Метод

$$\mathcal{L} = \mathcal{L}_{\text{Pos}}^N(f) + \mathcal{D}_{\text{KL}}(f || \mathcal{N}(\mathbf{0}, \mathbf{I}))$$





# Классификация изображений: метод

## Цель

Сравнить работу модели с  $\mathcal{L}_{\text{N-pair}}(f)$ ,  $\mathcal{L}_{\text{Neg}}(f)$  и  $\mathcal{L}_{\text{Pos}}(f)$  на задаче классификации на датасете MNIST10.

## Метод

- ▶  $\mathcal{T}$  – семейство аугментаций.
- ▶ Семплируются 2 аугментации  $t, t' \sim \mathcal{T}$ , применяются к каждому объекту.
- ▶ Обучается сеть-энкодер  $f(\cdot)$  и сеть-проекция  $g(\cdot)$  для максимизации соответствия представлений.

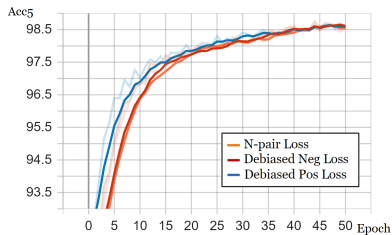
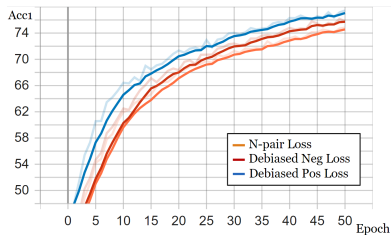
## Метрики

В качестве метрик берётся Accuracy и Top-k-accuracy:

$$Acc_1 = \frac{TP + TN}{TP + FP + TN + FN}, \quad Acc_k = \frac{1}{n} \sum_{i=1}^n [y_i \in \hat{y}_i^k]$$

# Классификация изображений: результаты

	$\mathcal{L}_{\text{N-pair}}$	$\mathcal{L}_{\text{Neg}}$	$\mathcal{L}_{\text{Pos}}$
Acc1	74.84	75.81	<b>77.45</b>
Acc5	98.56	98.56	<b>98.58</b>



# Задача ответов на вопросы по изображению

## Цель

Сравнить работу модели с  $\mathcal{L}_{\text{Pos}}(f)$  и  $\mathcal{L}_{\text{N-pair}}(f)$  на задаче VQA на датасете MS COCO.

## Модель

В качестве базовой модели берётся TCL<sup>2</sup> с  $\mathcal{L}_{\text{N-pair}}$  для сравнений эмбеддингов вида изображение-изображение, изображение-текст, текст-текст. Модель состоит из визуального энкодера ViT-B/16 и текстового энкодера BERT-base.

## Метрика

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

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

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

	$\mathcal{L}_{\text{N-pair}}$	$\mathcal{L}_{\text{Pos}}$
Accuracy	66.29	<b>67.23</b>

---

<sup>2</sup>Yang et. al., 2022.

## Пример работы моделей на задаче VQA



Рис.: Вопрос: «What is the child eating?». Ответ обеих моделей одинаковый и входит в список верных: «donut».

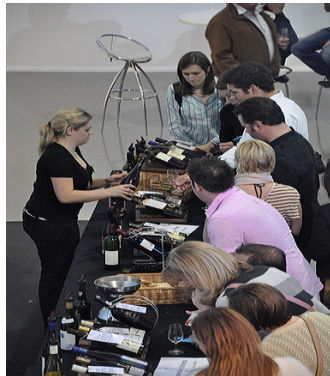


Рис.: Вопрос: «What kind of event are the people involved in?». Ответ модели с  $\mathcal{L}_{N\text{-pair}}$ : «party» неверный. Ответ модели с  $\mathcal{L}_{Pos}$ : «wine testing» входит в список верных.

## Выносятся на защиту

1. Исследованы три вида функции потерь в задаче восстановления изначального распределения методом обучения сравнениями: классический  $N$ -pair loss и его модификации, устраняющие смещение вследствие наличия ложноположительных и ложноотрицательных элементов выборки.
2. Предложена функция потерь, устраняющая смещение при ложноположительных элементах. Доказана её сходимость к классическому  $N$ -pair loss и свойство максимизации нижней границы взаимной информации.
3. Проведено сравнение трёх функций потерь в задаче классификации и VQA, а также изучена способность предложенной функции потерь восстанавливать изначальное распределение на примере двумерного пространства.