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

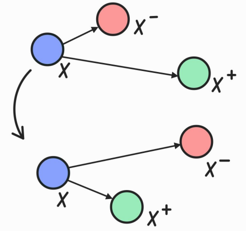
Требуется найти оптимальную функцию потерь, устраняющую смещение, и оценить её способность восстанавливать исходное распределение в пространстве представления.

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.



$\mathbf{x}$ – векторное представление объекта.
$\mathbf{x}^{+}$ – вектор схожего объекта.
$\mathbf{x}^{-}$ – вектор отличного объекта.
Задача обучения сравнениями:

$$\mathrm{dist}(f(\mathbf{x}), f(\mathbf{x}^{+})) \to \min_{f}, \quad \mathrm{dist}(f(\mathbf{x}), f(\mathbf{x}^{-})) \to \max_{f}.$$

Методы обучения сравнениями приводят к наличия смещения в распредлении данных:

$$\mathrm{dist}(f(\mathbf{x}), f(\mathbf{x}^{+}_{\text{wrong}})) \not\to \min_{f}, \quad \mathrm{dist}(f(\mathbf{x}), f(\mathbf{x}^{-}_{\text{wrong}})) \not\to \max_{f}.$$

Смещение ложноотрицательных элементов

Классическая функция потерь, не учитывающая смещения:

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

Устранение смещения ложноотрицательных элементов:

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

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

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

Chuang C. et.al. Debiased contrastive learning, 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 $N \to \infty$ несмещённая функция потерь стремится к функции потерь, учитывающей наличие ложноположительных элементов:

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

$$\xrightarrow{N \to \infty} \mathbb{E}_{\substack{\mathbf{x} \sim p \\ \mathbf{x}^{-} \sim p_{\mathbf{x}}^{-}}} \left[-\log \frac{R}{R + N \mathbb{E}_{\mathbf{x}^{-} \sim p_{\mathbf{x}}^{-}} e^{f(\mathbf{x})^T f(\mathbf{x}^{-})}} \right],$$

где

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

Смещенная функция потерь

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

Финальная оценка:

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

где P_{emp}, P^-_{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)

При минимизации $\mathcal{L}^N_{\text{Pos}}$ увеличивается нижняя граница совместной информации между положительной парой $I(\mathbf{x}, \mathbf{c})$, то есть:

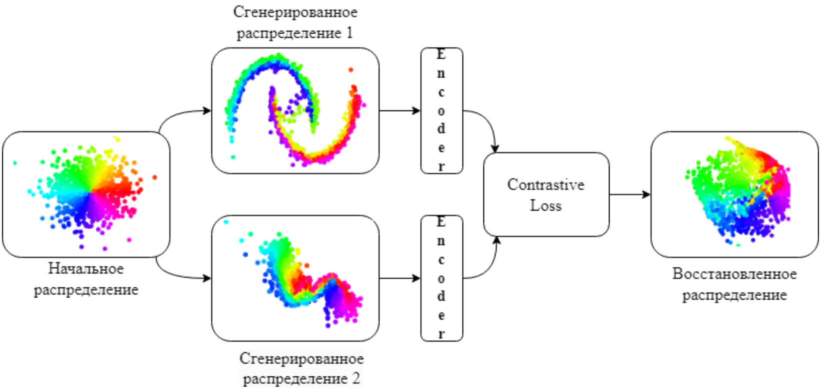
$$I(\mathbf{x}, \mathbf{c}) = \sum_{\mathbf{x} \in X, \mathbf{c} \in C} p(\mathbf{x}, \mathbf{c}) \log \frac{p(\mathbf{x}|\mathbf{c})}{p(\mathbf{x})} \geq$$

$$\geq \mathbb{E}_{\mathbf{x} \sim p} \log \left((N+2) \frac{p(\mathbf{x}|\mathbf{c})}{p(\mathbf{x})} \right) - \mathcal{L}^N_{\text{Pos}}$$

Эксперимент на искусственных данных

Оценка и визуализация способности функции потерь $\mathcal{L}_{\text{Pos}}(f)$ восстанавливать начальное распределение.

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



Классификация изображений на MNIST10

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.

Алгоритм

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

Метрики

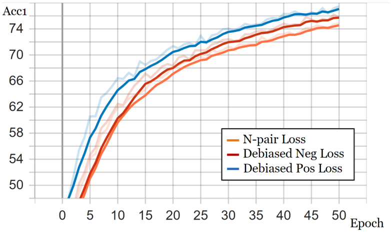
В качестве метрик берётся Accuracy и Top-k-accuracy:

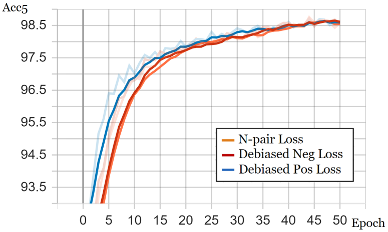
$$Acc_1 = \frac{TP + TN}{TP + FP + TN + FN}, \quad Acc_k = \frac{1}{n}\sum_{i=1}^{n}[y_i \in \hat{y}_i^k]$$

Результаты классификации

Лучшую точность показывает устранение смещения вследствие наличия ложноположительных элементов. Следовательно, задача чувствительна к аугментации изображений.

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





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 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:

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

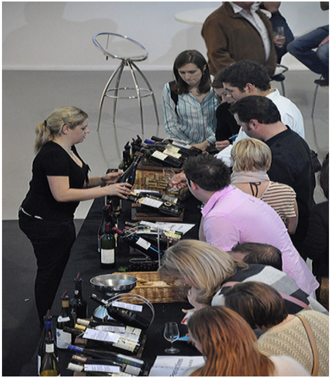
Результаты VQA для $\mathcal{L}_{\text{N-pair}}$ и $\mathcal{L}_{\text{Pos}}$

	$\mathcal{L}_{\text{N-pair}}$	$\mathcal{L}_{\text{Pos}}$
Accuracy	66.29	**67.23**

Yang J. et. al. Vision-language pre-training with triple contrastive learning, 2022.

Пример работы моделей на задаче VQA

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



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

Выносится 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.