

Модификация триплет лосса для задач глубинного метрического обучения

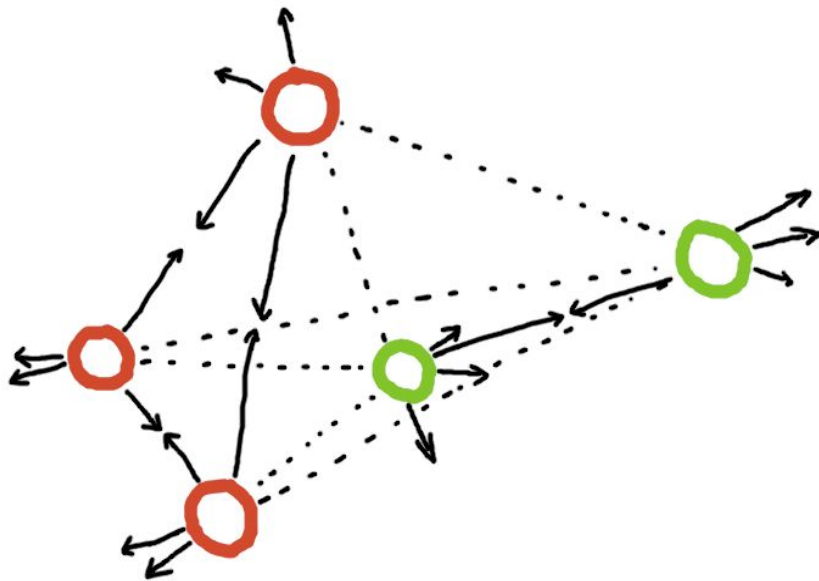
Исследовательский проект

Студент: Федоров Никита Максимович, БПМИ202

Научный руководитель: Качан Олег Николаевич,
приглашенный преподаватель ФКН

Постановка задачи метрического обучения

Модель $f(\theta, x)$ возвращает эмбединги. Хотим, чтобы величина $d(f(\theta, x), f(\theta, y))$ была мала для похожих объектов x и y и велика для разных (d – метрика расстояния в пространстве эмбедингов)



Триплетная функция потерь

Чаще всего используются **контрастивные функции потерь**.

Примером служит **триплет лосс**:

$$\mathcal{L}_{tri}(a, p, n) = \max(0, m + d(a, p) - d(a, n))$$

где m – гиперпараметр, a – якорь, p – позитив, n – негатив.

Актуальность задачи

Глубинное метрическое обучение позволяет:

- решать задачи поиска по изображению и идентификации человека по фото
- решать произвольные задачи ранжирования
- улучшать структуру эмбедингов в признаковом пространстве

Лучшие модели, появившиеся в последние годы, используют нетривиальные методы для достижения высокого качества. В то же время триплетная функция потерь понятна и проста в реализации.

Задачи курсового проекта

- произвести честный сравнительный анализ триплетной функции потерь с лучшими существующими решениями
- модификация формулы триплетной функции потерь

Обзор существующих методов

Лучшее решение на большинстве датасетов метрического обучения в статье^[1]

Эксперименты проводились на датасете **CUB-200-2011** (11788 картинок 200 различных видов птиц)

HyperViT: CMC@1 = **0.856**

[1] Aleksandr Ermolov. «Hyperbolic vision transformers: Combining improvements in metric learning». 2022.

Метрика CMC@1

Модели метрического обучения принято сравнивать по метрике:

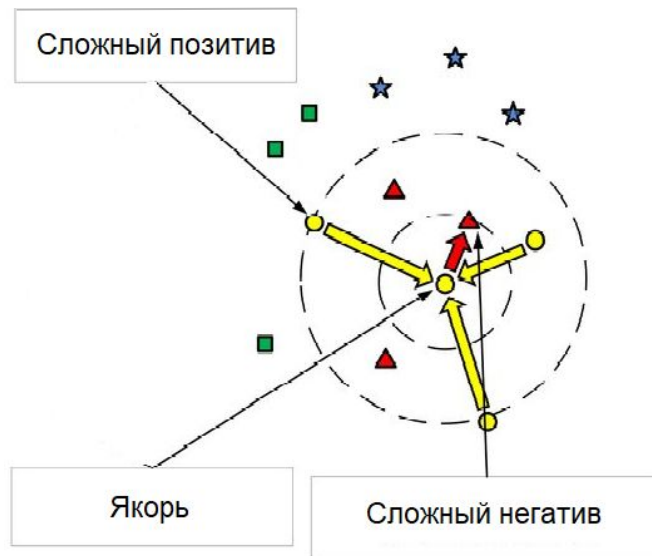
$$CMC@1 = \begin{cases} 1 & \text{ближайший эмбе́ддинг из того же класса, что и картинка-запрос} \\ 0 & \text{иначе} \end{cases}$$

Майнинг

Это процесс формирования троек (якорь, позитив, негатив) по батчу. Наиболее эффективен майнер сложных позитивов и негативов.

Сложный позитив - самый дальний от якоря позитив

Сложный негатив - самый близкий к якорю негатив



Подбор начальной конфигурации нейросети

Претренированная модель дает CMC@1 = **0.8307**, обученная - **0.8531**

Основные гиперпараметры:

- ViT-S претренированный на ImageNet1k
- майнер сложных позитивов и негативов
- в батче 8 классов по 4 картинки каждого класса

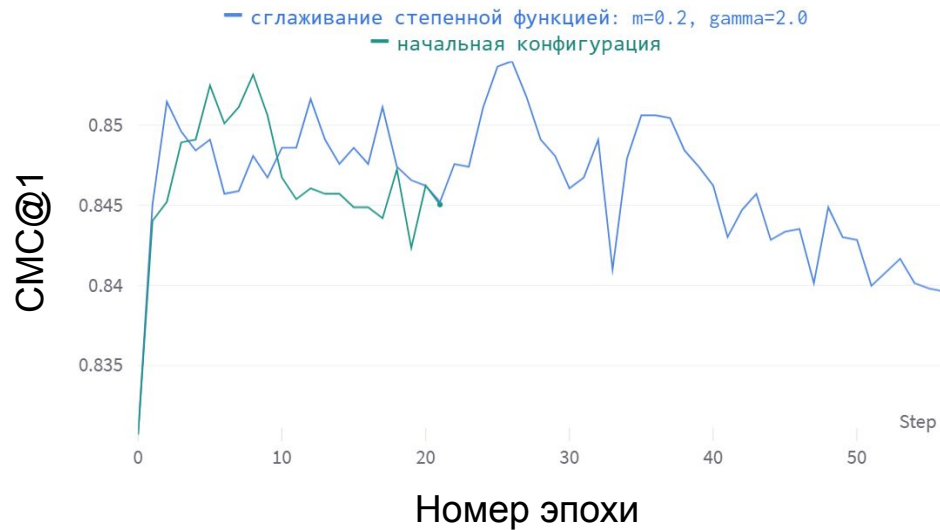


Сглаживание степенной функцией

Идея: замена $\max(0, \cdot)$ в формуле триплетной функции потерь на $x^\gamma \cdot [x > 0]$

Результат:

- 1) снижение переобучения
- 2) CMC@1 = **0.854**



Майнинг полусложных негативов и позитивов

Идея: майнить не самые сложные позитивы и негативы.

Результат:

1) снижение переобучения

2) CMC@1 = **0.8563**



Другие идеи модификации

1. Использование пороговой функции $(x \cdot [x > 0] \cdot [x < t])$ вместо $\max(0, \cdot)$
2. Увеличивающийся гиперпараметр отступа m
3. Кластерный майнер (якорь – средний эмбединг класса по батчу)
4. Учетывание расстояния между позитивом и негативом
5. Взвешивание расстояний, идея фокального лосса

Объединение модификаций

- Количество классов в батче увеличить до 24
- Сглаживание степенной функцией
- Майнинг первого по сложности негатива и второго по сложности позитива

Результат: CMC@1 = **0.8589**



Выводы

- классическая триплетная функция потерь показала результаты, сравнимые с лучшими из существующих подходов
- за счет выбора стратегии майнинга и сглаживания триплетной функции потерь удалось заметно превзойти лучшее решение на датасете CUB-200-2011