

Анализ влияния выбора функции потерь на эффективность обучения по нескольким примерам

Курсовая работа

Студент: Кучеров Василий Дмитриевич, гр. 323

Научный руководитель: канд. физ.-мат. наук Буряк Дмитрий Юрьевич



КАФЕДРА
суперкомпьютеров
и квантовой
информатики

Москва, 2023

Обучение по нескольким примерам (FSL)

Сферы применения:

- Верификация по биометрии
- Распознавание новых команд для голосового управления
- Исследование лекарств: поиск аналогов по свойствам небольшого количества реально синтезированных молекул

Решения:

- Увеличение размера тренировочной выборки
- Сужение пространства поиска модели
- Оптимизация параметров или процесса обучения



Постановка задачи

Целью данной работы является исследование подходов к повышению эффективности обучения нейронных сетей при решении задачи FSL.

Постановка задачи

Рассмотрим методы, основанные на построении эмбедингов. Эффективность таких сетей применительно к формированию пространства эмбедингов существенно зависит от выбора функции потерь, поэтому проведем сравнительное исследование этих функций потерь.

- Провести обзор наиболее популярных функций потерь в области FSL, а также реализовать их.
- Выбрать метод оценки качества пространств эмбедингов.
- Выбрать задачу для проведения эксперимента и модель нейронной сети.
- Провести сравнительный анализ пространств эмбедингов, полученных с использованием выбранных функций потерь.

Функции потерь

Contrastive Loss

$\mathbf{x}_i, \mathbf{x}_j$ — два примера, относящиеся к классам y_i и y_j .

ϵ — гиперпараметр, определяющий нижнюю границу расстояния между объектами разных классов.

$$\mathcal{L}_{\text{cont}}(\mathbf{x}_i, \mathbf{x}_j, \theta) = \mathbb{1}[y_i = y_j] \|f_{\theta}(\mathbf{x}_i) - f_{\theta}(\mathbf{x}_j)\|_2^2 + \mathbb{1}[y_i \neq y_j] \max(0, \epsilon - \|f_{\theta}(\mathbf{x}_i) - f_{\theta}(\mathbf{x}_j)\|_2)^2$$

Triplet Loss

\mathbf{x} — базовый пример

\mathbf{x}^+ — положительный пример (того же класса, что и базовый)

\mathbf{x}^- — негативный пример (другого класса)

$$\mathcal{L}_{\text{triplet}}(\mathbf{x}, \mathbf{x}^+, \mathbf{x}^-) = \sum_{\mathbf{x} \in \mathcal{X}} \max(0, \|f(\mathbf{x}) - f(\mathbf{x}^+)\|_2^2 - \|f(\mathbf{x}) - f(\mathbf{x}^-)\|_2^2 + \epsilon)$$

Функции потерь

Lifted Structured Loss

\mathcal{N} – все негативные пары

\mathcal{P} – все положительные пары

$$D_{ij} = \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|_2 \quad \mathcal{L}_{\text{struct}}^{(ij)} = D_{ij} + \log \left(\sum_{(i,k) \in \mathcal{N}} \exp(\epsilon - D_{ik}) + \sum_{(j,l) \in \mathcal{N}} \exp(\epsilon - D_{jl}) \right)$$
$$\mathcal{L}_{\text{struct}} = \frac{1}{2|\mathcal{P}|} \sum_{(i,j) \in \mathcal{P}} \max(0, \mathcal{L}_{\text{struct}}^{(ij)})^2$$

N-pair Loss

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

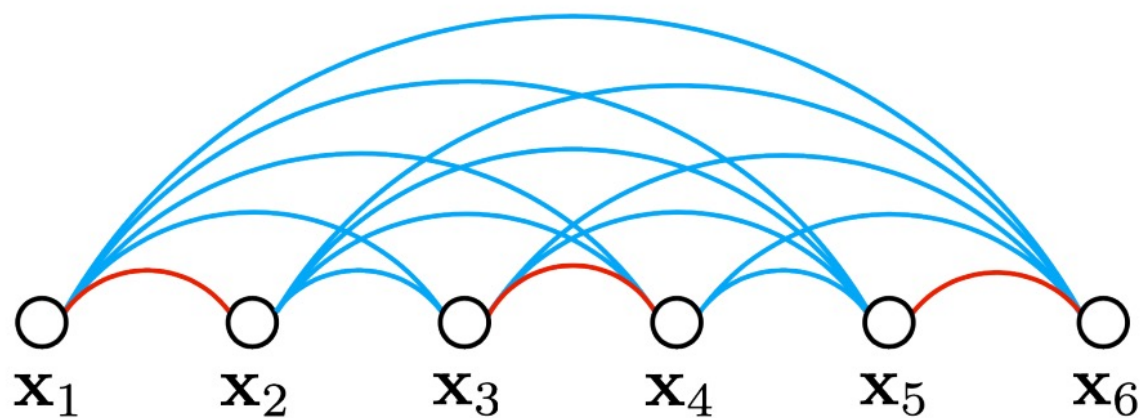
Функции потерь



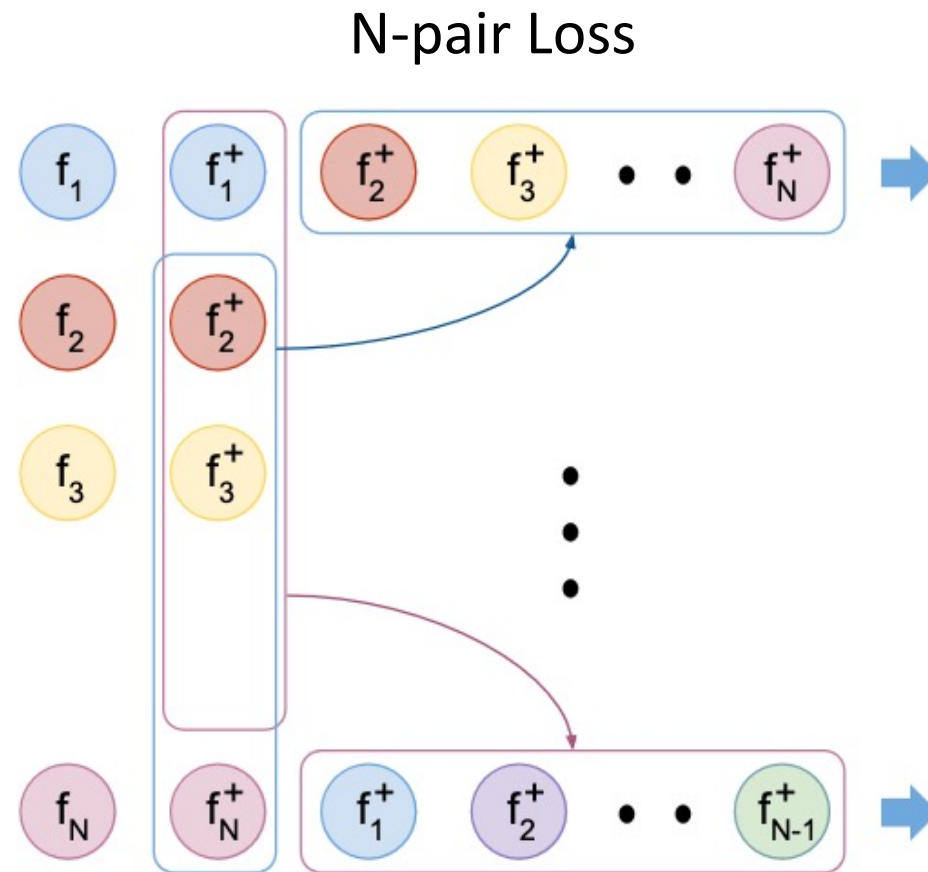
(a) Contrastive Loss



(b) Triplet Loss



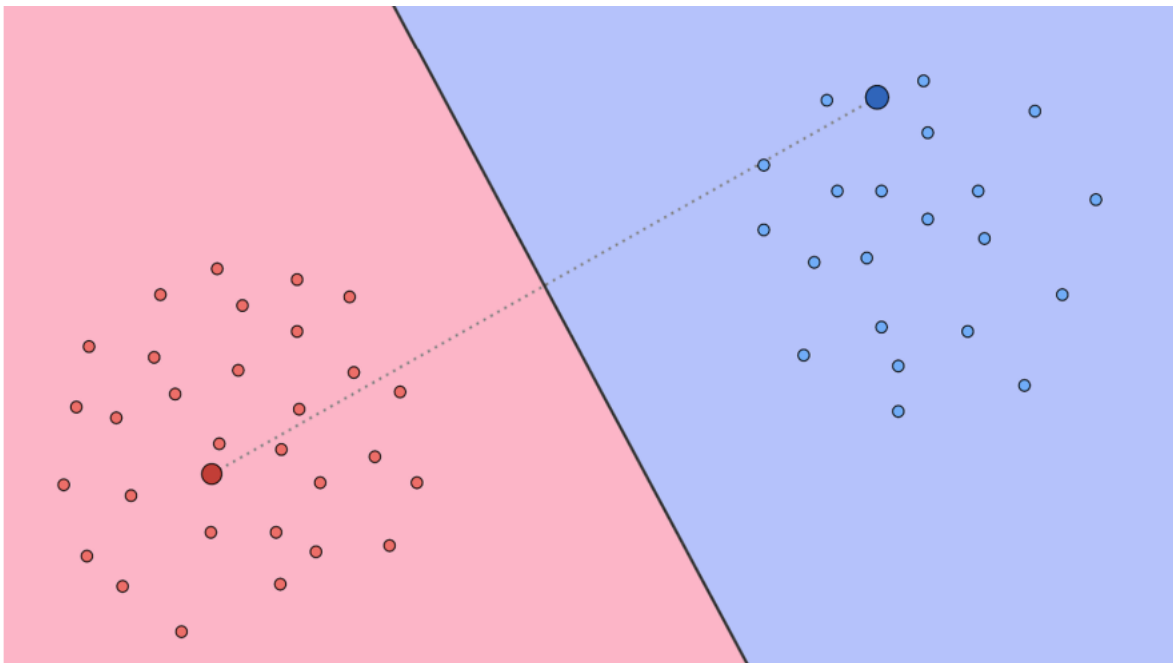
(B) Lifted Structured Loss



Кластеризация

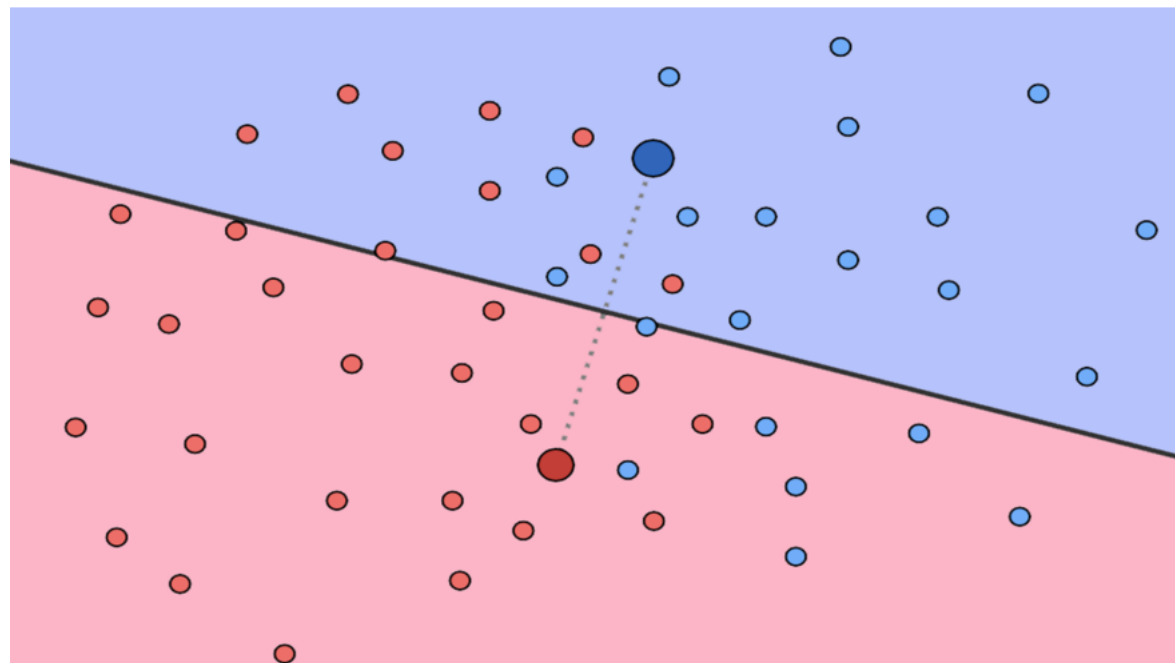
Отношение внутриклассовой к межклассовой дисперсии.

$$\frac{\sigma_{\text{within}}^2}{\sigma_{\text{between}}^2} = \frac{\frac{\sum_{i,j} \|\phi_{i,j} - \mu_i\|_2^2}{N}}{\frac{\sum_i \|\mu_i - \mu\|_2^2}{C}} = \frac{C}{N} \frac{\sum_{i,j} \|\phi_{i,j} - \mu_i\|_2^2}{\sum_i \|\mu_i - \mu\|_2^2}$$



Выражение показывает насколько разные гиперплоскости можно построить, выбирая различные пары объектов из двух классов:

$$R_{HV}(f_{\theta}(x_1), f_{\theta}(x_2), f_{\theta}(y_1), f_{\theta}(y_2)) = \frac{\|(f_{\theta}(x_1) - f_{\theta}(y_1)) - (f_{\theta}(x_2) - f_{\theta}(y_2))\|_2}{\|f_{\theta}(x_1) - f_{\theta}(y_1)\|_2 + \|f_{\theta}(x_2) - f_{\theta}(y_2)\|_2}$$



Задача распознавания ключевых слов

Google Speech V2

- 105 829 примеров 35 английских слов.
- Каждый пример – файл формата 16Khz WAV продолжительность в 1 секунду или меньше
- Использовался голос 2618 разных людей
- В комплекте с датасетом предоставлено несколько WAV файлов (около минуты длиной) с фоновым шумом

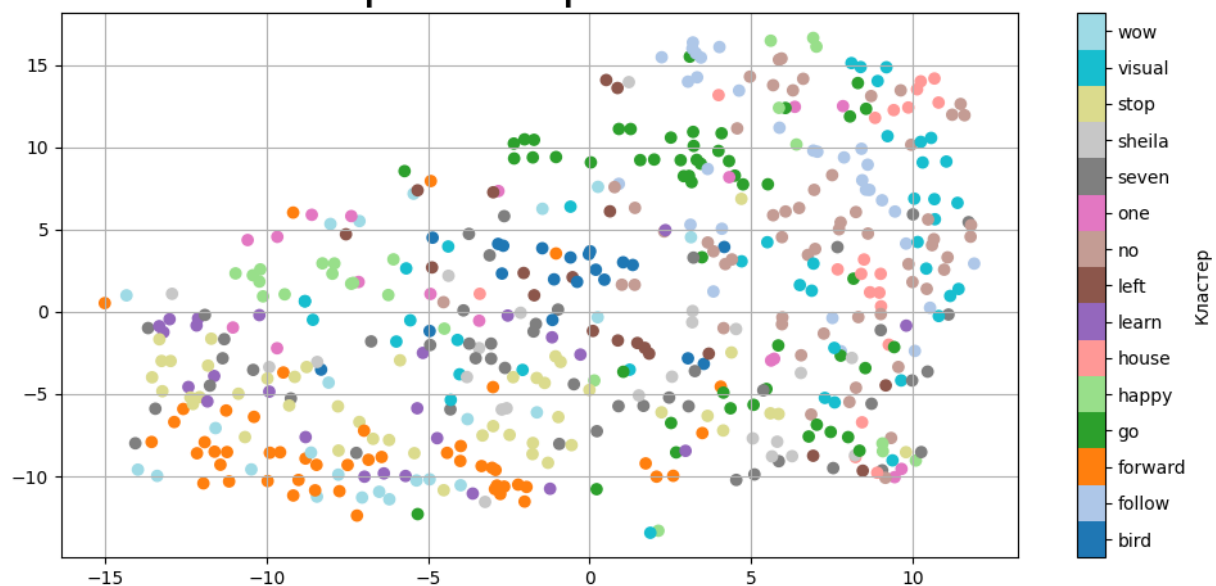
Word	Number of Utterances
Backward	1,664
Bed	2,014
Bird	2,064
Cat	2,031
Dog	2,128
Down	3,917
Eight	3,787
Five	4,052
Follow	1,579
Forward	1,557
Four	3,728
Go	3,880
Happy	2,054
House	2,113
Learn	1,575
Left	3,801
Marvin	2,100
Nine	3,934
No	3,941
Off	3,745
On	3,845
One	3,890
Right	3,778
Seven	3,998
Sheila	2,022
Six	3,860
Stop	3,872
Three	3,727
Tree	1,759
Two	3,880
Up	3,723
Visual	1,592
Wow	2,123
Yes	4,044
Zero	4,052

Результаты экспериментов

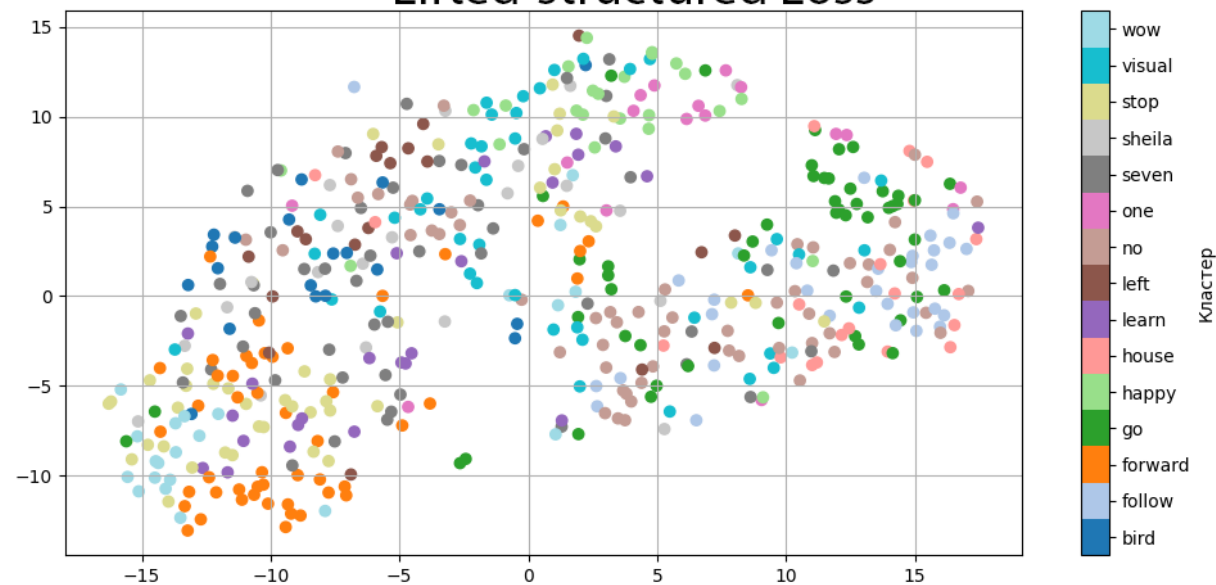
Функция потерь	FC Train	HV Train	FC Test	HV Test
Cross-Entropy	0.70	0.50	1.51	0.57
Triplet	0.44	0.45	1.22	0.58
Lifted Structured	0.32	0.42	0.93	0.54
N-pair	0.55	0.48	1.23	0.58

t-SNE

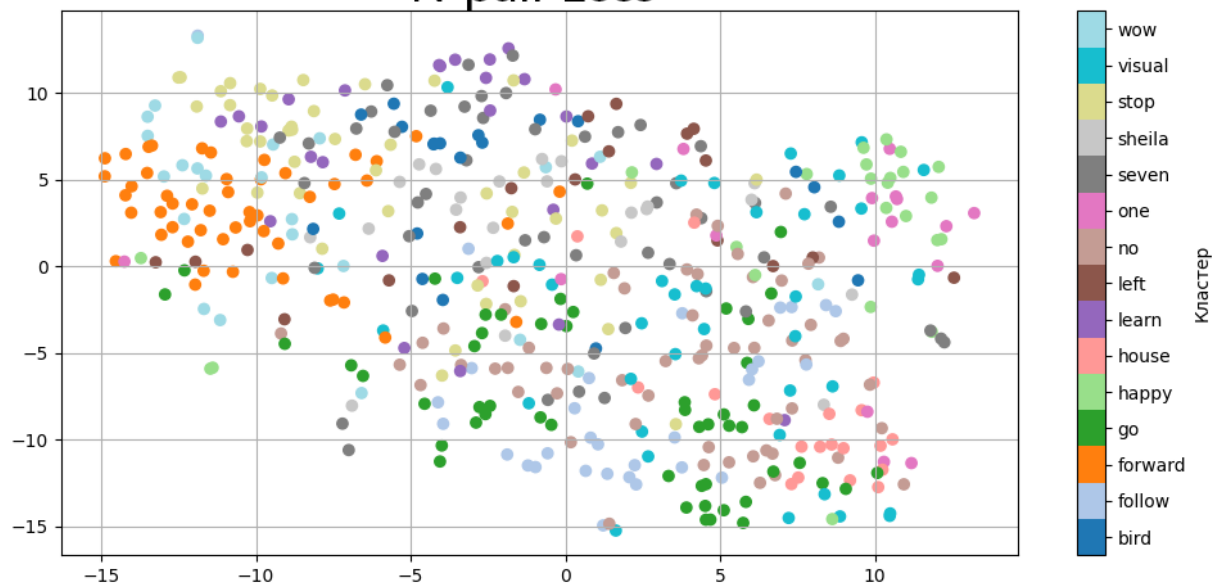
Кросс-энтропия



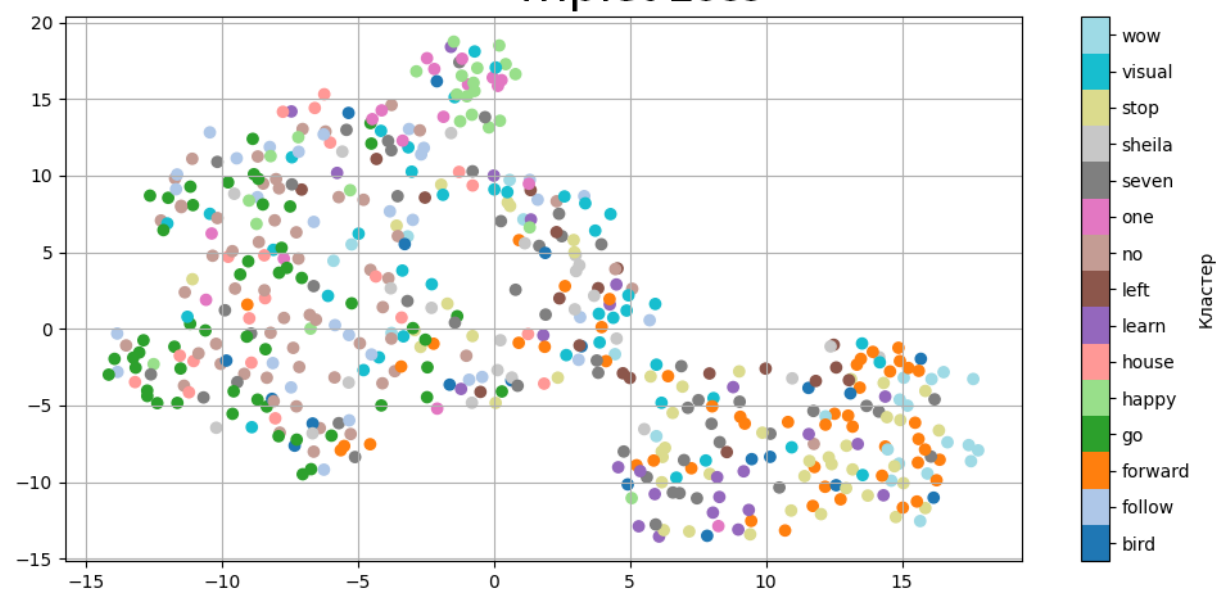
Lifted-structured Loss



N-pair Loss



Triplet Loss



Выводы

- В рамках работы было проведено исследование подходов к повышению эффективности обучения по нескольким примерам
- На языке Python с использованием библиотеки PyTorch были реализованы различные функции потерь и метрики кластеризации.
- На примере задачи распознавания ключевых слов на датасете Google Speech V2 был произведен сравнительный эксперимент выбранных функций потерь.
- В ходе эксперимента была выявлена наиболее эффективная функция по выбранным критериям – Lifted Structured Loss, которая показала на тестовом наборе результаты лучше: до 0.48 (31%) по критерию FC и до 0.04 (6%) по критерию HV

Спасибо за внимание!