

Обучение представлению групп точек данных

Парвиз Каримов

Научный руководитель: Исаченко Р.В.

МФТИ, 2023

Постановка задачи

Пусть дан датасет $\mathfrak{G} = \{(x_i, y_i)\}_{i=1}^n$, $x_i \in X$, $y_i \in \{1, \dots, K\}$. Составим из этих точек данных множества:

$$G_{j,k} = \{x_i | (x_i, y_i) \in \mathfrak{G} \wedge y_i = k \forall i\} : \forall j_1, j_2 G_{j_1,k} \cap G_{j_2,k} = \emptyset$$

Наша задача состоит в том, чтобы сопоставить каждой группе $G_{j,k}$ эмбединг $f_\theta(G_{j,k})$, представляющий собой информативное векторное представление $G_{j,k}$ (Representation Learning: A Review and New Perspectives).

Известные подходы

Представленная задача обычно решается в применении к задаче различения на уровне объекта с применением групповых характеристик.

Unsupervised Visual Representation Learning by Synchronous Momentum Grouping

Вместо классической для задачи контрастивного обучения функции

$$L_i = -\log \frac{\exp(\text{sim}(f_\theta(x_i^a), f_\theta(x_i^b)))}{\sum_j \exp(\text{sim}(f_\theta(x_i^a), f_\theta(x_j)))}$$

Предлагается учитывать групповые характеристики, используя

$$L_i = -\log \frac{\exp(\text{sim}(f_\theta(x_i^a), c_i^b))}{\sum_j \exp(\text{sim}(f_\theta(x_i^a), g_j))}$$

, где x_i^a - характеристики a -ого объекта, принадлежащего i -ому классу;
 c_i^b - групповая характеристика b -го объекта.

Известные подходы

GroupFace: Learning Latent Groups and Constructing Group-based Representations for Face Recognition

При определении объекта к группе с помощью $\arg \max_k p(G_k|x)$ с точки зрения сходимости (эмпирически) лучше приводить математическое ожидание к $\frac{1}{K}$:

$$\arg \max_k \frac{1}{K} (p(G_k|x) - \mathbb{E}[p(G_k|x)]) + \frac{1}{K}$$

The Group Loss for Deep Metric Learning

Принадлежность к группе можно получить на основе ковариационной матрицы, улучшая его с помощью некоторого итерационного процесса:

$$\begin{aligned} Q(t) &= ([X(t) \circ \Pi(t)] \mathbb{1}) \\ X(t+1) &= Q^{-1}(t)[X(t) \circ \Pi(t)] \end{aligned}$$

Результаты

Теорема

Пусть мы имеем оптимально обученную функцию представления объектов $f_\theta(x)$ с точки зрения Triplet loss-a, то есть для любого айтема x_a , его позитива x_p и негатива x_n верно, что

$$\exists m : \|f_\theta(x_a) - f_\theta(x_p)\| - \|f_\theta(x_a) - f_\theta(x_n)\| \leq m \quad \forall (a, p, n)$$

Рассмотрим группы G_{j_1, k_1} , G_{j_2, k_1} , G_{p_1, k_2} , в качестве эмбединга группы возьмём $f_\theta(G_{j, k}) = \frac{1}{|G_{j, k}|} \sum_{x \in G_{j, k}} f_\theta(x)$. Тогда

$$\|f_\theta(G_{j_1, k_1}) - f_\theta(G_{j_2, k_1})\| \leq 2 \max\{m, \max_{s_1 \in G_{j_1, k_1}, s_2 \in G_{p_1, k_2}} \|f_\theta(s_1) - f_\theta(s_2)\|\}$$

Дальнейшие планы

- ➊ Продолжить исследования в области изучения групповых характеристик для оптимально обученных с точки зрения некоторой функции потерь эмбедингов
- ➋ Рассмотреть разные групповые характеристики
- ➌ Провести эксперименты