Обучение представлениям коллекций данных

Парвиз Каримов

Московский Физико-Технический Институт (государственный университет)

Научный руководиель: Р. В. Исаченко

2024

Цель исследования

Мотивация

Современные подходы обучения представлениям используют представления коллекций на промежуточных этапах решения задачи. При этом, сами представления коллекций, их вариации и теоретические свойства рассматриваются крайне редко.

Цель работы

Рассмотрение подходов для составления векторных представлений коллекций и исследование их теоретических свойств.

Литература

Yoshua Bengio, Aaron C. Courville, and Pascal Vincent. Representation learning: A review and new perspectives.

Yonghyun Kim, Wonpyo Park, Myung-Cheol Roh, and Jongju Shin. Groupface: Learning latent groups and constructing group-based representations for face recognition.

Bo Pang, Yifan Zhang, Yaoyi Li, Jia Cai, and Cewu Lu. Unsupervised visual representation learning by synchronous momentum grouping.

Brenden M. Lake, Ruslan Salakhutdinov, and Joshua B. Tenenbaum. Human-level concept learning through probabilistic program induction.

Формальная постановка задачи

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

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

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

Формальная постановка задачи

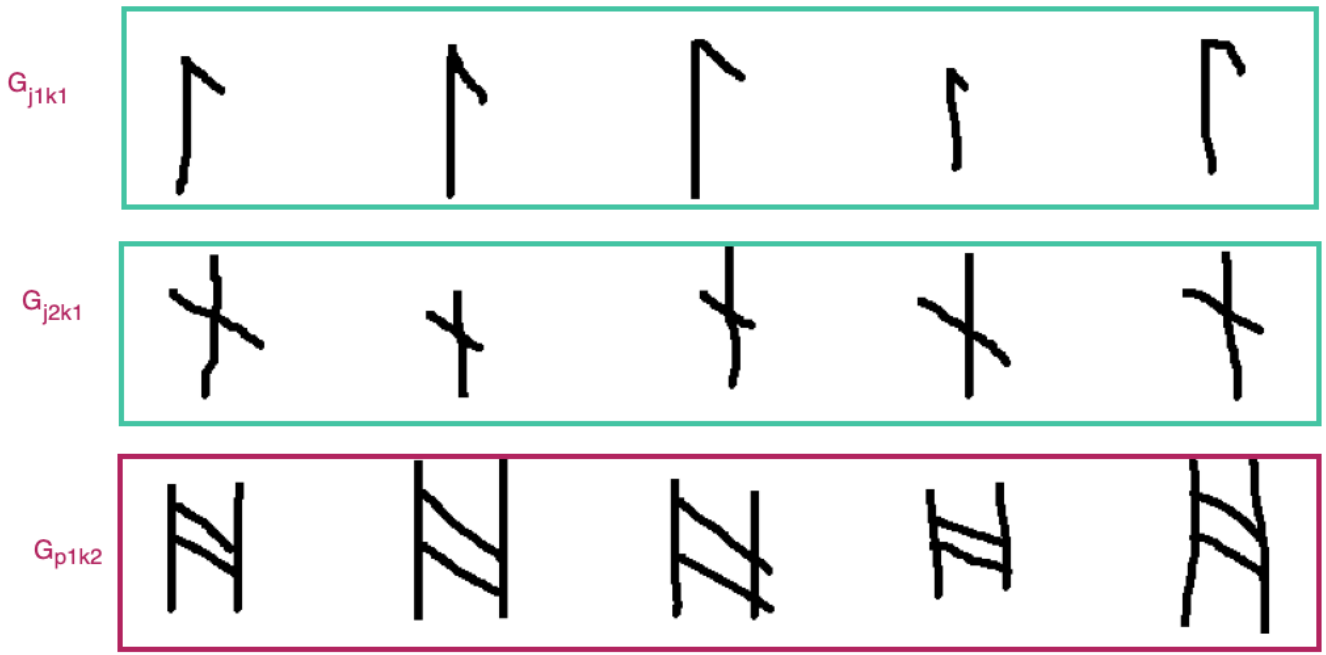

Рис.: Пример коллекций на датасете Omniglot. Группы G_{j_1,k_1}, G_{j_2,k_1} являются множествами букв в рамках одного алфавита, G_{p_1,k_2} является множеством некоторой буквы в рамках другого (отличного от k_1) алфавита.

Существующие методы

Instance level

Обучаются представления коллекций на уровне объектов исходной выборки, после чего из них некоторой аггрегирующей функцией (чаще всего центроид) получаются представления групп.

Group level

Представления коллекций, которые составлены из представлений объектов некоторой аггрегирующей функцией, обучаются совместно.

Существующие методы

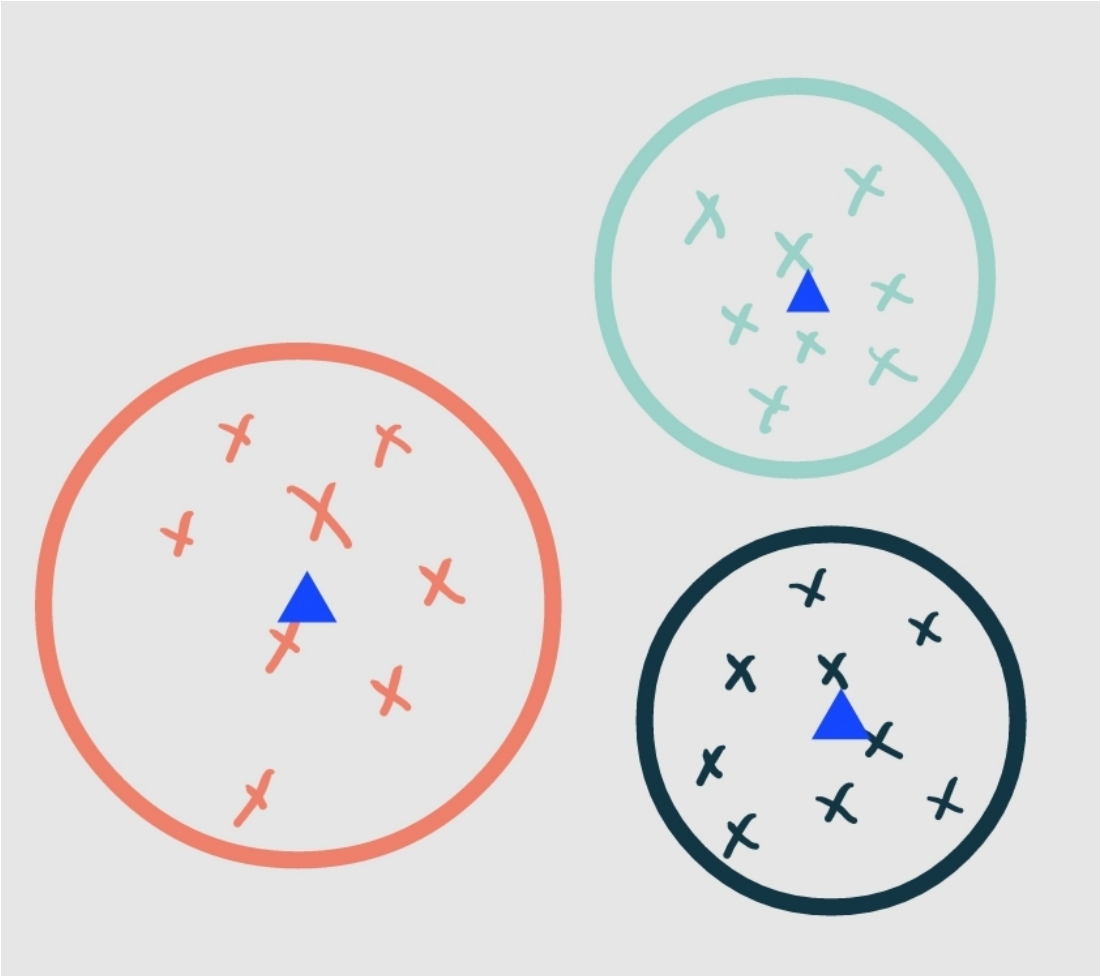

Рис.: Пример расположения представлений в $\mathbb{R}^2$. Треугольником изображен вектор центроида коллекции, разными цветами обозначены объекты, принадлежащие разным коллекциям.

Существующие методы

Method	top1 accuracy
instance level	
ReSSL	69.6
SimSiam	71.3
InfoMin Aug	73.0
group level	
PCL	67.6
DeepClusterV2	70.2
SMoG	73.6

Таблица: Сравнение разных способов обучения представлениям на датасете ImageNet. В качестве базовой модели для получения векторов представлений используется ResNet-50.

Предложенный метод

Задача решается посредством минимизации некоторой контрастивной функции потерь L - т.н. триплетной функции потерь

$$\min_{\theta} L = \min_{\theta} \sum_{(x_a, x_p, x_n)} (||f_\theta(x_a) - f_\theta(x_p)|| - ||f_\theta(x_a) - f_\theta(x_n)|| + m)_+,$$

где $x_a, x_p \in G_{j,k}, x_n \in G_{t,n}$.

После минимизации на объектном уровне представление коллекции $G_{j,k}$ проводится аггрегация значений на уровне группы, в работе

$$f_\theta(G_{j,k}) = \frac{1}{|G_{j,k}|} \sum_{x \in G_{j,k}} f_\theta(x).$$

Результаты

Теорема (Каримов П., 2024)

Пусть мы имеем оптимально обученную функцию представления объектов $f_\theta(x)$ с точки зрения триплетной функции потреь, то есть для любого элемента выборки 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 равномощные коллекции G_{j_1,k_1}, G_{j_2,k_1}, G_{p_1,k_2}, в качестве представления рассмотрим $f_\theta(G_{j,k}) = \frac{1}{|G_{j,k}|} \sum_{x \in G_{j,k}} f_\theta(x)$. Тогда

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

Доказательство

$$\|f_\theta(G_{j_1,k_1}) - f_\theta(G_{j_2,k_1})\| = \frac{1}{|G_{j_1,k_1}|}\left\|\sum_{x_i} f_\theta(x_i) \quad - \sum_{z_i} f_\theta(z_i)\right\| \leq$$

$$\leq \frac{1}{|G_{j_1,k_1}|}\sum_i \|f_\theta(x_i) - f_\theta(z_i)\| \leq \frac{1}{|G_{j_1,k_1}|}\sum_i (\|f_\theta(x_i) - f_\theta(w_i)\| + m) =$$

$$= m + \frac{1}{|G_{j_1,k_1}|}\sum_i \|f_\theta(x_i) - f_\theta(w_i)\| \leq m + \max_{x_i,w_i}\|f_\theta(x_i) - f_\theta(w_i)\| \leq$$

$$\leq 2\max\{m, \max_{x_i,w_i}\|f_\theta(x_i) - f_\theta(w_i)\|\}.$$

Результаты

Эксперимент

В качестве функции представления объекта обучается EfficientNet_b2. В процессе подбора триплетов используется hard-negative mining. В качестве датасета выбран Omniglot, группы $G_{j,k}$ - семплы буквы определённого алфавита, k - индекс алфавита.

$\|\|f_\theta(G_{j_1,k_1}) - f_\theta(G_{j_2,k_1})\|\|$	$\max \|\|f_\theta(s_1) - f_\theta(s_2)\|\|$
734.82	1750.37
280.37	1907.42
254.03	3338.06

Дальнешие планы

- Более точная верхняя граница для разницы между представлениями коллекций
- Нижняя границы между представлениями коллекций
- Результаты с выбором других функций представления коллекций и связанные с ними теоретические результаты.