Обучение представлению групп точек данных

Каримов П.Д.
Научный руководитель: Исаченко Р.В.

December 2023

Аннотация

В данной статье рассматривается задача сопоставления информативных векторных представлений групп данных. Исходный датасет состоит из пар (x_i, y_i), где $x_i \in X$ и $y_i \in \{1, ..., K\}$. Для достижения данной цели, мы рассматриваем оптимально обученные представления, которые позволяют преобразовать группы $G_{j,k}$ в векторные представления $f_\theta(G_{j,k})$.

1 Введение

Эффективность методов машинного обучения в значительной степени зависит от выбора представления данных (или признаков). на которых они применяются. По этой причине большая часть фактических усилий по внедрению алгоритмов машинного обучения направлена на разработку конвейеров предварительной обработки и преобразования данных, которые в результате чего создается представление данных, способное поддерживать эффективное машинного обучения. Такая разработка характеристик важна, но трудоемкое и подчеркивает слабость нынешних алгоритмов обучения: их неспособность извлекать информацию из данных. неспособность извлечь и систематизировать дискриминационную информацию из данных.

Эта статья посвящена обучению представлениям, т.е. обучению представлений данных, которые облегчают извлечение полезной информации при построении классификаторов или других предсказателей. В случае с вероятностными моделями хорошим представлением часто является такое, которое отражает апостериорное распределение базовых объясняющих факторов для наблюдаемых входных данных. Хорошее представление также полезно в качестве входных данных для контролируемого предиктора. Среди различных способов обучения представлениям в данной статье фокусируется на методах глубокого обучения: тех, которые формируются путем композицией множества нелинейных преобразований, с целью получения более абстрактных представлений.

2 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:

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

Наша задача состоит в том, чтобы сопоставить каждой группе $G_{j,k}$ эмбеддинг $f_\theta(G_{j,k})$, представляющий собой информативное векторное представление $G_{j,k}$. Определение "информативного"в задаче обучения представлениям обычно формулируется под конкретную задачу(Representation Learning: A Review and New Perspectives), обычно полагают, что в рамках такого представления "близкие"в каком-то смысле объекты находятся в пространстве представлений близко, а "далёкие далеко.

3 Существующие работы

Данная задача обычно применяется к задаче объектного различения (instance discrimination) для учёта групповой информации между объектами для улучшения качества модели. Ниже представлены общие описания рассмотренных статей.

3.1 Unsupervised Visual Representation Learning by Synchronous Momentum Grouping

Обычно это делается минимизацией лосса, который учитывает взаимодействия между айтемами:

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

Можно попробовать обобщить этот лосс до групп, чтобы получить фичи для групп:

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

Но если пытаться делать так - некоторые айтемы могут быть близки одновременно нескольким группам.

Если мы хотим, чтобы группы были таки существенно разные (с точки зрения айтемов, которые хочется этим группам соотнести) - стоит брать group-item лосс:

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

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

Таким образом, мы пытаемся приблизить семплы группы к самой группе и отдалить эти семплы от других групп.

Группы инициализируются, например, каким-нибудь алгоритмом кластеризации. Обновляются группы так:

$$c_i = \arg\min_{g_k} \mathrm{sim}(f_\theta(x_i), g_k)$$
$$\mathrm{g}_k \leftarrow \beta g_k + (1-\beta)\mathrm{mean}_{c_t = g_k} f_\theta(x_t)$$

Такой способ позволяет через эмбеды групп пропускать градиенты, и этим отличается от аналогов, которые приводятся в статье.

Поскольку составленные таким образом группы могут сколлапсировать из-за того, что на каждом шаге мы движемся по батчу - авторы предлагают периодически перегруппировывать центроиды.

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

В этой статье авторы предлагают к эмбедам инстансным добавлять прямо групповые фичи (явно), представляя специфическую архитектуру сети.

В качестве лосса берут сумму классификационного (CE-лосс) и репрезентационного (ArcFace-лосс) с какими-то весами.

Специфичным образом определяется принадлежность к группе, пытаясь адресовать неравномерность распределения по группам - не

$$\arg\max_k p(G_k|x)$$

, а

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

Перескоринг интуитивно обосновывается тем, что матожидание этой величины равно $\frac{1}{K}$.

3.3 The Group Loss for Deep Metric Learning

В этой статье авторы ставят в противовес классическому выбору лосса в Representation learning a.k.a. Contrastive/Triplet loss свой лосс, который каждому айтему в батче ставит группу. Делают они это вытаскиванием фичей из нейронки, подачей софтмакс-вероятностей как начальному приближению их классов и запускают некоторый итерационный процесс. По окончании берут CE-лосс и бегут backprop-ом.

4 План эксперимента

Целью эксперимента является сравнение различных методов составления эмбеддингов групп и, вместе с тем, проверка качества этих представлений на основе расстояния между ними.

Датасет Omniglot [Lake et al.(2015)Lake, Salakhutdinov, and Tenenbaum] содержит 1623 различных рукописных символа из 50 различных алфавитов. Каждый из 1623 символов был нарисован в режиме онлайн через Amazon's Mechanical Turk 20 разными людьми. Для этого датасета обычно в качестве аугментации данных применяют повороты на 90,180,270 градусов. Каждый алфавит в представленной постановке можно использовать как группу G, в качестве групп $G_{j,k}$ взяв каждую из аугментированных версий алфавита. Таким образом, мы получим необходимые нам группы.

В эксперименте предполагается рассмотреть 4 конфигурации:

- обучение на основе связей группа-группа с использованием центроида в качестве конечного эмбеддинга группы

- обучение на основе связей группа-объект с использованием центроида в качестве конечного эмбеддинга группы

- обучение на основе связей группа-группа с использованием медоида в качестве эмбеддинга группы

- обучение на основе связей группа-объект с использованием центроида в качестве конечного эмбеддинга группы

В качестве метрики качества будем рассматривать среднее нормы разности между эмбеддингами групп. В случае рассмотрения групп в рамках одного алфавита следует добавить так же верхнюю оценку на норму разности между ними (см. теорему (1)).

5 Результаты

В рамках представленной задачи можно посмотреть в сторону того, что будет, если мы оптимально потренируем представления на уровне айтемов, например:

Теорема 1 *Пусть мы имеем оптимально обученную функцию представления объектов $f_\theta(x)$ с точки зрения Triplet loss-а, 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 эмбеддинга группы возьмём $f_\theta(G_{j,k}) = \frac{1}{|G_{j,k}|} \sum_{x \in G_{j,k}} f_\theta(x)$. Тогда

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

Таким образом, выбор представления группы как среднего арифметического поайтемных эмбедов в случае, если $m < 0$, является в какой-то степени оправданным. Отметим, однако, что в теореме ничего не упоминается про расстояние с центроидом отрицательного примера ещё, в этом направлении ещё предстоят некоторые исследования.

Список литературы

[Lake et al.(2015)Lake, Salakhutdinov, and Tenenbaum] Brenden M. Lake, Ruslan Salakhutdinov, and Joshua B. Tenenbaum. Human-level concept learning through probabilistic program induction. volume 350, pages 1332–1338, 2015. doi:10.1126/science.aab3050. URL `https://www.science.org/doi/abs/10.1126/science.aab3050`.

[Bengio et al.(2012)Bengio, Courville, and Vincent] Yoshua Bengio, Aaron C. Courville, and Pascal Vincent. Representation learning: A review and new perspectives. volume 35, pages 1798–1828, 2012. URL `https://api.semanticscholar.org/CorpusID:393948`.

[Pang et al.(2022)Pang, Zhang, Li, Cai, and Lu] Bo Pang, Yifan Zhang, Yaoyi Li, Jia Cai, and Cewu Lu. Unsupervised visual representation learning by synchronous momentum grouping. In *European Conference on Computer Vision*, 2022. URL `https://api.semanticscholar.org/CorpusID:250490993`.

[Kim et al.(2020)Kim, Park, Roh, and Shin] Yonghyun Kim, Wonpyo Park, Myung-Cheol Roh, and Jongju Shin. Groupface: Learning latent groups and constructing group-based representations for face recognition. pages 5620–5629, 06 2020. doi:10.1109/CVPR42600.2020.00566.

[Elezi et al.(2019)Elezi, Vascon, Torcinovich, Pelillo, and Leal-Taixé] Ismail Elezi, Sebastiano Vascon, Alessandro Torcinovich, Marcello Pelillo, and Laura Leal-Taixé. The group loss for deep metric learning. In *European Conference on Computer Vision*, 2019. URL `https://api.semanticscholar.org/CorpusID:208527171`.