Self-supervised learning

Автор: Виданов Андрей
Руководитель: Барабанщикова Полина

12 мая 2023 г.

Аннотация

В последние годы все активнее развиваются техники обучения без учителя, которые не требуют предварительной разметки на данных. Для изображений является полезной задача создания представлений, поскольку они могут использоваться в различных задачах машинного обучения для повышения качества финальных моделей. Переход к подобным методам является перспективным, но сопряжен с рядом новых задач.Предлагается рассмотреть методы на основе многокомпонентной функции потерь, в частности VICReg и TiCo, а также их модификации. Эксперименты проводятся на наборе данных CIFAR10. В качестве финального классификитора используется простая линейная модель

Введение

В наше время все больше внимания приковывают техники самообучения [6]. В данной области ставится множество задач и применяются различные, в том числе генеративные, методы [4]. Как показывают эксперименты, применение классических для традиционного машинного обучения техник, таких как batch norm, может улучшать результаты до передовых в данной области знания [5].

Однако мотивы применения тех или иных подходов зачастую являются сугубо практическими (улучшение качества модели), но не теоретически обоснованными, что ставит под вопрос универсальность их применения для задач своего типа. Теоретические изыскания в данном вопросе активно проводятся [2], тем не менее множество работ сосредоточено на получении эффективных эвристик для генерации векторных представлений (эмбеддингов). Специфика задачи генерации эмбеддингов для картинок требует особых методов. В частности, для задач связанными с изображениями нет методов для создания отрицательных пар (пара изображений разных классов [7]), обучаясь на которых модель могла бы улавливать чем отличаются кардинально разные картинки. Избежать данной проблемы можно спроектировав модели, не требующие негативных пар. Модель минимизирует расстояние между эмбеддингами похожих картинок. Но данные методы склонны к вырождению, иначе говоря, сталкиваются с коллапсом(явление, когда модель сопоставляет всем картинкам одно и то же представление). В результате чего стоит крайне аккуратно подбирать лосс-функции [2], или проектировать модель [1], [3].

Приденные выше техники эффективны на практике, но теоретическая мотивация их использования до сих пор остается непрозрачной. Задачей исследования является нахождения зависимостей между применениями различных техник и качеством получаемых эмбеддингов, а также риска столкнуться с коллапсом. Предполагается, что модификации лосс-функции приведут к улучшению финальных показателей в прикладных задачах.

Измеримыми целями исследования явлеются показатели финальнового качества классификации, полученные в ходе применения сгенерированных эмбеддингов. В частности, метод генерации представлений обучается на тренировочной выбоке, далее на представлениях на тренировочной выборке обучается простой линейный классификатор, после чего на тестовой часте данных классификатор предсказывает класс изображения, затем считается точность предсказания.

Постановка задачи

Рассматривается задача построения по объекту вектора, f действует из пространства картинок в пространство векторов $f : \mathbb{R}^{c\times h\times w} \to \mathbb{R}^d$,(c - количество цветовых каналов, h - высота, w - ширина исходного изображения),
который, в некотором смысле, ее описывает.

Рассмотрим структуру работы моделей данного типа. На вход модели подается картинка $\boldsymbol{x}$. По ней строятся два видоизмененных изображения $\boldsymbol{y_1}, \boldsymbol{y_2}$. Например, применяются сдвиги, повороты, размытие, сжатия вдоль осей). Далее к ним применяется обучаемая модель $f_{\boldsymbol{\theta}}(\boldsymbol{y_1}), f_{\boldsymbol{\theta}}(\boldsymbol{y_2})$(где $\boldsymbol{\theta}$ вектор параметров модели). После чего применяется проектор $p_{\boldsymbol{\theta}}$

$$\boldsymbol{z_i} = p_{\boldsymbol{\theta}}(f_{\boldsymbol{\theta}}(\boldsymbol{y_i})),$$

который также может зависеть от обучаемых параметров, и производится сравнение с помощью функции потерь L, которая минимизируется при обучении. Для подсчета функции потерь получившиеся векторы собираются в матрицы, которые обозначим через $\boldsymbol{Z} = [\boldsymbol{z_1}, ..., \boldsymbol{z_n}]$ и $\boldsymbol{Z'} = [\boldsymbol{z_1}, , , \boldsymbol{z_n}]$

$$L(\boldsymbol{Z}, \boldsymbol{Z'}) = \lambda s(\boldsymbol{Z}, \boldsymbol{Z'}) + \mu[v(\boldsymbol{Z}, \boldsymbol{Z'})] + \nu[c(\boldsymbol{Z}, \boldsymbol{Z'})],$$

где $s(\boldsymbol{Z}, \boldsymbol{Z'})$ — отвечает за инвариантность получаемых представлений,
$v(\boldsymbol{Z}, \boldsymbol{Z'})$ — отвечает за дисперсию представлений,
$c(\boldsymbol{Z}, \boldsymbol{Z'})$ — отвечает за ковариацию

Для обучения модели ставится задача оптимизации, а именно

$$\boldsymbol{\theta}^*_L = \arg\min_{\boldsymbol{\theta}} L(\boldsymbol{Z_\theta}, \boldsymbol{Z'_\theta}|\boldsymbol{X})$$

Качество получаемых эмбеддингов оценивается благодаря испльзованью их в качестве параметров для простой линейной модели классификации

$$min_{\varphi} CE(h_{\varphi}(f_{\boldsymbol{\theta}^*_L}(\boldsymbol{X})), y) \xrightarrow{L} min,$$

где CE - кросс-энтропия;
$h_{\varphi} : \mathbb{R}^d \to \mathbb{R}$ - линейная модель классификации с ветором параметров $\boldsymbol{\varphi}$;
y - истинный класс.

План экспериментов

Для экспериментов используется датасет CIFAR10. Набор данных CIFAR10 состоит из 60000 цветных изображений 32x32 в 10 классах, по 6000 изображений в каждом классе. Есть 50000 обучающих изображений и 10000 тестовых изображений.

Модель состоит из базовой части, в качестве которой используется ResNet50, головы - BarrowTwin [8] проекция. Для проверки качества полученных эмбеддингов используется линейная модель классификации. В качестве критерия качества модели классификации используется Accuracy. Ожидается, что качество классификации улучшается после процедуры самообучения.

Предварительный отчет

Тестирование модели не выявило логических противоречий с утверждениями ранее изложенными в работе Процедура самообучения способствует высокой точности классификации даже при использовании простой линейной модели.

Предлагаемый метод

В модели VicReg используется довольно простая функция для слагаемого отвечающего за ковариацию. Данный факт стал основным источником вдохновения для следующей идеи:
В качестве слагаемого контроля ковариации использовать Transformation Invariance and Covariance Contrast (TiCo) подход для функции потерь для VicReg

Стандартный подсчет ковариации:

```
cov_x = (z_a.T * z_a) / (N - 1)
cov_y = (z_b.T * z_b) / (N - 1)
dcov_x = diag(cov_x)
dcov_y = diag(cov_y)
covcomponent = div(sum(dcov_x**2)) + div(sum(dcov_y**2))
```

Заменить на

$B = (z_1.T \times z_1)/n$
$C = \beta C + (1 - \beta)B$
$covcomponent = mean(sum(((z_1 \times C) * z_1), dim = 1))$
Засчет пересчета ковариации для каждого батча ожидается повысить итоговое качество, т.к. регуляризация станет более адаптивной

Итоговая точность предсказаний(Accuracy) на задаче классификации		
TiCo Loss	Custom Loss	VicReg Loss
0.3046875	0.3779296875	0.421875

Таким образом предложенный метод продемонстрировал средние результаты между использованием оригинальной функции потерь VicReg и аналогичной функции модели TiCo.

Заключение

В данной работе была предложена новая функция потерь. Но, к сожалению, данный подход не продемонстрировал увеличение целевой метрики по сравнению с базовой моделью.

Список литературы

[1] Xinlei Chen Kaiming He - "Exploring Simple Siamese Representation Learning"

[2] Adrien Bardes, Jean Ponce, Yann LeCun - "VICREG: VARIANCE-INVARIANCE-COVARIANCE REGULARIZATION FOR SELF-SUPERVISED LEARNING"

[3] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, Stephane Deny ´

[4] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang, Jie Tang - "Self-supervised Learning: Generative or Contrastive"

[5] Abe Fetterman Josh Albrecht - Understanding Self-Supervised and Contrastive Learning with "Bootstrap Your Own Latent"(BYOL)

[6] Самообучение. Проблематика и постановка задачи

[7] Задача классификации по представлениям

[8] Zbontar, Jure, et al. "Barlow twins: Self-supervised learning via redundancy reduction."International Conference on Machine Learning. PMLR, 2021.