

Self-supervised learning

Автор: Виданов Андрей
Руководитель: Барабанщикова Полина

25 апреля 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 классификатора используется простая линейная модель

Введение

В последние годы все больше внимания развитию техник 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 практически (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].

Приданные 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.

Предлагаемые изменения в функции потерь являются точечными из-за чего получится лучше определить их влияние на результат модели.

Измеримыми целями исследования являются численные показатели финального качества классификации, полученные в ходе применения сгенерированных эмбеддингов. В частности, метод обучается на тренировочной выборке, далее на представления на тренировочной выборке обу-

чается простой линейный классификатор, после чего на тестовой части 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 $\mathbf{f} : \mathbb{R}^{c \times h \times w} \rightarrow \mathbb{R}^d$ (где 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 \mathbf{x} . По ней строятся два видоизмененных (например, применяются сдвиги, повороты, размытие, сжатия вдоль осей) изображения $\mathbf{y}_1, \mathbf{y}_2$. Далее к ним применяется обучаемая модель $\mathbf{f}_\theta(\mathbf{y}_1), \mathbf{f}_\theta(\mathbf{y}_2)$ (где θ вектор параметров модели). После чего применяется проектор \mathbf{p}_θ

$$\mathbf{z}_i = \mathbf{p}_\theta(\mathbf{f}_\theta(\mathbf{y}_i)),$$

который также может зависеть от обучаемых параметров, и производится сравнение с помощью функции потерь L , которая минимизируется при обучении. Для подсчета лосс-функции получившиеся векторы собираются в коллекции, которые обозначим через $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_n]$ и $\mathbf{Z}' = [\mathbf{z}'_1, \dots, \mathbf{z}'_n]$

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

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

Для обучения модели ставится задача оптимизации, а именно

$$\theta_L^* = \arg \min_{\theta} L(\mathbf{Z}_\theta, \mathbf{Z}'_\theta | \mathbf{X})$$

Качество получаемых эмбедингов оценивается благодаря использованию их в качестве параметров для простой линейной модели классификации

$$\min_{\varphi} CE(h_{\varphi}(f_{\theta_L^*}(\mathbf{X})), y) \xrightarrow{L} \min$$

, где CE - кросс-энтропия;

$h_{\varphi} : \mathbb{R}^d \rightarrow \mathbb{R}$ - линейная модель классификации с вектором параметров φ ;

y - истинный класс.

План экспериментов

Для экспериментов используется датасет "CIFAR10". Набор данных CIFAR-10 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 проекция. Для проверки качества полученных эмбедингов используется линейная модель классификации. В качестве метрики качества модели классификации используется Ассигасу. 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.

Предлагаемые методы

В качестве слагаемого контроля ковариации использовать Transformation Invariance and Covariance Contrast (TiCo) подход для функции потерь для VicReg

Стандартный подсчет ковариации:

$$\begin{aligned} cov_x &\leftarrow (z_a.T \times z_a)/(N-1); \\ cov_y &\leftarrow (z_b.T \times z_b)/(N-1); \\ dcov_x &\leftarrow diag(cov_x); \\ dcov_y &\leftarrow diag(cov_y); \\ covcomponent &\leftarrow div(sum(dcov_x^2)) + div(sum(dcov_y^2)) \end{aligned}$$

Заменить на

$$\begin{aligned} B &= (z_1.T \times z_1)/n \\ C &= \beta C + (1 - \beta)B \\ covcomponent &= mean(sum(((z_1 \times C) * z_1), dim = 1)) \end{aligned}$$

Засчет пересчета ковариации для каждого батча ожидается повысить итоговое качество, т.к. регуляризация станет более адаптивной

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