

Исследование контекста статьи

Bootstrap Your Own Latent A New Approach to Self-Supervised Learning

Авторы статьи:

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, Michal Valko

Автор исследования:

Сухоросов Алексей

Препринт выложен на arXiv в середине 2020 года, а конце 2020 года статья прошла на NIPS в oral виде. Статья заинтересовала сообщество, поскольку предложила неожиданное и неинтуитивное решение известной проблемы, добивающееся SOTA. Вскоре после того, как статью выложили начали появляться обзоры статьи на youtube каналах и в твиттере.

Статья написана четырнадцатью авторами из deepmind, пять указаны как основные авторы, участвовавшие в равной степени. У большинства авторов в публикациях можно встретить статьи по Reinforcement Learning (например, Deep Reinforcement Learning and the Deadly Triad) и общие исследования архитектур нейронных сетей (например FiLM: Visual Reasoning with a General Conditioning Layer и Can Recurrent Neural Networks Warp Time?). Наверное наиболее важной предшествующей статьёй можно назвать Bootstrap Latent-Predictive Representations for Multitask Reinforcement Learning. 6 из 7 авторов данной статьи являются и авторами статьи по BYOL, при этом авторы утверждают что методы в целом похожи. Кроме того, авторы указывают в статье, что метод сглаженного обновления параметров target сети взят из области Reinforcement Learning'a, поэтому вероятно статья является скорее крайне удачной попыткой применить накопленный опыт авторов в смежной области, нежели закономерным результатом предыдущих исследований.

Из наиболее значимых источников можно выделить несколько статей. Авторы ссылаются на статью Continuous Control With Deep Reinforcement Learning утверждая, что именно в ней был представлен метод обновление target сети параметрами online сети (с экспоненциальным скользящим средним). Кроме того, в источниках встречается ранее упомянутая статья Bootstrap Latent-Predictive Representations for Multitask Reinforcement Learning.

У статьи есть непосредственное продолжение от тех же основных авторов: BYOL works even without batch statistics. В этой статье авторы опровергают гипотезу неявных negative examples. Кроме того, авторы применили идеи из статьи к обработке видео в Broaden Your Views for Self-Supervised Video Learning. Вообще статья достаточно популярная: 648 цитирований по данным google scholar.

По self-supervised representation learning регулярно выходит много статей, но в основном методы используют negative examples. Примерно одновременно с BYOL появилась статья по модели MoCo, которая на тот момент уступала по качеству на

ImageNet'е, но последние её модификации добиваются лучшего, чем у BYOL качества. Кроме того, в начале 2021 вышла ещё одна статья по self-supervised representation learning: Unsupervised Learning of Visual Features by Contrasting Cluster Assignments, в которой представлен метод SwAV так же превосходящий качество BYOL. На сегодняшний день SOTA является модель EsViT (Efficient Self-supervised Vision Transformers for Representation Learning), вышедшая в середине 2021 года.