

Рецензия на **Bootstrap Your Own Latent A New Approach to Self-Supervised Learning** от Павлова Вадима(181)

Главным результатом статьи является новый подход к обучению в режиме self-supervised (далее BYOL), который не требует подбора негативных примеров. Идея заключается в том, чтобы поддерживать две сети - online и target, обучая только online сеть, а target обновляя скользящим средним online сети, используя при этом L2 loss. Эксперименты авторов показывают что BYOL дает качество лучше, чем другие self-supervised модели, а также является более устойчивым к выбору аугментаций.

Можно выделить несколько преимуществ статьи:

1. BYOL действительно позволяет обучать представления картинок в режиме self-supervised без использования негативных примеров, и при этом не сходиться к тривиальному решению. Эта особенность позволяет избавиться от некоторых проблем предыдущих подходов - необходимость аккуратного выбора негативных примеров и большого размера батча.
2. Авторы проводят много экспериментов, сравнивая с лучшими существующими на момент выхода статьи решениями, а также показывают влияние различных гиперпараметров и частей модели на итоговое качество.
3. Актуальность задачи обучения векторных представлений картинок без разметки, ровно как и данной работы, не вызывает вопросов.

Недостатки работы:

1. Отсутствие теоретической обоснованности. В первой версии статьи авторы вообще никак не объясняли, почему их метод не сходится к тривиальному решению, потом добавили некое интуитивно-эвристическое рассуждение
2. Вопросы к приведенным в работе числам. Непонятно, насколько значимо улучшение в качестве, если репродукция SimCLR приносит больше процентных пунктов, чем сам BYOL
3. Не все эксперименты согласуются с последующими работами, например, в статье **[Exploring Simple Siamese Representation Learning]** утверждают, что экспоненциальное сглаживание для target сети не является обязательным, в то время как авторы BYOL говорят, что без этого ничего не учится
4. Воспроизводимость - полностью представлен псевдокод, но нет полной официальной реализации, а запустить сторонний оказалось очень тяжело

Статья написана понятно, все эксперименты описаны четко, по возможности с интуицией, которая за ними стоит, что позволяет чувствовать себя с авторами на одной волне.

Интересные моменты из рецензий на NeurIPS 2020:

1. Все рецензенты указали на отсутствие теоретической обоснованности

2. Двое рецензентов отметили недостаток сравнения с моделями MoCo и Mean Teacher

Я ставлю статье оценку 8, я уверен в оценке на 4/5.