

Momentum Residual Neural Networks

(Michael E. Sander, Pierre Ablin, Mathieu Blondel, Gabriel Peyre)

1. Содержание и вклад.

Статья затрагивает актуальную тему о количестве памяти, требуемого для обучения ResNet с backpropagation, которая растет линейно с глубиной модели. Авторы предлагают модификацию (Momentum ResNet), которая меняет forward rule модели, добавляя к ней momentum, и делает ее обратимой. Также они показывают, что Momentum ResNet можно интерпретировать как дифференциальное уравнение второго порядка, а добавление импульса увеличивает аппроксимационные возможности Momentum ResNets (можно выучить любое линейное отображение с точностью до мультипликативного множителя, в отличие от ResNet).

2. Сильные стороны.

- В статье представлены актуальные, применимые на практике результаты. Авторы предоставили пакет Pytorch, который позволяет получить Momentum-аналог любой модели ResNet с небольшим дообучением, что является значимым результатом в задачах fine-tuning, для которых количество памяти является узким местом.
- Все выводы подкреплены обширными теоретическими и экспериментальными исследованиями. В частности, показано преимущество Momentum ResNet над ResNet и существующими инвертируемыми архитектурами, благодаря аппроксимирующей способности модели, возникающей при добавлении импульса.

3. Слабые стороны.

- Недостаточно подробное для воспроизведения описание эксперимента с облаками точек, но предоставлен код эксперимента, в котором есть все нужные гиперпараметры.
- В ключевых экспериментах о сравнении качества и потребляемой памяти Momentum ResNet и ResNet представлены результаты только для ResNet101 и его Momentum-аналога для различных наборов данных. Очень не хватает аналогичного эксперимента для ResNet различной глубины хотя бы на одном наборе данных.
- Статья содержит немало орфографических и лингвистических ошибок, затрудняющих восприятие.

4. Качество текста.

За исключением орфографических и лингвистических ошибок, затрудняющих понимание, статья написана последовательно, очень доходчиво и подробно.

5. Воспроизводимость.

К статье прилагается репозиторий на github с открытым кодом, а именно предоставлен пакет для PyTorch, который позволяет получить Momentum-аналог любого ResNet (также в текущей версии можно получить Momentum-аналог трансформера и, скорее всего, полученный метод можно будет применить к любой модели, содержащей residual блоки). Все ключевые эксперименты описаны с подробным описанием гиперпараметров обучения и воспроизводимы, кроме эксперимента с облаком точек, но для него имеется код со всей нужной информацией.

6. Дополнительные комментарии, предложения по улучшению и вопросы к авторам.

Следующие эксперименты могут сделать экспериментальное исследование в статье более полным:

- Эксперимент по исследованию качества и количества потребляемой памяти Momentum ResNet по сравнению с ResNet для моделей различной глубины, а не только для ResNet101.
- Эксперимент по исследованию времени обучения Momentum ResNet по сравнению ResNet (насколько изменение forward rule модели замедляет/ускоряет время обучения?)

Также исследования предложенного метода можно продолжить, проведя эксперименты из статьи для других архитектур, содержащих residual блоки (например, для трансформеров).

7. Оценка статьи по критериям НИПСа.

- Оценка: 9 из 10
- Уверенность: 5 из 5