

GradInit: Learning to Initialize Neural Networks for Stable and Efficient Training

Chen Zhu, Renkun Ni, Zheng Xu, Kezhi Kong, W. Ronny Huang, Tom Goldstein

Автор рецензии: Михаил Ким, 182

В статье предложен автоматический метод инициализации весов, метод учитывает специфику оптимизатора, величину шага и данные, но при этом является model-agnostic. Предложенный метод инициализации стабилизирует процедуру обучения и ускоряет сходимость.

Сильные стороны:

1. Предложен эффективный алгоритм решения оптимизационной задачи
2. Предложен автоматический метод подбора гиперпараметров
3. С помощью большого количества экспериментов с разными моделями было показано, что:
 1. Метод уменьшает дисперсию градиентов
 2. Ускоряет сходимость
 3. Позволяет избавиться от различных методов стабилизации обучения: warmup schedule, normalization layers

Актуальность

С усложнением моделей и методов в Deep Learning часто возрастает значимость инициализации, а также сложность процесса обучения. Поэтому эффективный алгоритм инициализации крайне важен для развития области в целом. Также стоит отметить, что большинство предшествующих методов не учитывают специфику задачи (данные) и(или) специфику оптимизатора и трудно применимы по причине большой вычислительной сложности, предложенный же метод лишён этих недостатков.

Слабые стороны:

1. Слабо исследованы случаи, когда метод неэффективен

2. Эксперименты покрывают небольшой набор задач: классификация и машинный перевод
3. Не исследована применимость для GAN, Inverse RL и других моделей, в которых стабильность обучения критична

Идеи в статье изложены доходчиво, однако некоторые графики довольно сложно разглядеть. Код метода доступен, в статье перечислены гиперпараметры, необходимые для воспроизведения экспериментов.

Оценка: 8

Уверенность: 5