

Название статьи (авторы статьи): Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time (<https://arxiv.org/pdf/2203.05482.pdf>)

Автор обзора-рецензии: Иванов Семен

При проведении исследования, мы предлагаем ответить вам на следующие вопросы:

1. Опишите суть работы в паре предложений, выделите ее основной вклад.

Предлагается ансамблировать модели, обученные с разными гиперпараметрами, но одной и той же инициализацией, путем усреднения их весов. Таким образом, мы получим ту же модель с точки зрения вычислительных затрат, но более устойчивую и лучшего качества в сравнении с просто лучшей моделью на валидации.

В статье рассматриваются разные способы отбора моделей для ансамбля и подбора весов для их усреднения

2. Когда написана работа? Опубликована ли она на какой-то конференции? Кто ее авторы, есть ли у них другие схожие работы? Подумайте как авторы пришли к идее статьи -- может быть это прямое улучшение их предыдущей работы, может это похоже на случайную находку

<https://icml.cc/media/icml-2022/Slides/16990.pdf>

Работа опубликована весной 2022 и принята на ICML2022 в качестве доклада

У работы целых 11 авторов и заявлено, что это совместная работа шести организаций: University of Washington, Columbia University, Google Brain, Meta AI и Tel Aviv State University. Не совсем понятно, каким образом они объединились, но большинство из авторов работают или получают PhD в University of Washington.

Если посмотреть на прошлые публикации авторов, то можно выделить следующие основные направления: оптимизация нейронных сетей и исследование их свойств в общем; cv, image-text

3. Какие из статей в списке ссылок (или почему-то не из списка, hello Mr. [Schmidhuber](#)) оказали наибольшее влияние на данную работу? Можно ли выделить какие-то 1-3 статьи, которые можно назвать базовыми для этой работы? Опишите в чем связь с этими работами (без математики, просто суть).

<https://arxiv.org/abs/2008.11687> "What is being transferred in transfer learning?" – в данной статье был получен результат, что pre-trained сети дообученные с разными гиперпараметрами остаются примерно в одном "бассейне" на плоскости функции потерь и лежат достаточно близко с точки зрения значения весов модели. Авторы ссылаются на данный текст, чтобы обосновать адекватность усреднения весов разных моделей с одинаковой инициализацией

<https://arxiv.org/abs/1803.05407> "Averaging Weights Leads to Wider Optima and Better Generalization" – схожая идея, но здесь усредняются веса одной модели, но на разных шагах обучения; в текущей статье же усредняются веса разных моделей с одинаковой pre-trained инициализацией

Ну и множество старых работ про то, что ансамблирование моделей помогает улучшить качество. Например, [https://link.springer.com/chapter/10.1007/3-540-45014-9\\_1](https://link.springer.com/chapter/10.1007/3-540-45014-9_1)

4. Кто цитирует данную статью? Есть ли у этой работы прямые продолжения, которые стоит прочесть тем, кто заинтересовался этой работой?

У этой работы есть прямое продолжение: <https://www.mdpi.com/2673-2688/3/4/48> – здесь проверяют возможность усреднения весов моделей, обученных на разных подмножествах данных (если вкратце, то качество на CIFAR10 получилось улучшить, но результаты стали менее стабильными с точки зрения соответствия улучшению на валидации улучшению на тесте)

Также есть ряд статей, которые цитирую данную работу. Например, здесь предложенный метод применяют к задаче классификации кожного рака:

<https://www.sciencedirect.com/science/article/pii/S0959804922004129#bib22>

5. Есть ли у работы прямые конкуренты (которые, например, вышли одновременно с работой или еще по каким-то причинам не вошли в предыдущие два пункта)? Опишите как соотносится данная работа с этими конкурентами (без математики, просто суть).

Не нашел

6. Опишите сильные, на ваш взгляд, стороны работы. Стоит обратить внимание на корректность утверждений в работе, значимость и новизну вклада, актуальность для исследовательского сообщества, понятность текста и воспроизводимость результатов (чтобы хорошо это сделать полезно набраться контекста из предыдущих пунктов: изучить prior work, продолжения, может даже посмотреть код/поговорить с хакером).

Из плюсов работы можно выделить:

Простая и легко реализуемая идея; наличие кода и весов полученных моделей

Вклад: простой способ улучшить качество модели, удалось улучшить SOTA

7. Опишите слабые, на ваш взгляд, стороны работы, обращая внимание на те же моменты, что и в предыдущем пункте.

Рассматривается только задача классификации

9. Попробуйте на основе результатов статьи предложить исследование, не проведенное к текущему моменту, или идею применения в индустриальных приложениях.

Можно попробовать иные способы взвешивания, например, обратно пропорционально лассо или норме градиента. Еще можно попробовать заранее отфильтровывать плохие модели

Еще можно попробовать “тупой” learned soup, где мы считаем, что выходы модели, полученной линейной комбинацией весов, – это просто линейная комбинация выходов отдельных моделей (возможно получится менее ресурсозатратным способом подобрать веса)

10. Может вы просто еще что-то любопытное про статью найдете =)