mixup: BEYOND EMPIRICAL RISK MINIMIZATION

Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz

Содержание и вклад

Авторы статьи предложили метод тренировки моделей, при котором модель берет на вход не просто элемент из набора данных и соответствующий класс (авторы статьи используют onehot encoding), а линейную комбинацию двух произвольных элементов и соответственно классов. Таким образом, по сути, mixup -- аугментация. При этом авторы наблюдают повышение обобщающей способности модели, снижение влияния неправильных классов в наборе данных, повышение устойчивости к атакам и более стабилизированное обучение ганов.

Сильные стороны

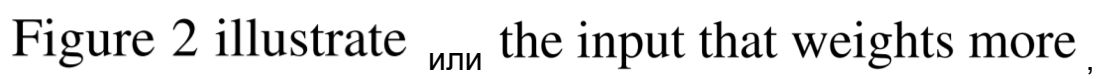
- Предложена альтернатива аугментациям, не зависящая от набора данных
- Сам метод очень прост в имплементации
- Все гиперпараметры для воспроизведения результатов указаны в статье
- Очень много экспериментов как на разных датасетах (включая даже разные типы данных, например изображения и звуки), так и для разных параметров, причем для разбора разнообразных предположений (запоминание классов, устойчивость)
- Авторы исправили все недочеты, на которые им указали рецензенты
- SOTA

Слабые стороны

- Не совсем понятно, насколько актуальна задача предсказания элементов in-between data
- Хотя в финальной версии статьи есть теоретическое обоснование того, как именно можно интерпретировать переход к миксапу (vicinal distribution), не очевидно, почему от этого должно стать лучше

Насколько хорошо написана статья

Статья написана понятным языком, все части связаны, легко следить за ходом мысли. Есть пара орфографических ошибок вроде



однако они не затрудняют прочтение.

Воспроизводимость

Воспроизводимость является одной из сильнейших сторон этой статьи, так как сама идея выражается буквально двумя строчками кода, а все гиперпараметры указаны.

Дополнительные комментарии

Мне статья очень понравилась, было понятно и просто читать. Даже матчасть, единственное, что может показаться неочевидным человеку, который плох в математике, на самом деле была совсем несложной и даже интересной.

Пожалуй, единственное, что мне показалось немного странным, был упор на то, что они улучшают качество моделей на пространстве между элементами набора данных. Когда речь идет конкретно о борьбе против атак, такие аугментации разумны, но просто улучшать качество в пространстве между данными мне не кажется актуальной задачей. К тому же был включен график, который демонстрировал ровно то, что у них, по сравнению с обычным обучением, все лучше на этом пространстве, что было и так очевидно. Но на самом деле это мелочь, хорошо, что они так подробно описывают все происходящее.

Рецензии с openreview

Оценки были 6-7-7.

Рецензенты указывали на нехватку бейзлайнов, теоретических обоснований местами, отсутствие кода и другие, более конкретные небольшие замечания; однако авторы ответили на все вопросы, значительно дополнив статью.

Оценка по шкале НИПС

Оценка 9, уверенность 4