

## Рецензия на Classifier-Free Diffusion Guidance

Рахманов Сергей, БПМИ192

Classifier-Free Diffusion Guidance – статья, которая вышла в сентябре 2021 года. В ней описывается принцип, который позволяет генерировать с заданным условием данные при помощи диффузионной модели без обучения дополнительного классификатора. Эта работа внесла достаточно большой вклад в развитие глубинного обучения, нашлось много применений описанной идеи, среди которых самое известное – качественные Text2Image модели на основе диффузии.

На момент выхода статьи самыми передовыми методами, реализующие генерацию изображения по текстовому описанию являлись DALL-E (OpenAI) и CogView (Ding et al.). Это авторегрессионные модели, и они не основывались на диффузии. Но тогда генеративные диффузионные модели уже существовали (Nichol & Dhariwal), которые способны создавать солидные по качеству изображения, сравнимые по метрикам со StyleGAN2. В той же статье от Nichol и Dhariwal предложили метод для условной генерации, подразумевающий дополнительное обучение классификатора. Это работающий, но довольно неуклюжий и затратный метод.

Спустя несколько месяцев после выхода статьи о генеративных диффузионных моделях на Openreview публикуется статья Classifier-Free Diffusion Guidance от Jonathan Ho и Tim Salimans. Эта работа проходит на воркшоп NeurIPS 2021. Оба автора работают в Google Brain, являются PhD. Jonathan Ho получил степень в UC Berkeley, Tim Salimans – в Erasmus University Rotterdam. Они и ранее исследовали глубинное обучение: обучение без учителя, обучение с подкреплением, различные генеративные модели. В последнее время их работа сконцентрирована на исследовании диффузионных моделей. У Ho индекс Хирша равен 26, у Salimans – 29.

У работы есть достаточно много сильных сторон. Предлагаемая идея простая в реализации, элегантная и практичная. В статье подробно описан механизм работы идеи, представлены схема на псевдокоде и математический аппарат, стоящий за концептом. Также авторы провели серию экспериментов, которые наглядно показывают результаты применения guidance с различными значениями без классификатора.

Главным минусом статьи является отсутствие прилагающегося кода. Нельзя попробовать воспроизвести эксперименты авторов, сверить результаты и провести свои опыты. Также в таблице сравнения метрик с другими методами авторы привели диффузионные модели с классификатором только с одним показателем guidance.

Идея, предложенная в статье, нашла широкое применение в различных последующих исследованиях, связанных с диффузионными моделями. Самое первое и известное использование – это генерация картинок по тексту. Imagen (Google Brain), DALL-E 2 (OpenAI), Stable Diffusion (CompVis) используют classifier-free guidance. Кроме того, в соответствующих статьях имеются отдельные абзацы, посвященные этой технике. Примечательно, что авторы Classifier-Free Diffusion Guidance являются соавторами Imagen, и в целом это объясняет, например, отсутствие кода у статьи.

Помимо Text2Image, бесклассификаторный guidance нашел применение и в других областях: начиная от дифференциально приватных диффузионных сетей (Dockhorn et al.), заканчивая использованием диффузионных моделей в мета-обучении (Nava et al.). Всего у статьи около ста цитирований.

Некоторые моменты авторы в своей статье не раскрыли, и можно провести исследования. Например, в работе указан только положительный guidance, когда мы хотим усилить влияние условия для генерации. Но что, если наоборот штрафовать повышение условной вероятности (например, переставив  $w$  и  $1-w$  в формуле)? Будем ли мы получать семантически обратные изображения? Также можем ли мы перемешать несколько условий с разными весами?

Кроме того, в работа показана только работа с изображениями. Можно попробовать применить classifier-free guidance в диффузионных моделях, генерирующие текст (DiffusionLM) и звук (DiffWave).