

МОДИФИЦИРОВАННЫЙ МЕТОД ОЦЕНКИ КАЧЕСТВА ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ НЕЙРОННЫХ СЕТЕЙ

© 2020 А. Д. Обухов[✉]

*Тамбовский государственный технический университет,
ул. Советская, 106, 392000 Тамбов, Российская Федерация*

Аннотация. Актуальной проблемой является оценка эффективности и качества произвольных генеративно-сопоставительных нейронных сетей (GAN). В настоящее время приемлемые результаты получены только для сетей, генерирующих изображения определённого формата и размеров. Остальные типы и структуры данных возможно оценивать исключительно экспертным методом, недостатком которого является субъективность, низкая производительность, невозможность автоматизации обработки больших объемов информации. В ходе анализа существующих подходов к оценке эффективности GAN определены наиболее распространенные метрики, среди которых необходимо отметить Inception Score и Fréchet Inception Distance. Однако, данные метрики не могут быть использованы для оценки GAN, генерирующих объекты, отличных от изображений определённого формата. Это обусловлено спецификой нейронных сетей, используемых для расчета данных метрик. Поэтому в работе предлагается модифицированный метод оценки качества GAN, основанный на метрике Inception Score, отличающийся использованием произвольного классификатора для расчета расстояния Кульбака-Лейблера, что обеспечивает возможность анализа качества генерации произвольных объектов. Представлено математическое описание методов оценки качества GAN и внесенных модификаций. Проведены практические эксперименты на нескольких известных наборах данных: графических (MNIST) и численных (Human Activity Recognition Using Smartphones, Epileptic Seizure Recognition). Для каждого набора осуществлена апробация разработанного модифицированного метода оценки GAN. Полученные результаты подтверждают возможность применения метода для оценки произвольных наборов данных. Использование полученных теоретических и практических результатов при реализации и обучении GAN позволит повысить качество сгенерированных нейронными сетями образцов, автоматизировать процесс их оценки.

Ключевые слова: машинное обучение, нейронные сети, генеративно-сопоставительные сети, оценка качества нейронных сетей, качество генерации информационных объектов, Inception Score, Fréchet Inception Distance.

ВВЕДЕНИЕ

Технологии машинного обучения и нейронные сети значительно расширили возможность автоматизации процессов анализа, обработки и генерации информации. Одним из активно развивающихся направлений является разработка и применение ге-

неративно-сопоставительных нейронных сетей (GAN). Процедура их обучения заключается в «соревновании» между генератором (создающим новые информационные объекты) и дискриминатором (оценивающим эти сгенерированные объекты на правдоподобность или наличие определенных признаков) [1]. Данные сети успешно применяются для генерации изображений, видео, текста.

Одной из актуальных задач при реализации и применении GAN является проблема

[✉] Обухов Артем Дмитриевич
e-mail: obuhov.art@gmail.com



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.

оценки качества сгенерированных данных. Одним из первых способов ее решения является экспертная оценка сгенерированных объектов, насколько они соответствуют реальным. С другой стороны, такой подход занимает много времени, не может использоваться в процессе обучения автоматически, отличается субъективностью и полностью зависит от квалификации эксперта.

Поэтому большую актуальность имеет задача определения качества GAN и сгенерированных с ее помощью объектов. В данном исследовании ставится следующая задача: осуществить анализ существующих методов оценки качества GAN и модифицировать их с целью использования в произвольных GAN, работающих с различными типами и структурами данных.

1. АНАЛИЗ СОСТОЯНИЯ ВОПРОСА ОЦЕНКИ КАЧЕСТВА GAN

Применение GAN позволяет решить широкий ряд задач по генерации и обработке информации. Например, повысить качество изображений и видео [2], сформировать фотореалистичные изображения на основе некоторых правил или признаков [3]. В сфере робототехники [4] нейронные сети данного типа применяются как в машинном зрении (для обработки и повышения качества), так и для решения задач планирования и прогнозирования, управления. GAN также используются для обработки изображений для последующего анализа в различных предметных областях, например, в медицине [5] или физике [6]. Технология GAN постоянно развивается, что приводит к расширению спектра решаемых задач.

Существенным недостатком GAN является высокая сложность их обучения. Так как в процессе обучения идет «соревнование» двух нейронных сетей, задача перестает быть поиском экстремума функции потерь (например, MSE), а превращается в поиск седловой точки [7]. Также, если в процессе обучения дискриминатор станет слишком эффективным, он будет возвращать значения очень близкие к 0 или к 1, так что генератор будет

испытывать трудности в чтении градиента. Наоборот, если генератор будет превосходить дискриминатор, он будет постоянно использовать недостатки дискриминатора и формировать неправильные выходные данные. Для устранения этой проблемы проведена значительная работа по повышению устойчивости процесса обучения GAN [8].

Рассматривая подходы к оценке GAN, необходимо отметить работу [9], которая содержит сравнительный анализ 24 количественных и 5 качественных метрик. Несмотря на такое количество различных подходов, на данный момент нет единого мнения относительно того, какая метрика лучше всего отражает качество моделей GAN и должна использоваться для объективного сравнения моделей.

Необходимо отметить, что центральное внимание в GAN и оценке их эффективности уделяется измерению качества сгенерированных изображений [10]. Однако, как было отмечено выше, возможности GAN не ограничены лишь генерацией и преобразованием изображений. Их применение возможно для формирования и обработки различных информационных объектов.

Среди существующих метрик эффективности GAN можно выделить две ключевых, получивших наибольшее распространение: Inception Score (IS) и Fréchet Inception Distance (FID). Кратко рассмотрим особенности каждой из них.

Метрика Inception Score предложена авторами работы [11] и является одним из способов объективной оценки качества сгенерированных изображений, что подтверждается многочисленными исследованиями. IS основана на использовании предварительно обученной модели нейронной сети глубокого обучения Inception v3 [12]. Данная модель используется для классификации сгенерированных изображений, после чего оценивается вероятность принадлежности изображений каждому из 1000 классов. Прогнозы суммируются и образуют итоговую оценку IS. Начальная оценка принимает значения от 1 до N, где N – количество классов.

Метрика FID [13] заключается в сравнении признаков сгенерированных и реальных изображений. В FID также используется обученная сеть Inception v3, однако, с исключением последнего слоя, что позволяет использовать для оценки не метки классов, а специфические признаки изображений, полученные из значений функций активации предпоследнего слоя модели. Для реальных и сгенерированных изображений рассчитывается многомерное нормальное распределение на основе среднего значения и ковариации активаций предпоследнего слоя. Расстояние между двумя распределениями определяется как FID. Низкое значение FID соответствует изображениям высокого качества и наоборот. Добавление шума и искажение изображений повышает FID, что подтверждает зависимость между величиной FID и качеством изображений.

Анализ показал, что существенным ограничением применения метрик IS и FID для оценки GAN является их ориентирование на анализ изображений. Для черно-белых изображений и цветных изображений произвольного размера применение IS и FID возможно за счет преобразования исходных данных. Однако, подобные преобразования вносят искажения в структуру данных при изменении количества цветовых каналов и масштабировании.

Для объектов, представленных векторами целых или вещественных значений произвольной размерности, использование IS и FID невозможно. Нейронная сеть Inception V3, применяемая для расчета IS и FID, не поддерживает входные данные произвольных форматов и ориентирована на анализ изображений размером от 75×75 до 299×299 пикселей. Даже ее модификация с добавлением дополнительных слоев на входе не позволит получить корректный результат: в структуре Inception V3 присутствует большое число сверточных слоев Convolution, MaxPooling и других, которые не обнаружат в исходных данных тех признаков, что можно выявить на изображениях. Следовательно, такая сеть не сможет сформировать корректный выходной вектор признаков. Классификация данных с

использованием такой сети также не будет оптимальной.

В ходе анализа специфики метрик IS и FID, а также их программных реализаций, получено, что модернизация метрики FID для оценки произвольных данных будет затруднена следующим фактором: необходимо обучить нейронную сеть-классификатор с заданным количеством классов и распределить исходные данные по ним. Если данные не размечены (не маркированы), то необходимо использовать методы кластеризации или автоэнкодеры. Для извлечения вектора признаков в первом случае необходимо удалить последний слой (что для текущих реализаций методов кластеризации, например, K-means, затруднено), во втором — использовать значение промежуточного слоя между декодером и энкодером, которое не во всех случаях может однозначно характеризовать признаки объектов.

Метрики IS и FID успешно реализованы в рамках различных библиотек (Keras, Tensorflow, Pytorch и т.д.) [14]. Однако, данные методы позволяют осуществить анализ лишь цветных изображений, так как именно на них обучена сеть Inception V3, используемая при расчете IS и FID. Кроме того, ряд исследователей, например [9, 10], отмечают, что IS недостаточно совершенна и обладает рядом существенных недостатков, что подтверждает актуальность задачи дальнейшего исследования и улучшения метрик, повышения их объективности, нахождения новых подходов к их использованию для решения задач оценки качества нейронных сетей.

2. ПОСТАНОВКА ЗАДАЧИ

Пусть задан вектор исходных данных $X = (x_1, \dots, x_n)$ произвольной размерности XN , для которого необходимо получить выходной вектор $Y = (y_1, \dots, y_n)$ такой же размерности. Для решения этой задачи требуется выбрать структуру и обучить GAN NN_{GAN} , осуществляющую отображение $X \rightarrow Y$, причем, если исходные данные принадлежат некоторому классу ($X \in Z_1$), то и выходные данные также входят в него: $Y \in Z_1$. Для получения опти-

мального результата решения задачи необходимо определить такую структуры нейронной сети, при которой метрика ее качества достигает наилучшего значения:

$$IS(NN_{GAN}) \rightarrow \max, \quad (1)$$

где IS — расчет метрики IS для произвольной сети NN_{GAN} .

Для определения метода расчета метрики IS для произвольных векторов данных X, Y необходимо разработать метод, позволяющий определить параметры нейронной сети NN_{GAN} , а также осуществить оценку качества сгенерированных образцов информационных объектов.

3. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

В рамках данного исследования предлагается разработать метод оценки качества GAN, основанный на модифицированной метрике IS и использовании вместо Inception V3 нейронной сети классификации (кластеризации) с заданными параметрами и архитектурой. Сначала рассмотрим формализацию расчета классической метрики IS.

Высококачественная GAN должна генерировать изображения x , относящиеся с высокой условной вероятностью $p(y|x)$ к определенному классу y . При соблюдении этого требования суммарная вероятность всех изображений будет иметь низкую энтропию. Также необходимо обеспечить разнообразие генерируемых изображений, а не только соответствие их заданному классу. Для выполнения этого условия предельный интеграл должен обладать высокой энтропией:

$$\int p(y|x = GAN(z)) dz. \quad (2)$$

Получаем два ключевых условия, обеспечивающих высокое качество GAN. Для их объединения разработчики метрики IS предлагают вычисление расстояния Кульбака-Лейблера [15] $KL(C||M)$, где C — условное распределение, M — предельное распределение.

Вычисление начальной оценки для набора сгенерированных изображений включает определение условной вероятности для каж-

дого изображения $p(y|x)$. Далее рассчитывается предельная вероятность как среднее значение условных вероятностей для изображений в группе $p(y)$. Тогда для расчета IS для каждого изображения x необходимо вычислить KL_x по следующей формуле:

$$KL_x = p(y|x) \cdot (\log(p(y|x)) - \log(p(y))). \quad (3)$$

Затем KL_x суммируется по всем изображениям и усредняется по всем классам N . Полученное среднее значение KL_i используется для получения итоговой оценки:

$$IS = \frac{\sum_{i=1}^N e^{\overline{KL}_i}}{N}. \quad (4)$$

Внесем следующие модификации, которые позволят применять метрику IS для произвольных исходных данных. Для этого необходимо осуществить переход от сети Inception V3 к произвольному классификатору для вычисления IS при определении $p(y|x)$ в формулах (2) и (3). Возможны два варианта:

Для реальных данных существуют метки классов, всего N классов. Обучается нейронная сеть классификации NN_C , выполняющая отображение $X \rightarrow \{y_k | k = 1..N\}$. Тип сети зависит от исходных данных: для изображений используются последовательность сверточных слоев, для произвольных векторов с численными значениями — плотные многослойные сети, для классификации временных рядов или лексем — рекуррентные сети.

Для реальных данных не существует меток классов, либо представлены данные единственного класса ($N = 1$). Тогда задается произвольное число категорий (кластеров) $N > 1$, и обучается алгоритм кластеризации NN_C на основе К-средних, осуществляющий отображение $X \rightarrow \{y_k | k = 1..N\}$.

Таким образом, указанные модификации в рамках рассмотренного метода оценки качества GAN позволят применять метрику IS для произвольных наборов данных независимо от их структуры и типа.

4. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Для проверки адекватности предлагаемого метода проведен ряд исследований на двух

типах данных: в первом случае осуществляется генерация размеченного набора данных — изображений MNIST, во втором — исследуются неразмеченные датасеты Human Activity Recognition Using Smartphones (HAR, 7767 тренировочных наборов размерностью 561 значений, содержит сведения о человеческой деятельности с использованием набора данных смартфонов [16]) и Epileptic Seizure Recognition (ESR, 11500 записей со 179 значениями [17]). Схема эксперимента устроена следующим образом: подбирается приемлемая структура GAN, после чего осуществляется исследование зависимости значений модифицированной метрики IS от времени обучения (количества эпох). Дополнительно проверяется зависимость значений метрики от параметра N .

В ходе первого эксперимента построена многослойная сверточная GAN. Так как набор MNIST размечен на 10 классов, то $N = 10$. Обучение GAN осуществляем в течение 8 эпох на наборе из 50000 тренировочных изображений. Результаты представлены на рис. 1.

Из графиков IS и FID можно отметить, что они достаточно адекватно отражают прогресс в обучении (рис. 2): слева расположены

изначально нечеткие изображения с шумом (при IS = 5.13), справа — итоговый, более качественный результат (при IS = 7.12)

Второй эксперимент связан с генерацией численных данных на основе неразмеченных датасетов. Поэтому в соответствии с предлагаемым методом необходимо осуществить кластеризацию исходных данных по N категориям. Сначала рассмотрим обучение и оценку GAN для датасета HAR.

Экспериментальные исследования показали высокую метрику IS для GAN при генерации неразмеченных данных произвольной размерности. Для IS проведен ряд опытов с различным количеством N категорий: от 10 до 1500 (рис. 3). Процесс обучения GAN осуществлялся в течение 1000 эпох, каждые 10 эпох производился замер IS при различных N .

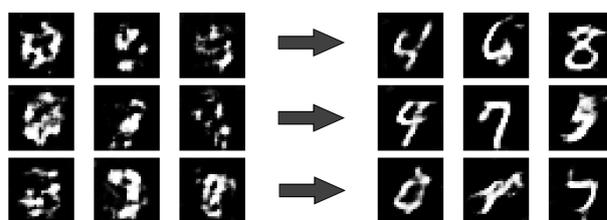


Рис. 2. Сгенерированные объекты набора данных MNIST (на 1 и 8 эпохах)

[Fig. 2. Generated objects of the MNIST dataset (at 1 and 8 epochs)]

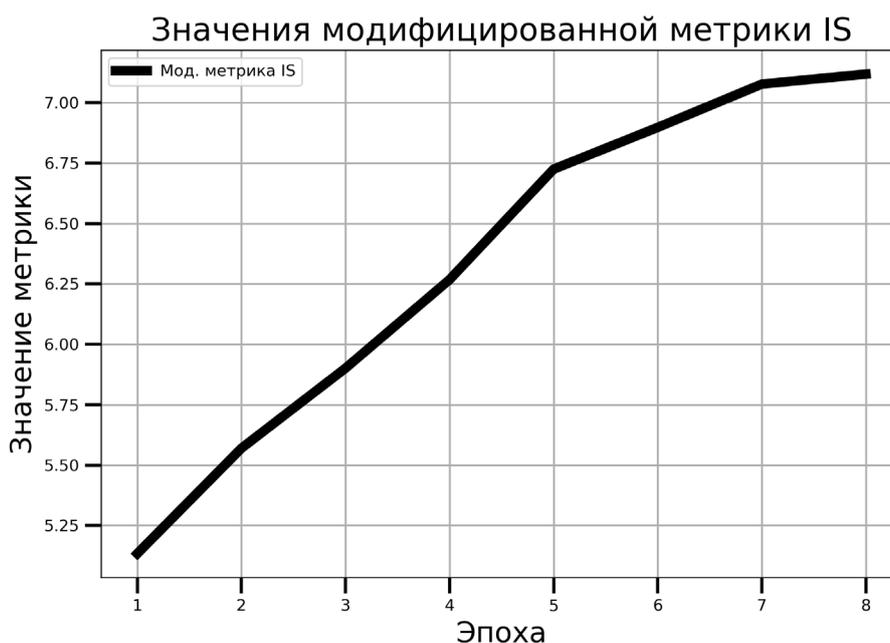


Рис. 1. Процесс обучения GAN на наборе данных MNIST

[Fig. 1. GAN training process on the MNIST dataset]

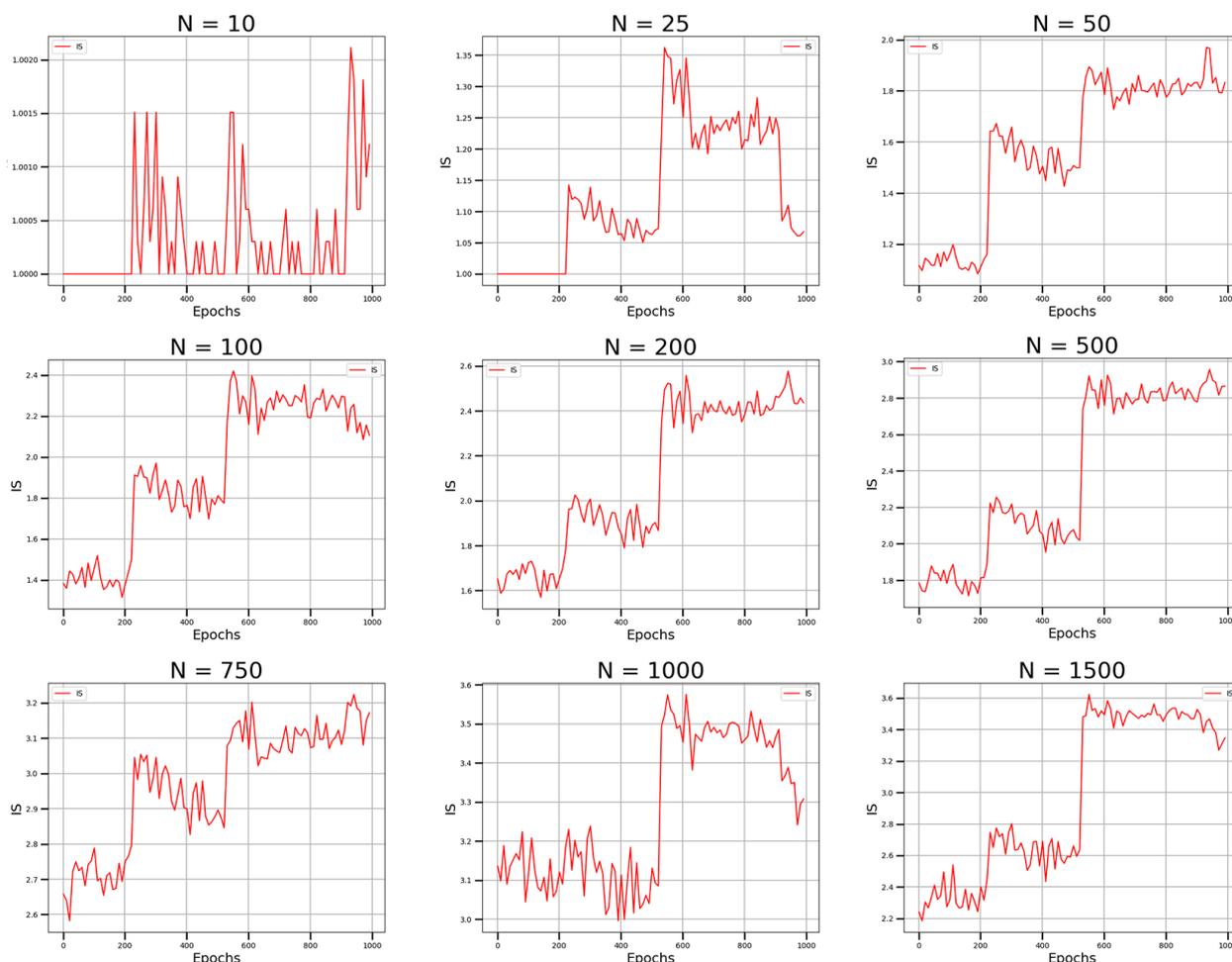


Рис. 3. Замеры IS (при различных N) для обучения GAN на наборе данных HAR
 [Fig. 3. Measurements of IS (with different N) for training GAN on the HAR dataset]

Это позволило избежать погрешности, возможной при повторном обучении GAN.

Полученные результаты привели к следующим выводам: при $N \leq 25$ процесс обучения оценивается некорректно. IS в этих диапазонах либо колеблется в пределах одной величины, либо падает, что не позволяет использовать ее при оценке качества GAN. С другой стороны, при $N \geq 25$ процесс обучения отражен корректно: с ростом количества итераций оценка IS растет вместе с уменьшением ошибок дискриминатора и генератора. Причем, корректность оценки IS не падает даже при $N > 1500$. Однако, отметим, что при $N > 500$ увеличение значения IS весьма незначительно. Это можно объяснить избыточностью кластеров, имеющих небольшие отличия между собой. Избыточность кластеров также приводит к увеличению вычислительной нагрузки. Таким образом, луч-

шие результаты получены при $N \approx 500$. Выдвинем следующую гипотезу: IS для произвольного неразмеченного набора данных $X = \{x_i\}$, где каждый x_i представлен XN элементами, определяется на основе распределения данных на N кластеров, где N выбирается на основе условия:

$$1 + \frac{XN}{20} \leq N \leq 1 + XN. \quad (5)$$

Полученные результаты удовлетворяют данному условию. Левая и правая часть двойного неравенства (5) получено на основе анализа экспериментальных данных (представленных ранее на рис. 3). Причем, правая часть неравенства обусловлена снижением вычислительной нагрузки и на точность определения IS влияния не оказывает.

Для проверки выдвинутой гипотезы о величине N дополнительно проведен эксперимент на наборе данных ESR (11500 записей со

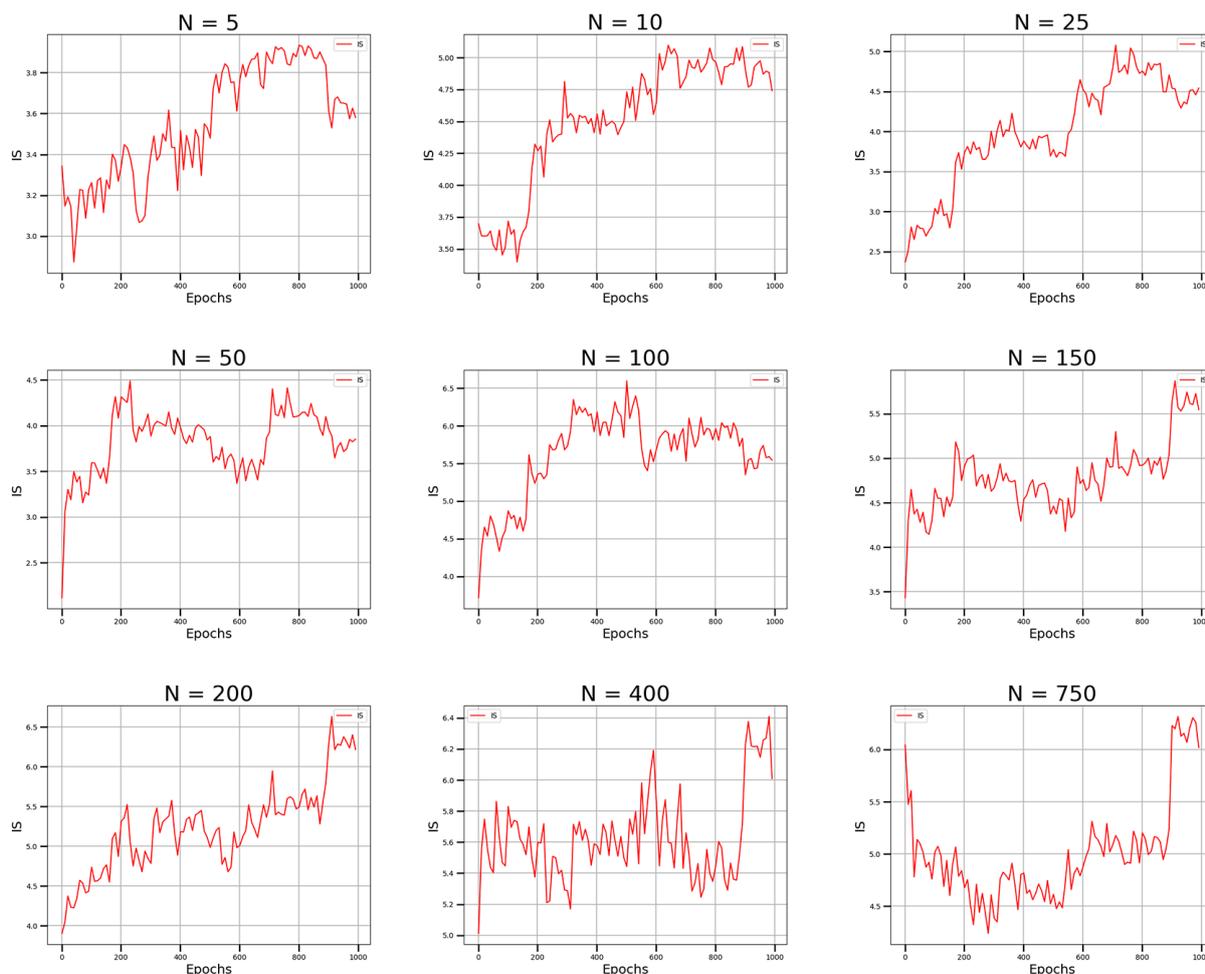


Рис. 4. Замеры IS (при различных N) для обучения GAN на наборе данных ESR [Fig. 4. Measurements of IS (with different N) for training GAN on the ESR dataset]

179 значениями). На основе условия (5) количество кластеров для неразмеченных данных должно находиться в интервале $10 \leq N \leq 180$. Результаты эксперимента (рис. 4) подтверждают выдвинутую гипотезу. При $N < 10$ величина оценки IS значительно падает. Оптимальная величина достигается при $N \rightarrow 180$.

Подведём итог по расчету IS для произвольных данных: размер N для размеченных данных соответствует количеству классов, для неразмеченных выбирается исходя из условия (5).

Таким образом, предложенный метод оценки качества GAN для произвольных данных был успешно апробирован в ходе трех экспериментов на датасетах MNIST, HAR и ESR. Для каждого эксперимента применялась модифицированная метрика IS.

В ходе первого эксперимента на размеченном наборе данных MNIST получены удов-

летворительные результаты. В ходе обучения GAN качество сгенерированных образцов значительно улучшено. Поведение метрики IS соответствует запланированному в поставленной задаче (1): ее значения увеличиваются с ростом качества образцов.

Второй и третий эксперимент показали, что для неразмеченных данных выбор количества кластеров N должен осуществляться в соответствии с условием (5). Это позволяет метрике IS соответствовать прогрессу в обучении GAN. Также необходимо отметить, что условием (5) задано необязательное ограничение на максимальную величину N , после которой увеличение количества кластеров нецелесообразно с точки зрения вычислительной нагрузки. Рекомендуется выбор числа кластеров приблизительно равный размеру вектора входных данных XN .

ЗАКЛЮЧЕНИЕ

В работе рассматривается задача оценки качества GAN для произвольных наборов данных. В настоящее время такие популярные метрики как IS и FID применяются исключительно для оценки качества сгенерированных изображений. В данном исследовании разработан метод оценки качества произвольных GAN на основе модификации метрики IS, что позволило применить их для различных наборов данных, а не только изображений. В ходе экспериментальных исследований доказана эффективность метода для размеченных и неразмеченных данных. Полученная модифицированная оценка IS может использоваться при сравнении качества произвольных GAN.

Новизна разработанного модифицированного метода оценки качества GAN заключается в использовании произвольного классификатора для расчета расстояния Кульбака-Лейблера, что обеспечивает возможность анализа качества генерации произвольных объектов.

Дальнейшие исследования в данном направлении будут связаны с разработкой методов модификации метрики FID с целью ее использования при оценке качества GAN.

Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ в рамках гранта Президента РФ МК-74.2020.9.

КОНФЛИКТ ИНТЕРЕСОВ

Автор декларирует отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

СПИСОК ЛИТЕРАТУРЫ

1. Сухань, А. А. Генеративно-сопоставительные нейронные сети в задачах определения трендов / А. А. Сухань // Московский экономический журнал. – 2019. – №. 6. – С. 180–191.
2. Wang, X. Esrgan: Enhanced super-resolution generative adversarial networks / X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy // Proceedings of the European Conference on Computer Vision (ECCV). – Munich, 2018. – P. 1–16.
3. Yi, R. ApdrawingGAN: Generating artistic portrait drawings from face photos with hierarchical GANs / R. Yi, Y. J. Liu, Y. K. Lai, P. L. Rosin // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2019. – P. 10743–10752.
4. Литвинов, О. В. Применение генеративно-сопоставительных сетей в робототехнике / О. В. Литвинов, А. В. Бахшиев, И. С. Фомин // Неделя науки СПбПУ (Санкт-Петербург, 19–24 ноября 2018 г.). – Санкт-Петербург, 2018. – С. 57–60.
5. Iqbal, T. Generative adversarial network for medical images (MI-GAN) / T. Iqbal, H. Ali // Journal of medical systems. – 2018. – Т. 42. – № 11. – P. 231.
6. Paganini, M. Accelerating science with generative adversarial networks: an application to 3D particle showers in multilayer calorimeters / M. Paganini, L. de Oliveira, B. Nachman // Physical review letters. – 2018. – Т. 120. – № 4. – P. 042003.
7. Nagarajan, V. Gradient descent GAN optimization is locally stable / V. Nagarajan, J. Z. Kolter // Advances in neural information processing systems. – 2017. – P. 5585–5595.
8. Heusel, M. Gans trained by a two time-scale update rule converge to a local nash equilibrium / M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter // Advances in neural information processing systems. – 2017. – P. 6626–6637.
9. Borji, A. Pros and cons of gan evaluation measures / A. Borji // Computer Vision and Image Understanding. – 2019. – Т. 179. – P. 41–65.
10. Shmelkov, K. How good is my GAN? / K. Shmelkov, C. Schmid, K. Alahari // Proceedings of the European Conference on Computer Vision (ECCV). – Munich, 2018. – P. 213–229.
11. Salimans, T. Improved techniques for training gans / T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen // Advances in neural information processing systems. – 2016. – P. 2234–2242.
12. Szegedy, C. Rethinking the inception architecture for computer vision / C. Szegedy, V. Van-

houcke, S. Ioffe, J. Shlens, Z. Wojna // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – P. 2818–2826.

13. Kokate, P. An Empirical Comparison of Generative Adversarial Network (GAN) Measures / P. Kokate, A. D. Joshi, P. S. Tamizharasan // Advances in Communication and Computational Technology. – Springer, Singapore, 2016. – P. 1383–1396.

14. Ramasubramanian, K. Deep learning using keras and tensorflow / K. Ramasubramanian, A. Singh // Machine Learning Using R. – Apress, Berkeley, CA, 2019. – P. 667–688.

15. Нагибин, С. Я. Выбор нейронных сетей для оценки и прогнозирования рисков

промышленной безопасности предприятий / С. Я. Нагибин, С. Е. Левин, Д. И. Лоскутов, В. В. Шилов // Национальная безопасность России: актуальные аспекты. – 2019. – С. 37–45.

16. Anguita, D. A public domain dataset for human activity recognition using smartphones / D. Anguita, A. Ghio, L. Oneto, X. Parra, J. L. Reyes-Ortiz // Esann. – Bruges, 2013. – P. 437–442.

17. San-Segundo, R. Classification of epileptic EEG recordings using signal transforms and convolutional neural networks / R. San-Segundo, M. Gil-Martín, L. F. D'Haro-Enríquez, J. M. Pardo // Computers in biology and medicine. – 2019. – Т. 109. – P. 148–158.

Обухов Артем Дмитриевич — канд. техн. наук, доцент кафедры «Системы автоматизированной поддержки принятия решений» Тамбовского государственного технического университета.

E-mail: obuhov.art@gmail.com

ORCID: <https://orcid.org/0000-0002-3450-5213>

DOI: <https://doi.org/10.17308/sait.2020.3/3044>

Received 04.09.2020

Accepted 30.09.2020

ISSN 1995-5499

A MODIFIED METHOD FOR ASSESSING THE QUALITY OF GENERATIVE ADVERSARIAL NEURAL NETWORKS

© 2020 A. D. Obukhov✉

*Tambov State Technical University
106, Sovetskaya Street, 392000 Tambov, Russian Federation*

Annotation. Evaluation of the efficiency and the quality of arbitrary generative adversarial neural networks (GAN) is a pressing problem. Currently, acceptable results have only been obtained for networks that generate images of a certain format and size. Other data types and structures can be assessed exclusively by an expert method, the disadvantage of which is their subjectivity, low productivity, and the impossibility of automating the processing of large amounts of information. The analysis of existing approaches to assessing the effectiveness of GAN identified the most common metrics, among which Inception Score and Fréchet Inception Distance should be noted. However, these metrics cannot be used to estimate GANs generating objects other than images of a certain format. This is due to the specifics of neural networks used to calculate these metrics. Therefore, the paper proposes a modified method for assessing the quality of the GAN, based on the Inception Score metric, which is characterized by the use of an arbitrary classifier to calculate the Kullback-Leibler distance, which makes it possible to analyse the generation quality of arbitrary objects. The article presents a mathematical description of the methods for assessing

the quality of the GAN and the modifications introduced. Practical experiments were carried out on several well-known datasets: graphic datasets (MNIST) and numerical (Human Activity Recog-

✉ Obukhov Artem D.
e-mail: obuhov.art@gmail.com

dition Using Smartphones, Epileptic Seizure Recognition) datasets. For each set, the developed modified GAN estimation method was tested. The obtained results confirm the possibility of using the method to evaluate arbitrary data sets. The use of the obtained theoretical and practical results in the implementation and training of GAN will improve the quality of samples generated by neural networks and automate the process of their assessment.

Keywords: machine learning, neural networks, generative adversarial networks, assessment of the quality of neural networks, the quality of generation of information objects, Inception Score, Fréchet Inception Distance.

CONFLICT OF INTEREST

The author declare the absence of obvious and potential conflicts of interest related to the publication of this article.

REFERENCES

1. *Anguita D., Ghio A., Oneto L., Parra X. & Reyes-Ortiz J. L.* A public domain dataset for human activity recognition using smartphones. In *Esann*. 2013, April. Vol. 3. P. 437–442.
2. *Borji A.* Pros and cons of gan evaluation measures. *Computer Vision and Image Understanding*. 2019. 179. P. 41–65.
3. *Heusel M., Ramsauer H., Unterthiner T., Nessler B. & Hochreiter S.* Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in neural information processing systems*. 2017. P. 6626–6637.
4. *Iqbal T. & Ali H.* Generative adversarial network for medical images (MI-GAN). *Journal of medical systems*. 2018. 42(11). P. 231.
5. *Kokate P., Joshi A. D. & Tamizharasan P. S.* An Empirical Comparison of Generative Adversarial Network (GAN) Measures. In *Advances in Communication and Computational Technology*. P. 1383–1396. Springer, Singapore.
6. *Litvinov O. V., Bahshiev A. V., Fomin I. S.* Primenenie generativno-sostyazatel'nyh setej v robototekhnike [Application of generative adversarial networks in robotics] *Nedelya nauki SPb-PU. Sankt-Peterburg*. 2018. P. 57–60. (in Russian)
7. *Nagarajan, V. & Kolter J. Z.* Gradient descent GAN optimization is locally stable. In *Advances in neural information processing systems*. 2017. P. 5585–5595.
8. *Nagibin S. Ya., Levin S. E., Loskutov D. I., Shilov V. V.* Vychor nejronnyh setej dlya ocenki i prognozirovaniya riskov promyshlennoj bezopasnosti predpriyatij [Selection of neural networks for assessing and predicting industrial security risks of enterprises]. *Nacional'naya bezopasnost' Rossii: aktual'nye aspekty*. 2019. P. 37–45. (in Russian)
9. *Paganini M., de Oliveira L. & Nachman B.* Accelerating science with generative adversarial networks: an application to 3D particle showers in multilayer calorimeters. *Physical review letters*. 2018. 120(4). 042003.
10. *Ramasubramanian K. & Singh A.* Deep learning using keras and tensorflow. In *Machine Learning Using R*. 2019. P. 667–688. Apress, Berkeley, CA.
11. *Salimans T., Goodfellow I., Zaremba W., Cheung V., Radford A. & Chen X.* Improved techniques for training gans. In *Advances in neural information processing systems 2016*. P. 2234–2242.
12. *San-Segundo R., Gil-Martín M., D'Haro-Enríquez L. F. & Pardo J. M.* Classification of epileptic EEG recordings using signal transforms and convolutional neural networks. *Computers in biology and medicine*. 2019. 109. P. 148–158.
13. *Suhan' A. A.* Generativno-sostyazatel'nye nejronnye seti v zadachah opredeleniya trendov [Generative adversarial neural networks in the problems of determining trends]. *Moskovskij ekonomicheskij zhurnal*. 2019. 6. P. 180–191. (in Russian)
14. *Szegedy C., Vanhoucke V., Ioffe S., Shlens J. & Wojna Z.* Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. P. 2818–2826.
15. *Wang X., Yu K., Wu S., Gu J., Liu Y., Dong C. & Change Loy C.* Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018. P. 1–16.

16. Yi R., Liu Y. J., Lai Y. K. & Rosin P. L. Ap-drawingGAN: Generating artistic portrait drawings from face photos with hierarchical GANs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019. P. 10743–10752.
17. Zhang Z., Yang L., Chen L., Liu Q., Meng Y., Wang P. & Li M. A generative adversarial network-based method for generating negative financial samples. International Journal of Distributed Sensor Networks. 2020. 16(2). P. 213–229.

Obukhov Artem D. — PhD in Technical Sciences, Associate Professor, Department of Automated Decision-Making Support Systems, Tambov State Technical University.

E-mail: obuhov.art@gmail.com

ORCID iD: <https://orcid.org/0000-0002-3450-5213>