

КРИТЕРИИ ОЦЕНКИ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ СЕТЕЙ ДЛЯ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ В МОДЕЛЯХ МАШИННОГО ОБУЧЕНИЯ

Аннотация. В статье представлен подход к оценке эффективности генеративно-состязательных сетей для предварительной обработки данных. Рассмотрены и классифицированы некоторые метрики оценки правдоподобности сгенерированных данных, перечислены распространенные методы оценки предсказательной силы модели, а также уделено внимание применимости модели машинного обучения для решения задач в предметной области.

Ключевые слова: генеративно-состязательные сети, оценка эффективности, машинное обучение, преобработка данных, искусственные нейронные сети.

Введение. Понятие эффективности — сложная философская категория, не имеющая однозначного определения. Одним из подходов к пониманию ее сущности заключается в выделении трех аспектов [1; 937]: реализация целей при приемлемом отношении затрат и результатов, соотношение затрат и результатов и достижение эффекта, которые, на наш взгляд, применительно к использованию генеративно-состязательных сетей (GAN) для предварительной обработки данных, представляют собой иерархическую систему, где:

- реализация целей при приемлемом отношении затрат и результатов — это приемлемость решения задачи предметной области, для которой разрабатывалась модель машинного обучения;
- соотношение затрат и результатов — метрики ее точности;
- достижение эффекта заключается в том, насколько правдоподобно были сгенерированы данные.

Проблема исследования. Возрастающая популярность генеративно-состязательных сетей для решения различных задач порождает вопрос о целесообразности их широкого применения, что делает оценку эффективности GAN крайне актуальной проблемой. Основной сложностью в этом вопросе является невозможность использования функции потерь при обучении генератора, и, следовательно, отсутствие способа объективно оценить ход обучения и качество модели по потерям. Это приводит к необходимости оценивать качество сгенерированных данных, предназначенных для решения задач в конкретной предметной области. Проблема в контексте вопроса предварительной обработки данных усугубляется тем, что большинство ее алгоритмов ориентировано на структурированные (табличные) данные, а разработка методики оценки эффективности GAN производилась с учетом специфики традиционной сферы их применения — изображений, что существенно ограничивает или делает невозможным применение некоторых показателей.

Исходя из вышесказанного, целью нашей работы является построение методики оценки эффективности применения генеративно-состязательных сетей для предварительной обработки данных в моделях машинного обучения (МО), для чего необходимо решить следующие задачи:

1. Рассмотреть существующие подходы к оценке эффективности.
2. Оценить применимость предлагаемых в них метрик для различных типов данных.

3. Сформировать трехуровневую систему критериев оценки эффективности применения генеративно-состязательных сетей для предварительной обработки данных.

Материалы и методы. Для изучения проблемы было рассмотрено 7 русских и 9 зарубежных источников по теме, которые представилось возможным разделить на три неравные группы: первые две, сопоставимые по количеству публикаций, посвящены теоретическому анализу методов оценки эффективности GAN [2-7] и проведению практических экспериментов с последующей оценкой результата [8-15]. В третьей же, представленной всего двумя работами [16-17], авторами производится усовершенствование существующих методов.

Результаты. Трехуровневую систему оценки эффективности GAN для предварительной обработки данных графически можно представить следующим образом (рис. 1)

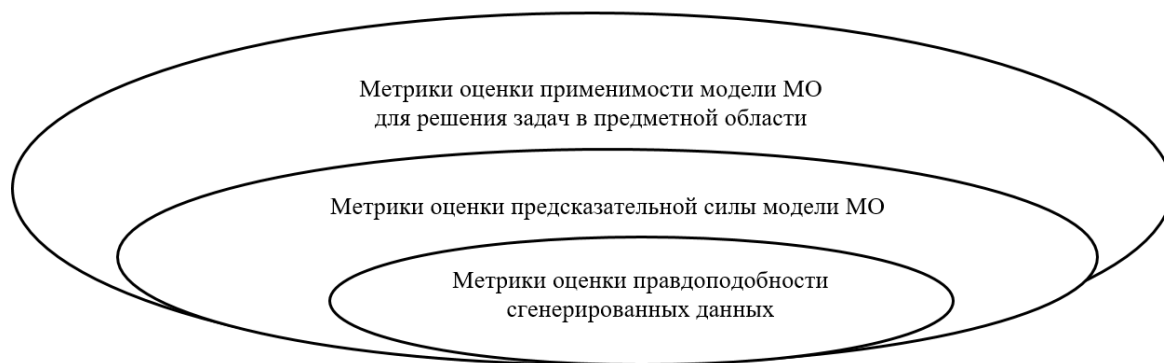


Рис. 1. Трехуровневая система оценки эффективности данных

Рассмотрим подробнее предлагаемые для каждого уровня метрики.

1. Метрики оценки правдоподобности сгенерированных данных

Все их многообразие мы разделили по двум основаниям классификации: по степени участия человека и по характеру оценки (табл. 1).

Таблица 1

Классификация метрик оценки правдоподобности сгенерированных данных

	Субъективные метрики	Субъективно-объективные метрики	Объективные метрики
Жестко классифицирующие метрики	Rapid Scene Categorization Preference Judgment	HYPE Neuroscore	
Коллективные метрики		Birthday Paradox Test (Парадокс дня рождения)	Generative Adversarial Metric (GAM) Normalized Relative Discriminative Score (NRDS)
Вычисляемые метрики			Inception Score (IS) Начальное расстояние Фреше (FID)

Самым простым и интуитивно понятным способом является классификация изображений респондентом на реальные и сгенерированные на основании визуальной оценки. Применение метода Rapid Scene Categorization заключается в максимально быстром выборе для

изображения соответствующей метки, а при применении Preference Judgment предлагается выбрать из двух сгенерированных на основе оригинала изображений наиболее близкое к нему.

Сравнительно новым подходом является оценка достоверности изображений через психофизиологическое исследование человеческого восприятия, когда аппаратура фиксирует бессознательную реакцию человека (Метрики HYPE и Neuroscore).

Также изрядной долей субъективности обладает метрика «Парадокс дня рождения», предполагающая ручную проверку предварительно отобранных автоматически пар изображений на дубликаты. Несмотря на то, что эти группы методов довольно эффективны при анализе изображений, перспективы их универсального использования выглядят весьма сомнительными в силу гораздо меньшей визуальной выразительности иных типов данных.

Для объективных метрик качество сгенерированных данных оценивается путем вычислений. Самыми часто используемыми из них являются Inception Score (IS) и начальное расстояние Фреше (FID), которые полагаются на предобученный на ImageNet классификатор. В основе первого метода лежит идея, что изображения должны строго принадлежать какому-либо классу, но при этом быть максимально разнообразными в его рамках. Второй метод заключается в сравнении признаков сгенерированных и реальных изображений. Исключение из Inception v3 последнего слоя позволяет оценивать специфические признаки изображений. 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.

Несмотря на то, что данные метрики используются для оценки изображений, было доказано [16; 104], что для метрики IS возможно использовать произвольный классификатор, что позволит оценивать различные типы данных, и аналогичная гипотеза была выдвинута в отношении метрики FID.

Что касается коллективных метрик, то метрика GAM заключается в организации соревнований между двумя GAN путем их обмена генераторами и дискриминаторами, а NRDS сравнивает несколько генеративно-состязательных сетей, базируясь на идее, что чем лучше сгенерированные образцы, тем больше эпох обучения будет необходимо, чтобы отличить их от реальных.

2. Метрики оценки предсказательной силы модели машинного обучения

Для оценки предсказательной силы моделей существует множество метрик, однако для сравнения GAN и других методов машинного обучения наиболее популярными являются следующие (табл. 2):

Таблица 2

Метрики оценки предсказательной силы модели машинного обучения

Для задач регрессии	Средняя абсолютная ошибка (MAE) Средняя квадратичная ошибка (MSE) Среднеквадратичная ошибка (RMSE) Нормированная среднеквадратичная ошибка (NRMSE)
Для задач классификации	Точность (Accuracy) Полнота (Recall) Специфичность (Specifity) F1 Score ROC-кривая Коэффициент Джини CAP-кривая (Cumulative Accuracy Profile)

3. Метрики оценки применимости модели машинного обучения для решения задач в предметной области

Целесообразность применения модели машинного обучения оценивается по соотношению результатов (качества модели) и затрат (необходимых для обеспечения ее работы ресурсов). Наиболее универсальные метрики оценки качества модели — это величина ошибки на тестовой выборке, производительность и внутренняя устойчивость модели, а также ее способность к обучению и самообучению. С точки зрения затрат можно отметить критерии: время работы и обучения модели, требования к емкости вычислительных ресурсов, объему информации, квалификации пользователей и финансовые затраты.

Заключение. Таким образом, существующие подходы к оценке эффективности GAN можно структурировать в систему в соответствии с ключевыми аспектами понятия «эффективность», что, на наш взгляд, позволит проводить более качественную и полную оценку и будет полезным для практического применения в исследованиях применимости генеративно-состязательных сетей. Вместе с тем, нельзя не отметить основную проблему данного подхода: большинство популярных методов оценки правдоподобности сгенерированных данных предназначено для изображений, поэтому поиск универсальных метрик для решения данной задачи, а также выявление и оценка силы связи между эффективностью на разных уровнях представляют собой дальнейшие направления для работы в целях совершенствования механизма оценки.

СПИСОК ЛИТЕРАТУРЫ

1. Штеле Е.А. К вопросу о понятии «эффективность» / Е.А Штеле, О.Б. Вечерковская. — Текст: электронный // Экономический анализ: теория и практика. — 2017. — № 5 (464). — URL: <https://cyberleninka.ru/article/n/k-voprosu-o-ponyatii-effektivnost> (дата обращения: 13.03.2024).
2. Гусева П.В. Оценка эффективности нейронной сети для анализа тональности русскоязычных текстов / П.В. Гусева. — Текст: непосредственный // Инновации. Наука. Образование. — 2021. — № 47. — С. 1526-1530.
3. Малахов Ю.А. Анализ и применение генеративно-состязательных сетей для получения изображений высокого качества / Ю.А. Малахов, А.А. Андросов, А.В. Аверченков. — Текст: непосредственный // Эргодизайн. — 2020. — № 4 (10). — С. 167-176.
4. Степанова А.А. Подходы к оценке эффективности нейросетевых технологий / А.А. Степанова. — Текст: непосредственный // Стратегия опережающего развития: сборник статей VI Международной школы молодых ученых в сфере экономики на Юге России, Сочи, 25–28 сентября 2019 года. — Сочи: АНО «Научно-исследовательский институт истории, экономики и права», 2019. — С. 113-115.
5. Alqahtani H. An analysis of evaluation metrics of GANs / H. Alqahtani, M. Kavakli, K. A. Gulshan. — URL: <https://clck.ru/3AX66f> (date of the application: 27.04.2024). — Text: electronic.
6. Borji A. Pros and Cons of GAN Evaluation Measures / A. Borji. — URL: [arXiv:1802.03446](https://arxiv.org/abs/1802.03446) (date of the application: 27.04.2024). — Text: electronic.
7. Borji A. Pros and cons of GAN evaluation measures: New developments / A. Borji // Computer Vision and Image Understanding. — Vol. 215. — 2022. — URL: <https://www.sciencedirect.com/science/article/pii/S1077314221001685> (date of the application: 27.04.2024). — Text: electronic.
8. Brownlee J. How to Evaluate Generative Adversarial Networks / J. Brownlee. — URL: <https://clck.ru/3AX5r8> (date of the application: 27.04.2024). — Text: electronic.

9. Hui J. GAN — How to measure GAN performance? / J. Hui. — URL: <https://jonathan-hui.medium.com/gan-how-to-measure-gan-performance-64b988c47732> (date of the application: 27.04.2024). — Text: electronic.
10. Загинайло М.В. Оценка эффективности различных методов обучения искусственных нейронных сетей / М. В. Загинайло, В.А. Фатхи. — Текст: непосредственный // Инновации. Наука. Образование. — 2021. — № 35. — С. 442-447.
11. Толстых А.А. Сравнение эффективности методов изменения скорости обучения искусственных нейронных сетей в различных задачах классификации / А.А. Толстых, А.Н. Голубинский. — Текст: электронный // МНИЖ. — 2022. — № 7-1 (121). — URL: <https://cyberleninka.ru/article/n/sravnenie-effektivnosti-metodov-izmeneniya-skorosti-obucheniya-iskusstvennyh-neyronnyh-setey-v-razlichnyh-zadachah-klassifikatsii> (дата обращения: 27.04.2024).
12. Фомина Е.С. Оценка эффективности применения специализированных нейронных сетей для повышения разрешения изображений, получаемых при дистанционном зондировании Земли / Е.С. Фомина. — Текст: электронный // Системы управления, связи и безопасности. — 2023. — № 3. — URL: <https://cyberleninka.ru/article/n/otsenka-effektivnosti-primeneniya-spetsializirovannyh-neyronnyh-setey-dlya-povysheniya-razresheniya-izobrazheniy-poluchaemyh-pri> (дата обращения: 27.04.2024).
13. Carrasco S. On the evaluation of Generative Adversarial Networks / S. Carrasco, S. Majchrowska. — URL: <https://towardsdatascience.com/on-the-evaluation-of-generative-adversarial-networks-b056ddcd3a> (date of the application: 27.04.2024). — Text: electronic.
14. Xu Q. An empirical study on evaluation metrics of generative adversarial networks / Q. Xu, G. Huang, Y. Yuan [et al.]. — URL: [arXiv:1806.07755](https://arxiv.org/abs/1806.07755) (date of the application: 27.04.2024). — Text: electronic.
15. Yang S. Evaluating the performance of generative adversarial network-synthesized periapical images in classifying C-shaped root canals / S Yang, K. Kim, E. Ariji [et al.]. //Sci Rep 13, 18038 (2023). — URL: <https://www.nature.com/articles/s41598-023-45290-1> (date of the application: 27.04.2024). — Text: electronic.
16. Обухов А.Д. Модифицированный метод оценки качества генеративно-сопоставительных нейронных сетей / А.Д. Обухов. — Текст: непосредственный // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. — 2020. — № 3. — С. 97-107.
17. Narteni S. Explainable Evaluation of Generative Adversarial Networks For Wearables Data Augmentation / S. Narteni, V. Orani, E. Ferrari [et al.]. — URL: <https://www.techrxiv.org/users/693109/articles/682929-explainable-evaluation-of-generative-adversarial-networks-for-wearables-data-augmentation> (date of the application: 27.04.2024). — Text: electronic.