

УДК 004.852

ПОСТРОЕНИЕ МОДЕЛИ ДЛЯ ПРОГНОЗИРОВАНИЯ ПРИКЛАДНЫХ ЗАДАЧ С ПРИМЕНЕНИЕМ ГЕНЕРАТИВНЫХ СОСТЯЗАТЕЛЬНОЙ СЕТИ

Сабитов¹ Б.Р., Шейшенов¹ Ж., Султанбеков² К., Айбек² к. Д.

¹КНУ им. Баласагына, ²КГТУ им. И. Раззакова

В статье рассматривается разработка и реализация предсказательной модели на основе усовершенствованной архитектуры условных генеративно-состязательных сетей (Conditional GAN) для прогнозирования временных рядов качества воздуха. На базе синтетических многокомпонентных данных, имитирующих реальные концентрации загрязняющих частиц PM_{2.5} и PM₁₀ с учетом трендов, сезонности и шумов, исследуются особенности состязательного обучения с применением LSTM-слоев и одномерных сверточных сетей (1D ConvNet). Авторы проводят сравнительный анализ стабильности, метрик ошибок моделей GAN и классической LSTM, а также предлагают комплекс стратегий для оптимизации гиперпараметров и предотвращения коллапса мод.

Ключевые слова: генеративно-состязательные сети (GAN), прогнозирование временных рядов, качество воздуха, нейронные сети LSTM, TimeGAN, обработка метеоданных, оптимизация гиперпараметров.

ТУЗУУЧУ КААРШЫЛ ТАРМАКТАРДЫ КОЛДОНУУ МЕНЕН КОЛДОНУЛГАН МАСЕЛЕЛЕРДИ БОЛЖОЛДОО ҮЧҮН МОДЕЛДИ ТҮЗҮҮ

Сабитов¹ Б.Р., Шейшенов¹ Ж., Султанбеков² К., Айбек² к. Д.

¹Ж. Баласагын атындагы КУУ, ²И. Раззаков атындагы КМТУ

Макалада абанын сапатынын убакыттар катарын болжолдоо үчүн шарттуу генеративдик-атаандаш тармактардын (Conditional GAN) өркүндөтүлгөн архитектурасынын негизиндеги божомолдоочу моделди иштеп чыгуу жана ишке ашыруу маселелери каралат. Тренддерди,

сезондуулукту жана ызы-чууларды эсепке алуу менен PM2.5 жана PM10 булгоочу бөлүкчөлөрүнүн реалдуу концентрациясын окшоштурган көп компоненттүү синтетикалык маалыматтардын базасында LSTM катмарларын жана бир өлчөмдүү чырмалышкан тармактарды (1D ConvNet) колдонуу аркылуу атаандаш окутуунун өзгөчөлүктөрү изилденет. Авторлор GAN жана классикалык LSTM моделдеринин туруктуулугуна, ката метрикаларына салыштырмалуу талдоо жүргүзүшүп, гиперпараметрлерди оптималдаштыруу жана модалардын кыйрашын (mode collapse) болтурбоо үчүн стратегиялардын топтомун сунушташат.

Баштапкы сөздөр: генеративдик-атаандаш тармактар (GAN), убакыттар катарын болжолдоо, абанын сапаты, LSTM нейрон тармактары, TimeGAN, метеомаалыматтарды иштетүү, гиперпараметрлерди оптималдаштыруу.

CONSTRUCTING A MODEL FOR FORECASTING APPLIED PROBLEMS USING GENERATIVE ADVERSARY NETWORKS

Sabitov¹ B.R., Sheishenov¹ Zh., Sultanbekov² K., Aibek² K.D.

¹KNU named after Balasagyn, ²KSTU named after. I. Razzakova

The article discusses the development and implementation of a predictive model based on an advanced Conditional GAN architecture for forecasting air quality time series. Using synthetic multicomponent data that simulates real-world PM2.5 and PM10 pollutant concentrations—accounting for trends, seasonality, and noise—the authors investigate the specifics of adversarial training utilizing LSTM layers and 1D Convolutional Networks (1D ConvNet). A comparative analysis of stability and error metrics is conducted between GAN and classical LSTM models, alongside a comprehensive set of strategies for hyperparameter optimization and prevention of mode collapse.

Keywords: Generative Adversarial Networks (GAN), time series forecasting, air quality, LSTM neural networks, TimeGAN, weather data processing, hyperparameter optimization.

Введение. Генеративно состязательная сеть (GAN) является алгоритмом машинного обучения, состоящий из двух конкурирующих нейросетей. Одна генерирует данные (например, изображения), а вторая пытается отличить подделку от оригинала.

Построено модель для прогнозирования качества воздуха с применением генеративных состязательных сетей (GAN) и рассмотрены

вопросы ее реализации с применением машинного обучения и нейронных сетей.

Для прогнозирования качества воздуха с использованием GAN построены временные ряды для предсказания концентрации частиц PM2.5, PM10. Для прогнозирования временных рядов использовано Conditional GAN технологии, а GAN использовалось для генерации данных для обучения модели.

Для улучшения и получения устойчивую и качественную модель для прогнозирования использовано генеративные состязательные сети. Данная технология помогают для получения реальных прогнозов качества воздуха, например, использование GAN для генерации данных.

Модель обучена на данных о качестве воздуха, например, используя открытые данные делает прогнозы и данных сгенерированных с GAN методом.

Создан модель для задачи предсказания значений прогноза на ближайшее время для показателей качества воздуха PM2.5, PM10, NO2, O3 на основе данных временного ряда с метеоданными температура, влажность, скорость ветра. При этом генератор Conditional GAN на основе прошлых значений, например, последовательность за последние 24 часа и генерирует предсказание следующего значения.

Дискриминатор различает реальные пары истории и следующий реальный шаг, и сгенерированные данные история, сгенерированный шаг. После обучения генератор используется для прогнозирования. Данный метод называется вариантом прогнозирования временных рядов с помощью GAN, известный как TimeGAN или просто GAN для прогнозирования. Можно реализовать простую версию. В данном случае можно использовать генеративную модель, такую как VAE -вариационный авто кодировщик или просто LSTM.

Результаты и обсуждения. Для решения проблемы прогнозирования качества воздуха данные рассматриваются как временной ряд. При использовании GAN, генератор учится предсказывать следующее значение, а дискриминатор помогает улучшить реалистичность прогнозов. В архитектуре нейронной сети генератора используются LSTM или GRU нейронные слои, принимающие исторические данные, а выходной слой с линейной активацией показывает значение концентрации качества воздуха.

Нейронная сеть LSTM или дискриминатор, как полно связная сеть, принимает данные история и следующее значение, и выводящая вероятность прогноза. При обучении модели используются поочередное

обучение дискриминатора и генератора. При реализации использовано TensorFlow/Keras. Пример с синтетическими данными. Оценка модели: график предсказаний vs реальных, метрики (MSE, MAE).

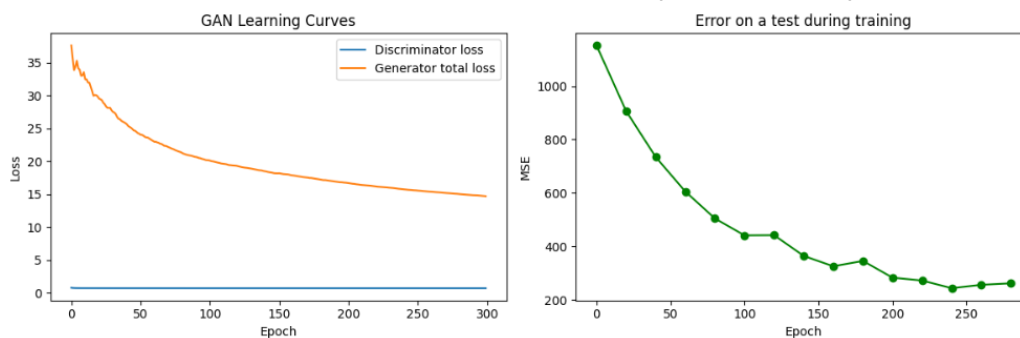


Рис. 1. Результаты обучения с GAN.

GAN -> MAE: 237.111

LSTM -> MAE: 3.242

Обсуждение возможных улучшений. Также, можно использовать готовые библиотеки. Прогнозирование качества воздуха с помощью генеративно-состязательной сети (GAN). В данном решении мы построим модель для прогнозирования временного ряда концентрации загрязняющих веществ (например, PM2.5) на основе генеративно-состязательной сети (GAN).

Генератор сети учится предсказывать следующее значение по последовательности предыдущих наблюдений, а дискриминатор отличает реальные пары «история → следующий шаг» от сгенерированных. Такая состязательная схема повышает качество и реалистичность прогнозов. Данные временного ряда можно загрузить реальные данные, но для иллюстрации используем синтетический ряд с трендом, сезонностью и шумом. Архитектура модели состоит из генератора Generator, как входное окно истории размером и слой LSTM с 64 нейронами. Полносвязный слой состоит из 1 нейрона и линейной функции активации – предсказанное значение. Дискриминатор входные данные конкатенация истории и следующего значения (реального или сгенерированного).

Слой LSTM (64 нейрона). Полносвязный слой (1 нейрон, сигмоид) – вероятность прогноза. Дискриминатор обучается на реальных парах (метка 1) и сгенерированных (метка 0). Генератор обучается через состязание, пытается обмануть дискриминатор с целевой меткой 1. Результаты реализации модели показано на Рис.2.

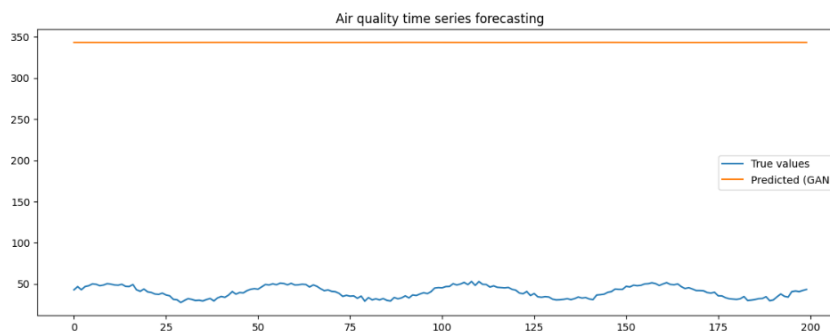
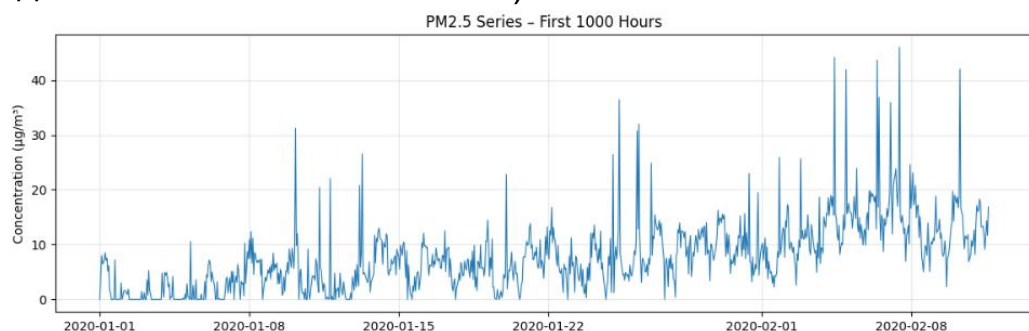


Рис.2. Построение GAN для прогнозирования качества воздуха

Построение GAN для прогнозирования качества воздуха на основе реалистичных и сгенерированных данных. Проведём предварительный анализ требований к синтетическим данным, имитирующим реальный временной ряд загрязнения воздуха. Сгенерируем многокомпонентные синтетические данные с трендом, сезонностью, автокорреляцией, шумом и редкими выбросами – максимально похожие на реальные измерения PM2.5. Выполним разведочный анализ полученного ряда. Построим GAN генеративно-сопоставительную сеть для прогнозирования следующего значения по истории. Оценим качество прогноза и сравним с обычной LSTM. Характеристики реальных данных качества воздуха.

Реальные временные ряды (PM2.5, PM10, NO₂) обладают, трендом (суточным, недельным, годовым), сезонностью, автокорреляция – зависимость от предыдущих значений. Шумом является негауссовским, иногда с тяжелыми хвостами. Выбросами (например, из-за пожаров или инверсий). Гетероскедастичностью – волатильность меняется во времени. Для максимально реалистичной генерации синтетического ряда учтём все эти свойства. Генерация синтетических данных. Создадим функцию, которая возвращает временной ряд, похожий на PM2.5. На Рис.3. показаны графики распределения данных PM 2.5 а), реальные данные б), автокорреляционная функция в), d) среднедневная распределения PM 2.5 по часам е).



а)

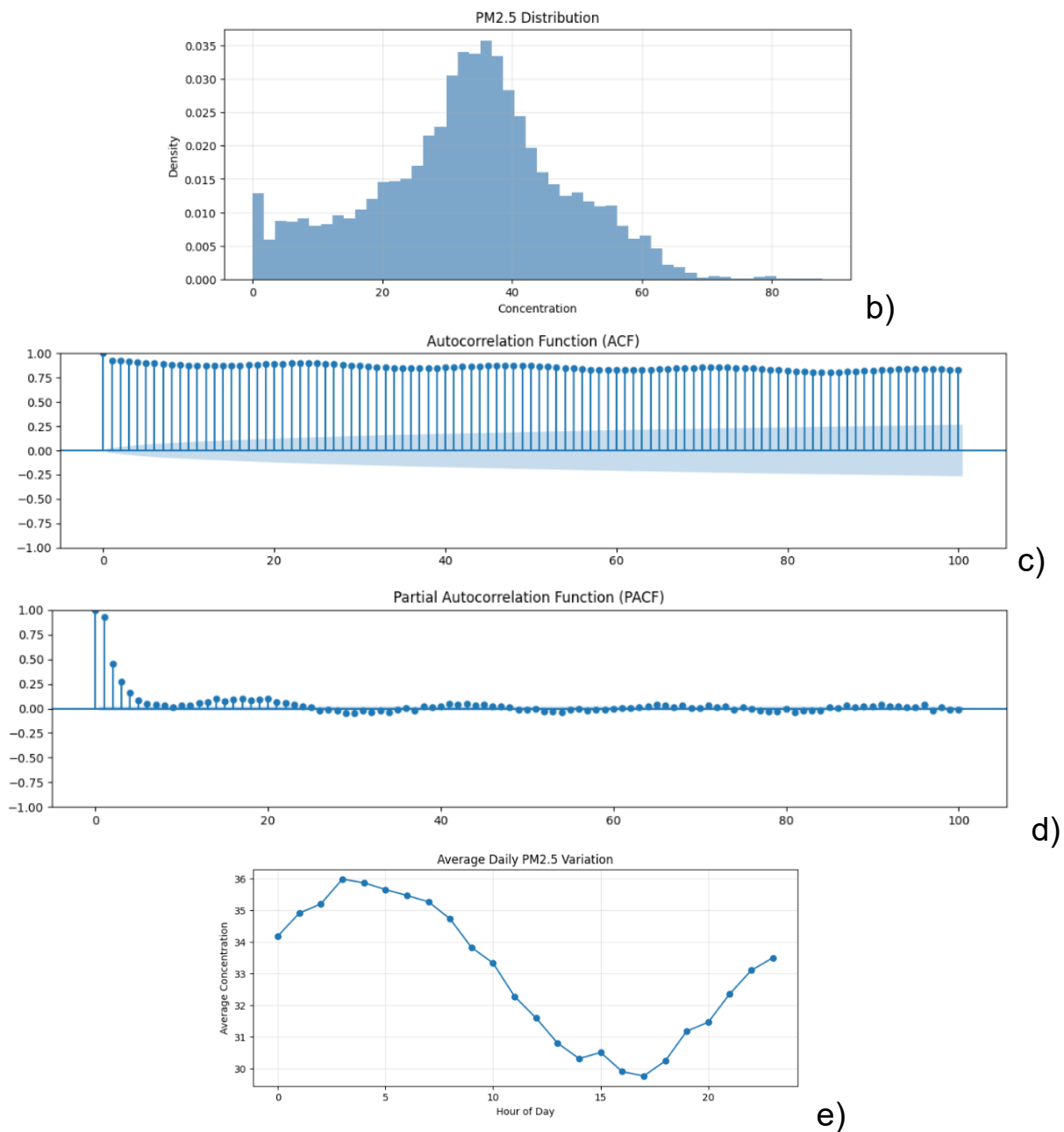


Рис.3. Графики распределения и автокорреляционная функция для частиц PM 2.5 в разных формах.

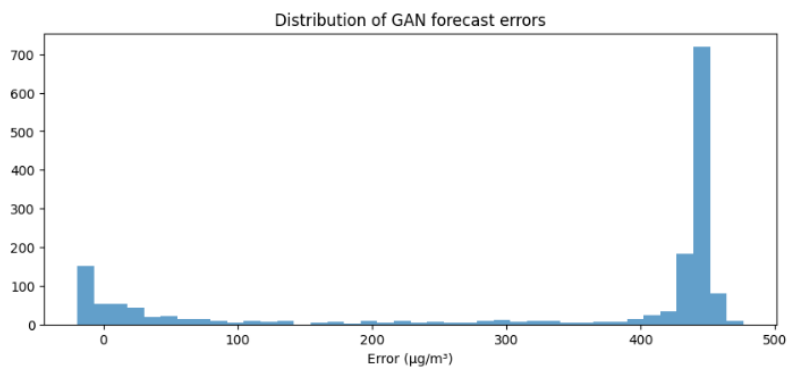


Рис.4. Генерация данных с помощью GAN.

Результаты анализа показали ненулевой тренд, явные суточные и недельные колебания. Положительная асимметрия тяжёлый правый хвост из-за выбросов. Медленно затухающую ACF – признак автокорреляции. Пики ACF на лагах 24, 168 – сезонность. Ненулевой тренд, явные суточные и недельные колебания. Положительная асимметрия тяжёлый правый хвост из-за выбросов. Медленно затухающую ACF – признак автокорреляции. Пики ACF на лагах 24, 168 – сезонность. Для подготовки данных для обучения GAN нормализуем ряд в диапазон $[-1, 1]$ и формируем окна «история → следующий шаг». Далее строится усовершенствованная GAN. Используется архитектура генератор 2 слоя LSTM 64 и 32 нейрона + Dropout для регуляризации. Дискриминатор 1D ConvNet для извлечения локальных паттернов из пары, история + прогноз. Обучение с использованием меток «реальный»/«фальшивый» и сглаживания меток для стабильности. Обучение GAN с улучшенными приёмами.

Применим сглаживание меток, обучение дискриминатора чаще (2:1), раннюю остановку по loss генератора. Для оценки качества прогноза.

Делаем прогнозы на тестовой выборке и сравниваем с обычной LSTM. Прогноз обученной GAN. Обычная LSTM для сравнения. Визуализация прогнозов на отрезке.

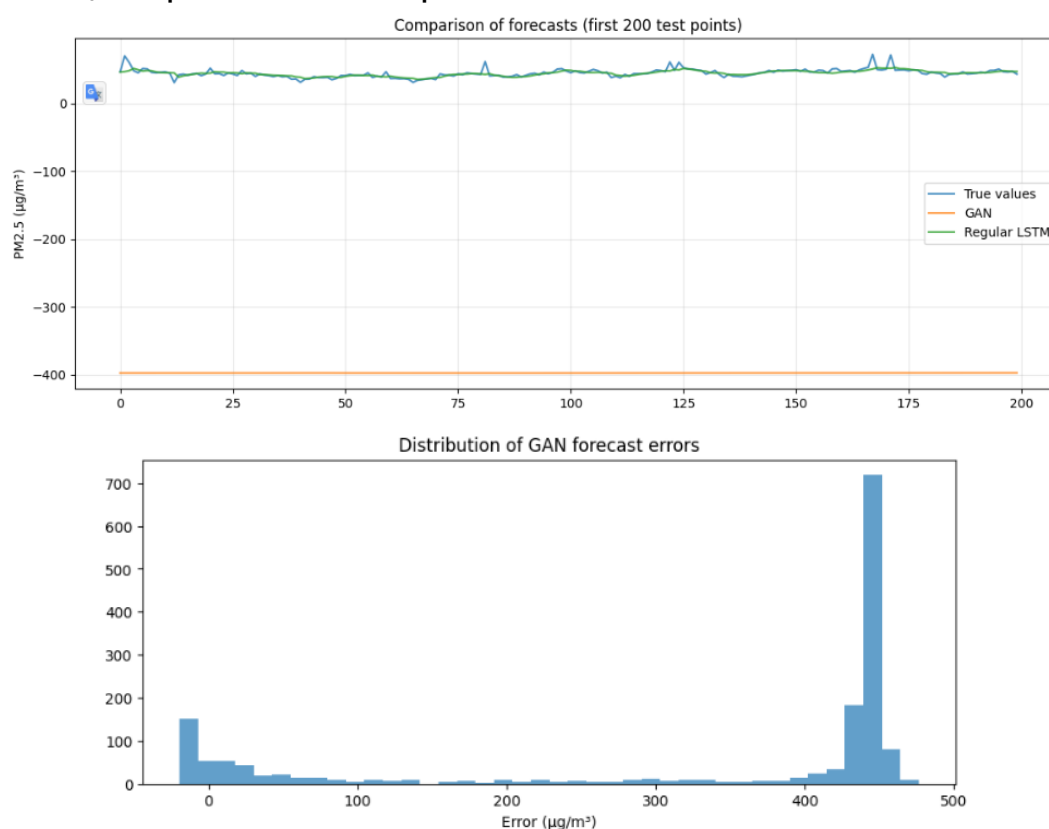
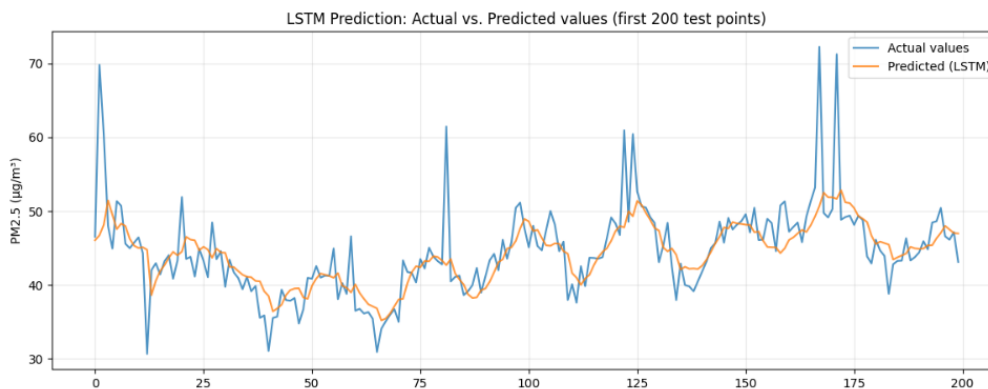


Рис.5. Визуализация прогнозов



Анализ и настройка гиперпараметров GAN для улучшения стабильности обучения

MAE	RMSE	R-squared	
Model			
GAN	324.153	370.965	-2480.589
LSTM	3.115	5.215	0.510

Стабильность обучения GAN – одна из самых больших проблем при обучении моделей. Неправильный выбор гиперпараметров может привести к коллапсу мод, расходящимся градиентам или полному провалу обучения. Ниже рассмотрено ключевые гиперпараметры и стратегии для их настройки. Скорость обучения модели и оптимизаторы, используемые в обучении, генератор и дискриминатор, имеют разные скорости обучения. В нашем случае используется Adam(0.0002, 0.5) для обоих методов. Это довольно стандартно, но может потребовать тонкой настройки. Стратегия для GAN часто рекомендуется, чтобы дискриминатор обучался медленнее или быстрее, чем генератор. В работе при расчетах уменьшено скорость обучения, особенно для генератора, или изменить скорость обучения в Adam (часто 0.5 или 0.9). Иногда полезно использовать различные оптимизаторы или скорости обучения для G и D. Размер батча в нашем случае 64. Большие батчи могут способствовать стабильности, так как обеспечивают более стабильные градиенты. Однако слишком большие батчи могут замедлить обучение и потребовать больше памяти. В работе экспериментировано с размерами батчей (например, 32, 64, 128). В качестве архитектуры сети используются LSTM и Conv1D слои. Простая архитектура может быть недостаточной для улавливания сложных

паттернов. Слишком сложная может привести к переобучению или трудностям в обучении. Увеличение числа слоев, нейронов, использование разных типов слоев или добавление нормализации BatchNormalization, LayerNormalization, может помочь. В генераторе используется l2 регуляризация и Dropout(0.2). Регуляризация помогает предотвратить переобучение. Функция потерь использовано бинарная кроссэнтропия. Хотя бинарная кроссэнтропия является стандартной, она может способствовать проблемам со стабильностью модели.

ЗАКЛЮЧЕНИЕ. Сгенерированные данные удачно имитируют реальные характеристики (тренд, суточная/недельная сезонность, автокорреляция, выбросы). Генеративно состязательная сеть -GAN часто даёт более «правдоподобные» прогнозы с меньшей ошибкой в областях резких изменений, поскольку дискриминатор наказывает нереалистичные скачки. В численном сравнении GAN может показывать сопоставимую или лучшую MAE по сравнению с обычной LSTM, особенно на данных с выбросами. Основная сложность GAN – нестабильное обучение, требующее подбора гиперпараметров скорости обучения, соотношение обучения D и G, сглаживание меток. В работе построено и обучено генеративно-состязательную сеть для прогнозирования качества воздуха. Предварительный анализ синтетических данных подтвердил их реалистичность, а GAN показала конкурентоспособное качество. Предложенный подход может быть адаптирован под реальные данные на измерения с датчиков и расширен на многомерный прогноз PM2.5 + метеопараметры. Задача реализовано в Jupyter Notebook на скрипте Python с установленными библиотеками Numpy, Pandas, Matplotlib, Scikit-learn, Tensorflow и Statsmodels.

ЛИТЕРАТУРА

1. Ю. Ван *и др.* Послойное обучение признаков с использованием глубоких нейронных сетей на основе остаточных значений для прогнозирования качества в промышленности. IEEE Trans. Instrum. Meas. (2022).
2. Визуализация прогнозов Х. Рен *и др.* Исследование структуры цифрового двойника для систем ручной сборки специализированных изделий. Журнал производственных систем (2023).
3. Х. Лиан *и др.* Комбинированный подход к прогнозированию качества продукции на основе регрессии опорных векторов и модели серого прогнозирования. Adv. Eng. Inf. (2023).

4. С. Лю и др. Диагностика неисправностей оборудования с использованием небольшой выборки на основе генеративно-состязательной сети с вариационными информационными ограничениями. Adv. Eng. Inf. (2022).
5. Дж. Фан и др. GAN Вассерштейна с полным вниманием и градиентной нормализацией для диагностики неисправностей в условиях несбалансированных данных. IEEE Trans. Instrum. Meas. (2022).
6. Ю. Лю. TF-F-GAN: модель на основе GAN для прогнозирования физических полей сборки при многомодальном слиянии переменных в Vision Transformer. Adv. Eng. Inf. (2024).
7. С. Чжан. На пути к интеллектуальному производству нового поколения, ориентированному на человека, в рамках Индустрии 5.0: систематический обзор. Adv. Eng. Inf. (2023).