

## **АРХИТЕКТУРА GAN ПРИ СОЗДАНИИ ВРЕМЕННЫХ РЯДОВ**

**Цель и задачи:** поиск нейросетевого способа генерации новых данных временного ряда для решения задач соответственно генерации новых данных временного ряда, а также прогнозирование данных временного ряда.

Получение корректных данных для разного рода задач всегда являлось краеугольным камнем обучения нейронных сетей. Архитектура GAN предлагает решение данной проблемы.

В ходе решения задач прогнозирования и поиска аномалий во временных рядах мы иногда сталкивались с тем, что данных не хватает. Например, слишком маленькая последовательность данных для обучения сетей, решающие вышеуказанные задачи прогнозирования и поиска аномалий или необходимы другие данные, на которых можно проверить работоспособность сетей, найти некоторые возможные слабые места сети. Иногда вовсе нет данных. Сделать собственные данные программным способом, не собрав их с каких-либо процессов, довольно проблематично, поскольку функции для создания случайных значений в различных языках программирования носят псевдослучайный характер. Из-за этого последовательности начинают, так скажем, повторяться – они становятся периодическими. А такое свойство есть не у всех временных рядов. Что же делать?

### **Знакомство с GAN. Назначение и краткая история GAN**

К нам на помощь могут прийти генеративно-сопоставительные сети. Generative Adversarial Networks (GAN) являются одним из самых инновационных и перспективных подходов в области генерации данных. Они позволяют создавать новые различные данные. В данной статье мы рассмотрим применение GAN для генерации временных рядов. Однако для начала лучше раскроем назначение GAN и расскажем краткую историю о сети.

Назначение GAN заключается в создании новых данных, которые могут быть использованы для различных задач, таких как генерация изображений, звуков, текстов и т. д. Но GAN может быть использована для улучшения качества существующих данных, а также для создания новых данных, которые могут быть использованы в различных приложениях, например, компьютерном зрении, обработке естественного языка, машинном обучении, медицине, рекламе и так далее.

GAN были предложены в 2014 году Ианом Гудфеллоу и его коллегами. Эта модель машинного обучения состоит из двух нейронных сетей: генератора и дискриминатора. Генератор создает синтетические данные, а дискриминатор пытается отличить эти данные от реальных данных.

Впервые GAN были использованы для генерации изображений, но со временем они были адаптированы для генерации других типов данных, включая текст и временные ряды.

Как легко заметить, очень много областей применения GAN, поэтому и не мудрено, что и видов GAN будет немало. Существуют обычная GAN, используемая для генерации изображений и других типов данных, Conditional GAN – модель, которая может генерировать данные на основе определенных условий. Например, она может генерировать изображения определенного класса или текст на определенную тему. Также есть CycleGAN, преобразующая данные из одного домена в другой; Wasserstein GAN, использующая расстояние Вассерштейна вместо функции потерь для улучшения качества генерируемых данных; Progressive GAN, использующая постепенное увеличение разрешения изображений для создания более высококачественных изображений; StyleGAN, генерирующая изображения с высоким разрешением и контролирующая стиль и содержание изображения отдельно друг от друга. Также есть виды, представляющие для нас повышенный интерес. Это TimeGAN и GAN-AR, которые можно использовать для генерации данных временных рядов.

### Общее устройство и математическая модель GAN

Давайте разберёмся, как устроена GAN. GAN состоит из двух нейронных сетей: генератора и дискриминатора. Генератор создает новые данные, а дискриминатор определяет, насколько эти данные похожи на реальные данные. Обе сети обучаются параллельно, причем генератор стремится создать данные, которые обманут дискриминатор, а дискриминатор стремится правильно классифицировать данные.

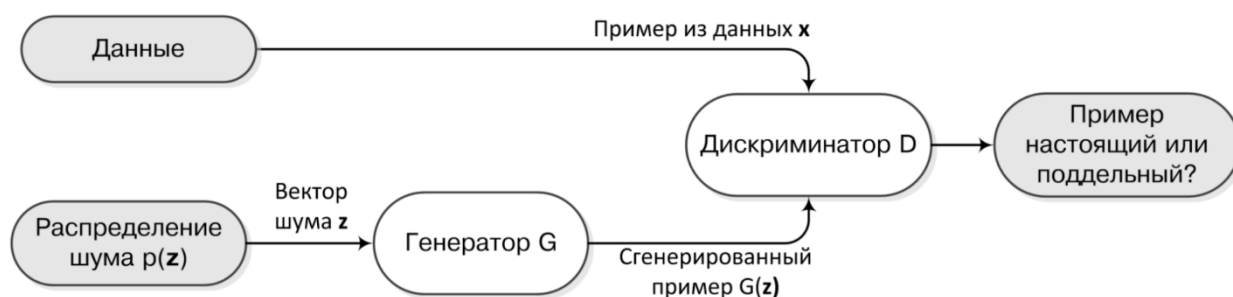


Рисунок 1 – Устройство GAN

Математическую модель GAN можно описать следующим образом: пусть есть две функции – генератор  $G$  и дискриминатор  $D$ . Генератор  $G$  принимает на вход случайный шум  $z$  и генерирует новые данные  $x$ . Дискриминатор  $D$  принимает на вход данные  $x$  и выдает вероятность того, что эти данные являются реальными. Обучение GAN происходит через минимизацию функции потерь, которая состоит из двух частей: первая часть отвечает за ошибку генератора, а вторая часть – за ошибку дискриминатора. Цель генератора – максимизировать ошибку дискриминатора, а цель дискриминатора – минимизировать ошибку.

Математическая модель GAN может быть дополнена различными слоями нейронных сетей, такими как сверточные и рекуррентные слои, чтобы адаптироваться к конкретной задаче. Например, для генерации временных рядов могут использоваться рекуррентные слои для учета зависимостей во времени.

## **Кратко о временных рядах и их генерации**

Теперь непосредственно перейдём к нашей теме. GAN также может быть использована для генерации временных рядов, что и является нашим целевым интересом, которые могут быть использованы в различных задачах, таких как прогнозирование цен на акции, трафика на дорогах и т. д. Для генерации временных рядов с помощью GAN используется специальный тип, называемый GAN-AR или TimeGAN.

Однако давайте для начала кратко разберёмся, что вообще такое временные ряды и почему их генерация так важна. Временные ряды – это последовательность значений, которые изменяются во времени. Они широко используются в различных областях, таких как экономика, финансы, климатология, биология и т. д. Генерация временных рядов является важной задачей, поскольку она может помочь в прогнозировании будущих значений и выявлении скрытых зависимостей.

### **GAN-AR**

Теперь рассмотрим две модели: TimeGAN и GAN-AR. Начнём с GAN-AR.

GAN-AR (Generative Adversarial Networks for Autoregressive Time Series) – это метод генерации данных временных рядов, который сочетает в себе две технологии: генеративно-сопоставительные сети (GAN) и авторегрессионные модели (AR).

Идея создания GAN-AR возникла в 2017 году у группы исследователей из университета Карнеги-Меллон (CMU) и компании Facebook AI Research (FAIR). Они заметили, что существующие методы генерации данных временных рядов, такие как LSTM и ARIMA, имеют свои ограничения, такие как ограниченность в генерации длинных последовательностей и склонность к переобучению.

Для создания GAN-AR исследователи использовали двухэтапный подход. На первом этапе они обучили GAN на реальных данных временных рядов, чтобы создать сгенерированные данные, которые максимально соответствуют реальным данным. На втором этапе они использовали AR-модель для прогнозирования будущих значений временного ряда на основе сгенерированных данных.

Одним из ключевых преимуществ GAN-AR является его способность генерировать длинные последовательности данных временных рядов, что делает его более гибким и универсальным по сравнению с другими методами. Кроме того, GAN-AR показал высокую точность прогнозирования будущих значений временного ряда на основе сгенерированных данных.

В целом, GAN-AR является перспективным методом генерации данных временных рядов, который может быть использован в различных областях, таких как финансы, экономика, медицина и промышленность.

### **TimeGAN**

А теперь поговорим о TimeGAN.

Применение GAN для генерации временных рядов началось с работы Юна и др. в 2019 году. Они предложили модель, названную TimeGAN, которая использует сверточные и рекуррентные нейронные сети для генерации временных

рядов. TimeGAN состоит из четырех компонентов: Embedder, Generator, Supervisor и Discriminator.

TimeGAN была протестирована на нескольких наборах данных, включая данные о погоде и финансовых данных. Результаты показали, что TimeGAN может создавать новые временные ряды, которые имеют сходство с реальными данными.

Для генерации временных рядов с помощью TimeGAN генератор принимает на вход случайный вектор шума и производит последовательность значений, представляющих временной ряд. Дискриминатор принимает на вход последовательность значений и определяет, является ли она реальной или сгенерированной. Процесс обучения включает в себя итеративное обновление генератора и дискриминатора до тех пор, пока генератор не производит временные ряды, которые невозможно отличить от реальных.

Поскольку TimeGAN является модификацией GAN-AR, то и TimeGAN обладает тем же ключевым преимуществом, что и GAN-AR, а именно способность генерировать длинные последовательности данных временных рядов и высокая точность прогнозирования будущих значений временного ряда на основе сгенерированных данных. Однако, помимо этого, в качестве преимущества TimeGAN можно выделить его способность генерировать временные ряды, обладающие долгосрочными зависимостями. Это достигается за счет использования авторегрессионных моделей, которые моделируют условное распределение вероятностей каждого временного шага, учитывая предыдущие временные шаги. Авторегрессионные модели широко используются в анализе и прогнозировании временных рядов, но они имеют ограничения в обработке сложных и нелинейных отношений между переменными. TimeGAN преодолевает эти ограничения, используя глубокие нейронные сети для моделирования сложных отношений между переменными.

### **Другие виды GAN, представляющие интерес**

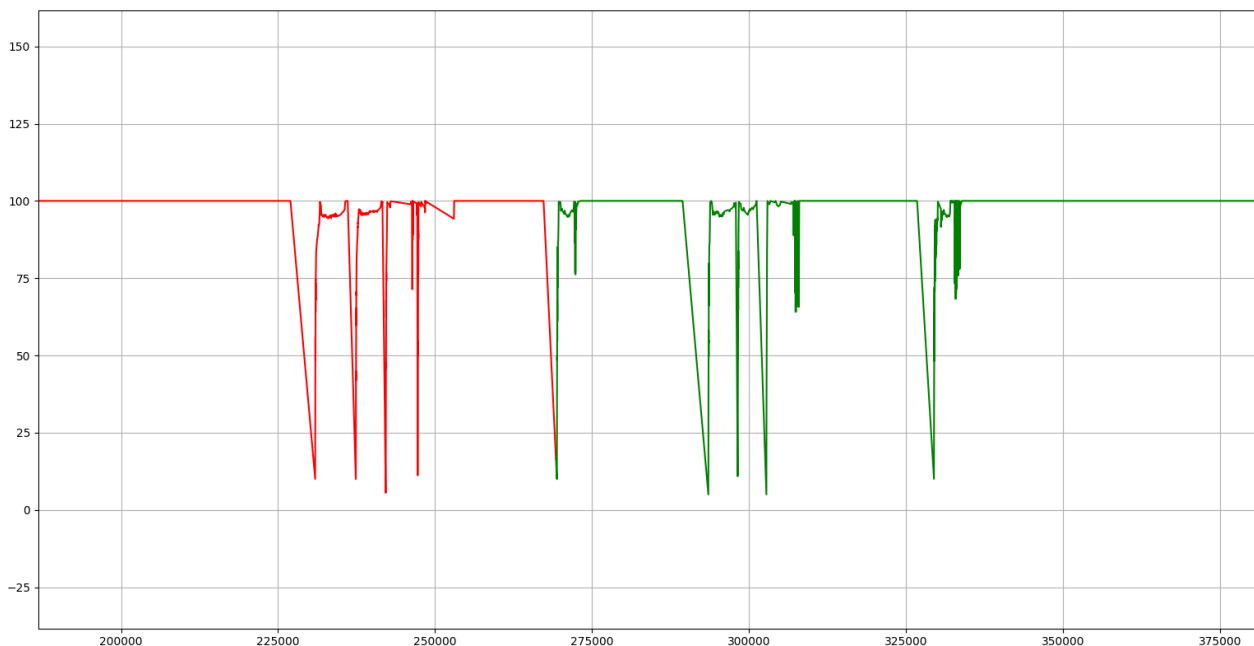
Другие исследователи также предложили свои модели для генерации временных рядов с помощью GAN. Например, модель GANomaly, предложенная Акчуриной и др., использует GAN для создания аномальных временных рядов. Это может быть полезно для обнаружения аномалий в данных и предотвращения возможных проблем.

Еще одной интересной моделью является TG-GAN. Эта GAN для генерации временных рядов, которые имеют различные уровни шума. Это может быть полезно для тестирования алгоритмов прогнозирования временных рядов на различных уровнях шума.

### **Практика**

Ранее было сказано, что нами решалась задачи прогнозирования временных рядов. И в ходе её решения мы столкнулись с большой проблемой: мы не могли по средством сети LSTM предсказывать на собственных предсказаниях LSTM, что ограничивает количество близких к действительности предсказаний. Однако, познакомившись с GAN, мы попробовали объединить две эти сети. Поскольку GAN генерирует похожие на реальные данных последовательности, то

это последовательности мы отправляли на вход LSTM, которая делала свои предсказания. Результат можно увидеть на рисунке 2, где красная линия – реальные данные, а зелёная линия – сгенерированные данные.



*Рисунок 2 – Результаты работы TimeGAN*

## **Вывод**

GAN является мощным инструментом для генерации новых данных, в том числе временных рядов. Они могут быть использованы в различных областях, таких как экономика, финансы, климатология и биология. Модели, основанные на GAN, показывают обещающие результаты в генерации временных рядов, и будущие исследования могут расширить их применение в других областях. Также не забываем, что, опять же, генерация данных временных рядов с помощью GAN является важной задачей в машинном обучении и анализе данных. Несмотря на то, что этот подход все еще находится в стадии исследований, в чём можно убедиться, посмотрев на года их разработок, он уже показал потенциал для создания новых методов генерации данных временных рядов.

## **Список цитированных источников**

1. Фостер, Д. Генеративное глубокое обучение. Творческий потенциал нейронных сетей / Д. Фостер – СПб. : – Питер, 2020. – 336 с.
2. Брюс, П. Практическая статистика для специалистов Data Science / П. Брюс, П. Гедок, Э. Брюс. – 2-е изд., перераб. и доп. – СПб. : БХВ-Петербург, 2021. – 352 с.
3. A TimeGAN Application for Generating Time Series Related to Climate Prediction – Ca' Foscari University of Venice Department of Environmental Sciences, Informatics and Statistics. – 2021. – 42 с.