

АНАЛИЗ ВОЗМОЖНОСТЕЙ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ГЕНЕРАЦИИ ФОТОРЕАЛИСТИЧНЫХ ИЗОБРАЖЕНИЙ

Россия, г. Пенза, Пензенский государственный технологический университет

The article discusses the possibilities of using neural networks to generate images, music, speech or text, as similar as possible to what a person creates. In particular, the application of an applied adversarial network (GAN) to achieve the goal. A comparative analysis of discriminatory and generative models of neural networks is carried out and their effectiveness in solving various problems, such as classification, image recognition, generation of video, text, etc., is analyzed. The detailed structure of the generative adversarial network (GAN) is also considered and a study is made of the methodology for training neural networks for the generator and discriminator.

Введение. С каждым днем сфер применения искусственного интеллекта становится все больше. Это связано, прежде всего, с активными исследованиями применения нейронных сетей и машинного обучения в различных сферах человеческой деятельности, а также разработкой новых вычислительных устройств с поддержкой NPU модулей для работы с нейронными сетями.

Одной из активно исследуемой области за последние дни является применение генеративные состязательные сети (GAN) в сфере генерации изображений, музыки, речь или текст, максимально схожими с тем, что создаёт человек.

Генеративно-состязательная сеть – это нейронные сети, которые обучаются имитировать заданное распределение данных. Данный метод был впервые описан Яном Гудфеллоу и его коллегами в статье *NeurIPS* 2014года [1].

GAN состоит из двух нейронных сетей, одна обучена генерировать данные, а другая обучена отличать поддельные данные от реальных данных. Хотя идея структуры для генерации данных не новая, когда дело доходит до генерации изображений или видео, GAN даёт впечатляющие результаты:

- *CycleGAN* может передавать стили изображения выполняя ряд убедительных преобразований стилей в изображениях.
- *StyleGAN* создает человеческие лица.

Дискриминационные и генеративные модели

Большинство приложений с нейронными сетями реализуются при помощи дискриминационных моделей. В свою очередь генеративные состязательные нейронные сети, представляют собой класс генеративных моделей.

Дискриминационные модели – это модели, которые используются для решения задач контролируемой классификации или регрессии. Например, если стоит задача обучения модели классификации рукописных символов, то для этого можно прибегнуть к использованию маркированных наборов данных, которые будут содержать изображения рукописных символов и метки, которые указывают, какое изображение представляет каждая картинка [2].

В процессе обучения используется алгоритм для настройки параметров модели, цель которого в том, чтобы минимизировать функцию потерь. После завершения обучения модель можно использовать для классификации рукописных символов путем вероятности написанных символов (рисунок 1).

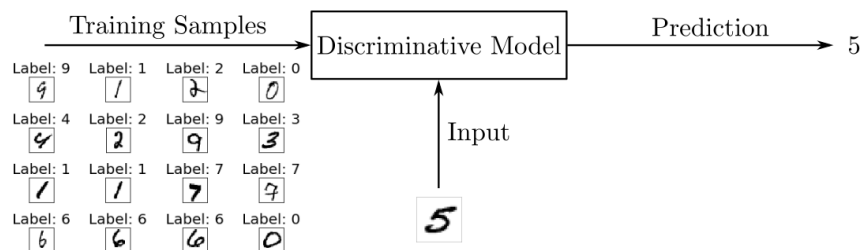


Рисунок 1 – Схема работы нейронной сети для классификации

Генеративно-сопоставительные сети (GAN) обучаются описывать то как генерируются наборы данных в виде вероятностной модели и применяя генеративные модели можно генерировать новые наборы данных.

Кроме того, дискриминантные модели данных используют метод обучения с учителем, а модели GAN могут использовать немаркированные наборы данных, что является формой обучения без учителя.

Используя тот же набор рукописных символов можно обучить генеративно-сопоставительную сеть генерировать новые рукописные символы. Для обучения необходимо реализовать алгоритм для настройки весовых коэффициентов модели, минимизировать функцию потерь и изучить распределение вероятностей обучающего набора. После чего полученная обученная модель может применяться для генерации новых образцов (рисунок 2).

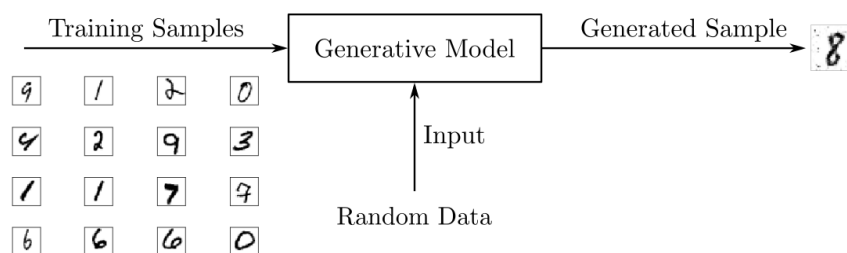


Рисунок 2 – Схема работы нейронной сети для генерации изображения цифр

Для генерации новых образцов данных моделью обычно применяют стохастический или случайный элемент, который влияет на возможный набор выборки данных, которые генерируются моделью.

Случайная выборка данных используется для управления получаемыми генерируемыми данными и получаются из векторов, которые представляют собой сжатую форму сгенерированных данных [3].

В последнее время к генеративно-сопоставительным сетям проявляется высокий интерес, хотя данный подход является не единственным который можно использовать для генерации набора данных, так же существуют такие подходы как:

- Машины Больцмана
- Вариационные автоэнкодеры
- Скрытые марковские модели
- Модели, предсказывающие следующее слово в последовательности, например, GPT-2.

Однако в последнее время генеративно-состязательные сети вызывают наибольший общественный интерес из-за впечатляющих результатов в области создания изображений и видео.

Архитектура генерирующих состязательных сетей

Генеративно-состязательные сети состоят из двух независимых нейронных сетей:

- Генератор (G)
- Дискриминатор (D)

Генератор (G) служит для оценки реалистичности генерируемых вероятностных выборок, чтобы по итогу генерации получились реалистичные данные.

В то время как дискриминатор (D) обучается оценивать вероятность того, что полученная выборка была взята из реальных данных, а не генератора (G) [4].

Данные нейронные генератор и дискриминатор обучаются соревноваться друг с другом, поэтому данная структура получила название генеративно–состязательная сеть. В данной сети генератор пытается с каждым разом все лучше и лучше обмануть дескриминатор, а тот в свою очередь пытается лучше отличать сгенерированные данные от реальных.

Для лучшего понимания работы генеративно–состязательной сети возьмём пример с набором данных, который состоит из двумерных выборок (x_1, x_2) , с x_1 в интервале от 0 до 2π и $x_2 = \sin(x_1)$ (рисунок 3).

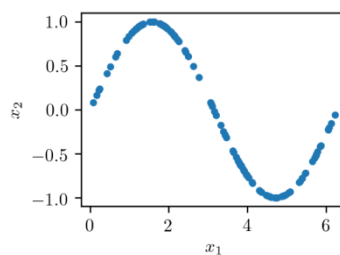


Рисунок 3 – Пример набора данных.

В данном наборе данных находятся точки (x_1, x_2) , которые располагаются вдоль синусоидальной кривой, которая имеет конкретное распределение. Общий вид структуры генеративно–состязательной сети для генерации $(\tilde{x}_1, \tilde{x}_2)$, схожих с образцом набором показан на рисунке 4.

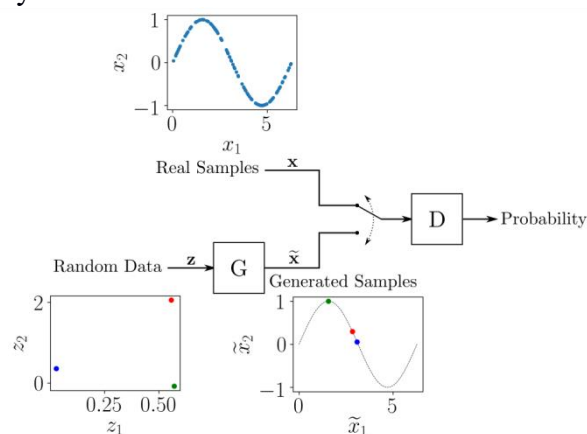


Рисунок 4 – структура GAN для генерации пар $(\tilde{x}_1, \tilde{x}_2)$.

На вход генератора G поступают случайные данные и его роль заключается в попытке создать новый набор данных, который будет максимально схож с реальными образцами данных. В приведенном примере на генератор поступают случайные пары

(z_1, z) , которые требуется преобразовать так чтобы они были максимально похожи на реалистичные образцы.

В качестве структуры нейронной сети для генератора G может выступать любая структура, что позволяет использовать такие структуры как сверточные нейронные сети (CNN), многослойный персептрон (MLP) или любую другую структуру нейронной сети, главное, чтобы размеры входа и выхода соответствовали размерам генерируемых данных и реальных данных [5].

Дискриминатор D работает как с реальными данными из обучающего набора, так и со сгенерированными данными генератором G , его главная задача заключается в оценке вероятности реалистичности поступающих данных. В ходе обучения дискриминатор D выводит 1 когда на него поступают реальные данные, а 0 когда сгенерированные.

Как и в случае со с генератором G можно выбрать произвольную структуру для нейронной сети дискриминатора D , главное, чтобы она соблюдала необходимые размеры входных и выходных данных. В данном примере используется ввод двумерной структуры.

Обучение генеративно–состязательной сети представляет собой минимакс игру для двух игроков, в которой дискриминатор D пытается минимизировать вероятность ошибки при определении различий между реальными данными и сгенерированными, а генератор G пытается максимально увеличить вероятность того, что дискриминатор D совершил ошибку [6].

Хотя набор обучающих данных, состоящий из реальных данных, не маркирован, процесс обучения дискриминатора D и генератора G выполняются под контролем. На определенных этапах обучения параметры дискриминатора D и генератора G обновляются.

Т.е. в исходной формулировке генеративно–состязательной сети параметры дискриминатора D обновляются k раз, генератора G обновляются только один раз для каждого шага обучения сети.

Для корректного обучения дискриминатора D на каждой итерации случайные реальные выборки маркируются как 1, а сгенерированные генератором G как 0. Такой подход позволяет использовать простую структуру контролируемого обучения для обновления параметров дискриминатора D для минимизации функции потерь (рисунок 5).

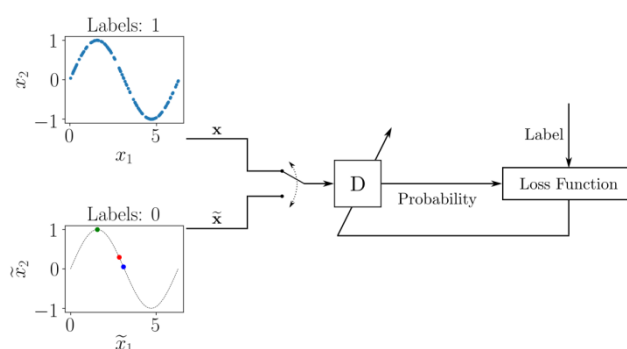


Рисунок 5 – структура GAN с контролируемым обучением.

Для каждого набора обучающих данных, содержащих маркированные реальные и сгенерированные образцы, обновляются параметры дискриминатора D , чтобы минимизировать функцию потерь. После обновления параметров дискриминатора D тренируют генератор G для получения более качественных образцов. Выход генератора G подключен к дискриминатору D , параметры которого остаются замороженными (рисунок 6).

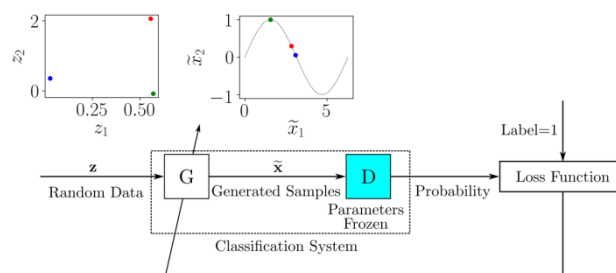


Рисунок 6 – схема тренировки модели GAN.

Можно представить систему, состоящую из генератора G и дискриминатора D , как единую систему классификации, которая принимает случайные выборки в качестве входных данных и проводит классификацию, которую в данном случае можно интерпретировать как вероятность.

Когда генератору G удастся обмануть дискриминатор D , вероятность результата на выходе приближается к 1. В данном случае можно использовать обычную структуру обучения с учителем: набор данных для обучения системы классификации, состоящей из генератора G и дискриминатора D , будет предоставлен случайными входными выборками, а маркировка, связанная с каждой входной выборкой, будет равна 1.

Во время обучения, когда параметры дискриминатора D и генератора G обновляются, ожидается, что сгенерированные выборки, предоставленные генератором G , будут больше походить на реальные данные, и дискриминатору D будет сложнее различать реальные и сгенерированные данные.

На основании полученных в статье результатов сделаем следующие выводы.

1. В статье описаны отличительные особенности дискриминационных и генеративных моделей нейронных сетей и проанализированы их эффективности в решении различных задач.
2. В статье описана архитектура генеративной состязательной сети (GAN), которая позволяет генерировать изображения, музыку, речь или текст, максимально схожие с тем, что создаёт человек.
3. Исследованы методики проведения обучения нейронных сетей для генератора и дискриминатора.

1. Generative Adversarial Nets [Электронный ресурс] // Advances in Neural Information Processing Systems 27 / Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza. URL: <https://papers.nips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf> (дата обращения: 10.04.2021).

2. Martyshkin A.I.; Markin E.I. Research of the Handwriting Recognition Methods // HELIX – 2020. – № 10(5). – С. 62-68

3. Бершадская Е.Г., Маркин Е.И., Мартышкин А.И. Методы идентификации личности по изображению лица // XXI век: итоги прошлого и проблемы настоящего плюс. – Т.9. №1 (49). С. 49-53

4. Блинов С.Е., Мартышкин А.И., Маркин Е.И. Метод поиска эффективного представления объектов социальной сети на основе данных из множества источников // XXI век: итоги прошлого и проблемы настоящего плюс – 2021. – № 1(53). – С. 62-68

5. From GAN to WGAN [Электронный ресурс] // Lil'Log. URL: <https://lilianweng.github.io/lil-log/2017/08/20/from-GAN-to-WGAN.html> (дата обращения: 10.04.2021).

6. Vaishnav C., Priyansh S., Manisha P., Gaurav K. Semantic Image Completion and Enhancement using Deep Learning // Cornell University – 2020. – с. 6