

УДК 004.8

Нейронная сеть для создания музыкальных композиций на основе условного вариационного автоэнкодера

Золотарев Артем Максимович

zolotarevam@student.bmstu.ru

Белов Юрий Сергеевич

ysbelov@bmstu.ru

КФ МГТУ им. Н.Э. Баумана, Калуга, Россия

Рассмотрено использование условного вариационного автоэнкодера (CVAE) для генерации оригинальных музыкальных композиций. Обсуждаются архитектурные особенности модели, включая кодировщик и декодировщик, которые используют методы репараметризации и блоки ResNet для улучшения обучения. Оценка точности модели выполняется с помощью KL-дивергенции, что позволяет отслеживать качество генерации музыки. Цель работы — исследовать возможности CVAE в музыке и обозначить направления для будущих исследований.

Ключевые слова: нейронные сети, музыка, условный вариационный автоэнкодер

В последние годы искусственный интеллект все активнее проникает в мир искусства, в том числе и в музыкальную сферу. Одним из интересных направлений является использование нейронных сетей для создания оригинальных музыкальных композиций. В данной статье рассмотрим условный вариационный автоэнкодер как мощный инструмент для генерации музыки.

Разработка архитектуры. Пусть x и z обозначают наблюдение и скрытую переменную. Сеть кодировщика определяет апостериорное распределение $q(z|x)$ для скрытого представления z на основе наблюдения x .

Сеть декодера вычисляет распределение $p(x|z)$, используя скрытую выборку z как входные данные. Скрытое распределение $p(z)$ моделируется стандартным нормальным распределением Гаусса. Для получения выборки z в процессе обучения извлекаем значения из скрытого распределения, заданного кодировщиком, что создает проблему для обратного распространения [1].

Чтобы решить эту проблему, применяется метод репараметризации: z аппроксимируется по формуле 1:

$$z = \mu + \sigma \odot \epsilon, \quad (1)$$

где μ и σ — параметры, полученные от декодера; ϵ — случайный шум из нормального распределения. Это позволяет передавать градиенты в кодировщик через μ и σ , сохраняя стохастичность через ϵ [2].

Для построения вариационного автокодировщика будет использован блок ResNet, который помогает предотвратить затухание градиентов в глубоких нейронных сетях [3].

Архитектура кодировщика изображена на рис. 1.

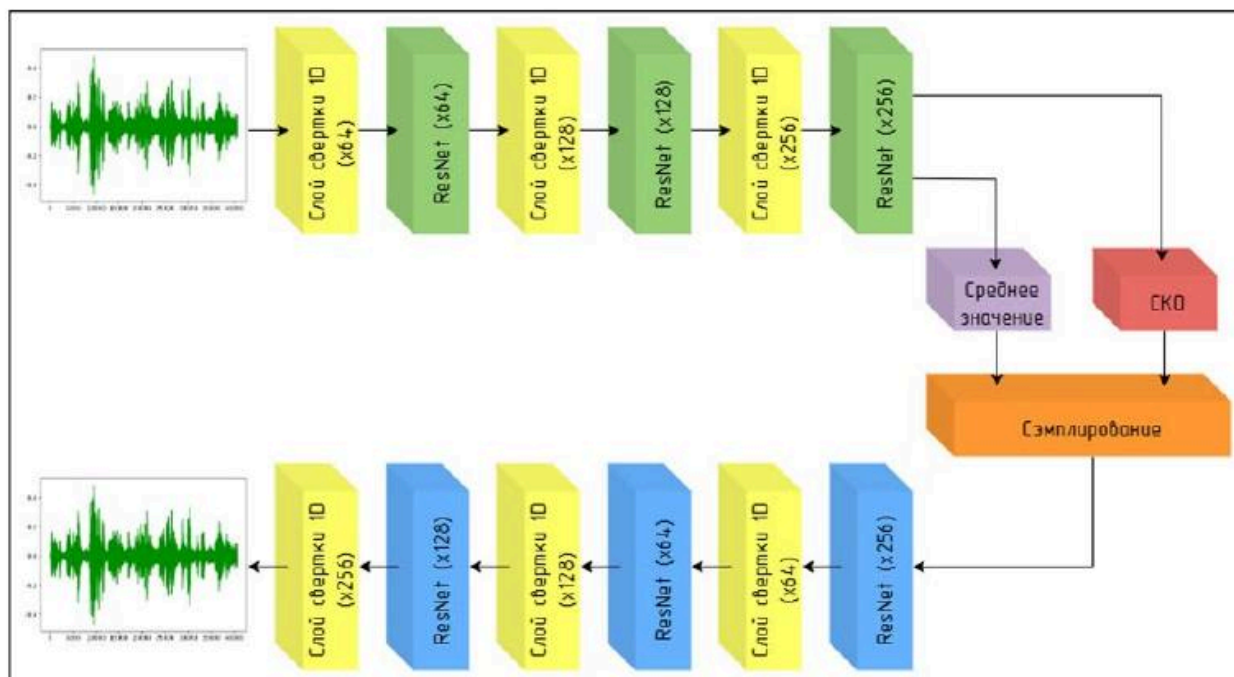


Рис. 1. Архитектура сверточного автокодировщика

Кодировщик начинается с входного слоя, принимающего данные размером (1, 150000), после чего следуют несколько слоев 1D свертки с увеличением числа фильтров, а также блоки ResNet для обработки данных. Эти слои последовательно уменьшают размерность и формируют скрытое представление, завершающееся полносвязным слоем.

Декодировщик принимает скрытые представления и сначала изменяет их форму, после чего последовательно используются блоки ResNet и транспонированные слои свертки 1D. Эти слои применяются для восстановления исходных данных, постепенно увеличивая размерность, и завершаются слоем с 90001 выходными фильтрами для получения окончательного результата.

Оценка точности работы модели. Для оценки точности созданного автокодировщика применялась дивергенция Кульбака — Лейблера (KL-дивергенция), которая показывает, насколько хорошо распределение вероятностей Q приближает распределение вероятностей P . Это достигается путем измерения кросс-энтропии с вычитанием энтропии.

В процессе обучения вариационного автоэнкодера кодировщик нацелен на изучение простого распределения $Q(z|X)$, максимально соответствующего реальному распределению $P(z|X)$. В данной работе KL-дивергенция используется для выявления различий между этими двумя распределениями (2). Таким образом, основной задачей функции VAE становится минимизация компонента KL-дивергенции [4].

$$DKL[Q(z|X)||P(z|X)] = E[\log Q(z|X) - \log P(z|X)] \quad (2)$$

На рис. 2 изображен график зависимости KL-дивергенции от эпохи обучения для классической музыки.

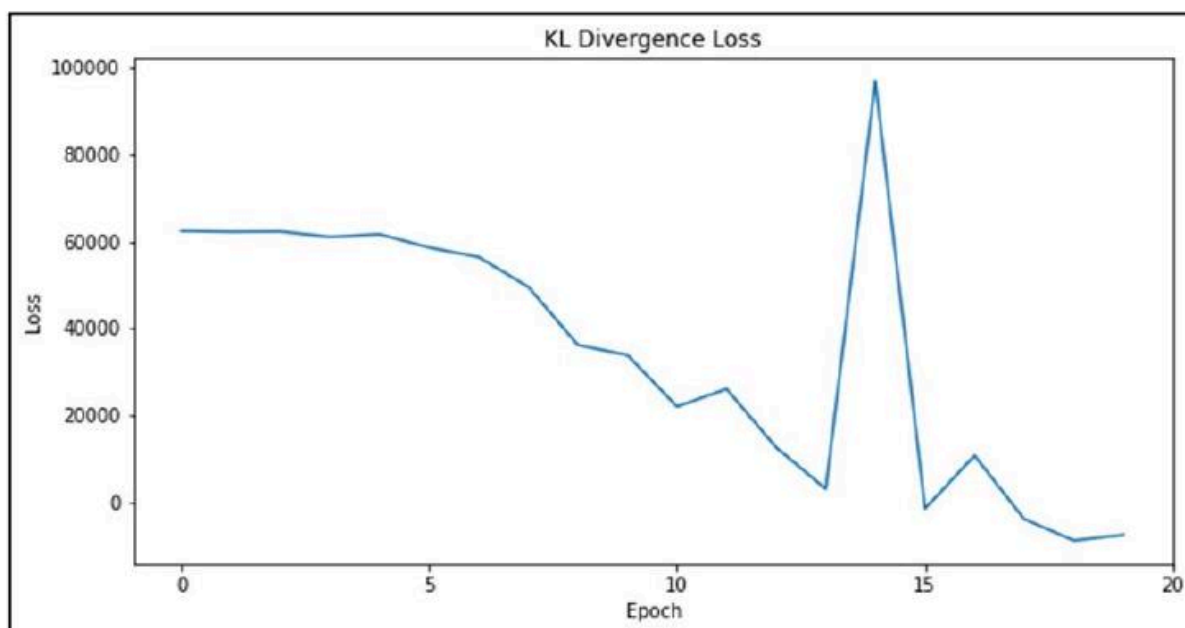


Рис. 2. График зависимости значений KL-дивергенции от эпохи обучения

График зависимости KL-дивергенции от числа эпох обучения подтверждает наличие четкой тенденции к снижению значений KL-дивергенции по мере продвижения обучения. Это говорит о том, что модель последовательно улучшает свои характеристики и все более точно воссоздает вероятностные распределения, что, в свою очередь, ведет к более качественной генерации музыки. Тем не менее, несмотря на эти положительные результаты, авторы столкнулись с проблемой присутствия низкочастотных шумов и искажений в конечных музыкальных произведениях. Это указывает на необходимость дальнейших доработок, как в архитектуре сети, так и в методах обработки и фильтрации аудиосигналов, чтобы минимизировать влияние нежелательных артефактов.

Авторы уверены, что дальнейшие улучшения и настройки архитектуры могут привести к значительным успехам в нейросетевой музыке и вдохновить будущие исследования, которые будут продолжать развивать и улучшать эту технологию.

Литература

- [1] Cao B., Fukumori T., Yamashita Y. Multi-Instruments Music Generation Based on Chord Input. IEEE 12th Global Conference on Consumer Electronics (GCCE), 2023, pp. 1082–1083.
- [2] Choi K., Hawthorne C., Simon I. Encoding musical style with transformer autoencoders. International Conference on Machine Learning, 2020, arXiv:1912.05537v2, pp. 1899–1908.
- [3] Subramani K., Rao P., D’Hooge A. Vapar Synth — A Variational Parametric Model for Audio Synthesis. ICASSP 2020 — 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, arXiv:2004.00001v1, pp. 796–800.
- [4] Asperti A., Trentin M. Balancing Reconstruction Error and Kullback — Leibler Divergence in Variational Autoencoders. IEEE Access, 2020, vol. 8, pp. 199440–199448.

Neural network for music composition generation based on conditional variational autoencoder

Zolotarev Artem Maximovich

zolotarevam@student.bmstu.ru

Belov Yuri Sergeevich

ysbelov@bmstu.ru

Kaluga Branch of BMSTU, Kaluga, Russia

The article discusses the use of a conditional variational autoencoder (CVAE) for generating original musical compositions. It addresses the architectural features of the model, including the encoder and decoder, which employ reparameterization techniques and ResNet blocks to enhance training. The model's accuracy is evaluated using KL-divergence, allowing for monitoring of the quality of music generation. The aim of the work is to explore the capabilities of VAE in music and outline directions for future research.

Keywords: neural networks, music, conditional variational autoencoder