

# АНАЛИЗ МЕТОДОВ НАХОЖДЕНИЯ СКРЫТОГО ПРЕДСТАВЛЕНИЯ МЕДИАФАЙЛОВ

*Рассматриваются основные методы нахождения скрытого представления медиафайлов, а также типы задач, для которых эти методы применяются.*

## ВВЕДЕНИЕ

В настоящее время актуально множество задач классификации и анализа медиафайлов. Для решения многих задач машинного обучения требуется большой тренировочный набор размеченных данных, на котором нужно обучать модель, потратив значительное время и вычислительные ресурсы.

Чтобы избежать повторного обучения моделей для каждой отдельной задачи, существуют способы нахождения скрытого представления медиафайлов. Они позволяют кодировать данные в универсальном виде, который можно использовать для решения, например, не одной конкретной задачи классификации, а нескольких без переобучения всей модели.

### I. АРХИТЕКТУРЫ НЕЙРОННЫХ СЕТЕЙ В РЕШЕНИИ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

Одним из способов получения скрытого представления является построение автокодировщиков на различных архитектурах нейронных сетей.

Выделяются виды нейронных сетей:

- сверточные;
- рекуррентные;
- трансформеры;
- wavenet.

Разные реализации используются для работы в основном с определенными типами данных. Одна из главных областей применения рекуррентных нейронных сетей – работа с языковыми моделями, в частности – анализ контекста и общей связи слов в тексте. В последнее время эти сети начали занимать нишу машинного перевода. Одним из крупнейших примеров является использование компанией Google рекуррентных нейронных сетей типа LSTM в системе Google Translate [1]. Однако они также применяются для кодирования и последующей генерации звука [2]. Сверточные сети широко применяются для всех видов сигналов. Wavenet используются в основном для синтезирования речи [3]. А трансформеры были изначально разработаны для машинного перевода и постепенно заменили рекур-

рентные нейронные сети в основных задачах нейролингвистического программирования [4]. Примеров применения данных нейронных сетей для обработки звука не так много (можно выделить Magenta – исследовательский проект Google AI – как один из наиболее значимых проектов, связанных с созданием музыки).

На сегодняшний день чаще всего применяются сверточные, рекуррентные и wavenet нейронные сети. В перспективе трансформеры могут позволить добиться более высокого качества обработки звука, хотя сейчас они остаются менее эффективными с точки зрения скорости вычислений.

## II. ВЫВОДЫ

Построение автокодировщиков с различными архитектурами и определение результатов их работы при условии одинакового количества операций с плавающей запятой в секунду может ответить на вопрос о наиболее эффективном методе работы со звуковым сигналом при решении ряда задач машинного обучения.

1. Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, Jeffrey Dean, Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, arXiv preprint:1609.08144v2 [cs.CL], 2016.
2. Zhichao Zhang, Shugong Xu, Tianhao Qiao, Shunqing Zhang, Shan Cao, Attention based Convolutional Recurrent Neural Network for Environmental Sound Classification, arXiv preprint:1907.02230v1 [cs.SD], 2019.
3. Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu, WaveNet: A Generative Model for Raw Audio, arXiv preprint:1609.03499v2 [cs.SD], 2016.
4. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, Attention Is All You Need, arXiv preprint:1706.03762v5 [cs.CL], 2017.

*Царева Вероника Викторовна, магистрант кафедры информационных технологий и управления БГУИР, veronika.tsariova@gmail.com.*

*Научный руководитель: Ревотюк Михаил Павлович, кандидат технических наук, доцент, rmp@bsuir.by.*