

УДК 004.032.26

**НЕЙРОННЫЕ СЕТИ: МАТЕМАТИЧЕСКИЕ МОДЕЛИ, ИСПОЛЬЗУЕМЫЕ
ДЛЯ РЕШЕНИЯ СЛОЖНЫХ ЗАДАЧ, ТАКИХ КАК РАСПОЗНАВАНИЕ
ОБРАЗОВ, КЛАССИФИКАЦИЯ И ПРОГНОЗИРОВАНИЕ И МЕТОДЫ
ОБРАБОТКИ И АНАЛИЗА ЗВУКОВЫХ ДАННЫХ, ВКЛЮЧАЯ
РАСПОЗНАВАНИЕ РЕЧИ, КЛАССИФИКАЦИЮ ЗВУКОВ, ВЫДЕЛЕНИЕ
ПРИЗНАКОВ.**



А.Х.Оразгелдиев
Преподаватель
кафедры
информационных
систем Института
телекоммуникаций и
информатики
Туркменистана
abdyorazow588@gmail.c
om



О.Д. Чарыева
Преподаватель
кафедры
информационных
систем Института
телекоммуникаций
и информатики
Туркменистана
caryevahadzat@gmai
l.com



Д. Я.Караджаева
Преподаватель
кафедры
информационных
систем Института
телекоммуникаций и
информатики
Туркменистана

А.Х. Оразгелдиев

Окончил Туркменский государственный университет имени Махтумкули Туркменистана. Область научных интересов сосредоточены на исследовании Нейронные сети, Распознавание речи

О.Д. Чарыева

Окончила Института телекоммуникаций и информатики Туркменистана. Область научных интересов сосредоточены на исследовании методов передачи знаний и обучения, направленных на улучшение производительности и решения задач в различных областях.

Д.Я. Караджаева

Окончила Туркменский государственный университет имени Махтумкули Туркменистана. Область научных интересов сосредоточены на исследовании обработка данных, большие базы данных

Аннотация: Методы обработки и анализа звуковых данных включают предварительную обработку, выделение признаков и классификацию аудиосигналов. Распознавание речи позволяет преобразовывать звук в текст с использованием акустических и языковых моделей. Классификация звуков применяется для идентификации различных типов аудиосигналов, включая окружающие шумы и эмоции в голосе. Современные технологии, такие как нейросети и трансформеры, значительно повышают точность анализа звука. Области применения включают голосовых ассистентов, системы биометрической идентификации и автоматическое создание субтитров.

Ключевые слова: Обработка звука, распознавание речи, классификация аудио, выделение признаков, нейронные сети, акустическое моделирование, языковая модель, преобразование Фурье, спектральный анализ, биометрическая идентификация.

Введение. Нейронные сети – это математические модели, вдохновленные структурой человеческого мозга и используемые для решения сложных задач, таких как

распознавание образов, классификация и прогнозирование. Они состоят из искусственных нейронов, которые соединены между собой и передают информацию друг другу, имитируя работу биологических нейронов.

Каждый искусственный нейрон принимает на вход несколько значений, обрабатывает их с помощью весов и функций активации, а затем передает результат дальше. Слои нейронной сети делятся на входной, скрытые и выходной. Входной слой принимает данные, скрытые слои выполняют вычисления и выявляют сложные закономерности, а выходной слой формирует конечный результат.

Распознавание речи – это процесс преобразования аудиосигнала в текст, который включает несколько шагов, таких как предварительная обработка звука, выделение акустических признаков, акустическое моделирование, языковое моделирование и декодирование. Распознавание в реальном времени требует оптимизированных алгоритмов и аппаратного ускорения, например, использования графических процессоров GPU и специализированных процессоров TPU.

Основные типы нейронных сетей включают полносвязные сети, сверточные нейронные сети CNN, рекуррентные нейронные сети RNN и трансформеры. Полносвязные сети соединяют каждый нейрон одного слоя с каждым нейроном следующего слоя и применяются для обработки числовых данных и классификации объектов. Сверточные нейросети предназначены для работы с изображениями и анализируют пространственные зависимости между пикселями, что делает их эффективными для распознавания лиц, медицинской диагностики и компьютерного зрения. Для классификации звуков используются методы глубокого обучения, такие как сверточные нейросети CNN и рекуррентные нейросети RNN, которые анализируют временные и спектральные характеристики сигнала.

Рекуррентные нейронные сети используются для работы с последовательностями данных, таких как речь, текст или временные ряды, поскольку они учитывают информацию из предыдущих шагов и могут запоминать контекст. Они применяются в системах машинного перевода, распознавании речи и прогнозировании временных рядов.

Трансформеры представляют собой более современный тип нейросетей, основанный на механизме самовнимания, который позволяет анализировать всю входную последовательность сразу, выявляя важные связи между различными частями данных. Они широко используются в обработке естественного языка, например, в чат-ботах, системах машинного перевода и генерации текста.

Обучение нейронных сетей происходит с использованием метода обратного распространения ошибки, который корректирует веса связей между нейронами с учетом разницы между предсказанным и реальным результатом. Для этого применяется градиентный спуск – метод оптимизации, позволяющий минимизировать ошибку модели путем пошагового обновления параметров.

Важную роль в обучении нейросетей играет функция активации, которая определяет, каким образом нейрон передает сигнал дальше. Популярные функции активации включают сигмоиду, гиперболический тангенс, ReLU и его модификации, такие как Leaky ReLU и ELU. Функция ReLU чаще всего используется в глубоких нейросетях, поскольку она помогает избежать проблемы исчезающего градиента и ускоряет обучение. Глубокое обучение – это направление машинного обучения, в котором используются многослойные нейросети с большим числом параметров. Такие сети способны извлекать сложные признаки из данных, что делает их эффективными в задачах компьютерного зрения, обработки речи, биоинформатики и других областях.

Архитектуры нейросетей могут быть как статичными, так и адаптивными. Например, сверточные нейросети адаптируют свои фильтры в процессе обучения, чтобы выявлять ключевые элементы изображения, такие как края, текстуры и формы, в то время

как рекуррентные нейросети адаптируют свою память, чтобы учитывать длинные зависимости в последовательностях.

Для повышения производительности и уменьшения переобучения используются методы регуляризации, такие как дропаут, нормализация весов и пакетная нормализация. Дропаут отключает случайные нейроны во время обучения, чтобы модель не запоминала данные слишком точно и могла лучше обобщать информацию. Пакетная нормализация стабилизирует распределение входных данных внутри сети, что ускоряет обучение и улучшает сходимость.

Обучение нейронных сетей требует значительных вычислительных ресурсов, поэтому часто используется аппаратное ускорение, такое как графические процессоры GPU и тензорные процессоры TPU. Эти устройства способны параллельно обрабатывать большие объемы данных, что делает обучение моделей значительно быстрее и эффективнее.

Современные нейросетевые архитектуры, такие как ResNet, EfficientNet и Vision Transformers, используются для сложных задач компьютерного зрения, включая медицинскую диагностику, автономное вождение и биометрическую идентификацию.

Применение нейронных сетей охватывает множество областей, включая финансы, где они используются для предсказания фондового рынка и выявления мошенничества, медицину, где они помогают анализировать снимки, находить патологии и разрабатывать персонализированные схемы лечения, а также производство, где они оптимизируют процессы и автоматизируют контроль качества продукции.

Нейросети активно используются в искусстве и развлечениях, например, для создания музыки, живописи и анимации, а также для генерации синтетических изображений и видео с высоким уровнем реалистичности.

Они также применяются в робототехнике, помогая автономным системам ориентироваться в пространстве, принимать решения и адаптироваться к изменяющимся условиям.

Обученные модели нейросетей могут использоваться повторно, благодаря чему создаются предобученные модели, такие как BERT, GPT и CLIP, которые можно дообучить для решения конкретных задач без необходимости начинать обучение с нуля.

Нейросети способны выявлять скрытые закономерности в данных, что делает их незаменимыми для анализа больших данных, обнаружения аномалий и построения рекомендательных систем.

Развитие технологий нейросетей привело к появлению мощных языковых моделей, таких как GPT-4, которые могут генерировать тексты, отвечать на вопросы, переводить языки и выполнять широкий спектр интеллектуальных задач.

Несмотря на свою эффективность, нейросети могут быть подвержены проблемам, таким как переобучение, нехватка данных и уязвимость к атакам, например, к атакам с использованием специально измененных входных данных, которые могут привести к неверным предсказаниям модели.

Для улучшения работы нейросетей разрабатываются методы интерпретируемости, такие как Grad-CAM и SHAP, которые позволяют визуализировать, какие именно признаки повлияли на решение модели, что особенно важно в критически важных областях, таких как медицина и право.

Методы обработки и анализа звуковых данных включают несколько ключевых этапов, начиная с преобразования звука в цифровую форму и заканчивая его интерпретацией с использованием искусственного интеллекта и машинного обучения.

Предварительная обработка звука включает удаление шума, нормализацию громкости и преобразование сигнала в удобное представление для дальнейшего анализа.

Выделение признаков – важный этап, который позволяет преобразовать звуковой сигнал в набор значимых параметров, таких как мел-кепстральные коэффициенты MFCC,

коэффициенты линейного предсказания LPC и спектральные признаки, которые используются для дальнейшей классификации и анализа.

Акустическое моделирование основано на использовании нейросетей или статистических моделей, таких как скрытые марковские модели НММ, которые помогают установить связь между звуковыми сигналами и фонемами.

Языковое моделирование применяется для предсказания наиболее вероятных последовательностей слов и построения осмысленного текста, используя вероятностные модели или нейросетевые архитектуры, такие как трансформеры.

Классификация звуков – это процесс, в котором звуковые данные разделяются на категории, например, идентификация типов окружающих звуков, детектирование тревожных сигналов или распознавание эмоций в голосе.

Одним из ключевых методов анализа звука является преобразование Фурье, которое позволяет разложить сигнал на составляющие частоты и представить его в виде спектрограммы.

Временные признаки, такие как продолжительность звука, энергия и тональность, помогают различать различные звуковые события и анализировать речь.

Спектральные признаки, такие как форма спектра, центроид и дисперсия, используются для анализа тембра звука и идентификации его источника.

Методы автоматического определения границ фонем и слов помогают улучшить точность распознавания речи, особенно в системах диалогового взаимодействия.

Генеративные модели, такие как вариационные автоэнкодеры и генеративные состязательные сети GAN, позволяют синтезировать звуки и улучшать качество речи в системах преобразования текста в речь.

Системы шумоподавления и улучшения качества звука используют нейросетевые алгоритмы для фильтрации фоновых шумов и улучшения разборчивости речи.

Распознавание эмоций по голосу основано на анализе интонации, ритма и спектральных характеристик, что позволяет использовать эти технологии в чат-ботах и виртуальных ассистентах.

Для анализа многоканальных аудиозаписей применяются алгоритмы локализации звука, которые позволяют определить источник звука в пространстве.

Современные системы анализа речи используют архитектуры трансформеров, такие как Wav2Vec и Whisper, которые позволяют достигать высокой точности при обработке звуковых данных.

Аудиобиометрия, основанная на анализе индивидуальных особенностей голоса, применяется в системах идентификации и аутентификации пользователей.

Обучение моделей обработки звука требует больших объемов данных, поэтому используются методы увеличения данных, такие как искусственное создание вариаций голосовых записей и смешивание фоновых шумов.

Применение искусственного интеллекта в анализе звука охватывает множество сфер, включая автоматические субтитры, системы управления голосом, медицинскую диагностику и криминалистику.

Заключение Будущее технологий обработки звука связано с развитием более точных и энергоэффективных моделей, способных работать в условиях реального мира с высокой точностью. Нейронных сетей связано с развитием новых архитектур, уменьшением энергопотребления, увеличением интерпретируемости моделей и созданием более универсальных систем, способных решать широкий спектр задач без необходимости узкоспециализированного обучения.

Список литературы

[1] Овездурдыева И, Мырадов М, Технологии работы с большими данными “big data”: сбор, хранение и обработка больших данных Научное издательство «Наука и мировоззрение» 2024 70-74 с

[2] Мурадов М Назарова С, Современные технологии распознавания речи Научное издательство «Наука и мировоззрение» 2024 301-305 с

[3] Мурадов М., Распознавание речи Сборник статей Международной научно-практической конференции Научный потенциал - 2023 Петрозаводск 2023 405-409 с

Авторский вклад

Оразгельдиев Абдымухаммет Ходжамбердиевич-занимался разработкой и применением нейросетевых моделей для распознавания речи и классификации аудиосигналов.

Чарыева Огулхажат Довледовна-внесла значительный вклад в исследование методов обработки звуковых данных, включая предварительную обработку и выделение признаков.

Караджаева Джемал Язырадовна-провела анализ современных технологий акустического и спектрального моделирования, а также их применения в биометрической идентификации.

NEURAL NETWORKS: MATHEMATICAL MODELS USED TO SOLVING COMPLEX PROBLEMS SUCH AS PATTERN RECOGNITION, CLASSIFICATION AND FORECASTING AND METHODS OF PROCESSING AND ANALYZING AUDIO DATA INCLUDING SPEECH RECOGNITION, SOUND CLASSIFICATION, FEATURE EXTRACTION.

A.Kh. Orazgeldiev
*Lecturer, Department of
Information Systems,
Institute of
Telecommunications and
Informatics of
Turkmenistan*

O.D. Charyeva
*Lecturer of the
Department of
Information Systems
of the Institute of
Telecommunications
and Informatics of
Turkmenistan*

D.Ya. Karadjaeva
*Lecturer, Department of
Information Systems,
Institute of
Telecommunications and
Informatics of
Turkmenistan*

Abstract: Methods for processing and analyzing audio data include pre-processing, feature extraction, and classification of audio signals. Speech recognition converts sound into text using acoustic and language models. Sound classification is used to identify different types of audio signals, including ambient noise and emotions in the voice. Modern technologies such as neural networks and transformers significantly improve the accuracy of sound analysis. Application areas include voice assistants, biometric identification systems, and automatic subtitling.

Keywords: Sound processing, speech recognition, audio classification, feature extraction, neural networks, acoustic modeling, language model, Fourier transform, spectral analysis, biometric identification