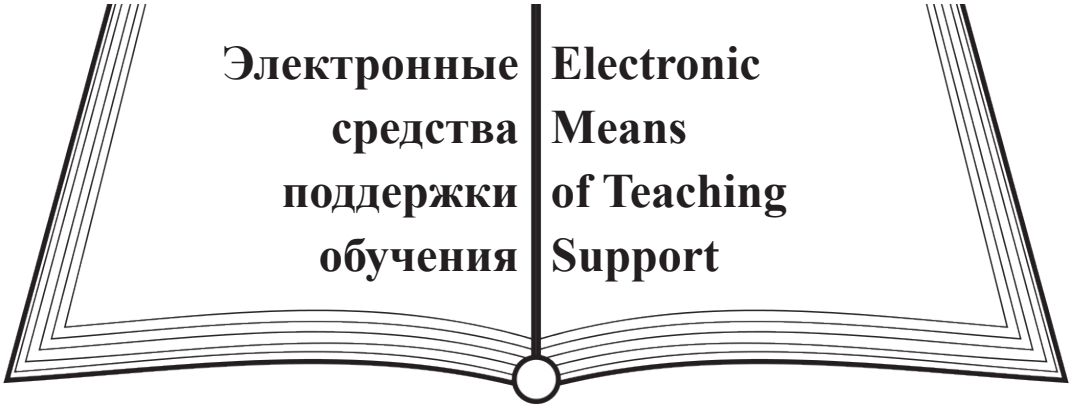


Научная статья
УДК 004.89
DOI: 10.25688/2072-9014.2023.66.4.06

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБРАБОТКИ ЗВУКОВЫХ СИГНАЛОВ В ОБРАЗОВАТЕЛЬНОМ ПРОЦЕССЕ

Виталий Алексеевич Кудинов[1] ✉,
Дмитрий Владиславович Водолад[2]

[1, 2] Курский государственный университет,
Курск, Россия

[1] kudinovva@yandex.ru ✉
[2] dima_v2014@mail.ru

Аннотация. Статья освещает отдельные аспекты обработки звуковых сигналов с использованием нейронных сетей в процессе образования. Рассматриваются шаги предварительной обработки данных, включая преобразование звуковых сигналов в числовой формат и их нормализацию. Особое внимание уделяется использованию спектральных представлений, таких как спектрограмма и мел-спектрограмма, и их роли в анализе звукового содержания. Подробно описывается процесс преобразования фреймов в спектральное представление, вычисления мел-частотных кепстральных коэффициентов (*англ.* mel-frequency cepstral coefficient, MFCC). В качестве иллюстраций представлены примеры спектрограмм и мел-спектрограмм. В целом статья предоставляет обзор основных особенностей обработки звуковых сигналов и является полезным ресурсом для исследователей и практиков в области анализа звука и обработки аудиоданных.

Ключевые слова: применение звуковых сигналов в процессе образования; нейронные сети; спектральные представления; спектрограмма; мел-спектрограмма; нормализация данных; образование.

Original article
UDC 004.89
DOI: 10.25688/2072-9014.2023.66.4.06

APPLICATION OF NEURAL NETWORKS FOR SOUND SIGNAL PROCESSING IN THE EDUCATIONAL PROCESS

Vitaly A. Kudinov[1] ✉,
Dmitry V. Vodolad[2]

[1, 2] Kursk State University,
Kursk, Russia

[1] kudinovva@yandex.ru ✉
[2] dima_v2014@mail.ru

Abstract. This article highlights some aspects of audio signal processing using neural networks in the educational process. The article discusses the steps of data preprocessing, including the conversion of audio signals into a numeric format and their normalization. Particular attention is paid to the use of spectral representations, such as the spectrogram and the chalk spectrogram, and their role in the analysis of sound content. The process of converting frames into a spectral representation, calculating the mel-frequency cepstral coefficients (MFCC) is described in detail. Examples of spectrograms and chalk spectrograms are presented as illustrations. In general, the article provides an overview of the main features of audio signal processing and is a useful resource for researchers and practitioners in the field of sound analysis and audio data processing.

Keywords: application of sound signals in the educational process; neural networks; spectral representations; spectrogram; chalk spectrogram; data normalization.

Для цитирования: Кудинов В. А. Применение нейронных сетей для обработки звуковых сигналов в образовательном процессе / В. А. Кудинов, Д. В. Водолад // Вестник МГПУ. Серия «Информатика и информатизация образования». 2023. № 4 (66). С. 67–77.

For citation: Kudinov V. A. Application of neural networks for sound signal processing in the educational process / V. A. Kudinov, D. V. Vodolad // MCU Journal of Informatics and Informatization of Education. 2023. № 4 (66). P. 67–77.

Введение

Изменения, происходящие в настоящее время в системе образования, указывают на потребность в создании эффективных компьютерных технологий для обучения. Так, системы автоматического распознавания речи могут быть использованы для разработки интеллектуальных обучающих программ, которые помогут учащимся совершенствовать навыки говорения и понимания на иностранных языках. Также звуковая информация

может быть применена для создания интерактивных заданий, требующих анализа звуковых сигналов, например в области музыки или звукозаписи.

Обработка звука способствует повышению доступности образования для различных групп обучающихся. Звуковые материалы могут быть использованы для создания аудиоуроков, аудиокниг и аудиовизуальных материалов, которые помогут обучающимся с нарушениями зрения и слуха получить качественное образование. Кроме того, звуковая информация может быть оцифрована и использована для создания аудиоколлекций и баз данных, предоставляющих доступ к разнообразным звуковым ресурсам и исследовательским материалам.

Одной из перспективных техник является использование в этой области нейронных сетей, которые позволяют анализировать и интерпретировать звуковую информацию, а также извлекать полезные данные для ее дальнейшего использования в образовательных процессах.

В данной статье будет представлен обзор особенностей обработки звуковых сигналов с использованием нейронных сетей и их применения в образовательном процессе. Будет рассмотрена предварительная обработка данных, включая преобразование звуковых сигналов в числовой формат и нормализацию. Особое внимание будет уделено спектральным представлениям, таким как спектрограмма и мел-спектрограмма, и объяснен процесс их получения. Кроме того, будут рассмотрены вычисление мел-частотных кепстральных коэффициентов (*англ.* mel-frequency cepstral coefficient, MFCC) и их важность в анализе звукового содержания. В заключение будет подчеркнута значимость нормализации данных для обеспечения стабильности и эффективности обучения нейронных сетей.

Методы исследования

Применение звука в образовательном процессе представляет собой значимую область исследований, предлагающую разнообразные возможности для обогащения и оптимизации учебного процесса. Звук как средство коммуникации и восприятия занимает важное место в образовательном процессе, способствуя более эффективной передаче информации и совершенствованию учебного опыта. Применение звукового оборудования, такого как динамики, микрофоны и звуковые системы, обеспечивает педагогическому персоналу возможность осуществления голосовой коммуникации, что способствует лучшему усвоению и запоминанию учебного материала. Аудиальные объяснения и лекции могут предоставить более увлекательное интерактивное обучение, что особенно важно для учащихся, ориентированных на аудиальный тип восприятия.

Звук также может быть использован для создания интерактивных образовательных заданий, где аудиофайлы и звуковые эффекты интегрируются в учебные материалы. Это позволяет обучающимся активно взаимодействовать с учебным контентом, развивая навыки аудирования, анализа и идентификации звуковых

элементов. Путем включения звуковых компонентов в учебные задания создается возможность более глубокого погружения обучающихся в изучаемый материал.

Применение звука в образовательной деятельности также связано с использованием современных технологий, таких как виртуальная реальность (*англ.* virtual reality, VR) и дополненная реальность (*англ.* augmented reality, AR). Эти технологии позволяют создавать иммерсивные образовательные среды, в которых звуковые эффекты, аудиальные сигналы и визуальные элементы помогают обучающимся более реалистично и глубоко погрузиться в учебный материал, что способствует лучшему запоминанию и пониманию изучаемой информации.

Кроме того, звук играет важную роль в музыкальном образовании, где он используется для развития музыкальных навыков, анализа музыкальных жанров и стилей. Звуковые технологии позволяют обучающимся изучать музыку, записывать и проигрывать музыкальные композиции, а также создавать собственные музыкальные произведения.

Таким образом, использование звука в процессе обучения предоставляет уникальные возможности для повышения качества образования, активного участия учащихся и развития различных навыков. Внедрение звуковых технологий обеспечивает улучшение коммуникации, взаимодействия и восприятия информации, что в итоге способствует более эффективному и привлекательному обучению.

Для понимания сущности обработки звуковой информации необходимо подробно рассмотреть отдельные аспекты этого процесса. Предварительная обработка данных в контексте обработки звуковых сигналов с использованием нейронных сетей является важным этапом, который включает несколько шагов для подготовки сигналов к входу в нейронную сеть. Каждый из этих шагов играет ключевую роль в обеспечении эффективности и точности обработки звуковых данных.

Преобразование аудиосигналов в числовой формат является первым важным шагом предварительной обработки данных в обработке звуковых сигналов с использованием нейронных сетей. Этот процесс позволяет перевести аналоговый звуковой сигнал в формат, который может быть обработан нейронной сетью.

Основной метод преобразования звуковых сигналов в числовой формат — аналого-цифровое преобразование (*англ.* analog-to-digital converter, ADC). ADC преобразует аналоговый сигнал, представляющий собой непрерывные изменения звукового давления во времени, в последовательность дискретных чисел, называемых отсчетами. Эти отсчеты показывают значения звукового сигнала на определенные моменты времени и являются основой для дальнейшей обработки в нейронной сети.

Процесс аналого-цифрового преобразования состоит из двух основных этапов: семплирования и квантования.

Семплирование — процесс выборки значений аналогового сигнала на определенных интервалах времени: звуковой сигнал берется с определенной частотой, которая называется частотой семплирования[1] [1].

Частота семплирования должна быть достаточно высокой, чтобы предотвратить потерю информации и артефактов при преобразовании аналогового сигнала в цифровой. Согласно теореме Найквиста – Котельникова – Шеннона (теореме о выборке) частота семплирования должна быть по крайней мере вдвое выше наивысшей частоты входного сигнала (теоретический предел Найквиста).

Квантование — процесс присвоения дискретных значений отсчетам звукового сигнала. Поскольку цифровые данные представляют собой конечный набор значений, необходимо округлить значения отсчетов до ближайшего допустимого дискретного значения.

Квантование определяется разрядностью ADC, которая определяет количество бит, используемых для представления каждого отсчета. Чем больше разрядность, тем больше возможных дискретных значений и тем выше разрешение преобразования.

Примером является 16-битное квантование, где каждый отсчет представляется 16-битным числом, имеющим 65 536 возможных значений (от –32 768 до 32 767). Таким образом, аналоговые значения звукового сигнала преобразуются в дискретные значения, которые могут быть представлены и обработаны нейронной сетью.

После преобразования звуковых сигналов в числовой формат применяется разбиение на небольшие фрагменты, которые называются фреймами (от *англ.* frame — кадр, рамка) или окнами. Разделение на фреймы позволяет учесть временные характеристики звуковых сигналов и обеспечить локальность анализа во времени.

Процесс разбиения на фреймы включает следующие шаги:

1. Определяется фиксированная длительность каждого фрейма, которая обычно выбирается в зависимости от особенностей обрабатываемых звуковых сигналов и требуемой разрешающей способности. Например, типичная длительность фрейма составляет от 10 до 30 миллисекунд.

2. Для обеспечения плавного перехода между фреймами обычно используется перекрытие. Это означает, что окно фрейма начинается незадолго до конца предыдущего фрейма, чтобы сохранить непрерывность и учесть информацию с перекрытием. Типичное значение перекрытия составляет около 50 % длительности фрейма.

3. На каждый фрейм применяется оконная функция, которая позволяет уменьшить артефакты, связанные с переходами на границах фреймов. Оконные функции, такие как как функции Хэмминга, Ханна или Блэкмэна-Харриса,

[1] Анализ аудиоданных с помощью глубокого обучения и Python (часть 1). URL: https://nuancesprog.ru/p/6713/ (дата обращения: 10.06.2023).

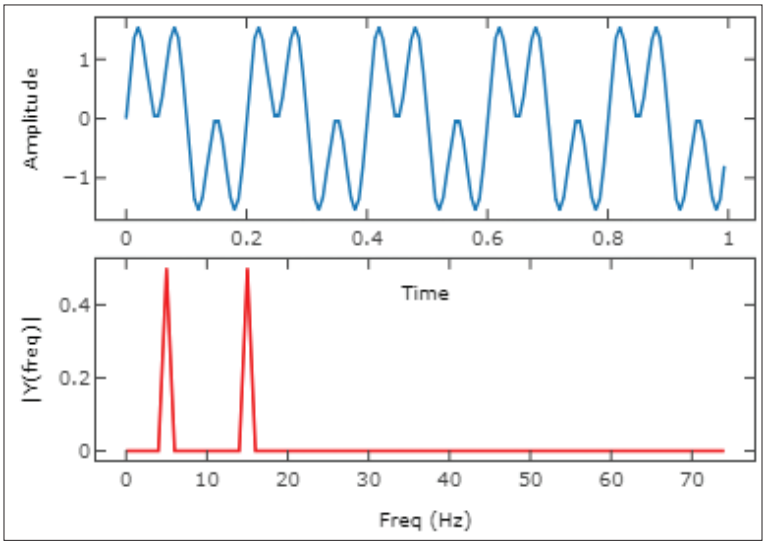
используются для сглаживания фрейма на его границах, что способствует уменьшению боковых лепестков в спектральном представлении.

Результаты исследования

Разделение звуковых сигналов на фреймы позволяет нейронной сети анализировать временные изменения и динамику звуковых данных. Каждый фрейм рассматривается независимо от остальных, что позволяет нейронной сети учитывать локальные характеристики и изменения в звуковых сигналах.

После разделения звуковых сигналов на фреймы следующим шагом в предварительной обработке данных является преобразование фреймов в спектральное представление — спектрограмму или мел-спектрограмму, — позволяющее нейронной сети анализировать частотные характеристики звуковых сигналов.

Спектрограмма и мел-спектрограмма являются графическими представлениями звуковых сигналов во времени и по частоте. Используется преобразование Фурье для вычисления спектральных составляющих звукового сигнала (рис. 1)[2].



Источник: https://machinelearningmastery.ru

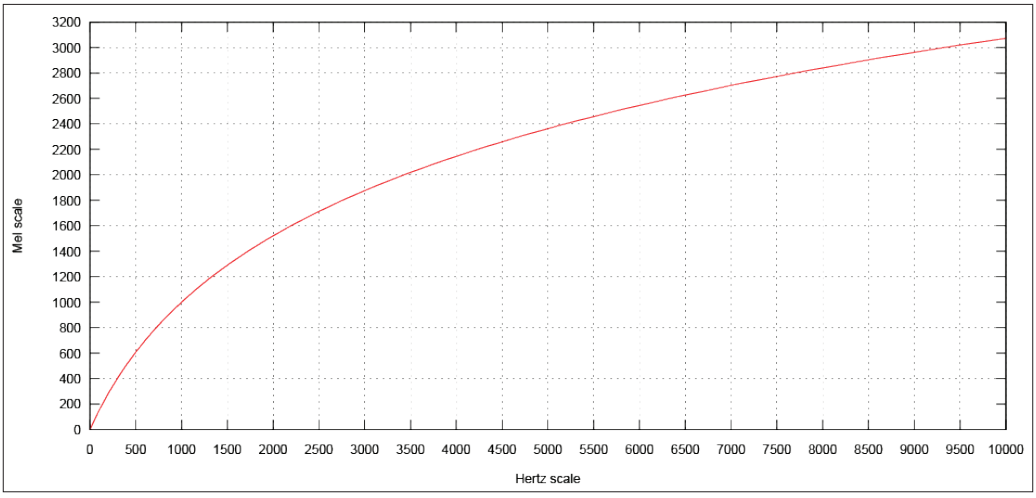
Рис. 1. Преобразование Фурье

Спектрограмма представляет собой двумерное изображение, где по горизонтальной оси отображается время, а по вертикальной — частота. Интенсивность цвета или яркость каждого пикселя в спектрограмме отражает амплитуду, или энергию соответствующей частоты, в каждый момент времени. Обычно спектрограмма строится путем разбиения звукового сигнала, а затем применяется

[2] Harmonic analysis and the Fourier Transform. URL: https://terpconnect.umd.edu/~toh/spectrum/HarmonicAnalysis.html (дата обращения: 18.05.2023).

преобразование Фурье к каждому фрейму. Полученные спектральные составляющие визуализируются с помощью цветовой гаммы: яркие области представляют высокую энергию на определенных частотах.

Мел-спектрограмма является вариантом спектрограммы, где частотная ось масштабируется по мел-шкале вместо линейной. Мел-шкала основана на восприятии человеком различных частот и позволяет лучше выделять различные акустические особенности (рис. 2). Контрольная точка между мел-шкалой и измерением нормальной частоты произвольно определяется путем присвоения перцептивной высоты тона 1000 мелов тону частотой 1000 Гц.



Источник: http://wiki-org.ru/wiki/Мел_(высота_звука)

Рис. 2. Мел-шкала

Приблизительно выше 500 Гц слушатели оценивают все большие интервалы как дающие одинаковые приращения высоты тона. Мел — психофизическая единица высоты звука, применяется в музыкальной акустике, шкала основана на сравнении высоты тона. Формула для преобразования f герц в m мелы выглядит следующим образом:

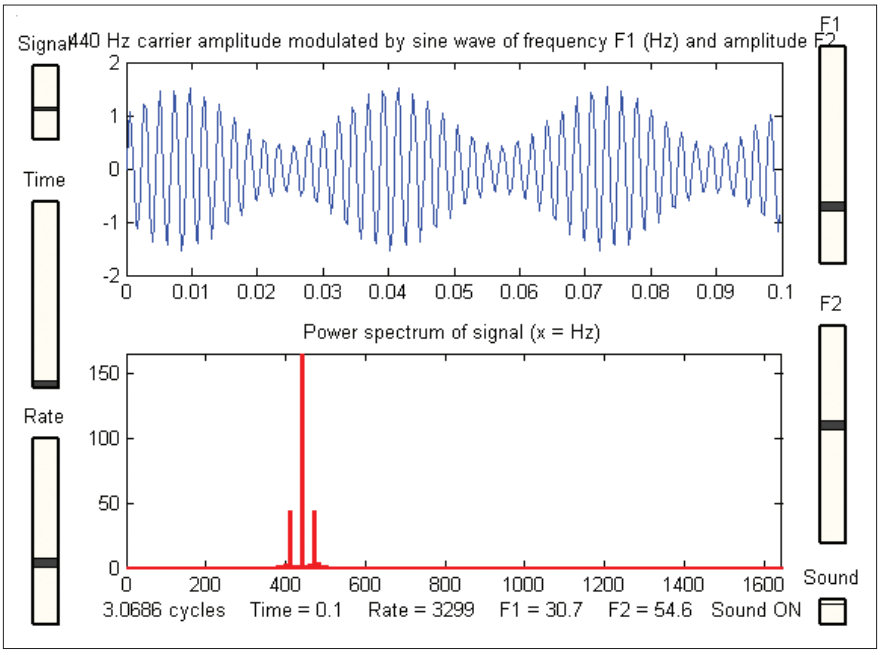
$$m = 2595 \log_{10}\left(1+\frac{f}{700}\right).$$

Для создания мел-спектрограммы применяется преобразование MFCC. Мел-спектрограмма представляет собой звуковой сигнал в спектральной форме на основе мел-шкалы частот. MFCC являются коэффициентами, которые описывают спектральные особенности звукового сигнала в мел-шкале.

Процесс MFCC состоит из нескольких этапов (приведены ниже).

1. *Вычисление энергии спектра.* Спектр мощности временного ряда описывает распределение мощности по частотным компонентам, составляющим этот сигнал. Согласно анализу Фурье, любой физический сигнал можно разложить на ряд дискретных частот (или спектр частот) в непрерывном диапазоне.

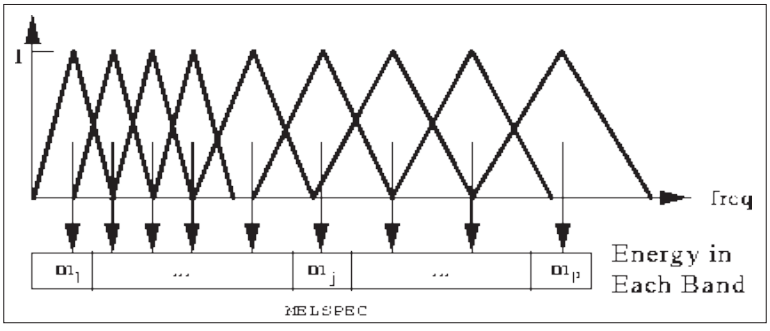
Статистическое среднее определенного сигнала, проанализированного с точки зрения его частотного содержания, называется его спектром. С применением преобразования Фурье к окну фрейма вычисляется спектральная плотность энергии, которая представляет собой распределение энергии сигнала в частотном диапазоне. На рисунке 3 представлен пример вычисления энергии спектра.



Источник: https://terpconnect.umd.edu/~toh/spectrum/AmpMod.GIF

Рис. 3. Вычисление энергии спектра

2. *Применение мел-фильтров.* Вычисленная энергия спектра проходит через набор мел-фильтров, которые имеют равномерное распределение на мел-шкале. Они имитируют способность различать частоты на основе мел-шкалы, которая более близка к восприятию звуков человеком. На рисунке 4 представлена шкала мел-фильтров.



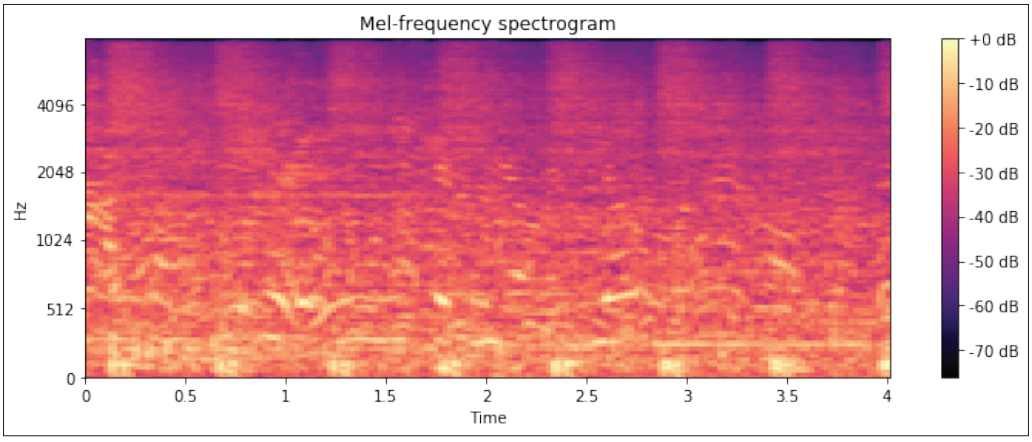
Источник: https://machinelearningmastery.ru

Рис. 4. Шкала мел-фильтров

3. *Логарифмирование.* После прохождения через мел-фильтры полученные значения подвергаются логарифмированию, которое выполняется для сжатия динамического диапазона и увеличения различимости низкоэнергетических компонент.

4. *Преобразование кепстральных коэффициентов.* Представляет собой преобразование Фурье. Позволяет описать спектральные характеристики сигнала в частотном диапазоне.

Процесс MFCC генерирует набор коэффициентов MFCC, которые представляют спектральные особенности звукового сигнала в мел-шкале. На рисунке 5 представлена мел-частотная спектрограмма семпла (аудиофайл) в наборе данных Urbansound8k.



Источник: https://www.pvsm.ru/python/269432

Рис. 5. Мел-частотная спектрограмма

Нормализация данных является важным шагом в предварительной обработке звуковых сигналов перед подачей их на вход нейронной сети.

Один из наиболее распространенных подходов к нормализации данных — это стандартизация. При стандартизации каждый спектральный коэффициент вычитается из его среднего значения и делится на стандартное отклонение по всему набору данных. Это приводит к тому, что коэффициенты имеют среднее значение, равное 0, и стандартное отклонение, равное 1. Такой подход обеспечивает более стабильное и сопоставимое распределение значений, что может улучшить обучение нейронной сети.

Другой подход к нормализации данных — это нормализация по минимуму и максимуму. При этом подходе значения спектральных коэффициентов масштабируются таким образом, чтобы они находились в определенном диапазоне, например от 0 до 1. Для этого каждый коэффициент вычитается из минимального значения и делится на разницу между максимальным и минимальным значениями в наборе данных.

В некоторых случаях может быть полезно нормализовать значения спектральных коэффициентов так, чтобы они находились в определенной шкале,

не обязательно от 0 до 1. Например, можно привести значения коэффициентов к шкале от –1 до 1 или к другому диапазону, соответствующему требованиям задачи.

Нормализация данных позволяет сделать значения спектральных коэффициентов более устойчивыми к изменениям и более пригодными для эффективного обучения нейронной сети, что позволяет сети лучше обобщать данные и делает обучение более стабильным. Кроме того, нормализация данных помогает избежать проблемы влияния различных масштабов значений на обучение сети, так как значения входных данных будут сопоставимыми и однородными.

Заключение

Применение звука в обучении играет значительную роль в обогащении и оптимизации учебного процесса.

Обработка звуковых сигналов с использованием нейронных сетей играет ключевую роль в эффективном и точном анализе звукового контента. Основные этапы предварительной обработки данных, такие как преобразование в числовой формат, разбиение на фреймы, спектральное представление и вычисление мел-частотных кепстральных коэффициентов, позволяют сетям лучше понимать и анализировать звуковые сигналы. Нормализация данных также играет важную роль, обеспечивая стабильность и эффективность обучения нейронных сетей.

Использование звука в образовательной деятельности открывает новые возможности для создания инновационных и точных решений в области обработки звука. Применение звуковых технологий в процессе обучения способствует улучшению коммуникации и восприятию информации. Таким образом, эффективное использование звуковых сигналов является важным аспектом, способствующим развитию современных образовательных практик и обогащению учебного опыта обучающихся.

Список источников

1. Игнатенко Г. С. Классификация аудиосигналов с помощью нейронных сетей / Г. С. Игнатенко, А. Г. Ламчановский // Молодой ученый. 2019. № 48 (286). С. 23–25.

References

1. Ignatenko G. S. Classification of audio signals using neural networks / G. S. Ignatenko, A. G. Lamchanovsky // Young scientist. 2019. № 48 (286). P. 23–25.

Статья поступила в редакцию: 20.06.2023;
одобрена после рецензирования: 04.09.2023;
принята к публикации: 11.09.2023.

The article was submitted: 20.06.2023;
approved after reviewing: 04.09.2023;
accepted for publication: 11.09.2023.

Информация об авторах / Information about authors:

Виталий Алексеевич Кудинов — доктор педагогических наук, профессор, профессор кафедры программного обеспечения и администрирования информационных систем, Курский государственный университет, Курск, Россия.

Vitaly A. Kudinov — Doctor of Pedagogical Sciences, Professor, Professor of the Department of Software and Administration of Information Systems, Kursk State University, Kursk, Russia.

kudinovva@yandex.ru ✉

Дмитрий Владиславович Водолад — студент факультета физики, математики, информатики, Курский государственный университет, Курск, Россия.

Dmitry V. Vodolad — Student of the Faculty of Physics, Mathematics, Computer Science, Kursk State University, Course, Russia.

dima_v2014@mail.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.