

Висновок. В ході виконання даної роботи проведено дослідження у сфері синтезу мовлення, проаналізовано методи синтезу голосового мовлення та знайдено рішення для даної проблеми. Ними виявилися саме нейронні мережі [1], які зарекомендували себе в багатьох сферах.

Для розробки інтелектуального синтезатора голосу використано навчену модель Tacotron 2. В якості навчального набору даних вибрано LJSpeech dataset, який містить 13100 аудіо записів. Результатом є готова система, яка може бути використана для генерації голосової інформації на основі текстової.

Список використаних літературних джерел

- [1] Що таке нейронні мережі та як вони працюють? Класифікація штучних нейромереж [Електронний ресурс] – <https://livingfo.com/shcho-take-nejronni-merezhi-ta-iak-vony-pratsiuiut>
- [2] Моделі комп'ютерного подання українського мовлення для проблеми конкатенативного сегментивного синтезу [Електронний ресурс] – <https://jrnl.nau.edu.ua/index.php/IPZ/article/view/3049/2985>

Андрій Юнко, Ігор Пірко*

магістрант кафедри КН, НЛТУ України, Львів, Україна.

* к.ф.-м.н., доцент кафедри комп'ютерних наук, НЛТУ України, Львів, Україна.

Інтелектуальна система класифікації аудіосигналів засобами машинного навчання

Анотація – У роботі проведено дослідження методів та підходів для аудіоаналізу. Розглянуто форми представлення аудіосигналу щодо його перетворення перед проведенням аналізу. Проаналізовано можливість виділення ознак з аудіосигналу у часовій та частотній областях. Наведено приклади найчастіше використовуваних та найбільш інформативних ознак. Описано способи обчислення, представлений їхній зовнішній вигляд у вигляді графіків.

Ключові слова – машинне навчання, класифікація аудіосигналів, Python, Librosa.

Сучасні комп'ютерні алгоритми, зокрема алгоритми машинного навчання дозволяють розвиватися області аудіоаналізу. Він включає автоматичне розпізнавання мови, цифрову обробку, класифікацію, пошук, виявлення і генерацію акустичних сигналів. Аналіз аудіоінформації все частіше стає завданням для різноманітних програмних засобів, що використовують математичний апарат прийняття рішень. У зв'язку з цим актуальним є розвиток методів машинного навчання в галузі аудіоаналізу. Глибокі нейронні мережі зарекомендували себе в розпізнаванні зображень і можуть бути застосовані до широкого спектру інших завдань, зокрема, до обробки аудіофайлів, що дозволить проводити класифікацію і класифікацію досліджуваних акустичних сигналів, може бути використано для аналізу зареєстрованих аудіосигналів в задачах акустичного контролю різних об'єктів.

Отримання ознак є важливим етапом як в аудіоаналізі, так і в завданнях розпізнавання образів та машинного навчання загалом. Потрібно отримати набір характеристик, які інформативно відбивають властивості вихідних даних, з набору даних, що цікавить. Це дозволяє зменшити розмірність даних. Після отримання необхідних ознак сигналу для їх подальшого використання проводиться нормалізація ознак. У цьому випадку застосовуються методи скорочення розмірності вектора ознак.

Аудіосигнали аналізуються за допомогою короткотривалої технології обробки, відповідно до якої аудіосигнал розбивається на вікна і аналіз виконується на основі цих кадрів. Під час короткочасної обробки фокусуються на невеликій частині (кадрі) сигналу, тобто на кожному етапі обробки аудіосигнал множиться на зсунуту віконну функцію певної тривалості.

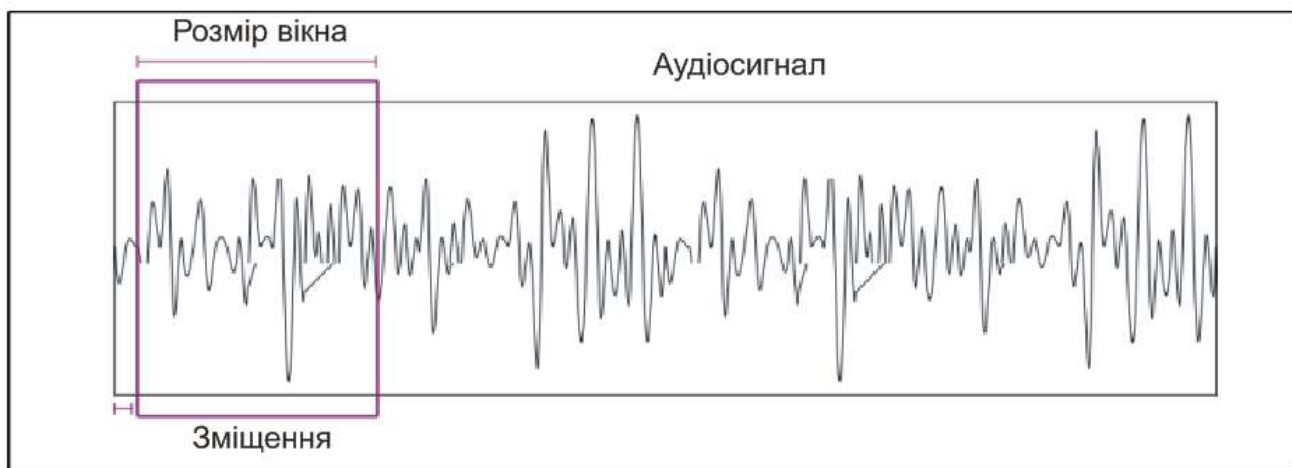


Рис. 1. Короткотривала обробка звукового файлу з допомогою вікна.

Пошук музичної інформації (MIR) – це наука про отримання інформації з музичних аудіосигналів. MIR – це галузь досліджень із багатьма реальними застосуваннями. Ті, хто бере участь у MIR, можуть мати освіту в галузі академічного музикознавства, психоакустики, психології, обробки сигналів, інформатики, машинного навчання, оптичного розпізнавання музики, обчислювального інтелекту. MIR використовується компаніями та науковцями для класифікації, маніпулювання та створення музики. Однією з класичних тем дослідження MIR є класифікація жанрів, що відносить музичний твір до одного із заздалегідь визначених жанрів, таких як класика, джаз, рок тощо. Класифікація настрою, класифікація виконавців, ідентифікація інструментів і тегування музики також є популярними темами.

Звук – це сигнал із набором певних параметрів: частота, смуга пропускання, пропускна здатність, децибели. Типовий звуковий сигнал може бути виражений як функція від амплітуди та часу.

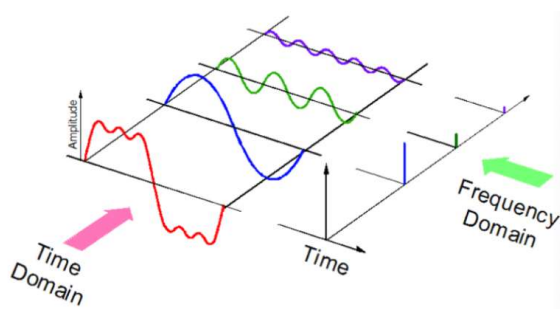


Рис. 2. Структура звукового файлу.

Для роботи зі звуком на Python існує ряд бібліотек, що підтримують базову функціональність. В роботі використовується бібліотека Librosa. Вона може працювати з будь-якими звуковими сигналами, але ця бібліотека орієнтована переважно саме на музику. Вона дозволяє створити повноцінну систему отримання музичної інформації (MIR).

Для отримання графіка амплітудної огинаючої аудіосигналу використовують спектрограму – візуальне подання спектра частот звукових чи інших сигналів, що змінюються з часом. Іноді їх називають сонограмами. На двовимірних графіках по першій осі задається частота, по другій – час. Для її створення на Python використовують метод `librosa.display.specshow` [1].

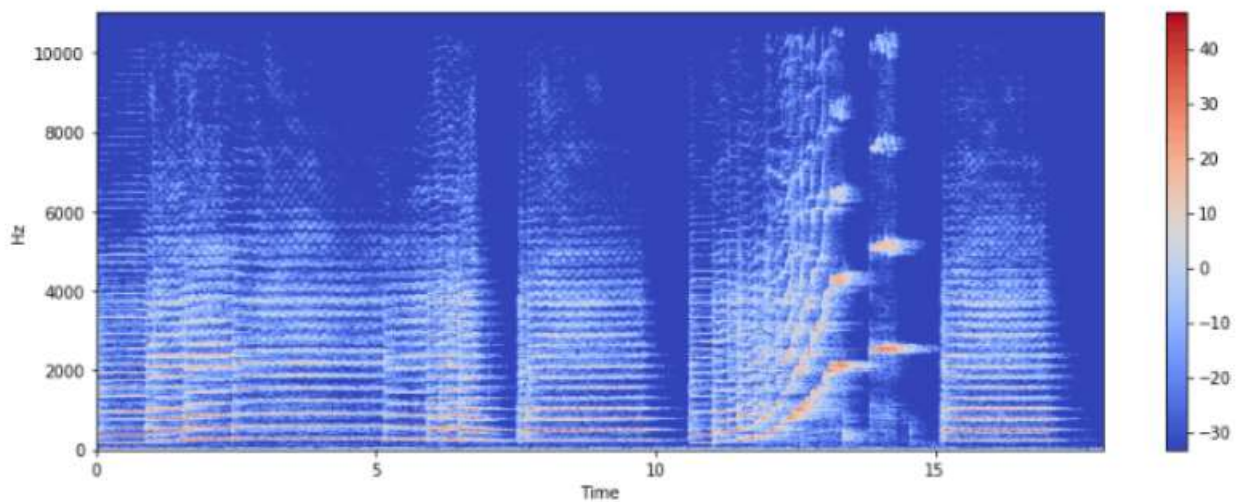


Рис. 3. Спектрограма аудіосигналу.

Кожний звуковий сигнал має багато параметрів, з яких потрібно отримати потрібні. Процес отримання інформації для аналізу називається витягуванням об'єктів (feature extraction). Основна ознака є частота переходу через нуль – це частота зміни знака сигналу, тобто частота, з якою сигнал змінюється з позитивного на негативний і навпаки. Ця функція широко використовується як для розпізнавання мови, так і для отримання музичної інформації.

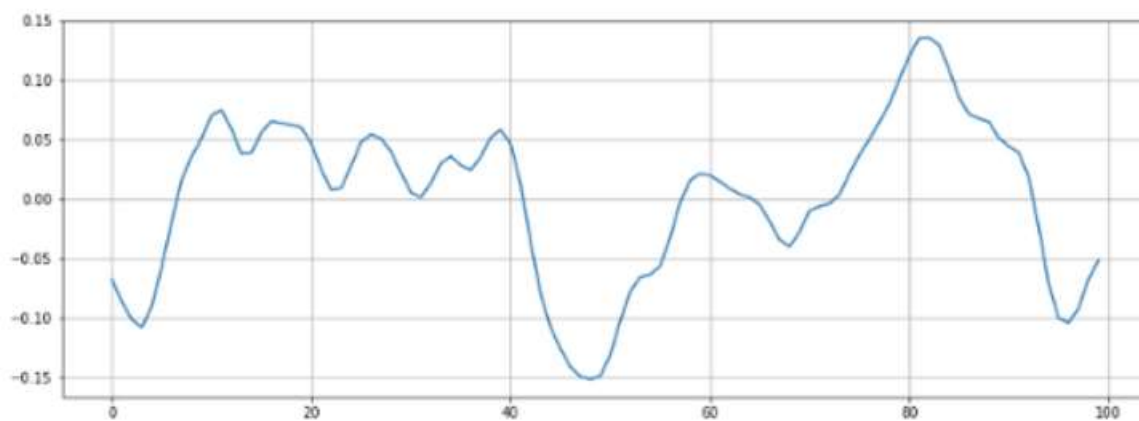


Рис. 4. Частота переходу нуля.

В роботі використано відомий набір аудіоданих, який складається з 1000 звукових доріжок довжиною 30 секунд і містить 10 жанрів: блюз, класика, кантрі, диско, хіп-хоп, джаз, регі, рок, метал та поп. У кожному жанрі є 100 звукових кліпів. Набір даних складається із записаних музичних аудіоданих 8 різних музичних інструментів. Для кожного інструменту семпли охоплюють весь набір тонів, що програватимуться в кожній октаві з різними рівнями сили (піано, форте) і тривалості. Крім того, набір даних також містить зразки, у яких на інструменті використовуються різні техніки гри, наприклад вібрато, тремоло, пічікато та понтічелло. Цей набір інструментів було обрано через високу якість музичних аудіосигналів, які охоплюють три сімейства інструментів: мідні, струнні та дерев'яні духові [2].

Для навчання моделі було обрано такі інструменти:

banjo – банджо;

cello – віолончель;

clarinet – кларнет;

english horn – англійський горн;

guitar – гітара;

trumpet – труба;

violin – скрипка.

Перед навчанням моделі класифікації необхідно перетворити необроблені дані зі звукових вибірок на більш осмислене представлення. Отримують з аудіофайлів всю необхідну інформацію: мел-частотні кепстральні коефіцієнти,

спектральний центроїд, частоту переходу через нуль, частоти колірності, спектральний спад частоти. Після цього можна використовувати існуючі алгоритми класифікації для розподілу музичних інструментів. Можна використовувати безпосередньо спектрограми і застосовувати моделі класифікації на них [3].

Index	Instruments	Samples
1	Banjo	23
2	Cello	166
3	Clarinet	131
4	English Horn	234
5	Guitar	29
6	Oboe	155
7	Trumpet	140
8	Violin	366

Кепстральні коефіцієнти MFCC кожного музичного аудіосигналу було виділено за допомогою бібліотеки Librosa. Для кожного аудіофайлу його MFCC усереднюються для отримання остаточного вектора ознак [4].

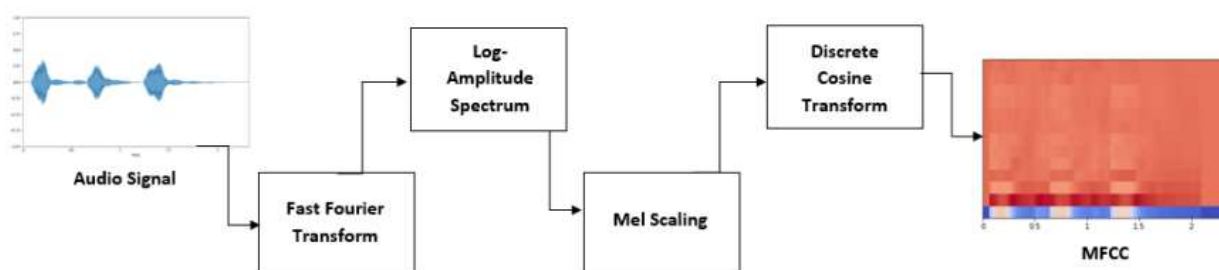


Рис. 5. Кроки для отримання ознак з аудіосигналу.

Висновки. В даній роботі для розпізнавання аудіосигналів музичних інструментів використано метод класифікації k-найближчих сусідів. Проведено порівняння різних прийомів розпізнавання музичного інструменту за звуками однієї ноти. Модель навчалася на музичних аудіофайлах музичних інструментів фіксованої довжини.

Список використаних літературних джерел

- [1] Мясников Л.Л., Мясникова Е. Автоматическое распознавание звуковых образов / – Л. Наука, 1970. – 183 с.
- [2] Аграновский А.В., Леднов Д.А. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов / Москва: Изд-во "Радио и связь", 2004. – 164 с.

- [3] Вологдин Э.И. Методы и алгоритмы обработки звуковых сигналов / – СПб.: СПбГУТ, 2009. – 96 с.
- [4] Петровский А.А. Анализаторы речевых и звуковых сигналов: Методы, алгоритмы и практика / – Минск: Бестпринт, 2009. – 456 с.

Roman Saveliev, Mykhailo Dendiuk*

Ph.D. student, Ukrainian National Forestry University, Lviv, Ukraine.
romansavelyev6@gmail.com.

*senior teacher of CS Department, Ukrainian National Forestry University, Lviv, Ukraine. <https://orcid.org/0000-0002-7631-022X>. dendyuk@nltu.edu.ua.

Application of synthetic data for analytics with generative ai and vector search capabilities

Abstract – This research investigates the transformative potential of Generative Artificial Intelligence (Generative AI) in the realm of data analytics within distributed systems. Specifically, we explore the utilization of Generative AI models, such as Generative Adversarial Networks, to generate synthetic datasets that closely mimic real-world data. Our focus extends to the innovative application of vector search within these synthetic datasets, examining its implications for data analytics and exploring the ethical considerations surrounding the use of synthetic data.

Keywords – Generative AI, Data Analytics, Distributed Systems, Synthetic Data, Vector Search

Generative AI, a groundbreaking technology, introduces a new paradigm in data analytics within distributed systems. The research centers around the application of Generative AI models, for instance Generative Adversarial Networks, to create synthetic datasets that closely mirror real-world data.

Generative AI addresses the constraint of limited access to diverse datasets by enabling the creation of synthetic data. The study delves into the technical intricacies of synthetic data generation, emphasizing the preservation of statistical characteristics from the original datasets.