Київський національний університет імені Тараса Шевченка

Факультет комп'ютерних наук та кібернетики

Кафедра обчислювальної математики

Кваліфікаційна робота на здобуття ступеня бакалавра

за спеціальністю 113 Прикладна математика на тему:

Дослідження алгоритмів розпізнавання звукових доріжок та визначення назви музичного твору по звуковій доріжці за допомогою технологій штучного інтелекту

Виконав студент 4го курсу
Залужний Юрій Андрійович

Науковий керівник:
асистент
Денисов Сергій Вікторович

Засвідчую, що в цій роботі немає запозичень з праць інших авторів без відповідних посилань.

Студент

Роботу розглянуто й допущено до захисту на засіданні кафедри обчислювальної математики
«29»травня 2023 р.,протокол № 8
Завідувач кафедри
С. І. Ляшко

РЕФЕРАТ

Обсяг роботи 54 сторінки, 9 ілюстрацій, 32 використаіх джерела.

Об’єктом дослідження є методи ідентифікації музичного твору за записом його відтворення. Було розглянуто два методи: порівняння відбитку запису відтворення з відомими відбитками та отримання та аналіз тексту з запису музичного твору.

Метою роботи є порівняння ефективності двох зазначених моделей та створення додатку, що за записом відтворення музичного твору ідентифікує його та надає користувачеві його назву.

У роботі було розглянуто існуючі методи ідентифікації записів відтворення музичних творів. Було реалізована алгоритми створення та порівняння відбитків записів та нейронну мережу, що за послідовністю слів у записі визначає назву музичного твору за допомогою мови програмування Python та її фреймворків.

ЗМІСТ

ВСТУП

У сучасному суспільстві люди використовують та споживають музику по-різному та з різних причин. Музика, як правило, є всюди: вона може звучати на задньому плані фільму у кінотеатрі, у популярних відеоіграх або бути частиною соціального медіа-поста. Музика є ключовим елементом у переконанні сердець і розумів людей, тому вона відіграє важливу роль у політиці і часто використовується як засіб впливу і переконання.

Із розвитком технологій способи використання та споживання музики розширюються. Історики встановили, що перші музичні інструменти були створені понад 30 000 років тому - ці інструменти використовувалися для створення обрядової музики, призначеної для релігійних цілей. З того часу музична сфера масово змінилася - від розвитку блюзу за рахунок міграції африканців до Сполучених Штатів, зростання протестів панків у 90-х роках, які використовували музику як спосіб вираження політичних питань, до появи революційного бізнесу Apple, який запропонував можливість зберігати велику кількість музики у кишенькових пристроях та використання музики на сучасних цифрових платформах, таких як YouTube, що дають можливість прослуховувати музику у будь-якому куточку світу.

На сьогоднішній день YouTube відповідає за 4% інтернет-трафіку у всьому світі. Цей сайт став платформою для безлічі безкоштовних відеоматеріалів, багато з яких супроводжуються музикою. Однак часто ці музичні твори використовуються без дозволу автора, що призводить до суттєвих матеріальних втрат. Отже, з’являється необхідність у автоматичному аналізі матеріалів з метою виявлення тих, що порушують авторські права.

У наш час музика лунає з усіх кутків. Часто трапляються ситуації, коли людина чує певну композицію і хоче дізнатися її назву, але не має можливості дізнатися що це за

твір. Тоді в нагоді стають додатки, що використовуючи мікрофон телефона, визначають трек та його автора.

Метою цієї роботи є порівняння різних методів визначення назви музичного твору за його записом та розробка додатку на базі мови програмування Python та її фреймворків із можливістю завантаження звукової доріжки, подальшій її обробці за допомогою алгоритму, і повернення користувачу результатів аналізу із зазначеним твором та його автором.

Результатом проведеної роботи буде готовий до подальшого використання додаток для аналізу та визначення музичних творів.

1. Історія появи та розвитку алгоритмів розпізнавання звуку

1.1 Передумови та поява алгоритмів розпізнавання звуку

Понад 120 мільйонів років тому розпізнавання звуків було примітивним інстинктом виживання ранніх ссавців, тим не менш, це найдоступніший сенсорний канал для нас у сучасну епоху. Індивідуальна перцепція одного і того ж музичного фрагменту може значно варіюватися від особи до особи. Проводиться велика кількість наукових досліджень, метою яких є глибокий аналіз та розуміння механізмів сприйняття звуків людиною. Отримана від зовнішнього середовища звукова інформація, що становить близько чверті від загальної кількості інформації, яку людина отримує, розпізнається та обробляється за допомогою слухової системи та вищих відділів головного мозку. Слухова система складається з периферійної частини і вищих відділів слухової системи. Наразі найбільш детально вивчені процеси, що відбуваються у периферійній частині. Периферична частина слухової системи традиційно розділяється на три компоненти: зовнішнє, середнє та внутрішнє вухо. Зовнішнє вухо формується вушною раковиною та слуховим каналом, який завершується тонкою мембраною, відомою як барабанна перетинка. Зовнішні вуха в поєднанні з головою створюють компоненти зовнішньої акустичної антени, яка взаємодіє з барабанною перетинкою та зовнішнім звуковим полем. Середнє вухо характеризується як повітряна порожнина, що взаємодіє з носоглоткою та євстахієвою трубою для регулювання атмосферного тиску. У результаті такого поєднання, при зміні атмосферного тиску, повітря має змогу витікати та проникати до середнього вуха, відповідно, барабанна перетинка не відзначає повільні зміни статичного тиску. Внутрішнє вухо розташоване в комплексі каналів у скроневій кістці, включаючи в себе орган рівноваги, відомий як вестибулярний апарат, та равлик. Вищі відділи слухової системи функціонують як специфічний логічний процесор, дозволяючи особі відокремлювати значущі звуки від шумів, класифікувати їх за певними характеристиками, використовуючи пам'ять для порівняння з вже наявними

акустичними образами, визначати їх інформаційну цінність та формулювати рішення щодо подальших дій.

Вивчення сприйняття людиною звуку є основою для розуміння, як ми взаємодіємо з нашим звуковим оточенням. В той же час, це поле знань стало каталізатором для розвитку технологій розпізнавання звуку. Furui надає всебічний огляд прогресу, досягнутого в технології розпізнавання мовлення і дикторів за останні 50 років. Бюхлер у 2002 році провів дослідження, де зосереджується на розробці автоматичної системи класифікації звуків для слухових апаратів, а Баунтуракіс у 2015 році досліджував алгоритми машинного навчання для розпізнавання звуків навколишнього середовища. Маклафлін описує систему класифікації звукових подій, яка порівнює зовнішні ознаки слухового образу із зовнішніми ознаками на основі зображення спектрограми, використовуючи машину опорних векторів і класифікатори глибоких нейронних мереж. Загалом, дослідження зазначених вчених демонструють, що алгоритми розпізнавання звуку еволюціонували від зіставлення звуків з шаблоном до статистичного моделювання, і від евристичної часової нормалізації до методів, заснованих на теорії ймовірності. У дослідженнях також підкреслюється важливість виділення ознак і використання методів машинного навчання для надійного розпізнавання звуку в шумному середовищі.

Диктофони, що були вперше винайдені Томасом Едісоном наприкінці XIX століття, відкрили нові можливості для запису мовлення і стали незамінними помічниками для лікарів та секретарів, які щоденно здійснювали велику кількість аудіозаписів. Проте значний прогрес у впровадженні технології розпізнавання мови було досягнуто лише в 1950-х роках. До цього часу існували тільки спроби запису мовлення, але не його інтерпретації. У пізнішій період, Одрі, машина розроблена у Bell Labs, демонструвала здатність розуміти цифри від 0 до 9 із точністю 90%, але лише тоді, коли до неї говорив її розробник. При спілкуванні з іншими особами цей показник коливався в межах 70-80%. Цей факт вказує на одну з постійних проблем в сфері розпізнавання мови - унікальність кожного голосу та велика варіативність

розмовної мови. Відмінно від писемного тексту, який володіє високим ступенем стандартизації, голосова мова суттєво варіюється в залежності від регіональних діалектів, темпу, інтонації, а навіть соціального статусу та статі виконавця, що робить масштабування будь-якої системи розпізнавання мови важким завданням. Олександр Вайбель, розробник машини Harpy в Університеті Карнегі-Меллона, котра могла розуміти більше 1 000 слів, ілюстрував цю складність на прикладах зі озвучуванням однакових слів у різних контекстах. До 1990-х років, навіть найпрогресивніші системи базувалися на методах шаблонного співставлення, де аудіохвилі трансформувалися у числовий формат та зберігалися для подальшого порівняння. Ці системи активувались лише при співпадінні вхідного звуку із збереженим шаблоном. Як наслідок, для успішного розпізнавання потрібно було вимовляти слова дуже чітко та повільно, а також уникати будь-якого шумового середовища.

IBM Tangora, що була розроблена середині 1980-х років і названа на честь Альберта Тангори, на той момент найшвидшого друкаря у світі, мала змогу адаптуватися до особливостей голосу конкретного користувача. Незважаючи на те, що вона на той час вимагала повільної, чіткої вимови та відсутності шуму в середовищі, впровадження прихованих марковських моделей забезпечило більшу гнучкість за рахунок кластеризації даних та передбачення наступних фонем на основі попередніх шаблонів. Незважаючи на те, що для кожного користувача було необхідно надати 20 хвилин навчальних даних у формі записаного мовлення, Tangora здобула здатність розпізнавати до 20 000 англійських слів та кілька повних речень. Існує визнання, що ефективність розпізнавання мови залежить від здатності адаптуватися до унікального стилю спілкування кожної окремої особи, але перехід до цієї моделі представляв собою значний виклик. Лише у 1997 році було представлено перший у світі безперервний розпізнавач мови, а саме програмне забезпечення Dragon's NaturallySpeaking: з його введенням зникла необхідність робити паузи між кожним окремим словом. Здатний розпізнавати 100 слів за хвилину, цей продукт зберігає

актуальність і сьогодні (хоча і в дещо модифікованій формі) та користується популярністю, зокрема серед лікарів для здійснення записів.

Машинне навчання, подібно до багатьох інших областей наукового прогресу, забезпечило більшість значних досягнень у сфері розпізнавання мови протягом цього століття. Google інтегрував передові технології з потужностями хмарних обчислень для обміну даними та покращення точності алгоритмів машинного навчання. Вершиною цього процесу стало впровадження програми Google Voice Search для iPhone у 2008 році. Базуючись на великому обсязі тестових даних, програма Voice Search демонструвала значне покращення точності порівняно з передовими технологіями розпізнавання мови. Використовуючи ці досягнення, Google інтегрував персоналізовані елементи в результати голосового пошуку, а також застосував ці дані для розробки алгоритму Hummingbird дозволяє отримати більш глибоке розуміння мови. Всі ці процеси були об'єднані в Google Assistant, сервіс, який в даний час використовується приблизно на 50% всіх смартфонів. Однак, Siri — розробка Apple в області розпізнавання голосу — стала першим продуктом, що привернув значну увагу громадськості. В результаті десятирічних наукових досліджень було досягнуто створення цифрового асистента, що грунтується на штучному інтелекті і який приніс значний внесок у покращення якості розпізнавання мови, надаючи йому людський відтінок. Дана технологія відкриває широкі перспективи для розвитку та вдосконалення взаємодії між людьми та машинами у сферах, що вимагають обробки та розуміння людської мови. Після появи Siri, компанія Microsoft представила Cortana, Amazon випустила Alexa, ініціюючи тим самим змагання між технологічними гігантами за лідерство в області розпізнавання мови.

Уже протягом сотень років людство приділяло значні зусилля для досягнення прогресу в галузі розвитку штучного інтелекту, зокрема в процесі вдосконалення автоматизованих систем. Одним із ключових завдань цього довготривалого процесу була реалізація здатності машин пройти той самий шлях, який середньостатистична людина успішно подолає лише за кілька років. Саме завдяки безперервному

дослідженню і розвитку інтелектуальних систем, вдалось досягти значних досягнень в цій сфері. Використовуючи сучасні методи машинного навчання та глибокого навчання, вдалось розробити алгоритми та моделі, які можуть ефективно виконувати завдання, які раніше вимагали значної людської праці і часу. Теперішні алгоритми машинного навчання можуть розпізнати та інтерпретувати людську мову з точністю, яка наближається до стовідсоткової. Методи, які стимулювали цей значний прогрес, так розроблені, що їх робота вже починає збігатися з механізмами обробки інформації людським мозком. Завдяки інтеграції хмарних технологій в обчислювальні системи та їх здатності розпізнавати голосові команди, вже успішно було впроваджено їх в мільйони пристроїв, надаючи можливість клієнтам або ж користувачам отримувати діалогові відповіді на різноманітні запити.

1.2 Наявні ідеї та методи розпізнавання звуку

У дослідженнях припускають, що штучний інтелект може допомогти з розпізнаванням звуку. Сін у 2016 році представив метод, що поєднує оптимальне вейвлет-пакетне перетворення та штучну нейронну мережу для розпізнавання якості звуку, який демонструє хорошу точність у розпізнаванні шумів транспортних засобів. Серезуела-Ескудеро того ж року досліджує можливості системи класифікації звуку, яка поєднує нейроморфну слухову систему для виділення ознак і штучну нейронну мережу для класифікації, причому нейронна мережа із відривом досягає кращої точності в розпізнаванні звуків за наявності білого шуму. Карбонелла пояснює, як методи штучного інтелекту, можуть бути корисними для розуміння мови, включаючи акустико-фонетичне декодування та фонологічну інтерпретацію. Маклафлін описує систему класифікації звукових подій, яка порівнює зовнішні ознаки слухового образу зі зовнішніми ознаками на основі спектрограми, використовуючи машину опорних векторів і класифікатори глибоких нейронних мереж, і зрештою демонструє хороші результати в розпізнаванні звукових подій в реальних шумних умовах.

Також у дослідженнях йдеться про те, що існують різні типи алгоритмів, які можна використовувати для розпізнавання звуку. Валеро порівнює методи

машинного навчання для розпізнавання звукового ландшафту, включаючи дерева рішень, алгоритм K-найближчих сусідів, моделі гаусових сумішей і нейронні мережі. Чмулик пропонує загальну систему розпізнавання звуку, яка використовує еволюційні алгоритми для вибору дискримінантних акустичних ознак. Баунтуракіс досліджує методи автоматичного розпізнавання та класифікації дискретних звуків навколишнього середовища, порівнюючи існуючі методи, які довели свою ефективність у розпізнаванні мови та музики. Бюхлер розробляє систему автоматичної класифікації звуків для застосування в слухових апаратах, використовуючи механізми аналізу слухової сцени для виокремлення слухових ознак. Загалом, зазначені дослідження та роботи свідчать про те, що методи машинного навчання та алгоритми оптимізації можуть бути використані для вибору і класифікації акустичних ознак для розпізнавання звуків, і що різні методи можуть бути більш придатними для різних типів завдань розпізнавання звуків.

Ефективність технології розпізнавання мовлення оцінюється за рівнем точності, тобто частотою помилок у словах (WER), і швидкістю. На частоту помилок може впливати низка факторів, таких як вимова, акцент, висота тону, гучність і фоновий шум. Досягнення людського паритету - тобто рівня помилок на рівні двох людей, які розмовляють - вже давно є метою систем розпізнавання мови. Для розпізнавання мови текстом і підвищення точності транскрипції використовуються різні алгоритми та обчислювальні методи. Нижче представлено короткі пояснення деяких з найбільш поширених методів:

- ***Обробка природної мови (NLP):*** хоча NLP не обов'язково є конкретним алгоритмом, що використовується для розпізнавання мови, але все ж таки, це область штучного інтелекту, яка фокусується на взаємодії між людьми і машинами за допомогою мови через усне мовлення та текст. Багато мобільних пристроїв включають розпізнавання мовлення у свої системи для здійснення голосового пошуку, наприклад, Siri, або для забезпечення більшої доступності текстових повідомлень.

- ***Приховані марковські моделі (HMM):*** приховані марковські моделі базуються на ланцюговій моделі Маркова, яка передбачає, що ймовірність певного стану залежить від поточного стану, а не від його попередніх станів. У той час як ланцюгова модель Маркова корисна для спостережуваних подій, таких як введення тексту, приховані марковські моделі дозволяють включити в імовірнісну модель приховані події, такі як мітки частин мови. Вони використовуються як моделі послідовностей у розпізнаванні мови, призначаючи мітки кожній одиниці - тобто словам, складам, реченням тощо - у послідовності. Ці мітки створюють відображення з наданими вхідними даними, що дозволяє визначити найбільш відповідну послідовність міток.
- ***N-грами:*** найпростіший тип мовної моделі (ММ), який призначає ймовірності реченням або фразам. N-грама - це послідовність з N слів. Наприклад, "order the pizza" - це триграма або 3-грама, а "please order the pizza" - 4-грама. Граматика та ймовірність певних послідовностей слів використовуються для покращення розпізнавання та точності.
- ***Нейронні мережі:*** нейронні мережі, які в основному використовуються для алгоритмів глибокого навчання, обробляють навчальні дані, імітуючи взаємозв'язок людського мозку за допомогою шарів вузлів. Кожен вузол складається з входів, ваг, зсуву (або порогу) і виходу. Якщо вихідне значення перевищує заданий поріг, він "вистрілює" або активує вузол, передаючи дані на наступний рівень мережі. Нейронні мережі вивчають цю функцію відображення через контрольоване навчання, підлаштовуючись на основі функції втрат через процес градієнтного спуску.
- ***Діаризація диктора (SD):*** алгоритми діаризації диктора ідентифікують і сегментують мову за ідентичністю мовця. Це допомагає програмам краще розрізняти людей у розмові і часто застосовується в кол-центрах, де розрізняють клієнтів і торгових агентів.

Усі вищезазначені дослідження та роботи свідчать про значний прогрес в області використання штучного інтелекту для розпізнавання звуку. Існують різні методи і алгоритми, такі як вейвлет-пакетне перетворення, машини опорних векторів, нейронні мережі, еволюційні алгоритми та інші, які були використані для розпізнавання різних акустичних проявів, включаючи шум транспортних засобів, чисті тона і дискретні звуки навколишнього середовища.

1.3 Сфера застосування

Алгоритми розпізнавання звуків мають широкий спектр застосувань. Кай обговорює алгоритм для виявлення фонових звуків, який може бути використаний у сфері безпеки, охорони здоров'я та робототехніки. Дослідники робототехніки намагалися імітувати природну слухову пильність у роботів. Наприклад, голова робота може повертатися до джерела звуку. У деяких системах відеоспостереження камери також можуть панорамувати і масштабувати джерело звуку вночі. Цей феномен називається режимом прослуховування "пасивне зондування". Ву та ін. розробили систему, яка може розпізнавати транспортні засоби на основі їхніх звукових сигнатур. Вони записали звуки різних транспортних засобів, побудували вектори ознак на основі спектрограм і принципу компонентного аналізу та класифікували вектори, визначивши евклідову міру відстані до центру кожного відомого класу. У зусиллях досягнення більшої ефективності слухової уваги, було вжито заходів щодо генерації ехо-сигналів шляхом активного випромінювання звуку. Цей процес, відомий вже як "активне зондування", застосовується, наприклад, ультразвуковим сенсорним масивом на автономних автомобілях з метою виявлення перешкод у режимі реального часу в найближчому оточенні. Багато тварин використовують активне зондування на основі звукових відлунь, так звану ехолокацію. Пошук музики онлайн - ще одна мотивація для розпізнавання звуків. Query by tapping (QBT) - це новий метод, заснований на ритмі пісень. Система фіксує основні елементи ритму, дозволяючи представити його в текстовій формі, яка має добре відпрацьовані алгоритми, що дозволяють толерантно ставитися до темпових

варіацій і помилок у вхідних даних. Принадність цього винаходу полягає в тому, що він не потребує спеціального апаратного чи програмного забезпечення - пристроєм введення є лише клавіша пробілу на клавіатурі комп'ютера. Класифікація аудіо жанрів також досліджується у роботах Цанетакіс, де вектор ознак будується з використанням статистичних характеристик для представлення "музичної поверхні" аудіо; вони також включають дискретне вейвлет-перетворення. Дослідник Фу фокусується на мобільному додатку для виявлення звукових подій, який можна використовувати для моніторингу в реальному часі. Наше життєве середовище містить багато типів звукових подій, які надають нам багато корисної інформації, що допомагає нам ідентифікувати та сприймати навколишнє середовище. Завдання виявлення звукових подій (SED - sound event detection) пропонується для того, щоб допомогти інтелектуальним пристроям розуміти звукові події та краще служити людям. Завдання SED включає локалізацію та класифікацію звукових подій, спрямовану на оцінку часу настання, зміщення звукових подій та прогнозування звукових подій до попередньо визначених типів. SED широко використовується. Наприклад, у сфері безпілотного водіння, якщо система може ідентифікувати звукові події транспортних засобів, що наближаються або віддаляються, це може зробити систему самокерованого водіння більш надійною. Залежно від того, як вирішувати проблему перекриття звукових подій в аудіо, завдання SED можна розділити на виявлення монофонічних звукових подій та виявлення поліфонічних звукових подій. Монофонічний SED виявляє лише найпомітнішу звукову подію в певний момент часу, тоді як поліфонічний SED може виявляти кілька звукових подій одночасно, що ближче до реального сценарію.

Нещодавно глибокі нейронні мережі продемонстрували хороші результати в задачі поліфонічної SED, які використовують енергетичні характеристики логарифмів мела-діапазону або кепстральні коефіцієнти мела-частоти (MFCC) в якості вхідних даних. Згорткові нейронні мережі (CNN) можуть використовувати просторово локальну кореляцію вхідних даних, а рекурентні нейронні мережі (RNN) можуть фіксувати довгостроковий часовий контекст аудіосигналу. В результаті використання

переваг обох підходів, згорткова рекурентна нейронна мережа (CRNN) забезпечила найсучасніші результати.

Потаміт розглядає сучасні підходи до автоматичного розпізнавання звуку та обговорює переваги цієї технології в різних сферах, включаючи розпізнавання мови, розпізнавання дикторів, моніторинг навколишнього середовища, біоакустичну ідентифікацію та музику. Завдяки Інтернету, мережевим технологіям і технологіям стиснення зараз доступні величезні обсяги музичних даних, і їхня кількість швидко зростає. Індексування, пошук і навігація в цих носіях здебільшого ґрунтуються на текстовій інформації, яка є неповною, виконується вручну і не може ефективно описувати зміст музичних творів. В останнє десятиліття аналіз і пошук музики став дуже активною сферою досліджень, і музичні онлайн-сервіси візьмуть на себе відповідальність відповідати на складні запити, які базуються на інтерпретації змісту музичного сигналу. Автоматичне вилучення і класифікація інформації про вміст аудіосигналів може зробити можливими величезну кількість нових застосувань. Інтелектуальна обробка музичних сигналів спрямована на такі застосування, як: класифікація музичних жанрів, транскрипція музики, розпізнавання виконавців, розпізнавання інструментів, індексування та пошук музичних даних, виявлення типу аудіосигналу (мова проти фонового шуму, розпізнавання музичних жанрів, стилів, настроїв, інструментів або, навіть більше, низькорівневих характеристик.

Баунтуракіс досліджує алгоритми машинного навчання для розпізнавання звуків навколишнього середовища, які можуть бути використані для семантики звукового ландшафту. На відміну від автоматичного розпізнавання мови (ASR) та пошуку музичної інформації (MIR), які досліджуються протягом тривалого часу,, розпізнавання звуків навколишнього середовища (ESR) отримало особливу увагу лише в останні десятиліття. Метою ESR є віднесення вхідних звуків навколишнього середовища до попередньо визначених категорій (або класів). Під звуками навколишнього середовища маються на увазі різноманітні звуки, як природні, так і штучні (тобто звуки, з якими людина стикається у повсякденному житті, окрім мови

та музики). У зв'язку з тим, що ESR все ще перебуває на стадії становлення, більшість дослідницьких спроб у цій галузі використовують методи, які спочатку були розроблені для розпізнавання мови та музики. Крім того, ESR може бути застосований у середовищі домашнього моніторингу, чи то для допомоги літнім людям, які живуть самотньо у власному будинку, чи то для "розумного" будинку. Техніка ESR також може бути використана для підвищення ефективності ідентифікації мовця та розпізнавання мови на фоні звуків навколишнього середовища.

Загалом, ці роботи показують, що алгоритми розпізнавання звуків мають широкий спектр застосувань, включаючи безпеку, охорону здоров'я, робототехніку, моніторинг у реальному часі, моніторинг навколишнього середовища, біоакустичну ідентифікацію, музику та семантику звукового ландшафту.

1.4 Висновки з 1 розділу

Впродовж багатьох століть людство зосереджувало великі зусилля на прогресі у розвитку штучного інтелекту, зокрема вдосконаленні автоматизованих систем. Однією з ключових мет цього тривалого процесу було створення машин, які здатні пройти той самий шлях, який людина може пройти за кілька років. Завдяки постійним дослідженням і розвитку інтелектуальних систем, досягнуті значні прориви в цій галузі. Сучасні методи машинного навчання та глибокого навчання дозволяють розробити алгоритми та моделі, що ефективно виконують завдання, які раніше вимагали багато людської праці і часу.

Використовуються різні методи і алгоритми, такі як вейвлет-пакетне перетворення, машини опорних векторів, нейронні мережі, еволюційні алгоритми та інші, для розпізнавання різних акустичних сигналів, включаючи шум транспорту, чисті тони і звуки навколишнього середовища. Вони також успішно використовуються для розпізнавання мовлення в реальних умовах з шумом у навколишньому середовищі. В цілому, ці методи підкреслюють потенціал штучного інтелекту для поліпшення технологій розпізнавання мови, хоча точність таких систем

ще може бути покращена, оскільки вони ще не досягли рівня людського розпізнавання звуку, і їх ефективність залежить від конкретного завдання.

Алгоритми розпізнавання звуків мають широкий спектр застосувань і можуть бути використані у безпеці, охороні здоров'я, робототехніці, відеоспостереженні, музичних додатках та інших. Дослідники активно вивчають імітацію природної слухової пильності та використання звукових сигнатур для розпізнавання транспортних засобів. Вони також розвивають системи для виявлення звукових подій у реальному часі, які можуть бути корисними для безпілотних автомобілів, спостереження за навколишнім середовищем та інших застосувань. Глибокі нейронні мережі, зокрема згорткові рекурентні нейронні мережі, показують добрі результати в розпізнаванні звукових подій, але їх використання на обмежених обчислювальних ресурсах може бути обмеженим. Однак, розвиток мобільних додатків для розпізнавання звуків на мобільних пристроях є важливим напрямом дослідження.

РОЗДІЛ 2. Алгоритми розпізнавання звуку

2.1 Способи запису та перетворення звукових доріжок

2.1.1 Історія звукозапису

Вперше звуковий запис було створено 12 серпня 1877 року, коли Едісон зафіксував мелодію "Mary Had a Little Lamb" на циліндрі фонографа.

У фонограф постійно вносилися вдосконалення. Олов'яна фольга, що спочатку покривала валик, незабаром змінилася тонким шаром воску. Проте домогтися якісного і довговічного запису не вдавалось.

У 1888 році німець Еміль Берлінер створив грамофон. Фонограма покривалася лаком, а за допомогою цинку, протравленого в хромовій кислоті, отримували гальванопластичні копії. Перші грамофонні платівки робили з целулоїду, а пізніше почали використовувати диски з шелаку, шпату та сажі, згодом і з синтетичними смолами. У пошуках нових можливостей, виконавці вирішили використовувати доступний матеріал - рентгенографічні плівки - для запису. Ці платівки отримали прозвище "записи на кістках".

Згодом було винайдено патефон. Завдяки певним поліпшенням, він був компактніший, що порівняно з грамофоном зробило його надзвичайно популярним та призвело до збільшення популярності пристроїв для прослуховування музичних записів. Більш того, ціни на платівки постійно знижувалися. Під час виготовлення використовували дешеві матеріали, і, нарешті, було прийнято рішення використовувати вініл.

Одночасно з винаходом фонографа, були зроблені перші спроби магнітного запису звуку. Реалізувати це вдалося датчанину Вальдемару Паульсену, який у 1898 році продемонстрував перший працездатний магнітофон, де сталевий дріт виступав як носій запису. Перші магнітофони не досягали високої якості запису, порівняно з

популярними грамофонами того часу. Однак з появою вакуумних електронних ламп все змінилося, і найкращі результати в магнітному звукозаписі були досягнуті з появою вдосконалених магнітних голівок, використанням підмагнічення та порошкової магнітної стрічки. Потім з'явилися касетні магнітофони, де обидві мініатюрні котушки з магнітною та порожньою стрічками були поміщені в спеціальну компактну касету.

З розвитком радіотехніки на ринок вийшли радіоли, програвачі та електрофони. Пружинний механізм був замінений електричним. Маючи той же розмір платівки, вони відтворювали звук зі швидкістю до 33 1/3 оборотів на хвилину, а збільшення щільності запису подовжило час використання.

У 1979 році компанії Philips та Sony представили свій абсолютно новий носій інформації - оптичний диск для запису і відтворення звуку. Починаючи з 1982 року запустилося масове виробництво компакт-дисків. Запис сигналів на оптичний диск виконується за допомогою лазерного променя за цифровим методом. Для зчитування лазерний промінь, сфокусований на доріжці, просувається по поверхні обертового диска і сканує записану інформацію.

На заміну компакт-дискам прийшов новий стандарт - DVD (Digital Versatile Disc). Його основною відмінністю є значно більша щільність запису інформації. Крім того, DVD-диск може мати один або два шари для запису. Загалом, DVD-стандарт має чотири модифікації:

DVD-5: Є одношаровим диском, який може вмістити до 4.7 гігабайт інформації.

DVD-9: Це двошаровий диск, що має місткість до 8.5 гігабайт.

DVD-10: Представляє собою двосторонній одношаровий диск з обсягом до 9.4 гігабайт (по 4.7 гігабайта на кожну сторону).

DVD-18: Це двошаровий двосторонній диск, який може вмістити до 17 гігабайт інформації (по 8.5 гігабайта на кожну сторону).

DVD-диски дозволяють записувати не тільки звук, але й відео та інші типи мультимедійної інформації, що зробило їх більш універсальними у порівнянні з компакт-дисками.

У наш час звук зберігається безпосередньо на картах пам’яті плеєрів, смартфонів та інших пристроїв. Крім того, швидкий розвиток технології жорстких дисків створив можливість зберігання не тільки звуку, а і відео.

2.1.2 Сучасні способи запису звуку

Існує багато загальновідомих способів запису звуку, кожен з яких має свої унікальні характеристики та застосування. Ось деякі з найпопулярніших методів:

1. Запис з мікрофона: Це найпоширеніший спосіб запису звуку. Мікрофон фіксує звукові хвилі та перетворює їх на електричні сигнали, які потім підсилюються і записуються на пристрій або комп'ютер. Існує безліч типів мікрофонів, включаючи динамічні, конденсаторні, рушійні та багато інших. Кожен мікрофон має свої характеристики звучання та використовується для різних завдань. Мікрофони можна підключати безпосередньо до аудіоінтерфейсу або записувального пристрою. Запис з мікрофона є універсальним і може використовуватися в різних середовищах, від професійних студій до домашніх записів.

2. Прямий вхід (DI): Цей метод використовується для запису звуків безпосередньо з аудіоінструментів, таких як електрогітара, бас-гітара або клавіатура. Сигнал з інструменту підключається безпосередньо до аудіоінтерфейсу, що дозволяє отримати прямий, неперетворений звук. Це корисно для збереження високої якості звуку та для подальшої обробки аудіосигналу в DAW.

3. Польовий запис: Цей метод використовується для запису звуків у реальних середовищах поза студією. Запис можна здійснювати за допомогою портативних аудіорекордерів або спеціалізованого обладнання, такого як стереомікрофони або параболічні мікрофони. Польовий запис часто використовується для захоплення звукового оточення, природних звуків, вуличного мистецтва або аудіо для фільмів та документальних фільмів.

4. Цифрові аудіостанції (DAWs): Це програмне забезпечення, яке використовується для запису, редагування та змішування звуку. DAWs надають широкий набір інструментів і функцій, таких як многодоріжковий запис, можливості реального часу для обробки звуку, ефекти та плагіни для покращення звукового матеріалу. Ви можете використовувати мікрофони або інші джерела запису, щоб записувати аудіосигнали безпосередньо у DAW та подальше редагувати їх.

5. Віртуальні інструменти: Вони є програмними інструментами або синтезаторами, які можна використовувати у DAW для створення та запису музики. Вони відтворюють звучання реальних інструментів, таких як фортепіано, скрипка або барабани, або створюють синтезовані звуки. Віртуальні інструменти можуть бути грати за допомогою MIDI-клавіатури або інших контролерів та дозволяють редагувати та маніпулювати звуком в різних способах.

2.1.3 Зберігання звуку

Під час збереження аналоговий звук перетворюється на цифровий одним з декількох методів. Одним з найпростіших методів перетворення є імпульсно-кодова модуляція (ІКМ), яка включає в себе отримання миттєвих значень рівня сигналу в певні моменти часу, що вимірюється аналого-цифровим перетворювачем (АЦП), шляхом поділу часу на рівні проміжки. Дельта-модуляція є одним з варіантів ІКМ, де на кожному кроці вхідний сигнал порівнюють з пилкоподібною напругою. Сигма-дельта модуляція - це метод представлення сигналу, який ґрунтується на принципі надлишкової

дискретизації і використовує формування шуму квантування для зниження рівня шуму.

Використання бітового коду має ряд переваг у таких випадках, як шифрування сигналу, цифрове підписування сигналу, передача кодованого сигналу, відновлення втрат, що виникають під час передачі через перешкоди, а також при інших задачах. Представлення аудіо у цифровому форматі дає можливість ефективно перетворювати вихідний матеріал з використанням спеціальних пристроїв та/або комп'ютерних програм, таких як звукові редактори. Це надає можливість широкого застосування аудіо в медіа-індустрії, повсякденному житті та промисловості.

Процес цифрового подання звукових коливань базується на наступних принципах:

1. Аналого-цифрове перетворення: Спочатку аналоговий звуковий сигнал перетворюється на цифровий формат за допомогою пристрою, відомого як аналого-цифровий перетворювач (АЦП). Цей процес включає квантування і дискретизацію сигналу для отримання послідовності цифрових значень.

2. Збереження цифрових даних: Отримані цифрові дані зберігаються на носії інформації, такому як магнітна стрічка (DAT), жорсткий диск, оптичний диск або флеш-пам'ять. Цей носій зберігає цифрову інформацію для подальшого відтворення.

3. Цифро-аналогове перетворення: Для прослуховування записаного звуку необхідно відтворити цифрові дані з носія. Це досягається за допомогою цифро-аналогового перетворювача (ЦАП), який перетворює цифровий сигнал на аналоговий. Отриманий аналоговий сигнал потім може бути відтворений через динаміки або навушники, що дозволяє прослуховувати записаний звук.

АЦП працює за наступним принципом:

1. Обмеження смуги частот: Аналоговий сигнал проходить через фільтр нижніх частот, який обмежує спектральні компоненти, частота яких перевищує половину

частоти дискретизації. Це здійснюється для уникнення ефекту аліасингу, коли відображення спектральних компонент вищих частот стає неправильним після дискретизації.

2. Дискретизація в часі: Аналоговий сигнал заміщується послідовністю його значень у дискретних моментах часу, що називаються відліками. Це виконується за допомогою пристрою вибірки-зберігання, який здійснює вимірювання аналогового сигналу на кожному дискретному моменті часу та зберігає ці значення.

3. Квантування: Отримані аналогові значення сигналу квантуються, тобто округлюються до певного кількості бітів. Це призводить до апроксимації аналогового сигналу з деякою похибкою, яка залежить від роздільної здатності АЦП.

4. Кодування: Квантовані значення перетворюються на цифровий код, який представляє ці значення у вигляді бінарного числа. Кодування може бути різним, наприклад, прямим бінарним кодуванням або використанням компресійних алгоритмів для зменшення обсягу даних.

Процес АЦП можна описати наступним чином: аналоговий сигнал розбивається на ділянки з використанням певної частоти дискретизації. Отриманий дискретний сигнал піддається квантуванню з певною розрядністю, після чого він кодується у послідовність символів. Для запису звуку з гарною якістю в діапазоні частот 20-20 000 Гц зазвичай використовується мінімальна стандартна частота дискретизації 44,1 кГц або вище (на сьогоднішній день також доступні АЦП і ЦАП з частотою 192,3 і 384,6 кГц). Для збереження звукозапису найчастіше використовується розрядність 16 біт, проте для збільшення діапазону та покращення якості можна застосовувати розрядність 24 та 32 біти.

2.2 Алгоритми порівняння звукових доріжок

Алгоритми порівняння звукових доріжок використовуються для визначення ступеня схожості чи відмінності між двома звуковими сигналами. Існує кілька різних методів і алгоритмів порівняння звукових доріжок, залежно від конкретних вимог і завдань. Деякі з них включають:

1. Кореляційний аналіз: Цей метод використовує кореляцію між двома звуковими сигналами для визначення ступеня їх схожості. Вимірюються кореляційні коефіцієнти між двома доріжками, і чим більше значення кореляції, тим більша схожість між ними.

2. Метод динамічного програмування: Цей метод розглядає звукові доріжки як послідовності елементів і використовує динамічне програмування для знаходження оптимального вирівнювання між ними. Він враховує різні аспекти звуку, такі як спектральна і енергетична інформація, для визначення ступеня схожості.

3. Метод векторного квантування: Цей метод представляє звукові доріжки у вигляді векторів і використовує методи квантування для порівняння цих векторів. Вимірюються відстані між векторами, і менша відстань вказує на більшу схожість між доріжками.

4. Метод спектрального аналізу: Цей метод використовує спектральну інформацію про звукові доріжки для порівняння. Вимірюються спектральні характеристики, такі як спектральна щільність чи спектральні коефіцієнти, і порівнюються між доріжками.

2.2.1 Кореляційний аналіз

Кореляційний аналіз є одним з методів порівняння звукових доріжок, що використовується для визначення ступеня схожості між двома звуковими сигналами. Основна ідея полягає в вимірюванні ступеня взаємозв'язку або схожості між двома сигналами шляхом обчислення кореляційних коефіцієнтів.

Кореляційний коефіцієнт відображає міру лінійної залежності між двома сигналами. Він вимірює ступінь схожості форми сигналів та зсуву між ними. Значення кореляційного коефіцієнта може варіюватися від -1 до 1. Значення близьке до 1 вказує на високу схожість між сигналами, значення близьке до -1 вказує на протилежну залежність, а значення близьке до 0 вказує на відсутність кореляції між сигналами.

Процес кореляційного аналізу включає наступні кроки:

1. Підготовка сигналів: Звукові сигнали, які потрібно порівняти, повинні бути підготовлені для аналізу, включаючи відповідну обробку та попередню обробку, таку як нормалізація або фільтрація.

2. Обчислення кореляційного коефіцієнта: Застосовується математична формула для обчислення кореляційного коефіцієнта між двома сигналами. Цей коефіцієнт може бути обчислений за допомогою різних методів, таких як метод Пірсона або метод Спірмена.

3. Аналіз результатів: Отримані значення кореляційного коефіцієнта можуть бути проаналізовані для визначення ступеня схожості між сигналами. Це може включати порівняння з пороговим значенням, що вказує на певний рівень схожості, або використання інших метрик для кількісної оцінки ступеня схожості.

Кореляційний аналіз є потужним інструментом для порівняння звукових доріжок і використовується в різних областях, включаючи музичну аналітику, виявлення плагіату, розпізнавання мови та в інших задачах, де необхідно порівняти схожість або ідентичність між звуковими сигналами.

2.2.2 Динамічне програмування

Один з підходів полягає у застосуванні динамічного програмування для обчислення міри подібності між двома звуковими доріжками. Для цього можна використовувати

алгоритм Needleman-Wunsch або Smith-Waterman, які походять від методів розв'язання задачи глобального та локального вирівнювання послідовностей.

Алгоритми Needleman-Wunsch та Smith-Waterman використовують матрицю, яка представляє собою двовимірний масив, де кожна комірка містить значення подібності або різниці між відповідними елементами двох послідовностей. За допомогою рекурентних формул ці матриці обчислюються зверху вниз або зліва направо. Потім можна використовувати цю матрицю для визначення найбільшої подібності між звуковими доріжками та отримання оптимального вирівнювання.

В результаті застосування методу динамічного програмування для порівняння звукових доріжок можна отримати значення подібності між ними, а також оптимальне вирівнювання, що показує, які частини доріжок співпадають або розходяться. Це може бути корисно для задач, таких як виявлення плагіату в музиці, пошук зразків у звуку або визначення схожості музичних фрагментів.

2.2.3 Векторне квантування

Метод векторного квантування є одним з підходів до компресії аудіоданих і може бути використаний для порівняння звукових доріжок.

Векторне квантування включає в себе розділення аудіосигналу на малий набір векторів, які називаються кодовими векторами. Кожен кодовий вектор представляє собою підмножину зразків аудіосигналу. Замість безпосереднього кодування кожного зразка, векторне квантування здійснює квантування кодових векторів. Кожен кодовий вектор замінюється індексом, який вказує на набір зразків, що його представляють.

Для порівняння звукових доріжок за допомогою векторного квантування, спочатку обидві доріжки піддаються процесу квантування за допомогою відповідного алгоритму векторного квантування. Потім порівнюються отримані індекси кодових

векторів. Якщо індекси збігаються, це вказує на схожість між доріжками в тих областях, які були використані для векторного квантування.

Метод векторного квантування може бути використаний для порівняння звукових доріжок, оскільки він дозволяє зберігати і порівнювати характеристики аудіосигналу за допомогою індексів кодових векторів. Це дає можливість виявити подібність аудіоданих навіть при зміні амплітуди або фазових характеристик сигналу. Однак, важливо враховувати, що векторне квантування є методом компресії і може призводити до втрати якості сигналу.

2.2.4 Спектральний аналіз

Метод спектрального аналізу є ефективним підходом до порівняння звукових доріжок. Він використовується для отримання і аналізу спектральних властивостей аудіосигналу.

Спектральний аналіз доріжки базується на перетворенні Фур'є, яке розкладає аудіосигнал на його складові частоти. Це дозволяє отримати спектрограму, яка відображає інтенсивність сигналу в залежності від частоти та часу. За допомогою спектрального аналізу можна виявити частотні характеристики сигналу, такі як основна частота, гармонічні складові, шумові компоненти та інші спектральні особливості.

Для порівняння звукових доріжок за допомогою спектрального аналізу, спочатку обидві доріжки піддаються спектральному аналізу. Потім порівнюються спектральні характеристики, такі як амплітудні спектри, спектральні форми, піки частот та їх інтенсивності. За допомогою порівняння спектральних характеристик можна визначити ступінь схожості між доріжками і виявити подібні музичні структури або звукові ефекти.

Метод спектрального аналізу є потужним інструментом для порівняння звукових доріжок, оскільки він дозволяє аналізувати спектральні властивості сигналу в деталях. Він може виявити навіть незначні зміни в спектральному складі доріжок, що робить його корисним для виявлення схожих музичних фрагментів, зразків або розпізнавання звукових ефектів.

2.3 Методи визначення музичних творів за записом

Визначення назви музичного твору за записом може бути складним завданням, але існують декілька основних алгоритмів та методів, які можуть бути використані для цієї мети.

2.3.1 Аналіз спектрограми

Спектрограма - це графічне зображення звукового сигналу, де по горизонтальній вісі відображається час, а по вертикальній - частота. Алгоритми аналізу спектрограми можуть виявляти характерні мелодійні або ритмічні структури, які можуть вказувати на назву пісні. Ось деякі кроки, які можна виконати для аналізу спектрограми з метою визначення назви музичного твору:

А) Отримання спектрограми: Звуковий сигнал треку спочатку перетворюється на спектрограму. Це можна зробити за допомогою алгоритмів, таких як швидке перетворення Фур'є (FFT), яке розбиває сигнал на його складові частоти.

Б) Аналіз частот: Детальне дослідження спектрограми може розкрити характерні частотні компоненти треку, такі як мелодійні лінії, акорди або ритмічні малюнки. Визначення унікальних частот або частотних шаблонів може допомогти встановити основні мелодійні елементи пісні.

В) Порівняння зі зразками: Можна створити базу даних спектрограм з відомими музичними творами і порівнювати аналізовану спектрограму з цими зразками. За допомогою алгоритмів порівняння, таких як шаблонне співставлення або

класифікація за допомогою машинного навчання, можна визначити схожість між аналізованим треком і відомими зразками.

2.3.2 Розпізнавання тексту

Розпізнавання мови може бути корисним інструментом для визначення назви музичного твору коли вокал присутній у записі. Однак, важливо зазначити, що розпізнавання мови не є безперечним методом і може бути обмеженим в своїй точності.

Існує кілька підходів до розпізнавання мови. Один з найпоширеніших методів - це використання моделей глибокого навчання, зокрема рекурентних нейронних мереж (RNN) або довготривалих нейронних мереж (LSTM). Ці моделі можуть бути навчені на великому обсязі даних мовного матеріалу для розпізнавання і класифікації мов.

Процес розпізнавання мови для визначення назви музичного твору може включати наступні кроки:

А) Підготовка даних: Запис музичного твору може бути розділений на короткі аудіофрагменти, наприклад, по 5-10 секунд, для подальшого аналізу.

Б) Перетворення аудіо в текст: Кожен аудіофрагмент може бути перетворений у векторне представлення або спектрограму за допомогою алгоритмів обробки сигналів. Потім, застосовуючи модель розпізнавання мови, текст може бути витягнутий з аудіо.

В) Визначення назви музичного твору: Після отримання тексту можна використати його для пошуку відповідної інформації, такої як назва пісні або виконавець, у базі даних музичних творів.

Важливо зазначити, що точність розпізнавання мови може залежати від якості запису, ясності мовлення, наявності шуму та інших факторів. Також, якщо вокал в записі

недостатньо виразний або використовується незвичайна мова, розпізнавання може бути ускладненим або неможливим.

2.3.3 Машинне навчання

Машинне навчання є потужним інструментом для визначення назви музичного твору. Існує кілька підходів до використання машинного навчання для цієї задачі. Найпоширеніші з них:

А) Класифікація з використанням векторів ознак: В цьому підході аудіофайли музичних творів можуть бути перетворені на вектори ознак, такі як спектрограми або характеристики часу-частоти. Потім, з використанням методів машинного навчання, таких як класифікація з використанням алгоритмів SVM (Support Vector Machines), нейронних мереж або ансамблів дерев рішень, модель може бути навчена визначати назви музичних творів на основі цих ознак.

Б) Використання глибокого навчання: Глибокі нейронні мережі, зокрема згорткові нейронні мережі (Convolutional Neural Networks, CNN) або рекурентні нейронні мережі (Recurrent Neural Networks, RNN), можуть бути використані для вирішення задачі визначення назви музичного твору. Ці моделі можуть бути навчені безпосередньо на аудіофайлах, використовуючи різноманітні архітектури, такі як варіації аудіо-CNN або LSTM моделі. Вони можуть автоматично вивчати репрезентативні ознаки з аудіо та визначати назви музичних творів.

В) Використання попередньо-навчених моделей: Існують попередньо-навчених моделі, такі як VGGish або OpenL3, які можуть бути використані для екстрагування аудіо-ознак з музичних творів. Ці моделі навчалися на великих наборах даних, що робить їх здатними до витягнення важливих ознак з аудіофайлів. Потім, на основі цих ознак, можуть бути застосовані методи класифікації для визначення назв музичних творів.

2.4 Опис основних алгоритмів роботи

2.4.1 Перетворення аналогового сигналу в дискретний

Перетворення аналогового сигналу в послідовність дискретних значень виконується за допомогою квантування. Цей процес включає апроксимацію значень аналогового сигналу до найближчих значень через рівні проміжки часу.

Основні етапи квантування аналогового сигналу:

1. Розділення амплітудного діапазону аналогового сигналу на скінчену кількість точок відліку (проміжків часу). Частотою дискретизації називають кількість відліків, які здійснюються на одиницю часу (секунду) при дискретизації аналогового сигналу. Вимірюється у герцах (Гц).

2. Знаходження значення аналогового сигналу в кожній точці відліку.

Результатом квантування стає послідовність дискретних значень, яка може бути збережена, оброблена або передана в цифровій формі.

Людське вухо може виявляти частоти приблизно від 20 Гц до 20 000 Гц. Внаслідок цього аудіо найчастіше записують з частотою дискретизації 44 100 Гц.

2.4.2 Перехід від часового домену до частотного

Звукові сигнали зберігається у часовому домені, тобто вони представлені як функції, що змінюються з плином часу. Вимірювання амплітуди сигналу відбувається на осі часу. Аналізуючи сигнали в часовому домені, можна спостерігати їх часові характеристики, такі як тривалість, періодичність, моменти зміни амплітуди тощо.

У частотному домені сигнали розглядаються в залежності від частоти. У цьому домені сигнали представлені у формі спектральних складових, які показують розподіл амплітуди сигналу за частотами. Аналізуючи сигнали в частотному домені, можна

виявити наявність різних частот у сигналі, їх амплітуди, фазові зміщення та співвідношення між різними частотами.

Перехід від часового домену до частотного домену і навпаки здійснюється за допомогою математичних операцій, таких як перетворення Фур’є, яке трансформує сигнал з часового домену в частотний домен, і обернене перетворення Фур’є, яке здійснює зворотний перехід.

2.4.3 Перетворення Фур’є

Перетворення Фур'є (Fourier transform) - це математичний метод, який перетворює функцію, залежну від часу, на функцію, залежну від частоти. Воно розкладає початкову функцію на суму гармонічних компонент з різними частотами, амплітудами і фазами. Це дає можливість виявити присутність і інтенсивність різних частот у вихідному сигналі.

Перетворення Фур'є може бути використане для аналізу сигналів у різних областях, таких як фізика, інженерія, обробка сигналів та інші. Воно дозволяє розкрити складові частоти в сигналі, виявити гармоніки, спектральні особливості і зв'язки між різними частотними складовими.

Перетворення відбувається за допомогою інтеграла Фур’є:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t}\, dt$$

У цій формулі, f(t) - це початкова функція, що залежить від часу t, F(ω) - її перетворення Фур'є залежне від частоти ω, та $e^{-i\omega t}$ - комплексна експонента, яка використовується для розкладу функції на гармонічні компоненти з різними частотами.

Дискретне перетворення Фур’є задається наступною формулою:

$$X_m = \sum_{n=0}^{N-1} x(n) e^{-i2\pi nm/N}$$

Де m – індекс перетворення в частотній області, n – індекс вхідного відліку, $x(n)$ – послідовність вхідних відліків, X_m – m-тий компонент ДПФ.

Для перетворення функції, що залежить від частоти, на функцію, залежну від часу, використовують обернене перетворення Фур'є:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{-i\omega t}\, dw$$

2.4.4 Швидке перетворення Фур'є

Швидке перетворення Фур'є - це алгоритм, який обчислює дискретне перетворення Фур'є послідовності або його обернене перетворення. Обчислення перетворення ДПФ безпосередньо за визначенням часто є надто повільним для практичного застосування. ШПФ швидко обчислює такі перетворення за допомогою декомпозиції матриці ДПФ. В результаті цей алгоритм здатний зменшити складність обчислення ДФТ з $O(N^2)$, яка виникає, якщо просто застосувати визначення DFT, до $O(N \log N)$,де N – розмірність даних. Різниця у швидкості може бути дуже суттєвою, особливо для великих наборів даних, коли N може сягати тисяч або мільйонів. Існує багато різних алгоритмів FFT, заснованих на широкому спектрі опублікованих теорій, від простої арифметики комплексних чисел до теорії груп і теорії чисел.

Алгоритм Кулі-Тьюкі є одним з найбільш поширених алгоритмів для обчислення швидкого перетворення Фур'є. Основна ідея алгоритму Кулі-Тьюкі полягає в рекурсивному розкладанні послідовності на дві підпослідовності з половинною довжиною. Застосовуючи рекурсію, цей процес продовжується досягнення базового випадку, коли довжина послідовності дорівнює 1 (в цьому випадку ДПФ буде

дорівнювати початковій послідовності). Потім виконується злиття підпослідовностей, що дає повний результат DFT.

Варіація radix-2 алгоритму Кулі-Тьюкі для $N = 2^m$:

Алгоритм розбиває послідовність значень x_n функції на дві частини: суму значень з парними індексами $(2m)$ та суму значень з непарними індексами $(2m + 1)$:

$$X_k = \sum_{m=0}^{\frac{N}{2}-1} x_{2m} e^{-\frac{2\pi i}{N}(2m)k} + e^{-\frac{2\pi i}{N}k} \sum_{m=0}^{\frac{N}{2}-1} x_{2m+1} e^{-\frac{2\pi i}{N}(2m)k}$$

Позначимо $E_k = X_{2m}$ та $O_k = X_{2m+1}$. $X_{k+N/2}$ також визначається з E_k та O_k.

$$X_{k+N/2} = \sum_{m=0}^{\frac{N}{2}-1} x_{2m} e^{-\frac{2\pi i}{N}(2m)k} - e^{-\frac{2\pi i}{N}k} \sum_{m=0}^{\frac{N}{2}-1} x_{2m+1} e^{-\frac{2\pi i}{N}(2m)k}$$

Тобто

$$X_k = E_k + e^{-\frac{2\pi i}{N}k} O_k$$

$$X_{k+N/2} = E_k - e^{-\frac{2\pi i}{N}k} O_k$$

2.4.5 Нейронна мережа LSTM

LSTM (англ. long short-term memory – довга короткотермінова пам'ять) - це архітектура рекурентної нейронної мережі. Мережі LSTM призначені для вирішення проблеми зникнення градієнту, яка виникає при навчанні глибоких нейронних мереж. Проблема зникнення градієнту полягає в тому, що градієнт стає дуже малим, коли він поширюється назад через шари мережі під час навчання, що ускладнює вивчення довгострокових залежностей.

LSTM були запропоновані З. Хохрайтером і Ю, Шмідгубером у 1997 році як рішення цієї проблеми. Вони включають клітини пам'яті та вентилі для вибіркового збереження або забуття інформації. Основною ідеєю за LSTM є введення клітини пам'яті, яка може зберігати інформацію протягом тривалих періодів часу, що дозволяє мережі виявляти довгострокові залежності.

Клітина пам'яті в LSTM має циклічну обробку, що дозволяє їй зберігати своє значення з плином часу. Вона також має три основні вентилі: вентиль введення, вентиль забуття та вентиль виведення. Ці вентилі регулюють потік інформації до, з і всередині клітини пам'яті.

1. Вентилі введення визначають, скільки нової інформації додається до клітини пам'яті.

2. Вентилі забуття контролюють, скільки інформації видаляється з клітини пам'яті.

3. Вентилі виведення регулюють, скільки інформації виводиться з клітини пам'яті до наступного шару мережі.

LSTM широко застосовуються в різних областях, таких як обробка природньої мови, розпізнавання мови, створення підписів до зображень та аналіз часових рядів. Вони відмінно підходять для моделювання та прогнозування послідовностей даних завдяки їхній здатності фіксувати довгострокові залежності.

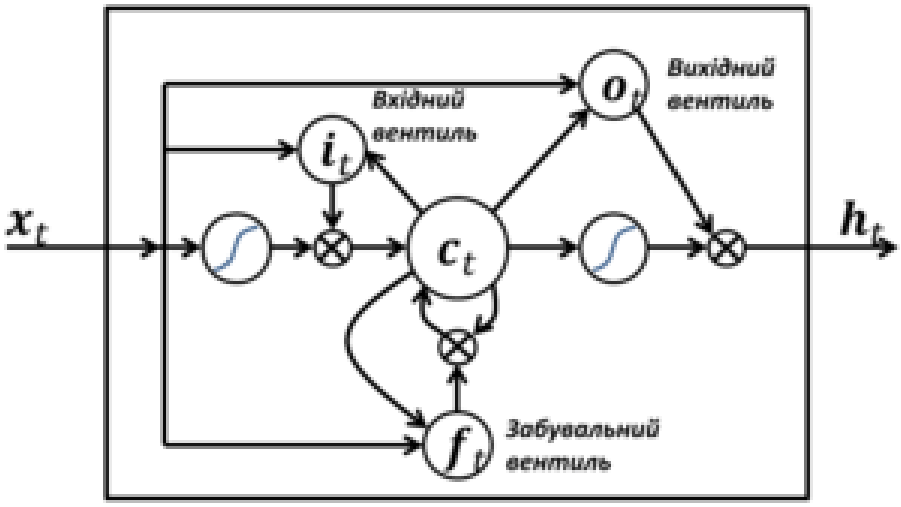


Рисунок 2.4.5.1 LSTM модель

Обчислення значення вузла традиційної LTSM проходить за наступним алгоритмом:

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$

$$h_t = o_t \circ \sigma_h(c_t)$$

Позначення:

$\circ$ – поелементний добуток

x_t – вхідний вектор значень

h_t – вихідний вектор значень

c_t – вектор стану комірки

W, U, b – Матриці та вектор параметрів

f_t, i_t, o_t – вектори вентилів: забувального, вхідного та вихідного відповідно

$\sigma_g, \sigma_c, \sigma_h$ – функції активації. Стандартний вибір: σ_g – сигмоїдна функція, σ_c – гіперболічний тангенс, σ_h - гіперболічний тангенс або ідентична функція.

РОЗДІЛ 3. Практична частина

Алгоритм був створений за допомогою мови Python. Python - це високорівнева, інтерпретована мова програмування, яка створена у 1991 році. Вона вирізняється своєю простотою, зрозумілістю синтаксису та читабельністю коду. Python є крос-платформеною мовою, підтримує об'єктно-орієнтоване програмування і має широкий набір стандартних та сторонніх бібліотек. Ця мова знайшла застосування в різних галузях, включаючи веб-розробку, аналітику даних, штучний інтелект та наукові обчислення. Завдяки активній спільноті розробників, Python постійно розвивається і набуває популярності в світі програмування.

Даними дослідження виступають музичні твори, назву яких треба знайти. Для аналізу було обрано базу даних, що містить по 100 творів наступних жанрів: блюз, класична музика, кантрі, диско, хіп-хоп, джаз, метал, поп, реггі, рок. Визначення назви музичного твору буде відбуватись двома методами: аналіз та порівняння відбитку твору з вже відомими відбитками з бази даних та аналіз тексту пісні та визначення за допомогою тексту назви твору з використанням нейронних мереж.

Програма використовує деякі бібліотеки із відкритим кодом мови Python для роботи із записом та аналізом звукової доріжки, машинним навчанням та алгоритмами аналізу тональності тексту, зокрема:

- PyAudio – бібліотека для роботи з аудіо в мові програмування Python. Вона надає можливості запису та програвання звукових файлів. PyAudio дозволяє контролювати параметри аудіо, такі як частота дискретизації та кількість каналів.

- NumPy (Numerical Python) – основна бібліотека для наукових обчислень в мові програмування Python. Вона надає потужні та ефективні методи для роботі з числовими данними, включаючи вектори, матриці та інші багатовимірні масиви. NumPy надає широкий спектр математичних функцій для виконання таких операцій

як векторизовані обчислення, обробка лінійних алгебраїчних операцій, трансформації Фур'є, генерація псевдовипадкових чисел і багато іншого.

- Matplotlib – бібліотека для візуалізації даних в мові програмування Python. Вона надає можливості для створення різноманітних типів графіків, діаграм, графіків розсіювання, контурних графіків, гістограм та багатьох інших візуалізаційних елементів.

- Spacy – відкрите програмне забезпечення для обробки природної мови в мові програмування Python. Він надає інструменти для виконання різних завдань, таких як токенізація тексту, виявлення частин мови, виявлення іменованих сутностей, вилучення відношень між сутностями, залежні синтаксичного аналізу та багато інших.

- TensorFlow – відкрите програмне забезпечення для числових обчислень та машинного навчання. Воно розроблене командою Google Brain і надає потужні інструменти для побудови та тренування широкого спектру моделей глибокого навчання.

- Keras – високорівнева бібліотека для глибокого навчання, яка працює поверх на основі TensorFlow. Вона надає зручний і простий у використанні інтерфейс для створення, тренування та оцінки моделей глибокого навчання.

- SQLite – вбудовувана реляційна база даних, яка надає легкий та швидкий спосіб збереження та отримання даних. Вона працює в межах одного процесу та не вимагає окремого сервера бази даних.

У цій роботі було розглянуто 2 методи визначення назви музичного твору на основі запису його відтворення: за допомогою порівняння його відбитку з відбитками у базі даних та визначення на основі тексту пісні.

3.1 Опис алгоритму визначення назви записаного музичного треку за допомогою бази даних відбитків відомих пісень

Загальну роботу цього алгоритму можна розділити на 2 частини:

1. Створення бази даних відбитків музичних творів.

2. Ідентифікація музичного твору по його запису.

3.1.1 Створення бази даних пісень

Для кожного треку було створено відбиток у базі даних SQLite наступним алгоритмом:

1. Зчитування даних. З файлу, що містить трек, зчитуються дані у 16-бітному форматі, отриманий результат представляється у вигляді списку чисел.

2. Отриманий список розбивається на частини задля точного аналізу вмісту пісень. Запис розділяється на частини по 44 частини на секунду.

3. Кожна з отриманих частин була переведена з часового домену до частотного домену з метою полегшення та пришвидшення аналізу вмісту музичного твору алгоритмом швидкого перетворення Фур’є.

4. Для кожного набору даних було визначено частоту з найбільшою амплітудою у межах інтервалів 40Гц – 80Гц, 80Гц – 120Гц, 120Гц – 180Гц, 180Гц – 300 Гц.

5. Задля швидкої ідентифікації треку значення найкращих частот було збережено в базі даних у вигляді ключів хеш-таблиці. В якості значення відповідно до ключа виступає назва треку та проміжок часу, в якому проводився аналіз частот.

3.1.2 Аналіз запису пісні

Для визначення назви треку за його записом проводяться наступні дії:

1. Запис музичного треку вбудованим мікрофоном за допомогою бібліотеки PyAudio. Результат представляється у 16-бітному форматі

2. Отримані дані розбиваються на частини аналогічно до їх оригіналу у базі даних.

3. Отримані дані за допомогою швидкого перетворення Фур'є переводяться з часового до частотного домену.

4. У кожній частині для кожного з основних інтервалів частот визначається частота з найбільшою амплітудою.

5. Отриманні значення частот хешуються для порівняння зі значеннями, збереженими в базі даних.

6. Отриманий список хеш значень порівнюється зі збереженими в базі даних значеннями за наступним алгоритмом: знаходиться запис у базі даних, різниця хеш значень якого з відповідними по часовому параметру хеш значеннями записаного музичного твору є мінімальною.

3.1.3 Результати використання першого методу

Для тестування було зроблено по 10 записів треків кожного жанру довжиною 5 секунд. Визначення назви музичного твору за допомогою алгоритму створення відбитків надає правильний результат у 67 відсотках випадків. Час обробки одного запису в середньому становить 4 секунди. Найкращий результат визначення назви можна побачити в наступних жанрах: рок (100%), поп (100%) та реггі (90%). Найгірші результати програма видає при аналізі треків жанрів блюз (40%), діско (40%) та класичної музики (20%).

Використання алгоритму хешування займає скорочує потрібний час аналізу приблизно в 3 рази зі збереженням рівня розпізнавання.

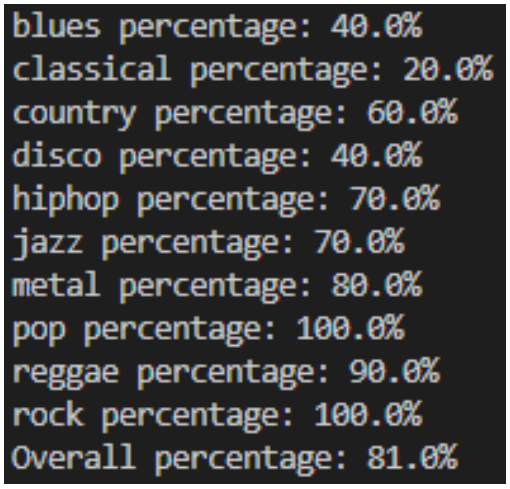


Рисунок 3.1.3.1 Результат обробки записів

3.2 Опис алгоритму визначення назви записаного музичного треку за допомогою аналізу тексту пісень

Алгоритм розпізнавання назви музичного твору за допомогою аналізу слів у його тексті поділяється на 2 частини:

1. Тренування нейронної мережі, що за текстом твору буде визначати його назву.

2. Ідентифікація музичного твору по його запису через розпізнавання тексту пісні.

3.2.1 Тренування нейронної мережі

Першим кроком створення нейронної мережі для передбачення назви пісні є створення набору тренувальних та тестових даних.

Спочатку для кожного тексту пісні проводиться токенізація – процес виділення слів із строки. Далі відкидаються усі слова довжиною менше трьох літер, та “стоп” слова ("the," "a," "an," тощо). З отриманого списку слів створюють списки довжиною 4 слова.

Для тренування маємо список послідовностей слів у якості тренувальних вхідних даних та відповідні назви пісень у якості результатів. 90% даних беруться в якості тренувального набору, решта – тестового.

Для обробки нейронною мережею слова трансформуються у векторне представлення за допомогою фреймворку spacy – кожному слову ставиться у відповідність вектор дійсних чисел довжиною 300.

Модель створеної за допомогою фреймворку Keras нейронної мережі:

1. LSTM шар. Приймає на вхід матрицю розміру (4, 300), вихідні дані такого ж розміру. Загальна кількість параметрів – 721200.

2. LSTM шар. Приймає на вхід матрицю розміру (4, 300), вихідні дані розміру (1, 300). Загальна кількість параметрів – 721200.

3. Шар повного з'єднання з сигмоїдною функцією активації. Приймає вхідний вектор даних розміру (300) та перетворює його на число з проміжку [0, 1]. Загальна кількість параметрів – 301.

Вхідні дані розбивають на пакети розміру 32. Кількість епох тренування – 100. У якості функції похибки було взято функцію середньої абсолютної похибки.

Результати навчання:

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 4, 300)            721200

 lstm_1 (LSTM)               (None, 300)               721200

 dense (Dense)               (None, 1)                 301

 lambda (Lambda)             (None, 1)                 0

=================================================================
Total params: 1,442,701
Trainable params: 1,442,701
Non-trainable params: 0

_________________________________________________________________
None

Epoch 1/100 loss: 8.9387 - custom_accuracy: 0.0377 - val_loss: 6.7961 - val_custom_accuracy: 0.0524
Epoch 10/100 loss: 1.4742 - custom_accuracy: 0.2839 - val_loss: 2.6847 - val_custom_accuracy: 0.1891
Epoch 20/100 loss: 0.8152 - custom_accuracy: 0.4807 - val_loss: 2.1975 - val_custom_accuracy: 0.2842
Epoch 30/100 loss: 0.5394 - custom_accuracy: 0.6359 - val_loss: 1.8672 - val_custom_accuracy: 0.4089
Epoch 40/100 loss: 0.3636 - custom_accuracy: 0.7912 - val_loss: 1.5711 - val_custom_accuracy: 0.4890
Epoch 50/100 loss: 0.2769 - custom_accuracy: 0.8749 - val_loss: 1.4634 - val_custom_accuracy: 0.5423
Epoch 60/100 loss: 0.2469 - custom_accuracy: 0.9105 - val_loss: 1.3677 - val_custom_accuracy: 0.5850
Epoch 70/100 loss: 0.2331 - custom_accuracy: 0.9214 - val_loss: 1.2943 - val_custom_accuracy: 0.6041
Epoch 80/100 loss: 0.1793 - custom_accuracy: 0.9606 - val_loss: 1.3498 - val_custom_accuracy: 0.6435
Epoch 90/100 loss: 0.2079 - custom_accuracy: 0.9447 - val_loss: 1.2731 - val_custom_accuracy: 0.6427
Epoch 100/100 loss: 0.1389 - custom_accuracy: 0.9754 - val_loss: 1.1785 - val_custom_accuracy: 0.6651
```

Рисунок 3.2.1.1 Тренування першої нейронної мережі

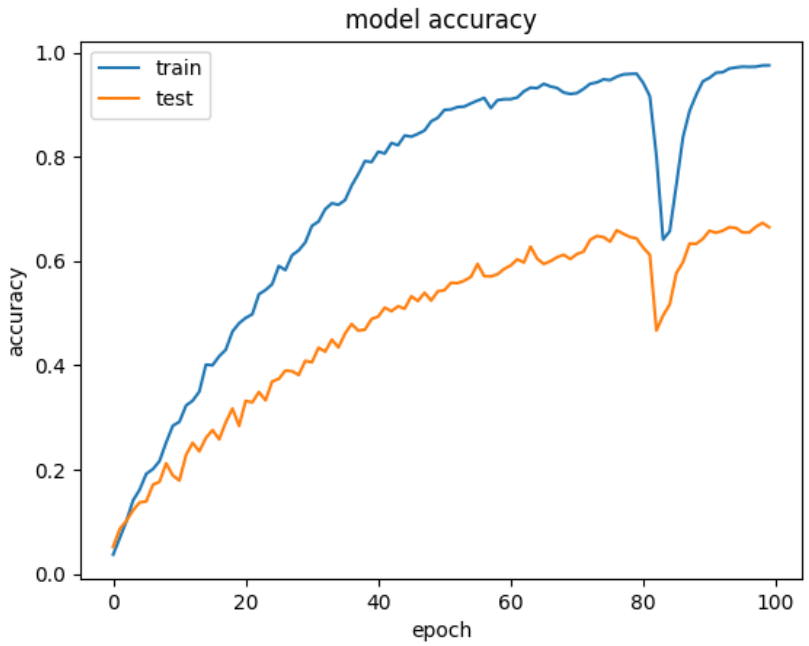


Рисунок 3.2.1.2 Точність першої нейронної мережі

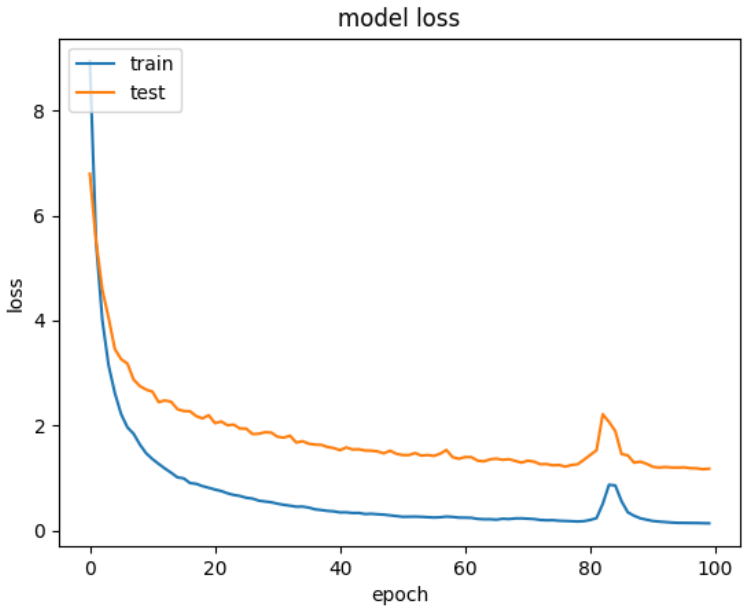


Рисунок 3.2.1.3 Похибка першої нейронної мережі

В результаті отримали мережу, що знаходить правильне значення для послідовності слів з тренувального сету з вірогідністю 97.5% та з тестового з вірогідністю 66.5%.

Також була протестована нейронна мережа, в якій був доданий LTSM шар, аналогічний до першого після нього:

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 4, 300)            721200

lstm_1 (LSTM)                (None, 4, 300)            721200

lstm_2 (LSTM)                (None, 300)               721200

dense (Dense)                (None, 1)                 301

lambda (Lambda)              (None, 1)                 0

=================================================================
Total params: 2,163,901
Trainable params: 2,163,901
Non-trainable params: 0
_________________________________________________________________
None

Epoch 1/100 loss: 9.0518 - custom_accuracy: 0.0365 - val_loss: 6.7433 - val_custom_accuracy: 0.0592
Epoch 10/100 loss: 1.1580 - custom_accuracy: 0.3684 - val_loss: 2.4291 - val_custom_accuracy: 0.2524
Epoch 20/100 loss: 0.5944 - custom_accuracy: 0.5837 - val_loss: 1.7342 - val_custom_accuracy: 0.4225
Epoch 30/100 loss: 0.3737 - custom_accuracy: 0.7786 - val_loss: 1.3812 - val_custom_accuracy: 0.5654
Epoch 40/100 loss: 0.2778 - custom_accuracy: 0.8847 - val_loss: 1.3156 - val_custom_accuracy: 0.5993
Epoch 50/100 loss: 2.5094 - custom_accuracy: 0.3739 - val_loss: 3.5868 - val_custom_accuracy: 0.3408
Epoch 60/100 loss: 0.2653 - custom_accuracy: 0.9190 - val_loss: 1.5211 - val_custom_accuracy: 0.6270
Epoch 70/100 loss: 0.1791 - custom_accuracy: 0.9659 - val_loss: 1.2851 - val_custom_accuracy: 0.6849
Epoch 80/100 loss: 0.1541 - custom_accuracy: 0.9734 - val_loss: 1.2570 - val_custom_accuracy: 0.7082
Epoch 90/100 loss: 0.1363 - custom_accuracy: 0.9748 - val_loss: 1.2031 - val_custom_accuracy: 0.7455
Epoch 100/100 loss: 0.1273 - custom_accuracy: 0.9780 - val_loss: 1.1457 - val_custom_accuracy: 0.7605
```

Рисунок 3.2.1.4 Тренування другої нейронної мережі

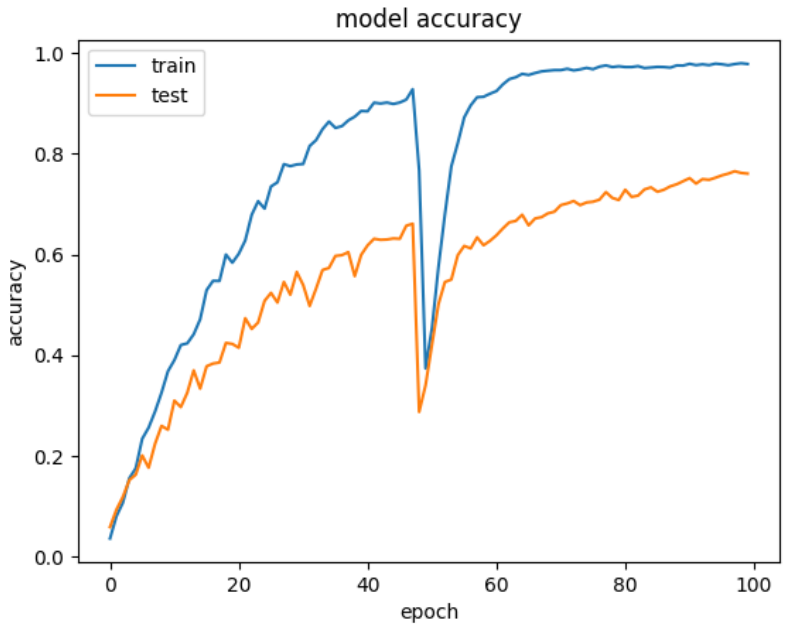


Рисунок 3.2.1.5 Точність другої нейронної мережі

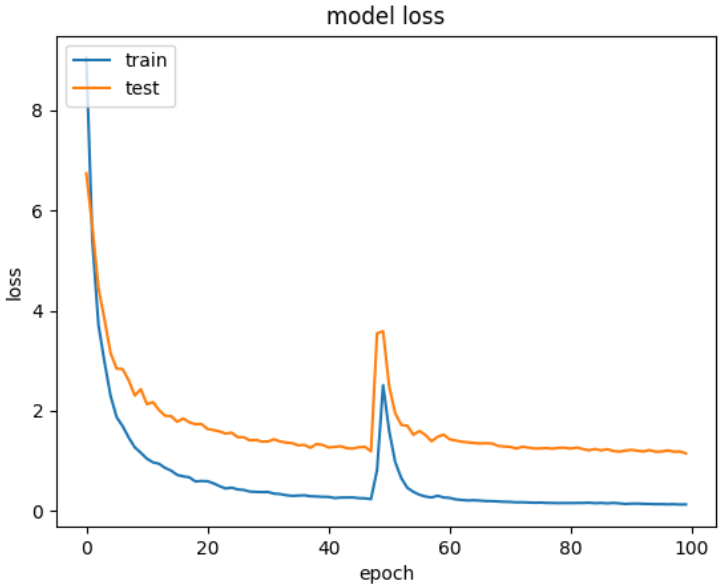


Рисунок 3.2.1.3 Похибка другої нейронної мережі

3.2.2 Аналіз запису музичного твору

Алгоритм визначення назви музичного твору за допомогою аналізу тексту полягає в наступному:

1. Запис музичного треку за допомогою вбудованого мікрофону.

2. Отримання слів пісні із запису за допомогою Google Cloud Speech API.

3. Визначення назви твору за допомогою другої натренованої нейронної мережі

3.2.3 Результати використання другого методу

Після навчання нейронна мережа визначала правильну назву твору за набором слів у 76% випадків. Однак через низький рівень якості аналізу тексту пісні з запису цей алгоритм аналізу музичного твору дав правильну відповідь лише для 8% записів. Крім того, даний метод немає сенсу використовувати для ідентифікації треків без слів, наприклад, класичної музики, та потребує запис треку з мінімальною кількістю сторонніх шумів.

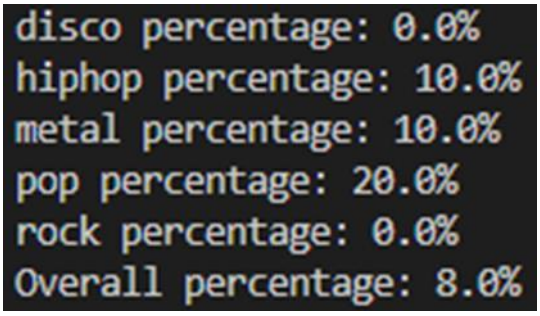


Рисунок 3.2.3.1 Результат використання другої моделі

Висновки

У наш час одним з основних напрямків розвитку штучного інтелекту є обробка звуку. В цій галузі штучний інтелект використовується починаючи з обробки запису з метою видалення сторонніх шумів, закінчуючи розпізнаванням людської мови та її генерацією. Дуже розповсюдженою у житті задачею є аналіз записів музичних творів з метою їх ідентифікаціїї, чи то людина хоче дізнатись назву треку, що їй сподобався, чи стрімінгова платформа має проаналізувати запис на предмет порушення авторських прав перед його публікацією.

У роботі було порівняно 2 методи ідентифікації музичного твору: на основі відбитків відомих пісень та розпізнаванням тексту пісні у записі. Для першого методу було створено базу даних з відбитками пісень та алгоритм порівняння відбитку запису з відомими. У другому методі була створена та натренована нейронна мережа, що за послідовністю слів з тексту пісні визначає назву твору. Другий метод виявився гіршим за перший не зважаючи на гарні результати роботи створеної нейронної мережі, оскільки для нього проблемою стало визначення тексту пісні з запису – стандартні засоби обробки природної мови не підходять, оскільки у розпізнаванні тексту пісень треба відділяти людську мову від музикального супроводження. Покращити результати цього методу можна створивши нову систему, натреновану спеціально для розпізнавання слів у музичному творі.

Список використаних джерел

1. FURUI, Sadaoki. Fifty years of progress in speech and speaker recognition. *The Journal of the Acoustical Society of America*, 2004, 116.4: 2497-2498.
2. BÜCHLER, Michael Christoph. *Algorithms for sound classification in hearing instruments*. 2002. PhD Thesis. ETH Zurich.
3. BOUNTOURAKIS, Vasileios; VRYSIS, Lazaros; PAPANIKOLAOU, George. Machine learning algorithms for environmental sound recognition: Towards soundscape semantics. In: *Proceedings of the Audio Mostly 2015 on Interaction With Sound*. 2015. p. 1-7.
4. MCLOUGHLIN, Ian, et al. Robust sound event classification using deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2015, 23.3: 540-552.
5. XING, Y. F., et al. Sound quality recognition using optimal wavelet-packet transform and artificial neural network methods. *Mechanical Systems and Signal Processing*, 2016, 66: 875-892.
6. CEREZUELA-ESCUDERO, Elena, et al. Sound recognition system using spiking and MLP neural networks. In: *Artificial Neural Networks and Machine Learning–ICANN 2016: 25th International Conference on Artificial Neural Networks, Barcelona, Spain, September 6-9, 2016, Proceedings, Part II 25*. Springer International Publishing, 2016. p. 363-371.
7. CARBONELL, Noëlle; HATON, Jean Paul; PIERREL, Jean-Marie. Artificial intelligence in speech understanding: Two applications at CRIN. *Computers and the Humanities*, 1986, 20.3: 167-172.
8. MCLOUGHLIN, Ian, et al. Robust sound event classification using deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2015, 23.3: 540-552.

9. VALERO, Xavier; ALÍAS, Francesc. Narrow-band autocorrelation function features for the automatic recognition of acoustic environments. *The Journal of the Acoustical Society of America*, 2013, 134.1: 880-890.
10. SHARAN, Roneel V.; MOIR, Tom J. An overview of applications and advancements in automatic sound recognition. *Neurocomputing*, 2016, 200: 22-34.
11. CAI, Yang. Instinctive computing. In: *Artifical Intelligence for Human Computing: ICMI 2006 and IJCAI 2007 International Workshops, Banff, Canada, November 3, 2006, Hyderabad, India, January 6, 2007, Revised Seleced and Invited Papers*. Springer Berlin Heidelberg, 2007. p. 17-46.
12. WU, Huadong; SIEGEL, Mel; KHOSLA, Pradeep. Vehicle sound signature recognition by frequency vector principal component analysis. In: *IMTC/98 Conference Proceedings. IEEE Instrumentation and Measurement Technology Conference. Where Instrumentation is Going (Cat. No. 98CH36222)*. IEEE, 1998. p. 429-434.
13. POTAMITIS, Ilyas; GANCHEV, Todor. Generalized recognition of sound events: Approaches and applications. *Multimedia Services in Intelligent Environments: Advanced Tools and Methodologies*, 2008, 41-79.
14. KOETSIER, Teun. On the prehistory of programmable machines: musical automata, looms, calculators. Mechanism and Machine theory, 2001, 36.5: 589-603.
15. CUMMISKEY, P.; JAYANT, Nikil S.; FLANAGAN, J. L. Adaptive quantization in differential PCM coding of speech. *Bell System Technical Journal*, 1973, 52.7: 1105-1118.
16. JANSSENS, Jelle; VANDACLE, Stijn; BEKEN, Tom Vander. The music industry on (the) line? Surviving music piracy in a digital era. *Eur. J. Crime Crim. L. & Crim. Just.*, 2009, 17: 77.
17. CROXTON, Frederick Emory. Applied general statistics. In: *Applied general statistics*. 1967. p. 754-754.

18. DIETRICH, Cornelius Frank. *Uncertainty, calibration and probability: the statistics of scientific and industrial measurement*. Routledge, 2017.
19. CORMEN, Thomas H., et al. *Introduction to algorithms*. MIT press, 2022.
20. KAMIEN, Morton I.; SCHWARTZ, Nancy Lou. *Dynamic optimization: the calculus of variations and optimal control in economics and management*. courier corporation, 2012.
21. SHEPPARD, William Fleetwood. On the Calculation of the most Probable Values of Frequency-Constants, for Data arranged according to Equidistant Division of a Scale. *Proceedings of the London Mathematical Society*, 1897, 1.1: 353-380.
22. BENNETT, William Ralph. Spectra of quantized signals. *The Bell System Technical Journal*, 1948, 27.3: 446-472.
23. STOICA, Petre, et al. *Spectral analysis of signals*. Upper Saddle River, NJ: Pearson Prentice Hall, 2005.
24. WELCH, Peter. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 1967, 15.2: 70-73.
25. BAILEY, David H.; SWARZTRAUBER, Paul N. A fast method for the numerical evaluation of continuous Fourier and Laplace transforms. *SIAM Journal on Scientific Computing*, 1994, 15.5: 1105-1110.
26. BOASHASH, Boualem. *Time-frequency signal analysis and processing: a comprehensive reference*. Academic press, 2015.
27. HEIDEMAN, Michael T.; JOHNSON, Don H.; BURRUS, C. Sidney. Gauss and the history of the fast Fourier transform. *Archive for history of exact sciences*, 1985, 265-277.
28. VAN LOAN, Charles. *Computational frameworks for the fast Fourier transform*. Society for Industrial and Applied Mathematics, 1992.
29. SCHMIDHUBER, Jürgen, et al. Long short-term memory. *Neural Comput*, 1997, 9.8: 1735-1780.

30. LI, Xiangang; WU, Xihong. Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition. In: *2015 ieee international conference on acoustics, speech and signal processing (icassp)*. IEEE, 2015. p. 4520-4524.
31. AMARI, S.-I. Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Transactions on computers*, 1972, 100.11: 1197-1206.
32. HOPFIELD, John J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 1982, 79.8: 2554-2558.

Відгук

на кваліфікаційну роботу бакалавра на тему:

«Дослідження алгоритмів розпізнавання звукових доріжок та визначення назви музичного твору по звуковій доріжці за допомогою технологій штучного інтелекту»

студента 4-го курсу факультету комп'ютерних наук та кібернетики

Київського національного університету імені Тараса Шевченка

Залужного Юрія Андрійовича

Обробка та розпізнавання звукових доріжок та визначення назви музичного твору по звуковій доріжці за допомогою різноманітних алгоритмів взагалі, та штучного інтелекту зокрема, є важливою задачею в сучасному світі - адже музика є невід'ємною частиною життя людини. Стрімке збільшення наявних даних та технічних можливостей збору, аналізу і порівняння аудіоінформації відкриває нові горизонти для її вирішення.

В дипломній роботі студент Юрій Залужний порівнює два методи: збереження унікальних відбитків треків в базі даних і створення нейронної мережі, що за текстом треку може визначити його назву. Перший метод включає в себе процес попередньої обробки аудіо, отримання відбитків, їх збереження і подальше порівняння з записаним звуком. Другий метод включає в себе використання нейронних мереж для визначення тексту пісні, а потім за допомогою цього тексту визначає назву треку.

Студент провів детальний теоретичний аналіз обох методів, їх переваг і недоліків. Також він виконав дуже докладний та ґрунтовний огляд підходів та історії проблеми автоматичного аналізу аудіо. В останній частині роботи автор описує проведені експерименти та наводить висновки. Проведена робота демонструє глибоке розуміння проблеми та здатність автора використовувати теоретичні знання в практичних задачах.

Можна зробити також зауваження — результати підходу на базі ШІ досить слабкі, тому було б цікаво провести експерименти з іншими архітектурами, і, можливо, покращеним набором ознак. Але це зауваження не зменшує загальної позитивної оцінки роботи.

Вважаю, що кваліфікаційна робота студента Юрія Залужного відповідає всім вимогам, які висуваються до бакалаврських робіт, і заслуговує на оцінку «відмінно», а її автор заслуговує на присвоєння кваліфікації бакалавра.

асистент кафедри обчислювальної
математики  Сергій ДЕНИСОВ

Рецензія

на кваліфікаційну роботу бакалавра на тему:

«Дослідження алгоритмів розпізнавання звукових доріжок та визначення назви музичного твору по звуковій доріжці за допомогою технологій штучного інтелекту»

студента 4-го курсу факультету комп'ютерних наук та кібернетики

Київського національного університету імені Тараса Шевченка

Залужного Юрія Андрійовича

У сучасному світі, де аудіо контент (як частина відео та окремо) стає все більш доступним та розповсюдженим, проблема аналізу звукових доріжок стає дедалі актуальнішою. Робота студента досліджує можливості автоматичного визначення назви музичного твору на базі двох підходів: збереження унікальних відбитків треків в базі даних та використання нейронних мереж для визначення тексту пісні та її назви.

Перша частина роботи присвячена історичному та теоретичному аспекту роботи з аудіо даними взагалі. Надаються відомості про теоретичну та практичну складову різних методів, вказуються їх переваги і недоліки. Студент чітко визначає проблему, поставлену задачу та описує методи, потрібні для виконання роботи.

В другій частині Юрій Залужний проводить обчислювальні експерименти на реальних прикладах з використанням власноруч розробленого програмного забезпечення та порівнює результати, отримані кожним з методів.

Зауважу, що хотілося б більше деталей щодо специфіки побудови нейронної мережі та її тренування, а також експериментів з більш сучасними архітектурами, наприклад, трансформерами. Але ці зауваження не знижують загальної позитивної оцінки роботи.

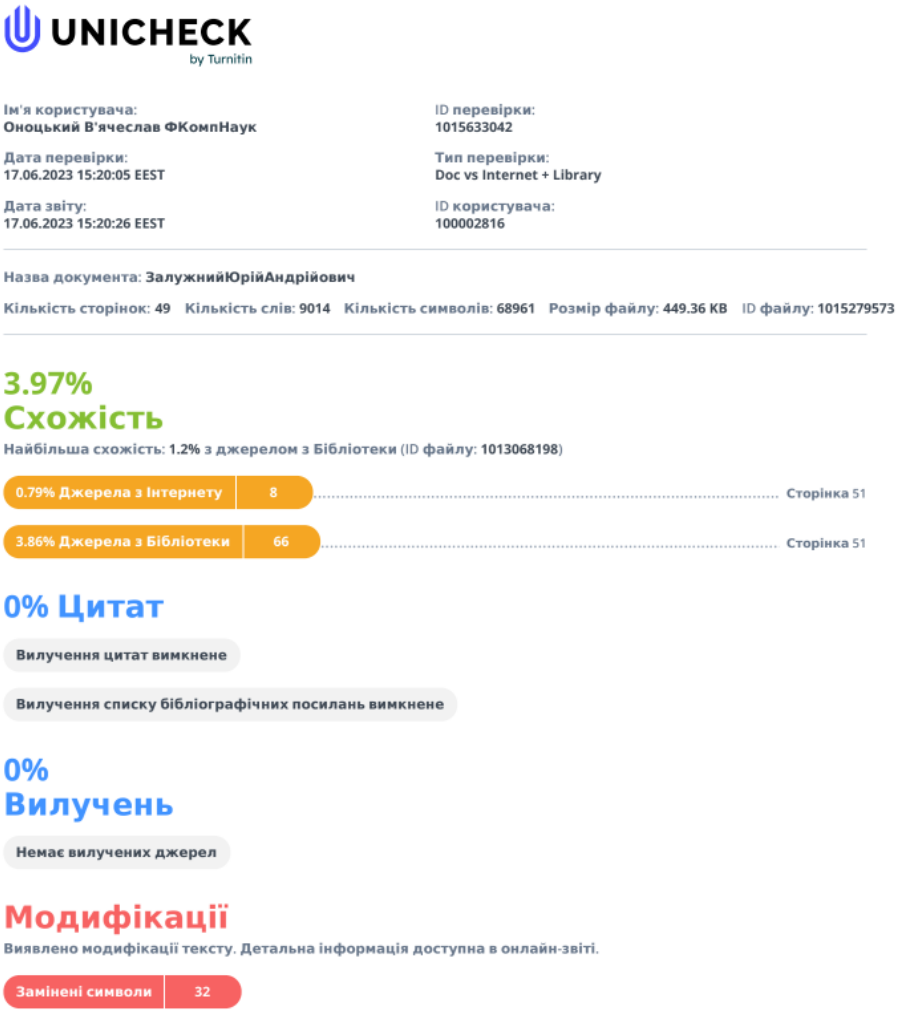
Вважаю, що кваліфікаційна робота студента Залужного Юрія відповідає всім вимогам, які висуваються до бакалаврських робіт, і заслуговує на оцінку «відмінно», а її автор заслуговує на присвоєння кваліфікації бакалавра.

кандидат технічних наук,
доцент  Катерина ГОЛУБЄВА

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

СИСТЕМА ЗАПОБІГАННЯ ТА ВИЯВЛЕННЯ АКАДЕМІЧНОГО ПЛАГІАТУ

Довідка про оригінальність кваліфікаційної роботи за освітнім рівнем бакалавр



Експертна оцінка роботи науковим керівником :

Робота студента 4-го курсу Залужного Юрія Андрійовича «Дослідження алгоритмів розпізнавання звукових доріжок та визначення назви музичного твору по звуковій доріжці за допомогою технологій штучного інтелекту» виконана самостійно, при цьому обсяг цитувань та запозичень становить 3.97% та не перевищує норму.

`

Науковий керівник: _____ Денисов С.В.
(підпис) (ПІБ)

Оператор: _____ Оноцький В.В.
(підпис) (ПІБ)