УДК 004.89

І. Ю. Шевченко
І. Р. Арсенюк

АНАЛІЗ ОСНОВНИХ ПІДХОДІВ ВИЯВЛЕННЯ СХОЖОСТІ МУЗИЧНИХ МЕЛОДІЙ У КОНТЕКСТІ ПОШУКУ МОЖЛИВОГО ПЛАГІАТУ

Вінницький національний технічний університет

Анотація
У роботі представлено аналіз основних підходів виявлення схожості музичних мелодій у контексті пошуку можливого плагіату. Даний аналіз є фундаментом для подальшої розробки вдосконаленного підходу виявлення схожості музичних мелодій.
Ключові слова: аналіз схожості мелодій, музичний плагіат, контрастивне навчання.

Abstract
The paper presents an analysis of the main approaches to detecting similarity between musical melodies in the context of searching for possible plagiarism. This analysis serves as a foundation for the further development of an improved approach to identifying melodic similarity
Key words: melody similarity analysis, musical plagiarism, contrastive learning.

Вступ

Натепер проблема музичного плагіату залишається однією з найгостріших як у правовому, так і в художньому контексті. У більшості випадків плагіат стосується не всього твору, а окремих коротких фрагментів – мелодій, гармонічних ходів або ритмічних мотивів, що тривають лише кілька тактів. Саме ці елементи формують впізнаваність композиції та її художню цінність.

Метою дослідження є аналіз основних методів виявлення схожості музичних мелодій, як фундамент для подальшої розробки вдосконаленного підходу виявлення схожості.

Результати дослідження

Аналіз показав, що існуючі підходи до аналізу музичної схожості можна умовно поділити на три групи:

- алгоритмічні. Наприклад, динамічне часове вирівнювання (Dynamic Time Warping, DTW);
- нейромережеві (згорткові нейронні мережі (Convolutional Neural Networks, CNN); рекурентні нейронні мережі (Recurrent Neural Networks, RNN);
- контрастивні. Наприклад, кероване контрастивне навчання (Supervised Contrastive Learning);

Порівняння основних підходів до виявлення музичної схожості наведено у таблиці 1.1.

Алгоритмічний підхід. Одним із найдавніших і найпоширеніших підходів аналізу схожості у часових послідовностях є динамічне часове вирівнювання. Він був запропонований ще у 1970-х роках для порівняння мовних сигналів, а згодом його почали застосовувати і до музичних даних. Його суть полягає у пошуку оптимального вирівнювання між двома часовими рядами, що мінімізує сумарну відстань між їх точками. DTW дозволяє ефективно порівнювати мелодії, що мають різний темп, довжину або локальні зсуви у часі, зберігаючи при цьому схожість загальної структури. Завдяки цьому підхід став основою для порівняння фраз у музичних творах, навіть якщо вони виконані з різною швидкістю [1].

Основний недолік DTW полягає у високій обчислювальній складності, що ускладнює його застосування до великих баз даних, а також у чутливості до структурних перестановок – зміни порядку нот або гармонічних фраз можуть суттєво вплинути на результат. Проте, DTW часто використовується як базовий еталон для порівняння ефективності сучасних методів, заснованих на нейромережевому чи контрастивному навчанні [2].

Таблиця 1.1 – Основні підходи до виявлення музичної схожості

Підхід	Сутність
На базі динамічного часового вирівнювання.	Вирівнює послідовності по часу з урахуванням змін темпу. Застосовується до MIDI.
На основі CNN.	Аналізують спектрограми для виявлення локальних ознак.
На основі RNN.	Моделюють часові залежності в нотних послідовностях.
На базі методу символічної багатовимірної подібності (Multi-dimensional Evaluation of Symbolic Music Fragments, MESMF).	Обчислює подібність MIDI-фрагментів з урахуванням висоти, тривалості та консонансів.
На базі паросполучення у двочастковому графі (Bipartite Graph Matching, BGM).	Представляє ноти або сегменти мелодій як вершини двох множин графа, а подібність між ними – як ребра з вагами. Використовує максимальне паросполучення для пошуку найкращих відповідностей між фрагментами.
На основі функції потрійних втрат (Triplet Loss).	Формує embedding-простір на основі трійок: якір, позитивний і негативний приклади.
На основі керованого контрастивного навчання.	Формують embedding-простір на основі парних прикладів «схожий/несхожий».

Нейромережеві підходи. Сучасні підходи до аналізу музичної схожості все частіше використовують нейронні мережі, які здатні автоматично навчатися релевантних ознак із даних без попереднього ручного проєктування. Одним із архітектурних рішень для роботи з аудіосигналами є згорткові нейронні мережі, які застосовуються для аналізу мел-спектрограм – двовимірних зображень, що відображають зміну частотних компонентів звуку у часі. Вони здатні автоматично виявляти локальні спектральні патерни, що відповідають таким характеристикам музики, як тембр, атака, паузи, гармонійна структура чи ритмічні повтори. Завдяки цьому CNN-моделі можуть розпізнавати подібність між аудіозаписами навіть тоді, коли виконання, інструменти або виконавці відрізняються [3].

Ключовою перевагою CNN є можливість інваріантності до невеликих часових і частотних зсувів, що робить їх особливо ефективними для задач виявлення музичної схожості та класифікації звуків. Проте недоліками цих мереж є непрозорість прийнятих рішень, тобто результати важко інтерпретувати з точки зору конкретних музичних елементів, а також чутливість до шумів та варіацій у записі (наприклад, до різниць у гучності чи якості мікрофонів), що може знижувати точність аналізу, що є критичним у юридичному контексті [3].

Рекурентні нейронні мережі. Для аналізу часових залежностей у музичних сигналах часто застосовують RNN, які здатні обробляти послідовні дані, запам’ятовуючи попередні стани, що робить їх придатними для моделювання мелодійних і ритмічних структур. Проте звичайні RNN страждають від проблеми зникання або вибуху градієнтів, через що їх ефективність падає під час роботі з довгими послідовностями. Для подолання цих обмежень була розроблена архітектура LSTM (Long Short-Term Memory), яка запроваджує спеціальні елементи пам’яті (memory cells) та ґейтові механізми (input, output, forget gates), що дозволяють зберігати інформацію протягом тривалого часу.

LSTM успішно застосовуються в задачах аналізу мелодій, ритмічних патернів, гармонічних переходів та генерації музики. Перевагою LSTM є здатність враховувати довгостроковий контекст – наприклад, залежності між мотивами на початку та кінці фрази. Це дозволяє моделі «розуміти» музичну логіку, а не лише локальні фрагменти, як у випадку CNN. Проте LSTM-моделі мають і певні недоліки: вони є ресурсомісткими (вимагають багато пам’яті та часу для навчання, особливо під час роботи з великими MIDI-датасетами); важко забезпечити інтерпретованість результатів, адже зв’язки між внутрішніми станами моделі не завжди мають очевидне музичне пояснення; існує ризик перенавчання при недостатній кількості різноманітних прикладів; під час роботи з великими наборами фрагментів LSTM часто поступається ефективністю контрастивним методам, які формують компактні векторні представлення для пошуку подібностей [4].

Для оптимізації обчислень іноді використовують спрощений варіант – керований рекурентний блок (Gated Recurrent Unit, GRU), який має менше параметрів, але зберігає здатність моделювати часові залежності. LSTM та GRU стали базою для багатьох сучасних систем музичного аналізу, які поєднують їх із CNN або контрастивними підходами, формуючи гібридні архітектури (наприклад, CRNN) [3].

Символічні підходи аналізу музики базуються не на аудіосигналі, а на структурованому по-данні нотних подій, наприклад у форматі MIDI (Musical Instrument Digital Interface). Вони дозволяють

безпосередньо працювати з параметрами, що описують музичну логіку – висотою звуку (pitch), тривалістю (duration), сильною долею (downbeat) та іншими характеристиками ритміко-мелодійної структури. Одним із найвідоміших підходів цього класу є Symbolic Melodic Similarity, що порівнює мелодії на основі інтервальних послідовностей, контурів та строкових моделей, дозволяючи оцінювати подібність між фрагментами за їх мелодичними характеристиками.

Перевагою Symbolic Melodic Similarity є стійкість до транспозицій і темпових варіацій, оскільки методи цього класу працюють з відносними мелодичними ознаками. Вони також ефективні для виявлення локальних збігів та повторюваних мотивів. Недоліком є обмежена здатність враховувати ритмічні та гармонійні аспекти, а також залежність точності від обраного способу представлення мелодії [5].

Графові підходи. Одним із сучасних символічних підходів базується на пошуку паросполучення у двочастковому графі (Bipartite Graph Matching, BGM), який представляє музичні фрагменти у вигляді вершин графа, а подібність між окремими нотами або сегментами – у вигляді вагових ребер [6]. Пошук збігів між двома мелодіями виконується шляхом максимального паросполучення у двочастковому графі (Maximum Bipartite Matching), що дозволяє знайти найкращу відповідність між частинами двох послідовностей. На відміну від MESMF, який оцінює схожість у векторному вигляді, метод BGM ураховує локальні відповідності та дозволяє виявляти плагіат навіть тоді, коли спільні фрагменти не є послідовними у часі.

Основним недоліком такого підходу є висока обчислювальна складність, яка зростає квадратично відносно кількості нот, тому метод переважно застосовується до коротких фрагментів.

Підхід на базі контрастивних методів навчання є одним із найсучасніших напрямів у задачах порівняння ознак і виявлення схожості між об’єктами. Їх основна ідея полягає у формуванні векторного простору ознак (Embedding Space), де схожі об’єкти розташовуються поруч, а відмінні – на значній відстані. Такі методи дозволяють моделі навчитися «відчувати» подібність не через пряме відтворення сигналу, а через відносне розташування зразків у просторі [2].

Одним із найпоширеніших представників цього класу є Triplet Loss [7]. Він формує трійки прикладів:

- Anchor (якір) – базовий об’єкт;
- Positive (позитивний приклад) – об’єкт того ж класу або подібний за змістом;
- Negative (негативний приклад) – відмінний за змістом або належить до іншого класу.

Метою є мінімізація відстані між парою *anchor–positive* та максимізація між *anchor–negative*. Таким чином, нейронна мережа поступово структурує простір так, щоб подібні мелодійні фрагменти опинялися поруч. Однак основним недоліком Triplet Loss є складність вибору оптимальних трійок – при великій кількості фрагментів кількість можливих комбінацій росте експоненційно, що збільшує час навчання і ризик локальних мінімумів [7].

Більш вдосконаленим варіантом є Supervised Contrastive Learning (SupCon) [2]. На відміну від Triplet Loss, SupCon використовує всі позитивні приклади в межах одного батчу як пари для поточного зразка, що забезпечує стабільніше навчання та ефективніше використання даних. Це особливо важливо для задач з обмеженими наборами фрагментів, як у випадку музичного аналізу.

У контексті музичного плагіату та аналізу мелодій SupCon дозволяє одночасно враховувати десятки схожих фрагментів, формуючи чітко розмежовані кластери у векторному просторі. Це підвищує точність пошуку схожих мелодій, зберігаючи при цьому інтерпретованість результатів, адже embedding-простір можна візуалізувати за допомогою методів зменшення розмірності (t-SNE, UMAP) [8].

Висновки

Здійснений аналіз сучасних підходів виявлення музичної схожості показав, що кожен з них має певні переваги та обмеження. Так, наприклад, алгоритмічні підходи забезпечують точне вирівнювання послідовностей у часі, проте мають високу обчислювальну складність і погано масштабуються для великих баз даних. Нейромережеві методи (CNN; LSTM) дозволяють автоматично виділяти ознаки з мелспектрограм, враховуючи контекст і тембр звучання, проте вимагають великих навчальних наборів і мають обмежену інтерпретованість результатів. Символічні підходи (наприклад, MESMF), ефективні для аналізу MIDI-представлень та інваріантні до транспозицій, але чутливі до ритмічних варіацій і потребують попередньої нормалізації даних. Графові підходи, (зокрема BGM), розглядають мелодії як набори елементів (нот або сегментів), між якими встановлюються зважені зв’язки подібності. Завдяки використанню оптимального зіставлення між двома множинами, вони забезпечують високу точність локального порівняння, проте мають значні обчислювальні витрати при роботі з великими наборами

фрагментів. Підходи на базі контрастивних методів (Triplet Loss, Supervised Contrastive Learning) демонструють найкращий баланс між точністю та інтерпретованістю: вони формують векторні простори, у яких подібні фрагменти розташовуються поруч, що полегшує пошук схожих мелодій навіть у великих базах даних.

Отже, найбільш перспективним напрямом для подальших досліджень може бути поєднання контрастивного навчання з нейромережевими методами побудови векторних представлень музичних даних, що надасть змогу моделі автоматично навчатися інформативним ознакам. Завдяки використанню нейромереж можна формувати простори ознак, у яких схожі фрагменти природно групуються за мелодійними та ритмічними характеристиками.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Bringmann K., Fischer N., van der Hoog I., Kipouridis E., Kociumaka T., Rotenberg E. Dynamic Dynamic Time Warping [Електронний ресурс]. – Режим доступу: https://arxiv.org/abs/2310.18128
2. Chen T., Kornblith S., Norouzi M., Hinton G. A Simple Framework for Contrastive Learning of Visual Representations [Електронний ресурс]. – Режим доступу: https://arxiv.org/abs/2002.05709
3. Choi K., Fazekas G., Sandler M. Convolutional Recurrent Neural Networks for Music Classification [Електронний ресурс]. – Режим доступу: https://arxiv.org/abs/1609.04243
4. X. Van Houdt G., Van de Wiele T., De Bruyn S. A Review on the Long Short-Term Memory Model [Електронний ресурс]. – Режим доступу: https://link.springer.com/article/10.1007/s10462-020-09838-1
5. Velardo V., Vallati M., Jan S. Symbolic Melodic Similarity: State of the Art and Future Challenges [Електронний ресурс]. – Режим доступу: https://direct.mit.edu/comj/article/40/2/70/1856306
6. X. He T., Liu W., Gong C., Yan J., Zhang N. Music Plagiarism Detection via Bipartite Graph Matching [Електронний ресурс]. – Режим доступу: https://arxiv.org/abs/2107.09889
7. Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering [Електронний ресурс]. – Режим доступу: https://arxiv.org/pdf/1503.03832
8. X. Van der Maaten L., Hinton G. Visualizing Data using t-SNE [Електронний ресурс]. – Режим доступу: https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

Шевченко Іван Юрійович – студент групи 2КН-24м, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: 01-24-184.stud@vntu.edu.ua

Арсенюк Ігор Ростиславович – кандидат технічних наук, доцент, доцент кафедри комп`ютерних наук, Вінницький національний технічний університет, м. Вінниця, e-mail: air@vntu.edu.ua

Shevchenko Ivan Yuriyovych – Student of the 2CS-24M group, Department of Intellectual Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: 01-24-184.stud@vntu.edu.ua

Arseniuk Igor Rostyslavovych – Associate Professor of the Computer Sciences Department, Vinnytsia National Technical University, Vinnytsia, e-mail: air@vntu.edu.ua