УДК 004.94 https://doi.org/10.15407/pp2026.03.111

Б.Є. Панченко, Т.О. Пасенченко

ВИЯВЛЕННЯ МУЗИЧНОГО ПЛАГІАТУ ЗА ДОПОМОГОЮ МОДЕЛЕЙ НЕЙРОННИХ МЕРЕЖ

Останніми роками створення та публікація музичних творів значно спростилися завдяки розповсюдженню цифрових технологій, генеративного штучного інтелекту та інструментів на його основі. Водночас задача виявлення музичного плагіату залишається важливою для правовласників, оскільки своєчасне виявлення елементів плагіату допомагає зберегти прибуток від реалізації авторських прав на твори. Наприклад, система YouTube Content ID дозволяє не лише виявляти та блокувати музичний плагіат, а й відраховувати прибуток з нього оригінальному правовласнику.

Методи виявлення музичного плагіату з використанням штучного інтелекту (а саме - моделей нейронних мереж) мають більшу гнучкість у порівнянні з класичними методами. Наприклад, можна знаходити плагіат при зміні музичного інструменту та аранжуванні, а також за наявності шумів. Недоліками цих методів є необхідність графічних процесорів для запуску нейронних мереж.

У статті наведено опис відмінностей аудіозапису, музичного аудіозапису та музичного твору. Описана задача виявлення музичного плагіату та деякі методи її розв'язання. Порівнюються класичні методи, що не використовують моделі нейронних мереж, та методи, що використовують такі моделі. Розглядається модель нейронних мереж MelodySim, що враховує подібність мелодій, та датасет, що використовувався для її тренування. Досліджується точність її передбачень на рівні сегментів шляхом створення допоміжного датасету та оптимізації порогу бінарної класифікації (decision threshold) пар сегментів задля підвищення значення метрики F1-Score.

Отримані результати можуть застосовуватися для покращення як передбачень MelodySim, так і архітектури самої моделі. Подібний до розглянутого підхід може бути використаний для покращення точності передбачень інших моделей зі схожою архітектурою.

Подальші дослідження можуть бути спрямовані на покращення моделі MelodySim та на розширення датасету.

Ключові слова: штучний інтелект, deep learning, MPD, MelodySim, бінарна класифікація.

B. Panchenko, T. Pasenchenko

MUSIC PLAGIARISM DETECTION USING NEURAL NETWORK MODELS

In recent years, the creation and publication of musical works have become significantly simplified thanks to the spread of digital technologies, generative artificial intelligence, and tools based on it.

At the same time, detecting musical plagiarism remains crucial for copyrights holders, as the timely detection of plagiarized elements helps to preserve revenue from the realization of copyrights on these works. For instance, the YouTube Content ID system not only detects and blocks music plagiarism but also redirects the revenue generated from it to the original copyright holder.

Methods for detecting music plagiarism which use artificial intelligence (specifically, neural network models) offer greater flexibility compared to classical methods. For example, plagiarism can be detected even when the musical instrument or arrangement changes, as well as in the presence of noise. The drawback of these methods is the requirement of graphics processing units to run the neural networks.

The article describes the differences between an audio recording, a musical audio recording, and a musical work. It outlines the problem of musical plagiarism detection and various methods for solving it. A comparison is made between classical methods that do not utilize neural network models and methods that incorporate them.

The MelodySim neural network model, which accounts for melodic similarity, is examined alongside the dataset used for its training. The accuracy of its predictions at the segment level is investigated by creating an auxiliary dataset and optimizing the binary classification decision threshold for segment pairs to increase the F1-Score metric value.

The obtained results can be applied to improve both the predictions of MelodySim and the architecture of the model itself. An approach similar to the one discussed can be used to improve the prediction accuracy of other models with a similar architecture.

Future research could be aimed at improving the MelodySim model as well as expanding the dataset.

Keywords: artificial intelligence, deep learning, MPD, MelodySim, binary classification.

Вступ

Кількість аудіозаписів, що створюються та стають доступними онлайн, наразі невпинно зростає. Так, згідно зі звітом Luminate [1], упродовж 2025 року до провайдерів цифрових послуг щоденно надходило понад 100 тисяч нових унікальних ідентифікаторів аудіозаписів ISRC (International Standard Recording Code), що на 7% більше аналогічного показника 2024 року.

Аудіозапис (у вигляді аудіофайлу) - це цифровий контейнер, що містить впорядкований масив даних про параметри оцифрованої звукової хвилі. Ця хвиля може містити довільні звуки, серед яких зустрічаються фіксації виконання музичних творів. Такі аудіозаписи для визначеності можна називати музичними аудіозаписами.

Під музичними творами часто розуміють продукт творчості композитора – паперові ноти чи вже цифрові файли від комп'ютерних програм для створення музики (наприклад, у форматах MIDI чи DAW). Музичні твори та файли, що є їхніми носіями, створюються за допомогою засобів музичної виразності, зокрема нот (висот, тривалості звуків та їхніх додаткових атрибутів) – послідовностей інтервалів між ними (мелодій), відстаней між звучанням нот у часі (з яких складається ритм та темп), логічних об'єднань звуків (гармоній), сили звучання (динаміки) тощо.

Створення музичного аудіозапису за допомогою комп'ютера може відбуватися безпосередньо шляхом генерації нот та файлів у форматі MIDI з подальшим синтезуванням аудіофайлу. Поширеними є також комбіновані підходи – аранжування (оркестрування) мелодії та мікшування (зведення) вже записаних звукових доріжок. Якщо перші методи були відомі ще у минулому сторіччі та не вимагали значних потужностей [2], сучасні часто базуються на використанні генеративних нейронних мереж [3]. Це дозволяє створювати музичні аудіофайли на основі текстових описів та оминати генерацію нот із їхнім озвучуванням. Існують також нейронні мережі, що дозволяють генерувати MIDI-файли за допомогою текстових описів [4].

Музичні аудіозаписи можуть містити як елементи оригінальних музичних творів, так і запозичені фрагменти інших. Якщо таке запозичення відбулося без згоди з правовласником, його називають плагіатом. Правовласникам при цьому важливо виявляти випадки плагіату та вжити відповідні дії для збереження свого прибутку.

Задача виявлення музичного плагіату (MPD – Music Plagiarism Detection) є спеціалізованою задачею пошуку музичної інформації (MIR - Music Information Retrieval). Вона полягає у виявленні в довільному аудіозаписі фрагментів музичних творів, захищених авторським правом. Причому, можуть виявлятися як скопійовані фрагменти аудіозаписів, так і плагіат на рівні мелодійного ряду. Останній варіант є більш цікавим, оскільки дозволяє виявляти менш очевидний музичний плагіат.

У цій роботі розглянуті класичні та нейромережеві методи розв'язання задачі MPD, зокрема розглянута модель MelodySim [5]. Виконана оптимізація порогу класифікації цієї моделі на рівні сегментів.

Огляд існуючих методів

Методи розв'язання задачі MPD можна поділити на дві групи – класичні, тобто без використання моделей нейронних мереж (neural network models), та на основі нейронних мереж.

Кожен з них спочатку перетворює вихідний аудіозапис чи його сегмент у певний набір ознак, які потім порівнюються з іншими наборами таких самих ознак, що можуть бути організовані у базу даних. Такі набори називають цифровими відбитками (digital fingerprints).

Цифрові відбитки мають відображати узагальнену інформацію про мелодійну та звукову складові аудіозапису, бути меншими за аудіозапис, легко обчислюватися та порівнюватися між собою. Останніми трьома властивостями вони нагадують хеш-суми (хоча, відбитки схожих аудіофайлів тут мусять теж бути схожими), тоді як узагальнена інформація про мелодійну складову нагадує інваріанти Заріпова [6].

Усі методи розв'язання задачі MPD містять два ключові алгоритми – для виокремлення цифрових відбитків та для їх порівняння між собою. Саме в цих алгоритмах можуть використовуватися нейронні мережі.

Класичні методи полягають у генерації наборів цифрових відбитків (digital fingerprints) для всіх аудіозаписів. Одним із перших таких методів був розроблений для застосунку Shazam [7]. Для формування цифрових відбитків у ньому використовуються мапи піків частот у спектрограмі – так звані карти сузір'їв (constellation maps). Також був запропонований індекс баз даних для ефективного пошуку таких відбитків.

Системи, побудовані на основі класичних методів, здатні обробляти значні об'єми нових даних (співставлення сотен тисяч нових аудіозаписів на день з наявними автентичними). Прикладом такої системи є YouTube Content ID [8], яка ще до розвитку технологій глибокого навчання (з 2007 року) відстежувала музичний плагіат на платформі YouTube.

Із розвитком технологій штучного інтелекту (ШІ), а саме глибокого навчання (deep learning) почали з'являтися методи розв'язання задачі MPD, що використовують моделі нейронних мереж. Ці методи потенційно здатні виявляти більш прихований плагіат, зокрема, виявляти складне аранжування (як, наприклад, зміну тональності), кавер-версії, фонові шуми тощо.

Суттєвим недоліком цих методів є значна обчислювальна складність тренування та застосування на реальних даних (інференсу) нейронних мереж. Зокрема, стає необхідним використання графічних процесорів (GPU - graphical processing unit).

Однією із сучасних моделей нейронних мереж для розв'язання задачі MPD є модель MelodySim [5]. Її особливістю є акцент на визначення плагіату не лише за аудіоподібністю сегментів аудіозаписів, а й за подібністю мелодій (melody similarity), що складають оригінальні музичні твори.

MelodySim як модель нейронних мереж є триплетною (сіамською) нейронною мережею. Під час тренування вона приймає трійки сегментів (в оригіналі розміром 10 секунд) аудіофайлів, де перші два сегменти є схожими один на одного, а третій – відмінний від них. Під час інференсу на вхід приймаються два відбитки сегментів аудіофайлів та повертається ймовірність їхньої схожості (тобто, можливого плагіату) в інтервалі $[0, 1]$.

Для того, щоб модель враховувала мелодійну інформацію, оригінальні сегменти аудіофайлів попередньо обробляються моделлю акустичного розуміння музики MERT [9] на 95 мільйонів параметрів (MERT-v1-95M), яка фактично використовується у ролі енкодера (encoder). Отримані подання (embeddings) і є еквівалентом цифрових відбитків із класичних методів виявлення музичного плагіату.

Таким чином, як два алгоритми у складі методу виявлення музичного плагіату використовуються дві моделі нейронних мереж: для генерування ознак застосовується модель MERT, а для їх порівняння між собою - модель MelodySim.

Для порівняння двох цілих аудіофайлів довільної довжини вони спочатку розбиваються на сегменти, а далі ці сегменти порівнюються попарно моделлю MelodySim. Будується матриця подібності, що складається з передбачень моделі для кожної пари сегментів та агрегується в єдиний висновок щодо взаємного плагіату згідно із двома пороговими значеннями (thresholds) та порогу класифікації подібності двох сегментів (частки подібних сегментів у рядках та стовпцях матриці) – чи є подібними ці сегменти?

Автори моделі MelodySim також розробили однойменний датасет (dataset) для її тренування, що складається з оригіналу та трьох версій композицій з датасету Slakh2100 [10]. Ці версії були згенеровані шляхом застосування таких аугментацій (augmentations) MIDI-файлів, як заміна інструменту, додання випадкової ноти, інверсія акордів та їхнє арпеджування, випадкове заглушення треків тощо. Після синтезування аудіофайлів також були застосовані аугментації на рівні всього файлу, а саме – зсув висоти звучання, зміна швидкості відтворення (темпу) та зсув часу відтворення.

Постановка задачі

Використовуючи модель MelodySim на однойменному датасеті авторами моделі був досягнутий відомий результат – на рівні аудіофайлів зважена метрика F1-Score стала дорівнювати 0.98. Проте сама модель працює лише на рівні сегментів – вона порівнює два сегменти та повертає ймовірність того, що вони подібні. Тобто розв'язується задача бінарної класифікації, де прикладами є пари сегментів, які необхідно віднести до одного із двох класів – схожість (плагіат) та несхожість (попарна оригінальність).

За замовчуванням, поріг класифікації (decision threshold) цієї моделі на рівні сегментів встановлений як 0.5. У публікації [5] автори моделі наголосили на тому, що майбутнє налаштування (tuning) цього порогового значення може більше збалансувати компроміс між влучністю (precision) та повнотою (recall).

Отже, задача полягає у налаштуванні (оптимізації) порогу класифікації для максимізації значень метрики F1-Score.

Для розв'язання задачі необхідно:

1. Згенерувати допоміжний датасет пар сегментів із основного датасету MelodySim.
2. Провести налаштування порогу класифікації на тренувальній частині (train split) цього датасету.
3. Виміряти значення метрики F1-Score на тестовій частині (test split) датасету.

Основна частина

Для формування допоміжного датасету з вихідного датасету MelodySim було відібрано 125 аудіофайлів (треків). Як зазначалося вище, кожен аудіофайл представлений оригіналом та трьома його версіями, котрі розміщені у відповідних папках. Усі аудіофайли розбиті на сегменти тривалістю 10 секунд. Саме на цій тривалості й тренувалася модель MelodySim.

Допоміжний датасет був розбитий на тренувальну та тестову частину, виходячи із співвідношення 80%/20%. Валідаційна частина у даному випадку не потрібна, бо оптимізується лише один параметр (поріг класифікації). Тренувальна частина допоміжного датасету була згенерована на основі перших 100 аудіофайлів, тестова частина – на основі 25 аудіофайлів, що залишилися.

Розмір всього датасету становить 5000 прикладів (samples), що є парами сегментів. Відповідно розмір тренувальної частини складає 4000 прикладів, а тестової – 1000. Датасет було згенеровано як збалансований, тобто на кожен позитивний приклад припадає рівно один негативний.

Позитивні приклади тренувального датасету генеруються наступним чином – з кожного із 100 аудіофайлів обирається по 10 довільних пар сегментів для кожної з 2 випадкових комбінацій версій. Тобто, всі позитивні приклади є парами сегментів зі спільним номером із конкретного аудіофайлу, в них відрізняються лише версії.

Негативні приклади тренувального датасету створюються так, щоб у кожній парі сегментів відрізнялися номери аудіофайлів, версії та номери сегментів.

Тестовий датасет формуються аналогічно тренувальному, відрізняються лише аудіофайли та їхня кількість.

Оптимальний поріг класифікації, що максимізує метрику F1-Score, визначався за допомогою підходу пошуку за сіткою (grid search) для підбору параметрів (parameter sweep). Інтервал пошуку – $[0, 1]$, кількість вузлів у сітці – 4000. Похибка становить 0.00025, що є цілком прийнятним значенням.

Для кожного з порогів класифікації були знайдені вектори передбачень, порівнюючи які з еталонними значеннями (ground truth), отримано матриці невідповідностей (confusion matrixes). З елементів цих матриць (TP, TN, FP, FN) можна знайти влучність (precision) та повноту (recall) для кожного значення порогу класифікації з сітки за наступними формулами:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}$$

Отримані значення зображені у вигляді кривої влучності-повноти (або PR-кривої) на Рис. 1.

Із використанням значень влучності та повноти для всіх значень порогу класифікації було знайдено оптимум, що максимізує метрику F1-Score. Він складає 0.9718 – відповідна точка відмічена на PR-кривій (Рис. 1).

Значення метрики F1-Score у разі цього порогу класифікації становить 0.7858. Ця метрика є середнім гармонійним влучності та повноти. А вони самі приймають наступні значення при оптимальному порозі:

$$Precision = 0.7935, Recall = 0.7783$$

На початку PR-кривої спостерігаються характерні коливання. Вони пов'язані з тим, що крива починає будуватися з високого (близького до 1) значення порогу класифікації, внаслідок чого кількість позитивних передбачень (істинних – TP та хибних – FP) дуже мала. Кожна зміна цих величин спричиняє коливання влучності, доки їхня сума (знаменник формули для знаходження влучності) не зросте достатньо, щоб нівелювати ці зміни.

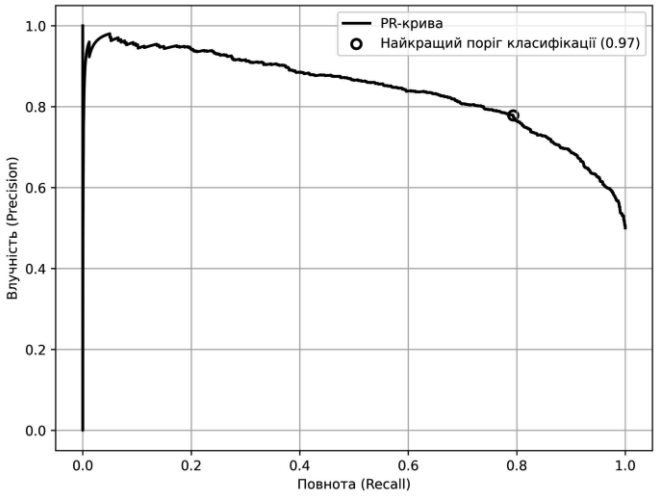

Рис. 1. PR-крива

Застосовуючи знайдене оптимальне значення порогу класифікації на тестовій частині допоміжного датасету знаходимо значення метрики F1-Score, що дорівнює 0.7955. Це й підтверджує коректність знайденого порогу класифікації.

Для того, щоб пересвідчитись у якості передбачень моделі із новим порогом класифікації, на основі даних з тестової частини побудуємо ROC-криву (Рис. 2). Значення площі під нею дорівнює $AUC = 0.8693$.

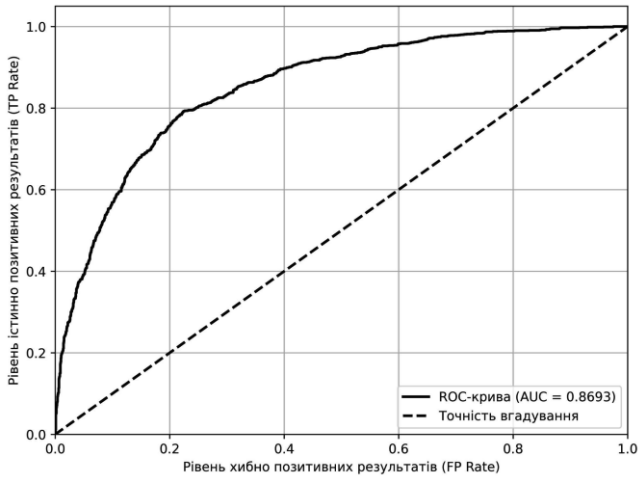

Рис. 2. ROC-крива

Отримані результати свідчать про здатність моделі досягати значно кращих результатів за оптимізації деяких гіперпараметрів, зокрема порогу класифікації. Проте досягнутої точності все ще недостатньо для промислового використання цієї моделі на рівні сегментів. Якщо на рівні аудіофайлів похибки порівняння сегментів можуть бути нейтралізовані під час агрегування матриці подібності, для покращення результатів на рівні сегментів, окрім оптимізації порогу класифікації, необхідно покращувати саму модель. Це можна зробити як удосконаленням її архітектури – наприклад, збільшенням її розміру, так і за допомогою розширення датасету й підвищення тривалості тренування.

Висновки

Отримані результати дозволяють точніше інтерпретувати передбачення моделі MelodySim на рівні сегментів. Це може бути корисним у використанні як цієї моделі, так і подібних моделей.

Можна виокремити два потенційні напрямки для подальших досліджень – покращення датасету (включення до нього різноманітніших даних чи експерименти із незбалансованими даними) та покращення моделі (удосконалення її архітектури та гіперпараметрів). Також перспективною може бути інтеграція подібних рішень із базами даних та корпоративними застосунками [11].

Література

1. 2025 Year-End Music Report. Luminate. 2026. URL: https://luminatedata.com/reports/yearend-music-industry-report-2025/
2. Зарипов Р.Х. Кибернетика и музыка. М.: Наука. 1971. 235 с.
3. Suno | AI Music Generator. URL: https://suno.com/
4. Roy A., Puri G., Herremans D. Text2midi-InferAlign: Improving Symbolic Music Generation with Inference-Time Alignment. arXiv. 2025. DOI: https://doi.org/10.48550/arXiv.2505.12669
5. Lu T., Geist C.-M., Melechovsky J., Roy A., Herremans D. MelodySim: Measuring Melody-aware Music Similarity for Plagiarism Detection. arXiv. 2025. DOI: https://doi.org/10.48550/arXiv.2505.20979
6. Зарипов Р.Х. Машинный поиск вариантов при моделировании творческого процесса. М.: Наука. 1983. 232 с.
7. Wang, A. 2003. An industrial strength audio search algorithm. In *Ismir* (Vol. 2003, pp. 7-13).
8. How Content ID works - YouTube Help. URL: https://support.google.com/youtube/answer/2797370?hl=en
9. Li Y., Yuan R., Zhang G., Ma Y., Chen X., Yin H., Xiao C., Lin C., Ragni A., Benetos E., et al. Mert: Acoustic music understanding model with large-scale self-supervised training. arXiv. 2023. DOI: https://doi.org/10.48550/arXiv.2306.00107
10. Manilow, E., Wichern, G., Seetharaman, P., & Le Roux, J. (2019). Cutting music source separation some slakh: A dataset to study the impact of training data quality and quantity. *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (pp. 45–49). IEEE. 2019. DOI: https://doi.org/10.1109/WASPAA.2019.8937170
11. Panchenko, B. and Pasenchenko, T. (2026) “Advantages and disadvantages of using incrementally maintained materialized views in enterprise applications”, *International Scientific Technical Journal "Problems of Control and Informatics"*, 71(1), pp. 59–67. DOI: https://doi.org/10.34229/1028-0979-2026-1-5

Дата першого надходження до видання: 26.05.2026
Внутрішня рецензія отримана: 17.06.2026
Зовнішня рецензія отримана: 21.06.2026
Дата прийняття статті до друку: 10.08.2026
Дата публікації: 29.08.26

Про авторів:

[1] *Панченко Борис Євгенійович*,
доктор фізико-математичних наук,
старший науковий співробітник
[1] *Panchenko Borys,*
PhD (physics & mathematics),
senior researcher
https://orcid.org/0000-0002-8085-4043

[2] *Пасенченко Томас Олексійович*,
аспірант
[2] *Pasenchenko Tomas*,
PhD student
https://orcid.org/0009-0007-8240-0372

Місце роботи авторів:

[1] Інститут кібернетики імені В.М. Глушкова НАН України
[1] V.M. Glushkov Institute of Cybernetics of the NAS of Ukraine
Тел.: +380(67)4493970
Email.: pr-bob@ukr.net

[2] Одеський національний університет імені І.І.Мечникова
[2] Odesa I.I.Mechnikov National University
Тел.: +380(68)7385496
Email: tomas.pasenchenko@gmail.com